HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Nat Neurosci. 2015 July ; 18(7): 1041–1050. doi:10.1038/nn.4041.

Representation of retrieval confidence by single neurons in the human medial temporal lobe

Ueli Rutishauser[1,2,3,4,5], **Shengxuan Ye**[1,4], **Matthieu Koroma**[1,6], **Oana Tudusciuc**[4,8], **Ian B. Ross**[7], **Jeffrey M. Chung**[2], and **Adam N. Mamelak**[1]

[1]Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles

[2]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles

[3]Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles

[4]Computation & Neural Systems Program, California Institute of Technology, Pasadena

[5]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena

[6]Departement de Biologie, Ecole Normale Superieure de Cachan, Cachan Cedex, France

[7]Department of Neurosurgery, Huntington Memorial Hospital, Pasadena

[8]Division of Humanities and Social Science, California Institute of Technology, Pasadena

Abstract

Memory-based decisions are often accompanied by an assessment of choice certainty, but the mechanisms of such confidence judgments remain unknown. We studied the response of 1065 individual neurons in the human hippocampus and amygdala while neurosurgical patients made memory retrieval decisions together with a confidence judgment. Combining behavioral, neuronal and computational analysis, we identified a population of memory-selective (MS) neurons whose activity signaled stimulus familiarity and confidence as assessed by subjective report. In contrast, the activity of visually selective (VS) neurons was not sensitive to memory strength. The groups further differed in response latency, tuning, and extracellular waveforms. The information provided by MS neurons was sufficient for a race model to decide stimulus familiarity and retrieval confidence. Together, this demonstrates a trial-by-trial relationship between a specific group of neurons and declared memory strength in humans. We suggest that VS and MS neurons are a substrate for declarative memories.

Introduction

Decisions are often accompanied by an assessment of how likely it is that a choice will be correct. Such confidence judgments are critical in complex environments where decisions

Author contributions

U.R. and A.M. designed the experiments; U.R. and O.T. performed experiments; U.R., M.K., and S.Y. performed analysis; A.M. and I.R. performed surgery; J.M.C. provided patient care. U.R. and A.M. wrote the paper. All authors discussed the results at all stages of the project.

need to incorporate future, not yet observed, outcomes based on previous actions, information, and outcomes. Determining whether a stimulus is novel or familiar is a complex decision involving the comparison of sensory information with internal variables. While the outcome is binary (familiar or not), in humans such memory retrieval decisions are typically accompanied by graded judgments of confidence. Such confidence judgments feel automatic and are often accurate[1–3]. Despite its ubiquity, the mechanism of confidence judgments in memory is not understood. One model proposes that confidence judgments require separate specialized processes that evaluate decisions after they have been made, thus drawing on metacognitive abilities that may be unique to humans[4]. In contrast, other models propose that an assessment of uncertainty is an integral and necessary part of any decision-making process itself[5]. Confidence can thus be assessed simultaneously and by the same process that makes the decision in the first place, a core concept of Bayesian models of decision-making[6]. While recent studies in non-human primates and rodents have provided evidence for the latter model during perceptual decisions[3, 7], nothing is known so far about how confidence judgments for memories are made. It has proven challenging to develop paradigms for animals to communicate an assessment of confidence in an experimental setting, a problem particularly acute for memories. Here, we take advantage of the availability of human neurosurgical patients for single-unit recordings to study this question.

The medial temporal lobe (MTL) is required to make declarative memory-based decisions[8] and populations of neurons in the MTL whose interaction is thought to underlie this ability have been identified. For example, the response of some neurons in the primate MTL is selective for visual categories or concepts[9–12]. Others signal whether a stimulus is novel or familiar[13–16], a response which can emerge after a single exposure.[13, 14]. Such memory-sensitive neurons represent a potential substrate for episodic memories by marking stimuli as either novel or familiar. If so, we hypothesize that their activity should correlate with memory strength and thus with confidence. In contrast, neurons not directly involved in memory retrieval, such as those representing visual features, should not correlate with memory strength.

Here, we used subjective confidence ratings made by subjects during a memory recognition task to identify groups of neurons that signaled memory strength. We make two key contributions. Firstly, we show that memory-selective and visually-selective neurons code orthogonal pieces of information about visual stimuli. Secondly, we show that only the activity of memory-selective neurons correlates trial-by-trial with memory strength. In contrast, the ability of visually selective neurons to differentiate different stimuli was not sensitive to memory strength.

