CJSJ

Decision Bias in Recognition Memory: How Memory-Selective Neurons Encode Criterion Shifts

Cloris Shi[1], Evan Layher[2]

Abstract— **Humans make recognition-based decisions by assessing stimuli familiarity while considering strategic biases. For example, in the legal system, people balance eyewitness memory against the consequences of mistakenly identifying the wrong suspect. Under uncertain memory, individuals can flexibly shift criteria to optimize decisions based on the situation. Prior single-neuron research has characterized memory-selective (MS) neurons that accurately distinguish new and familiar stimuli, but it remains unclear whether these neurons respond differently under various decision biases. Here, we recorded extracellular action potentials of single neurons across frontal and temporal cortices while subjects performed an image recognition task with criterion manipulations. Firing rate patterns of MS neurons were decoded using a Support Vector Machine (SVM) classifier to identify selectivity during pre-stimulus and stimulus periods. In all recorded brain regions, we identified MS, visually-selective, and criterion-selective neurons. Furthermore, MS neurons encoded criterion shifts and image categories, suggesting that memory is integrated in parallel with other stimuli within the single neurons. These findings reveal that MS neurons are influenced by decision biases and other stimulus features that encompass the nuances of memory-based decision-making.**

I. Introduction

Recognition memory allows us to identify familiar objects, individuals, and events. It is an important component of declarative memory, or memory that is selectively and consciously retrieved. Recognition memory is often impaired in individuals with amnesic or neurodegenerative conditions like Alzheimer's disease, so recognition tasks are routinely used for assessments of human cognition [1]. Under uncertainty, memory-related decisions rely on integrating information from two aspects: (1) strength of familiarity with the stimulus and (2) the external context in which the stimulus is presented [2]. In other words, memory recall is often integrated with metacognitive decision processes when people decide whether to report or withhold uncertain evidence of familiarity. Hence, decision-making requires the interplay of memory strength and context.

Signal detection theory, a mathematical model, provides a framework to quantify the distinction between the retrieval and the decisional components of recognition memory. In this model, a human indicates a "remember" response by making a high-confidence "old" judgment. Adaptive decision-making involves setting a decision criterion along the familiarity continuum, a threshold above which a human will classify an item as a prior encounter [4]. In an old-new recognition memory task, a liberal criterion accepts items as old despite little memory evidence, while a conservative criterion accepts fewer items as old and requires more memory evidence. Shifting criteria allows an agent to make optimal decisions in different scenarios.

The neuroscience of recognition memory has remained elusive. Neuroimaging studies of recognition memory have long overlooked the role of criterion-shifting, a flexible and highly individualistic characteristic that introduces response bias during evaluation [5]. In fact, participants engage in decision biases during tasks even in the absence of instruction. To better disentangle memory and decisional processes, it is necessary to control for decision biases in our experimental design through deliberate manipulations.

Variation in criterion-shifting between individuals suggests the existence of a biological basis encoding criterion. Some individuals will consistently maintain either a liberal or conservative criterion, while others are more flexible in adapting their criterion to optimally adhere to shifting priorities. Using a behavioral task to deliberately encourage criterion shifts, we have identified criterion-sensitive regions across widespread fronto-parietal regions by using functional magnetic resonance imaging (fMRI) [6]. Neuroimaging like fMRI uses hemodynamics, or changes in blood flow measured with blood-oxygen-level dependent (BOLD). Hemodynamics serves as a useful proxy because neuron activity increases with greater blood perfusion of the brain. However, neuroimaging displays the activity of populations of neurons, so improved techniques are needed to understand the dynamics of specific single neurons.

Single neuron recording provides much better temporal and spatial resolution than non-invasive neuroimaging or electroencephalograms. Prior single-neuron studies have suggested that single cells, or neurons in this case, can encode stimulus identity (visually-selective) and familiarity (memory-selective) with abstract concepts. Memory-selective neurons, which fire differently in response to previously seen concepts, are in the hippocampus and amygdala [7]. However, it remains unknown how neurons selectively integrate memory retrieval and decision biases, and whether both signals are engaged dynamically when required.