Results

Task and behavior

Subjects (44 sessions from 28 patients, see table S1 for demographics) performed a recognition memory test during which they rated 100 images as seen before or not[17]. Fifty of the images were familiar (shown ~30min before the task during a separate learning session), while the other 50 images were novel (stimulus type, “familiar” or “novel”). Images were presented for 1s each, and after a short delay subjects were asked to indicate

whether they had seen the image before (binary decision, “new” or “old”) together with a judgment of confidence in their decision (Fig. 1a). Each image belonged to one of five visual categories (cars, foods, people, landscapes, animals; see methods).

Subjects correctly identified 69±13% of familiar stimuli and reported 28±17% of novel stimuli as false positives (Fig. 1b). Confidence ratings were systematically related to accuracy (Goodman-Kruskal gamma correlation g=0.36±0.37, t-test vs chance p<1e-6). The higher the confidence, the better the accuracy (Fig. 1c–g). We computed a receiver operating characteristic (ROC) curve[18] for each session to quantify the relationship between accuracy and confidence (Fig. 1c). The average area under the curve (AUC) of the ROC was 0.75±0.08 (Fig. 1c,d). Different confidence ratings resulted in performance located in different locations within ROC space (Fig. 1c). The ROC was asymmetric (Fig. 1e, z-ROC slope 0.78±0.33, significantly less than 1, p<1e-18), as expected for declarative memories[19]. Subjects performed above chance at all levels of confidence and the majority of decisions were made with high confidence (Fig. 1f,g). Subjects assigned medium and low-confidences more rarely and with approximately equal likelihood (Fig. 1f). For a balanced statistical comparison between confidence levels with approximately equal trial numbers, we use two levels of confidence for the neuronal analysis: high and low. Trials with intermediate ratings were re-assigned a high-or low confidence rating depending on the proportion of trials (irrespective of performance) made with medium confidence (see methods). The resulting two confidence ratings were associated with different retrieval accuracy (Fig. 1h).

The decision time (DT, time from question onset till response) varied systematically as a function of confidence and accuracy (Fig. 1i–l; repeated measure ANOVA model, see methods). Correct high-confidence decisions were faster compared to low-confidence decisions (Fig. 1i, 1.54±0.11s vs. 2.49s ±0.20s, main effect of confidence $F_{1,30} = 25.74$, $P < 10^{-4}$; Fig. 1i shows pairwise comparisons). Correct familiar decisions were faster than correct novel decisions regardless of confidence (Fig. 1i). This was also true for incorrect trials: high-confidence incorrect decisions were faster than low-confidence incorrect decisions (Fig. 1j). Correct decisions were made with higher confidence than incorrect decisions (Fig. 1k, 1.95± 0.06s vs. 1.65±0.05s; main effect of correctness $F_{1,41} = 58.3$ $p<10^{-8}$, Wilcoxon signed-rank test correct vs. incorrect : p<2.74e-009). Also, correct decisions were made quicker than incorrect decisions for familiar stimuli (1.41±0.13s vs. 1.78±0.14s, significant interaction $F_{1,30} = 8.51$, $P < 0.05$, n = 31 subjects). Because incorrect decisions were made more slowly and with lower confidence, we matched the average confidence in correct and incorrect trials. We found that correct decisions are made faster even after matching confidence (incorrect vs. correct: 1.89s±0.15s vs. 2.21s±0.19s, Wilcoxon signed-rank test after matching for confidence p<0.01 see Fig. 1l). Together, this shows that subjects accurately assessed the quality of their memories (Fig. 1h) and the relationships between DT and confidence were as expected for declarative memory retrieval decisions[1].

We selected subsets of sessions for analysis based on behavioral metrics only. Two groups were selected: Group 1 (patients with above chance retrieval performance, n=38 sessions, AUC=0.81±0.10, g=0.39±0.29) and Group 2 (patients who were able to distinguish between high-and low confidence memories, 26 sessions, AUC=0.84±0.08, g=0.38±0.27).

Electrophysiology

We isolated 1065 putative single units from the amygdala and hippocampus in 44 sessions (on average 24 per session). Units were carefully isolated[17, 20] and recording and spike sorting quality were assessed quantitatively (Fig. S1). The average firing rate was 1.84 ±2.66 Hz (Table S2). Throughout the manuscript, we use the term "neuron" to refer to a putative single unit. Neurons were sensitive to the onset of visual stimuli as expected[9, 17]: 30% (321/1065) of the neurons responded when comparing baseline with post-stimulus periods ($p<0.05$, two-tailed t-test, 1s each). Note that the analysis that follows was not restricted to visually responsive neurons.

Single-neuron signatures of memory

We first tested whether the neuronal response following stimulus onset depended on whether the stimulus was novel (not seen before) or familiar (seen before) stimuli. We found that the response of 8.5% (81 out of 954, $p<1e-5$, Bernoulli; correct trials only in Group 1, n=38 sessions, see Table S1 and Fig. S6 for bootstrapped significance values) of all neurons differed between novel and familiar stimuli (see Table S2 for mean firing rates). This was true for both amygdala (43/577, 7.5%) and hippocampal (38/377, 10.1%) neurons. We will call such neurons memory-selective (MS)[13]. Similar to previous experiments[13, 14, 21], there were two types of MS neurons (Fig. 2 shows examples). The first had a higher firing rate to novel compared to familiar stimuli (45/81, Fig. 2a,b) whereas the second had an increased firing rate for familiar compared to novel stimuli (36/81, Fig. 2c,d). We will refer to these neurons as novelty- and familiarity-selective (NS and FS), respectively[13].

We next performed a single-neuron ROC analysis for every MS neuron and calculated its area under the curve (AUC). The AUC specifies the probability by which an ideal observer could predict the choice (novel or familiar) of a subject by counting spikes in an individual trial. Note that some studies refer to this metric as choice probability (CP)[22]. Only MS neurons from patients that were able to differentiate high from low confidences were considered (Group 2, 65 out of 664 units (9.8%) were MS units; Fig. 2e–h and Fig. S4p show example ROC curves). The average AUC for all MS neurons, considering all correct trials, was 0.64±0.04 (different from chance by design, as the neurons were selected to be different in the first place; what is important here is only the magnitude). We next computed AUC values using only high-or low confidence trials. Note that the selection of MS neurons does not consider confidence, making this comparison independent. AUC values were significantly larger for high compared to low confidence trials for all MS neurons together (Fig. 3a–c; 0.66±0.007 vs. 0.60±0.010; see legend for statistics) and for NS and FS neurons separately (Fig. 3d–e). This was true for both hippocampal and amygdala neurons, for neurons recorded from the left and right hemisphere only as well as when evaluating the differences using a bootstrap rather than parametric statistics (Fig S4; see legend for statistics). These differences could not be attributed to different units that might have been merged into one single cluster: the mean waveforms associated with each of the four trial types were indistinguishable (Fig. 2i–l). Comparing forgotten (false negatives, FN) trials with truly novel trials reveals an AUC larger than chance (Fig. 3c, 0.55±0.020, p=0.0048 vs. chance of 0.50) but significantly smaller than that for low-confidence correct decisions (0.60±0.010, p=0.0056). This indicates that MS neurons carried a memory signal that was

strongest for high confidence correct trials, intermediate for low-confidence trials and weakest for forgotten trials (Fig. 3c).

We performed a number of controls to exclude possible confounds. Using MS neurons from non-epileptic areas showed a similar difference (n=40, AUC 0.66±0.01 vs. 0.61±0.01, p=0.00066), as did using only neurons in epileptic tissue (later resected, AUC 0.67±0.01 vs. 0.61±0.02, p=0.0041). Equalizing the number of trials in the high-and low confidence groups did not change the result (AUC 0.67±0.01 vs. 0.60±0.02, p<4e-5). Finaly, randomly re-assigning confidences but keeping the novel/familiar labels intact abolished the high/low difference as expected (AUC 0.65±0.01 vs. 0.65±0.01, p=0.81; Fig. S4M–O shows bootstrap statistics).

We next compared the response patterns of FS and NS neurons. The previous ROC analysis is not sensitive to whether one or both terms constituting the difference are modulated. We thus next directly compared the normalized number of spikes fired by FS/NS neurons as a function of behavior. By design, FS and NS neurons responded maximally to familiar and novel stimuli, respectively (Fig. 3f–g). The response of FS/NS neurons differed significantly different between high-and low confidence trials, but only for the trial types to which the neurons increased their firing rate. Thus, the response of FS neurons differed between high- and low confidence trials only for familiar stimuli and vice-versa for NS neurons (Fig. 3f–g, see legend for statistics). Also, both FS/NS neurons decreased their firing rate to novel and familiar stimuli, respectively (Fig. 3h–i). The magnitude of this decrease, however, was insensitive to confidence. Thus, NS and FS neurons signal confidence asymmetrically because only the trial type to which they increase their firing rate relative to baseline is modulated by confidence. This conclusion relies on an absence of firing rate reduction below baseline, which is difficult to detect due to low baseline firing rates. However, note that this very problem would be faced by an imaginary downstream neuron receiving input from FS/NS neurons.

Single-neuron signatures of visual information

Each image shown belonged to one of five investigator–selected visual categories (cars, foods, people, landscapes, animals). The response of 17.5% (186/1065) of units was significantly modulated by category (1-way ANOVA, p<0.05, Fig. 4 shows examples), a proportion similar to what has been reported before[9] (see Table S1 and Fig. S6 for bootstrapped significance values). We refer to this group as visually selective (VS) neurons.