Here, we aimed to uncover the existence of criterion-selective neurons, which are cells that respond differently under liberal and conservative criteria. Then, we sought to understand how visually-selective and memory-selective neurons respond under controlled criterion-shifts to determine if memory and decisional aspects of recognition are processed in parallel.

II. Methods

A. Behavioral Task Design

Patients with drug-resistant epilepsy received electrodes for intracranial monitoring based on clinical criteria. Each macro-electrode contained eight 40 μm diameter microwires, from which broadband 0.1 to 9000 Hz extracellular signals were recorded at a 32 kHz sampling rate (ATLAS system, Neuralynx Inc.). Subjects (ages 24–65 (M=40); 5 female, 1 male, 1 non-binary) were 7 epilepsy patients from Cedars-Sinai Medical Center. The institutional review boards of Cedars-Sinai Medical Center and the California Institute of Technology approved all protocols.

[1] Cloris Shi is with Troy High School, 2200 Dorothy Ln, Fullerton, CA 92831 (corresponding author to e-mail: clorisfshi@gmail.com)
[2] Evan Layher is with the Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106 and Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048.

Participants performed a recognition memory task with criterion manipulations (Appendix 1A, B). Each subject viewed 256 images per session, which had 16 test blocks of 16 images retrieved from the LaMem database [8].

Each block began with a study phase. Subjects memorized a series of 32 images, shown once (16 images) or twice (16 images). Discriminability or difficulty of memory recall was considered "easy" for images displayed twice and "hard" for images displayed once.

In the test phase, subjects decided whether the image had been present in the study phase and how confident they were. Correct responses for old or new items are "hits" or "correct rejections," while incorrect responses for an old or new item are "false alarms" or "miss" respectively.

Two error schemes were implemented to alternatively encourage liberal or conservative criteria shifts. While implementing a liberal criteria, one minimizes misses, while under a conservative criteria, one minimizes false alarm. Subjects' performance were evaluated based on two parameters: (1) difficulty of the task, assessed by discriminability (d_a), and (2) strategy of the subject, assessed by criterion placement (c) and criterion shift (C).

A normal, equal-variance signal detection theory model evaluated task performance per individual (Appendix 1.C.). Discriminability (d_a), differences in discriminability across conditions (Δd_a), criterion placement (c_a), and criterion shifting (C) is computed across all conditions. Hit rate (*HR*) and false alarm rate (*FAR*) were computed from summation of the total hit (H), miss (M), correct rejection (CR), and false alarm (FA) rates within each trial:

$$HR = \frac{H}{H+M} \quad (1)$$

$$FAR = \frac{FA}{CR+FA} \quad (2)$$

$$d_a = \left(\frac{2}{1+s^2}\right)^{1/2}[z(HR) - sz(FAR)] \quad (3)$$

$$c_a = \frac{-\sqrt{2}s}{(1+s^2)^{1/2}(1+s)}[z(HR) + z(FAR)] \quad (4)$$

$$C = c_a(conservative) - c_a(liberal) \quad (5)$$

$$\Delta d_a = d_a(moderate) - d_a(low) \quad (6)$$

where z is the density of standard normal distribution and s is the standard deviation ratio between old item and new item distributions. We set s = 0.8, which is the mean ratio computed for recognition memory tests [9].

B. Spike Sorting

Extracellular action potentials from neurons were recorded in subjects during the task. Raw signals were bandpass filtered 300–3,000 Hz. Spikes were detected and sorted offline using OSort (v4.1), a semi-automatic, template-matching algorithm [10]. OSort used an adaptive threshold to rudimentarily filter putative neurons, but spikes could be interrupted by electrical noise or neurons firing in synchrony. Thus, manual offline sorting remained necessary with a qualitative rubric (Appendix 2.A). We recorded 955 single neurons in five brain regions: medial temporal lobe (MTL; n=334), medial frontal cortex (MFC; n=288), posterior temporal cortex (PTC; n=158), orbitofrontal cortex (OFC; n=104), and insular cortex centromedian thalamic nucleus (INS, CM; n=71) (Appendix 2.B).