The two populations were independent: 15/186 VS neurons were also MS neurons (8%) whereas 15/87 (17%) of MS neurons also distinguished categories (χ^2 test of independence, p=0.91; this also applies considering only neurons from Group 1 and 2 and when excluding neurons with firing rates <1 Hz). A small group of neurons (15/1065, 1.5%) were both MS and VS cells (see Fig. S5 for an example), a proportion larger than expected by chance (chance level 0.25%, p=0.001, Fig. S6) and compatible with independence of memory-and visual selectivity. In what follows, we analyze VS and MS neurons without excluding those that code for both.

Did the response of VS neurons depend on memory strength? To answer this question, we first identified the most and least preferred stimulus category for each VS neuron (i.e. the neuron in Fig. 4e best differentiates between animals and houses). We then used single-neuron ROC analysis to quantify how well the response of each VS neuron discriminated between these two categories for four different trial types: novel, familiar, high-and low confidence. Using only correct trials from neurons in Group 2 (128/664 were VS neurons, see Table S2) we found that AUC values did not differ as a function of confidence (Fig. 5a,c) or familiarity (Fig. 5b,d, see legend for statistics). The same conclusions hold when excluding low-firing rate neurons (Fig. S7). This shows that the ability of a VS cell to identify its preferred category did not depend significantly on stimulus familiarity or confidence. This conclusion relies on the absence of a significant difference, which does not exclude the possibility that our data does not have enough statistical power to detect an existing difference. However, note that using the same number of trials and time window, MS neurons showed a strong difference. Also, the pairwise comparison between the two conditions (high/low and new/old) is based on trials for which the neuron carried information to begin with (the preferred category), assuring that the individual AUC values were well above chance.

VS neurons discriminate before MS neurons

We next estimated the first point of time at which the response of VS and MS neurons differed between different visual categories and novel/familiar stimuli, respectively. We compared the cumulative sum of the spike trains, a method which provides an estimate of the differential latency of a neuron with millisecond precision[15] (see methods). The average differential latency of VS and MS neurons was 272ms and 461ms, respectively (relative to stimulus onset; Fig. 6a–b, see legend for statistics). Thus the response of MS neurons was delayed by 189ms relative to VS neurons.

Differential coding of visual category and memory

We next considered all recorded neurons together (n=664, Group 2). We fit a moving-window regression model for every single unit (using correct trials only) to estimate how much of the neuronal variability could be attributed to the factors visual category and familiarity. We estimated the effect sizes[23] by ω^2 as a function of time (see methods). The population conveyed information about both the visual categories and the familiarity of the stimuli (Fig. 6c). VS neurons signaled information earlier and did not provide novelty information (Fig. 6g). In contrast, MS neurons signaled information about the novelty of the stimulus but not its categorical identity (Fig. 6f). To analyze neuronal activity regardless of time, we averaged the effect size in a 1.5s time window starting 0.2s after stimulus onset. Units classified as MS and VS neurons tended to have high effect sizes only for novelty/familiarity or category, respectively (Fig. 6s). The effect sizes were not correlated, indicating that a neuron coded either familiarity/novelty or category, but not both (Fig. 6e). This was true for MS, VS and all other neurons (r=0.04, −0.003 and −0.008, respectively; all p>0.86, Fig. 6e). Thus, a neuron was informative about only one but not both of the variables. We also utilized a a regression model with an interaction term, which did not explain any additional variance (Fig. S3). Comparing the effect size between trials which were recognized with high-and low confidence revealed that the information conveyed by

MS neurons (Fig. 6h–i) was sensitive to subjective confidence whereas that by VS neurons was not (Fig. 6j–k). Note that the estimated effect size of a neuron did not depend on spike sorting quality (Fig. S1h–i).

Estimate of information content

What distinguishes a high from a low confidence memory? We used a population decoder to estimate the amount of information provided in single trials as a function of confidence and accuracy. The decoder had access to a pseudo-population of neurons and was trained and tested on subsets of independent trials. The resulting estimates are generalization errors, permitting comparisons such as whether training the decoder with a condition (i.e. high confidence) generalizes to other conditions (i.e. low confidence). Applying this method to all recorded VS/MS neurons revealed that visual information carried by VS neurons could be decoded earlier than memory information carried by MS neurons (Fig. 7a). This extends the earlier finding to single-trial decoding. To quantify the information available we used the mutual information (MI) between the spiking response and stimulus identity/familiarity (see methods). This again revealed an early-and late component that is carried by VS/MS neurons (Fig. 7b–c). We next trained a decoder that had access to all recorded neurons using only high confidence trials and tested its performance on both high-and low confidence trials (Fig. 7c,d). While this decoder based its decisions on neurons signaling high confidence memories, low confidence trials could still be decoded but the amount of information available was reduced by ~70% (Fig. 7e, middle; 0.14±0.04 vs. 0.04±0.02 bits). Thus, the population response identified for high confidence trials is still informative for low confidence memories. Training a decoder on all trials regardless of confidence and testing it on high and low confidence trials separately showed similar results (Fig. 7e; 0.15±0.03 vs. 0.05±0.02 bits). This result holds also when only considering MS neurons (Fig. 7e). We conclude that the amount of information available in the entire population, in bits, is ~3× higher for high compared to low confidence memories. We next estimated the MI during error trials. This revealed that when a stimulus was forgotten (false negative, FN), the spiking activity of MS neurons still contained information about the familiarity of the stimulus (Fig. 7f). While more than expected by chance (0.044 vs. 0.023 bits, 1.97× more information), this was less than that available for low confidence correct trials (Fig 7e). Forgotten trials thus form a continuum with the low-and high confidence correct trials, a property that is expected of a memory strength signal. Note that in contrast to MI, decoding accuracy cannot be used to compare amounts of information. Nevertheless, a similar qualitative pattern of readout ability was revealed by decoding accuracy (Fig. 7g,h).

Differences in electrophysiological signatures

We next compared the shape of the extracellular waveforms (EWs) associated with each neuron to investigate whether VS/MS cells might be physiologically different. The trough-to-peak time d (Fig. S2a) was bimodally distribution across all recorded neurons (Fig. S2a,b), indicating at least two types of EWs: short and long (mode 0.3ms and 0.8ms, Fig. S2b,c). Considering d separately for particularly well isolated MS and VS neurons (projection test distance >10 s.d.; all conclusions remain valid without this criteria) revealed that only the EWs of VS neurons were significant bimodally distributed (Fig. S2d, see legend for statistics). In contrast, 72% of all EWs of MS neurons were short (Fig. S2f). The

proportion of long and short EWs was significantly different for MS but not VS neurons (Fig. S2f, see legend for statistics). At the same time, both VS/MS neurons had low firing rates and did not differ according to other spike train metrics (CV2 and burst index, Tables S2 and S3). In conclusion, both MS and VS neurons had low firing rate but at the same time MS neurons had mostly short EWs. Based on this, we hypothesize that MS neurons are anatomically distinct from VS neurons (see discussion).

Decision making model

Is the information provided by MS neurons sufficient to decide both whether a stimulus is familiar as well as the confidence in that decision? To answer this question, we constructed a biologically plausible race model[24]. The model evaluates whether the difference D(t) between one FS and NS neuron is negative or positive (Fig. 8a). If positive, the accumulated evidence (EV) for the stimulus being familiar is increased and vice-versa for negative D(t). At the end of the trial the decision is familiar if $EV_{fam} > EV_{nov}$, and novel if otherwise. The confidence in the decision is proportional to the "balance of evidence" $\Delta E = |EV_{fam} - EV_{nov}|$[25]. We evaluated the performance of this model for all n=954 pairs of NS/FS neurons, separately for correctly recognized familiar (TP) and novel (TN) items (Fig. 8b–h). The model reliably distinguished between high and low confidence trials (Fig. 8c–f) and EV and ΔE were correlated with behavioral performance. The model's ability to distinguish between novel and familiar stimuli was better for high compared to low confidence trials (Fig. 8h). Also, ΔE was correlated trial-by-trial with confidence, both for behaviorally correct and incorrect trials (Spearman correlation 0.042±0.13, p<1e-20 vs. 0, n= 957 pairs and 0.047±0.17, p=0.0033, n=130 pairs). Of the two EV values, only the larger (the winner) correlated with confidence (0.05±0.13, p<1e-30) whereas the EV value of the smaller (looser) did not (0.002±0.16 p=0.68). We also used the model to evaluate the decision latency by setting, for each cell pair, a fixed decision threshold ΔE_{Th} (see methods). The first time when ΔE exceeded this threshold, the race was aborted and the latency noted. This model made decisions more quickly for trials that were made with high confidence (Fig. 8i) and made familiar decisions more quickly than novel decisions (Fig. 8j). This pattern is similar to that observed behaviorally (Fig. 1i). Together, this shows that a simple readout mechanism can reliably, and on single trials, make two decisions simultaneously using only information provided by MS neurons.