C. Statistical and Decoding Analysis

Neuron selectivity was determined by comparing firing rates between two 1s windows: (1) pre-stimulus phase defined -1000–0 ms before image onset and (2) stimulus phase from 200–1200 ms after image onset (t=0 ms). Recorded stimulus period is offset by 200 ms from image onset to compensate for the time neurons need to encounter and process visual stimuli [11].

We assessed four types of neuron selectivity: image category, criterion, memory, and confidence. A neuron was considered visually-selective (VS) if the firing rate differed across the four visual categories using a 1×4 analysis of variance (ANOVA) test at p<0.05. A memory-selective (MS) cell differentially fires when correctly classifying new and old stimuli; confidence-selective neurons respond differently in low and high confidence scenarios, as assessed using paired t-tests. A criterion-selective (CS) neuron fires distinctively in conservative and liberal criteria trials either in pre-stimulus or stimulus test phases.

Single-trial population decoding was performed on high-firing neural populations (>0.4 Hz) assembled across sessions. Firing rates for each cell were first detrended and then normalized. A 10-fold cross validation linear support vector machine (SVM) classifier for neural decoding was implemented in MATLAB (R2023a). The decoder randomly selected 80% of the neural population to analyze, and performance was tested on the remaining 20% to estimate the variance of random decoding. Time-resolved decoding was further performed on spike counts measured in a 200-ms moving window from -1000–1200 ms. The decoder's performance was evaluated against the 95th percentile of a null distribution. Performance was defined as average accuracy at selecting the correct category.

III. Results and Discussion

A. Task Performance

We sought to understand neuron responses to stimuli and behaviors. Discriminability manipulations successfully altered task difficulty between hard and easy discriminability conditions. Mean image discriminability (d_a) across hard discriminability trials (Mean (M)=1.16, Standard Deviation (SD) =0.943) remained significantly lower compared to easy ones (M=1.68, SD=1.18) (MΔ = -0.52, 95% CI [-0.66, -0.38]) confirming that viewing stimuli once versus two times during the study phase effectively modulated task difficulty for subjects.

Criterion (c) assesses the direction and magnitude of decision biases. Mean c in the conservative condition (M=0.066, SD=0.41) was significantly greater than the mean c in the liberal condition (M =-0.25, SD = 0.49), showing that the penalty system shifted criterion placement (MΔ = 0.32, 95% CI [0.26, 0.38]).

A criterion shift (C) between conservative and liberal sessions was more prominent across all subjects during uncertain memory. Mean C in hard discriminability conditions (M=0.40, SD=0.53) was significantly higher than mean C in easy discriminability ones (M=0.24, SD=0.53) (MΔ = 0.16, 95% CI [0.091, 0.23]). We conclude that participants rely more heavily on memory strength during easier tasks but use strategic criteria shifts when it is difficult to discriminate images.

A. *Neuron Selectivity*

Electrodes recorded a total of 955 single neurons and 2,917,842 action potentials. This was assessed using a standard analysis of variance (ANOVA) test that identified 787 (82%) selective neurons in the pre-stimulus (62%) and stimulus (38%) phases. Selective neurons are defined as neurons that exhibited statistically significant differences in firing rate pattern in response to various stimuli, such as image category, criterion placement, memory familiarity, and recollection confidence. In total, neurons encoded 2,351 selectivities, suggesting that one neuron could encode multiple selectivities.

Among stimulus-phase neurons, VS neurons (n=171), shown in Figure 1B, constituted the largest proportion (35%). Among VS cells, decoding analyses revealed that neurons were selective to memory (p<0.001) during the stimulus but not pre-stimulus phase. The concurrent processing of memory in VS cells suggests that the parameters of each image category are represented as a memorized concept. These findings matched concept cells that encode abstractions like the categories of our task [3]. As such, VS cells served as the control group on which decoding accuracy was evaluated for MS and CS neurons.