Discussion

We systematically compared two populations of neurons within the human MTL: VS and MS neurons. The former signaled information about the identity of the visual stimuli, whereas the latter signaled the familiarity of the stimuli. VS neurons discriminated between stimuli ~190ms earlier than MS neurons and only the activity of MS neurons was correlated with memory strength as expressed by a confidence judgment. Together, our result suggests that only MS neurons are directly involved in memory retrieval. The proportion of MS neurons identified here was similar to those identified before[13, 14, 26]. However, using confidence ratings revealed several important new aspects of these neurons. In particular, this revealed that NS and FS neurons coded information asymmetrically: their firing rate is only informative about the confidence of the trial types to which they increase their firing

rate (Fig. 3). In contrast, we show here that the activity of the VS neurons is not sensitive to memory strength and that they are functionally distinct from MS neurons. In addition, our data is an independent reproduction of the initial description of VS neurons[9]. 1.5% of all neurons qualified as both VS and MS neurons. While rare, our large dataset shows that the probabilities of a neuron to become VS or MS neuron are independent of each other. Such neurons have been hypothesized to represent a distributed sparse code for memories[27, 28], but due to their rarity it will be necessary to use closed-loop paradigms to investigate them systematically.

Our conclusions rest on single-neuron ROC analysis, a sensitive method to quantify the amount of information available in individual trials[29]. ROC analysis does not assume a particular distribution of the spike counts, which is important because spike counts are Poisson distributed. Using mutual information, we further estimated that the amount of information present in the population is about 3 times higher in a high relative to a low confidence trial. Note that low confidence decisions were nevertheless correct, thus what was missing was additional information required to reach a high confidence choice. Also, low confidence decisions were slower, a signature of recognition memory that has been observed even when not asking for a confidence[1].

Confidence judgments are subjective. Consequently the strength associated with a certain confidence varies between subjects. Our analysis, however, is insensitive to this because it relies on a within-neuron comparison between high-and low confidence trials. As a result, all that is required for our analysis to be valid is that subjects apply a threshold regardless of its value. For statistical reasons, we focused our analysis on two levels of confidence only. A third level is forgotten (FN) trials, which can be considered a “very low” confidence. Our results show that these three levels are represented by MS neurons. Clearly, subjects are capable of using more than two confidence levels[1] and it remains an open question whether each of these can be separated by MS neurons.

Could the neuronal differences between high-and low confidence be attributed to fluctuations in attention during retrieval? The specificity of the neuronal effects argues against this possibility, because a global attentional effect would affect all neurons equally. In particular, it would be expected to improve the reliability of visual category information[30]. Instead, here we found no difference in the coding reliability of VS neurons.

In psychology, global models of recognition memory[1, 31, 32] have as their underlying decision variable a familiarity or strength signal that pools memory strength among many associations or items. In these models, the familiarity signal itself does not contain information about the memory apart from signaling its familiarity. MS neurons had the same property and are thus candidates for the familiarity signal predicted by these models. This will make it possible to directly test key hypothesis made by these influential quantitative models of memory[32].

We used a simple integrator-type model to explore which decisions could be supported by the difference in firing rate between a pair of FS and NS neurons. Integration of the difference of two neurons with opposite tuning is statistically optimal in many situations[24].

Our model differs from drift-diffusion (DDM) models[24, 33] because it has two integrators, only one of which increases its value depending on the sign of the difference. FS/NS neurons are not anti-correlated (Fig. 3F–I), and thus the two integrators are not redundant as is assumed in DDM models. The difference of the two integrators is the "balance of evidence"[5, 7, 25]. In contrast, a standard DDM model has only one decision variable[34] and thus no mechanism for estimating the quality of a decision beyond the time taken to reach the decision threshold[3]. Here, we show that integration-to-bound decision models are applicable to memory-based decisions because this model can make confidence decisions based only on the activity of MS neurons. No human neurons that represent the difference FS-NS or the integrator values EV have yet been identified, but our model makes specific predictions that will facilitate their discovery. A key technique to identify signatures of evidence accumulation has been to present sensory stimuli of different strength[22, 35]. Here we relied on internal variability in memory strength only, but we expect that combing these two approaches will be an important future avenue.

Extracellular waveforms (EW) have been used to classify cells as inhibitory or excitatory[36–38], but no definitive data on the validity of this distinction exists for humans. The EW differs as a function of the location of the electrode relative to the cell, but since our electrodes were implanted blindly this is unlikely to account for the difference. Large pyramidal cells can have shorter waveforms compared to smaller pyramids[39] and in rats particularly short waveforms are hypothesized to be axonal activity[40]. Also, backpropagation of action potentials widens the EW[41] and the propensity for backpropagation varies between cell types. Consequently, an intriguing possibility is that MS cells are morphologically and/or physiologically different from VS cells but this hypothesis remains to be confirmed.

In addition to the hippocampus, we identified VS/MS cells in the amygdala, confirming previous reports of memory signals in the human amygdala[13, 14, 26]. While the amygdala is not necessary for declarative memory, it is crucial for many aspects of learning[42] and is sensitive to stimulus novelty[43]. Given this, it is not surprising that VS/MS cells are also present in the amygdala. We used natural scenes as stimuli, some with emotional content. It remains an open question whether MS cells in the amygdala are specifically modulated by the emotional content of the stimuli. It also remains an open question whether MS cells are modulated by recency rather than novelty. Lists of words are frequently used in recognition memory[1] tests, but most physiological studies so far have used natural scenes. Notably, a recent study utilizing words reported cells tuned to recently seen words but not broadly-tuned cells of the kind we report here[27].

Assessing the quality of one's own memory (an internal state) is thought to require metacognition[44], the existence of which in animals is debated[5, 45, 46]. While only humans can verbally declare their confidence, experiments with indirect measures reveal that several species can utilize a "don't know" option[3, 7, 47, 48] alone or in combination with post-decision wagering[3, 49] to prevent the learning of an association instead of a confidence judgment. The amount of effort expended has also been used to infer confidence[50]. Theoretically, degrees of uncertainty are central components of neural computation[56]. Together, there is thus emerging evidence that an assessment of uncertainty is an integral

part of neuronal decision making in general. Here, we have demonstrated that memory-selective neurons in humans carry a graded representation of memory strength that is reflected in the subjective confidence ratings made by the subjects.

Online Methods

Electrophysiology and electrodes

Broadband extracellular recordings were filtered 0.1Hz–9kHz and sampled at 32kHz (Neuralynx Inc). We recorded bilaterally from the amygdala and hippocampus (32 channels in total, see[17] for details). 1 microwire in each macroelectrode served as a local reference (bi-polar recording). Electrodes were localized based on post-operative MRI images[17]. Electrode locations were chosen according to clinical criteria alone. Only electrodes localized to the hippocampus or amygdala were included. Protocols were approved by the institutional review boards of the Cedars-Sinai Medical Center, Huntington Memorial Hospital and the California Institute of Technology.

Patients

28 patients who were evaluated for possible surgical treatment of epilepsy using implantation of depth electrodes volunteered for the study and gave informed consent. We evaluated all patients using standard neuropsychological tests (Supplementary Table 1). All included patients had clearly distinguishable spiking activity on at least one electrode in the areas of interest.

Task

Details of the task have been published previously[17]. The task consisted of two blocks: learning and retrieval, with a 15–30 min delay in between with a distractor task. During learning, 100 novel and unique images were shown. During recognition, a subset of 50 of these images were shown again (now familiar, “old”) together with 50 novel images (novel, “new”). Patients identified each image as novel or familiar on a 1–6 confidence scale (Fig. 1a). Only the data from the retrieval block of the task is reported here. Before the experiment, subjects performed a short training version of the same task but with different images. Some that performed multiple sessions of the task were recorded on different days with different sets of images. Images shown were 9°×9° deg in size. After offset of the image, the screen was blank and followed by the question screen 0.5s later (Fig. 1a) that was displayed till an answer was provided. Stimuli were photographs of natural scenes of five different visual categories (animals, people, cars/vehicles, outdoor scenes/houses and flowers/food items). There were the same numbers of images presented in each category. The task was implemented in MATLAB using the Psychophysics Toolbox[51].

Behavioral analysis

The decision time (DT) is the time between onset of the question screen and the button press. We excluded DTs>30s as well as those which are more than 3 standard deviations away from the mean (for each subject) for all DT analysis (Fig. 1h–k; 1.74%±1.00% of trials, ±s.d. across subjects, were removed). All DT comparisons were pairwise within-subject comparisons. We excluded sessions which did not contribute at least one data point

to each category of a comparison (number of sessions for Fig. 1i–l are 38, 42, 44, and 31, respectively). All findings reported in Fig. 1i–l remain when using all 44 sessions and non-paired statistics (not shown). To analyze behavioral performance and proportion of responses (Fig. 1b–h), all trials regardless of DT were included. Note that the proportion of responses (Fig. 1f) remains virtually unchanged when applying the same exclusion criteria as used for the DT analysis.

The association between confidence and retrieval accuracy was assessed using the Goodman-Kruskal gamma coefficient g[52], whose value is between −1…1. The relation V=0.5*g+0.5 converts g into the probability V that a confidence judgment is accurate[52]. On average, V=0.67±0.18 (±s.d.).

We used a 3-way repeated measure ANOVA with in-between factors memory (novel/familiar), confidence (high/low), and accuracy (correct/incorrect) to quantify the relationship with DT. The repeated factor was subject number. Pairwise post-hoc comparisons were done using a Wilcoxon signed-rank test.