MS neurons were identified during the stimulus phase of test periods. Their firing rates differ between trials where an accurate old or new response was given (Figure 1A). There were 18% of stimulus-phase neurons significantly (p<0.05) modulated by memory, similar to prior studies [12]. Memory-selectivity was decoded through a SVM classifier (p<0.001) across all defined brain regions but most appeared in the MTL (45%). This matches the current understanding of the MTL as the main site for the encoding and integration of memory. The decoding classifier revealed that MS neurons were also selective to confidence (p = 0.0035; Figure 2A), suggesting that memory may be encoded in gradations by familiarity. Visually-selective MS neurons constituted 30% of the MS population, and SVM decoding in Figure 2B supports that memory recall relies on simultaneous categorization (p < 0.001).

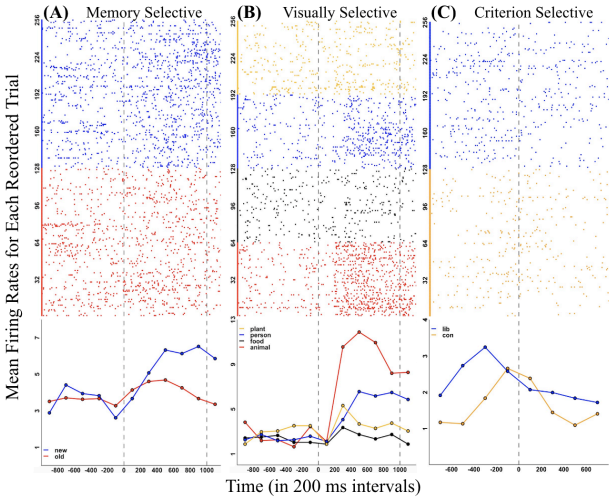


Figure 1. Example raster plots of mean firing rate of single neurons that are selective (p<0.05) to (A) memory (new > old) or (B) image category (plant > other categories) in stimulus period when image is shown, and (C) criterion (conservative > liberal) in the response period before decision is made.

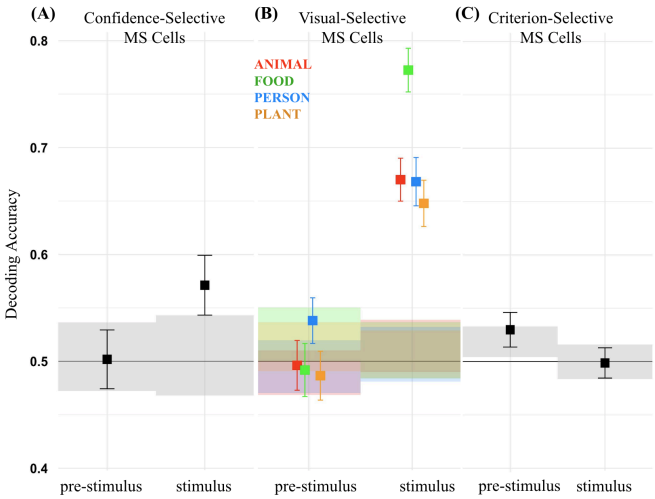


Figure 2. Significantly improved decoding performance suggests MS neurons encode confidence, visual category, and criterion. Decoding analysis depicts the mean of 100 iterations compared to the 95th percentile of a random (0.5) classifier (shaded). (A) Confidence selectivity in MS neurons (8%) during stimulus phase (p=0.0035). (B) Visual selectivity in MS neurons (30%) during stimulus phase (p<0.001). (C) Criterion-selectivity in pre-stimulus MS neurons (p=0.018).

Criterion-selective neurons, in Figure 1C, fire differently between liberal and conservative trials (p<0.001). They constituted 25% (193/787) of the total neuron population. During the pre-stimulus phase, criterion-selectivity within 17% of MS neurons was detected (Figure 2C), suggesting decision biases implemented before image onset and indicating the presence of anticipatory cells that filter memory recall. VS neurons, however, do not encode criterion, indicating a unique overlap of memory with decision bias. In the stimulus phase, criterion-selectivity was not significant among MS cells. While decision biases are implemented in the absence of image, criteria are not processed concurrently with memory in image presence.

IV. Conclusions

Single neuron actions and interactions are the building blocks of brain activity, and the clinical sequelae of most cognitive diseases stem from the dysfunction or failure of individual neurons. In our study on recognition memory, single neuron recording has helped elucidate how patients make context-informed decisions and recall information. To our knowledge, this paper is the first to characterize criterion-selective (CS) neurons. We identify that CS neurons influence memory recall when subjects set a decision bias before an image is seen. The encoding of criteria at the single-neuron resolution suggests a biological basis of individualistic decision thresholds. This finding supports the existence of criterion-related brain regions identified through neuroimaging by identifying a single-neuron basis of criterion-encoding [6]. As a result, our results also support the application of single-neuron recording to the monitoring of memory-related pathological development. Developing neural maps of decision-making pathways can support precision medicine, allowing for personalized treatment for neurodegenerative diseases. To this end, single unit recordings have been used to determine the structure of basal ganglia in Parkinson's patients,

creating a map of information flow throughout the brain and allowing physicians to monitor disease progression [13].

Further, our research also reveals that an individual neuron can encode multiple sources of information, defying categorization as selective to only one category of stimuli. We find that memory-selective neurons are influenced by criterion, confidence, and category, suggesting that single neurons are highly nuanced because they integrate various sources of information when processing memory signals. The potential of an individual neuron to harbor multiple varieties of stimuli provides an exciting opportunity to evaluate how such a synchronization of signals is affected by pathology. Thus, our research supports using single-neuron electrode recording to decipher the neural code underlying human behavior, understanding how cells respond to different stimuli.

There exists several limitations to our proposed single-neuron model of memory-based decision making. Electrode recording remains a solely clinical intervention, and it is unknown if our observed structure of memory pathways translates to a healthy population. Although we took care to record neurons from healthy brain regions of epileptic patients, it is possible that pathological tissue could be found at undetected sites. To this end, we seek to expand our sample from seven patients to a larger population. We further aim to replace the manual aspects of pre-processing data using OSort with machine-learning algorithms in neuron classification. Although it remains standard for a human evaluator to verify neurons using a qualitative rubric, we seek to completely automate neuron classification, allowing for single-neuron data to be quickly accessible after collection.

We also note that our dataset includes a disproportionately higher number of neurons across the frontoparietal lobe. In these regions, we found memory-selective neurons that responded when decision biases were implemented in recognition-based judgments. Thus, future research can explore increased electrode monitoring in the medial temporal and frontal lobes, regions in the brain where greater concentrations of neurons were identified. This may aid in localizing subdomains of memory or criterion activity and tracking memory-encoding pathways between neurons.

Already, single-neuron research has been used in humans to treat psychiatric and neurologic diseases. Deep brain stimulation (DBS), which uses electrodes targeting single neurons, has shown promise in treating chronic pain, motor disorders, Parkinson's disease, Alzheimer's disease, and depression. For these patients, single-neuron research can provide treatment tailored to an individual's criterion-shifting and memory networks. Thus, this study is positioned at the forefront of a paradigmatic transition towards a more highly-resolved and computationally-driven approach to studying macroscopic behaviors in human cognition.

V. Acknowledgements

We thank the Miller Lab and staff at Cedars-Sinai Epilepsy Monitoring Center for assistance. We thank Dr. Lina Kim of the UC Santa Barbara Research Mentorship Program for her guidance.

VI. References

1. Brown, M. W., Warburton, E. C., and Aggleton, J. P. (2010). Recognition memory: material, processes, and substrates. Hippocampus, 20(11), 1228-1244.
2. Fechner, H. B., Pachur, T., Schooler, L. J., Mehlhorn, K., Battal, C., Volz, K. G., & Borst, J. P. (2016). Strategies for memory-based decision making: Modeling behavioral and neural signatures within a cognitive architecture. *Cognition*, 157, 77-99.
3. Tajika, H. (2001). Recognition memory, psychology of.
4. Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., ... & Miller, M. B. (2012). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition*, 40, 1016-1030.
5. Layher, E., Dixit, A., & Miller, M. B. (2020). Who gives a criterion shift? A uniquely individualistic cognitive trait. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(11), 2075–2105. https://doi.org/10.1037/xlm0000951
6. Layher, E., Santander, T., Chakravarthula, P., Marinsek, N., Turner, B. O., Eckstein, M. P., & Miller, M. B. (2023). Widespread frontoparietal fMRI activity is greatly affected by changes in criterion placement, not discriminability, during recognition memory and visual detection tests. *NeuroImage*, *279*, 120307.
7. Minxha, J., Adolphs, R., Fusi, S., Mamelak, A. N., & Rutishauser, U. (2020). Flexible recruitment of memory-based choice representations by the human medial frontal cortex. *Science*, 368(6498), eaba3313.
8. Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. In Proceedings of the IEEE international conference on computer vision (pp. 2390-2398).
9. Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological review*, 99(3), 518.
10. Rutishauser, U., Schuman, E. M., & Mamelak, A. N. (2006). Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. *Journal of neuroscience methods,* 154(1-2), 204-224.
11. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? Neuron. 2012 Feb 9;73(3):415-34. doi: 10.1016/j.neuron.2012.01.010. PMID: 22325196; PMCID: PMC3306444.
12. Faraut, M., Carlson, A. A., Sullivan, S., Tudusciuc, O., Ross, I., Reed, C. M., ... & Rutishauser, U. (2018). Dataset of human medial temporal lobe single neuron activity during declarative memory encoding and recognition. *Scientific data*, 5(1), 1-11.
13. Boraud, T., Bezard, E., Bioulac, B., & Gross, C. E. (2002). From single extracellular unit recording in experimental and human Parkinsonism to the development of a functional concept of the role played by the basal ganglia in motor control. *Progress in neurobiology*, 66(4), 265-283.

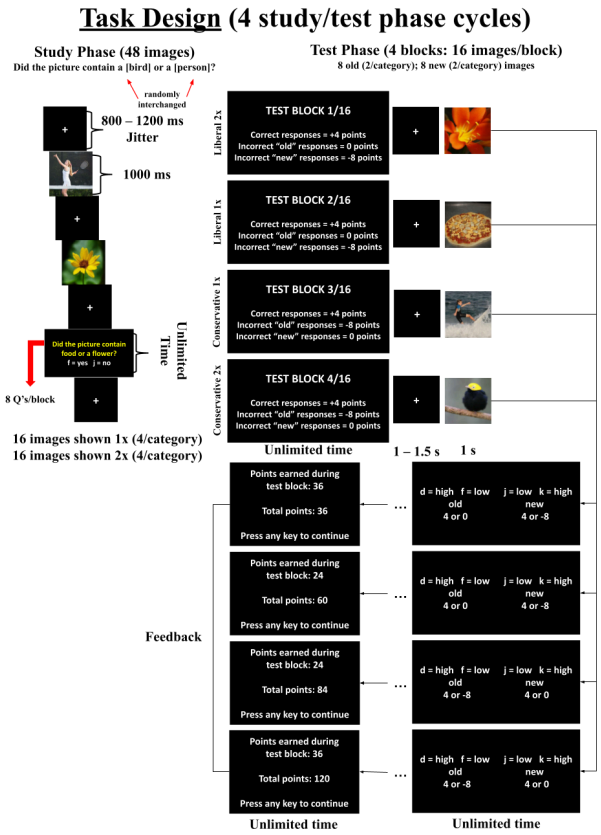
APPENDIX

1. Behavioral Tasks

A. Task Design

The task consisted of 4 study-test phase cycles. In each cycle, our 2 (discriminability condition: low vs. high) x 2 (criterion condition: conservative vs. liberal) factorial design created a total of 4 conditions, repeated 4 times for a total of 16 blocks.

Images were shown for 1000 ms, separated by a variable period. This was to prevent anticipatory memory signals in neurons in the case that we had used a predictable crosshair period. Questions (8 questions/block) assessing image category were interspersed between stimuli to ensure subjects remained attentive throughout the task.

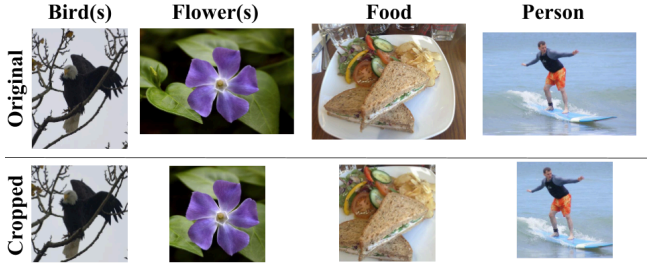


B. Image Selection and Modification

Images retrieved from the LaMem database were modified in two ways to ensure that memorability varied solely on image content [11]. First, images were center cropped to a consistent size (400x400 pixels) as dimensions may affect memorability to an unknown, albeit small, degree. Next, stimuli were selected if their memorability score ranged from 0.60–0.91 out of a 0–1 scale, ensuring that images displayed were memorable enough so participant memory capabilities could be adequately characterized.

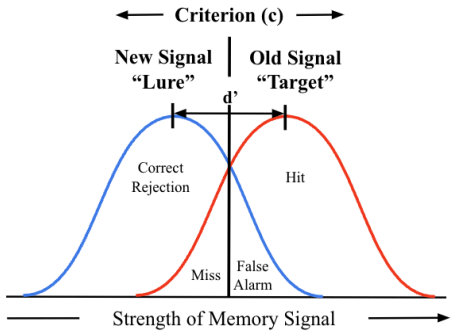
Images were assigned a memorability score from 0–1 (M=0.755) labeled by a convolutional neural network algorithm trained on human rankings. Memorability for selected task images (0.60–0.91, M=0.759) were equally stratified across four levels: 0.60–0.67, 0.68–0.75, 0.76–0.83, 0.84–0.91. We used non-overlapping image sets for patients completing multiple sessions.

During all test blocks, displayed images belonged evenly across subject category (4 images/category) within memorability levels (4 images/level). Below are sample images used in trials:



C. Behavioral Evaluation

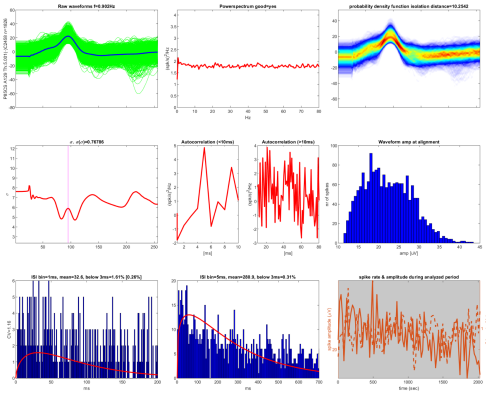
Performance on a behavioral task can be evaluated with several equations. A normally distributed, equal-variance signal detection theory model, shown below, was used to calculate hit rate (HR) and false alarm rate (FAR) per subject to characterize task performance per individual.



2. Neuron Spike Data Processing

A. Manual Spike Sorting Criteria

Neurons were evaluated on various criteria, including minimum firing rate, minimum amplitude, waveform coherence, power spectrum corruption, regularity of spiking rate, and firing frequency within a 3 ms interval. Here is a representative neuron.



B. Brain Regions and Specific Recording Sites

We aggregated different brain regions where electrodes were planted into five generalized brain domains. Electrodes were recorded in left (L) and right (R) sides of the brain, omitted below:

Medial Frontal Cortex	Medial Temporal Lobe	Posterior Temporal Cortex
Anterior cingulate cortex (ACC)	Amygdala (A)	Fusiform Face Area (FFS)
Supplementary motor association area (SMA)	Hippocampus (H)	Posterior temporal (PT)
	Parahippocampal gyrus (PHG)	
Orbitofrontal Cortex	**Other**	
Orbitofrontal (OFC, OF, OFI)	Insular cortex (INS)	
	Centromedian nuclei (CM)	