The behavioral ROC was calculated as a function of confidence as described previously[17]. The slope of a line fitted with least-squares regression to the z-transformed ROC was used to assess the degree of asymmetry of the ROC[53]. We reassigned the intermediate confidence level (2, 5) to either the low or high confidence level to collapse the 6 confidence levels to 4 levels. For every session, the intermediate confidence was assigned to either the low or high confidence group, based on which assignment produced a more equal proportion of high and low trials. This re-balancing was based on number of trials alone.

We assigned sessions to two groups. Group 1 consists of all sessions where patients performed at least 10% above chance. Group 2 is a subset of Group 1 and contains only sessions where patients accurately discriminated between high and low confidence memories (minimal accuracy for high 70% and low 55%). Using random subsets of 50% of the trials or only the first or second half of the trials resulted in identical group assignments.

Spike detection, sorting, and quality metrics

The raw signal was filtered with a zero-phase lag filter in the 300–3000Hz band and spikes were detected and sorted using the semiautomated template-matching algorithm OSort[20]. Channels with interictal epileptic activity were excluded. We computed several spike sorting quality metrics for all units (see Fig. S1): i) percentage of ISIs below 3ms was 0.24% ±0.45%, ii) the ratio between the peak amplitude of the mean waveform of each cluster and the standard deviation of the noise was 5.6±3.6 (peak SNR), iii) the pairwise projection distance in clustering space between all neurons isolated on the same wire was 16±11 (projection test[54]; in units of s.d. of the signal), iv) the modified coefficient of variation of variability in the ISI (CV2) was 0.93±0.21 (p=0.72, not significantly different from 1, as expected from a Poisson process), and v) the isolation distance[5556] (Fig. S1g; (n=746, median was 35.0; compare to Fig. S2b in[57] and Fig. 7 in[56]). The isolation distance quantifies, for every cluster, how far apart it is from the other clusters and the noise. We calculated the isolation distance in a 10 dimensional feature space[56] (Energy, peak amplitude, total area under the waveform and first 5 principal components of the energy

normalizes waveforms). To quantify whether our results depend on sorting quality, we correlated the effect size metric ω^2 with the isolation distance (Fig. S1h,i).

Selection of units

We counted spikes in a 200–1700ms window relative to stimulus onset. MS neurons were selected based on a significant difference between correctly identified novel and familiar stimuli in this period ($p<0.05$, two-tailed, boostrap comparison of means with 1000 runs). A MS neuron was FS if the mean if all familiar trials was larger than all novel trials and NS otherwise. VS neurons were selected using a 1×5 ANOVA with the factor visual category (1–5) based on the identical spike counts and with $p<0.05$.

Single-neuron analysis

We used non-overlapping bins of 250ms width. PSTH diagrams were smoothened, for display only, with a causal exponential kernel with λ=150ms. All analysis and statistics was based on un-smoothened data.

Single-neuron ROC analysis

Neuronal ROCs were constructed based on the spike counts in a 1.5s long window, starting 200ms after stimulus onset. We varied the detection threshold between the minimal and maximal spike count observed, linearly spaced in 25 steps. The AUC of the ROC was calculated by integrating the area under the ROC curve[18].

For MS neurons, ROC analysis was performed to quantify how well individual neurons distinguished between novel and familiar trials. Only neurons with at least 10 correct novel and familiar trials each were included in the ROC analysis. A separate ROC analysis was performed for high and low confidence trials. For confidence comparisons, only neurons that had at least 2 trials of each of the 4 confidence levels were included. To perform a fair comparison, only one of the two groups used for the ROC analysis was modified according to confidence while the other was kept constant. For FS neurons, the fixed group was all TN trials (regardless of confidence) which was compared with high-confident TP and low-confident TP trials separately. For NS neurons, the fixed group was all TP trials which were compared with high-confident TN and low-confident TN trials separately.

For VS neurons, we first identified, based on all trials regardless of behavior, a binary contrast (such as category 2 vs. 5, preferred vs. non-preferred) that a neuron distinguished best by testing all 10 possible contrasts and picking the one with the maximal AUC. We subsequently estimated the AUC for this best contrast using only novel, familiar, high, and low confidence correct trials.

Statistical comparisons between AUC values were made using two-tailed parametric tests (paired t-test and paired sign-tests, as indicated). For bootstrap comparisons, we performed B=1000 bootstrap runs to estimate the null distribution and estimated the p-value empirically by counting how many values in the null distribution were larger than the observed value. When no null distribution value exceeded the observed value, we set the p-value to 1/B.

To calculate a normalized firing rate (Fig. 3f–i), we divided the firing rate by the mean firing rate of the neuron in the entire task. For the cumulative distribution comparisons (Fig. 3f–i), we only included neurons that had at least 2 trials in each of the 6 behavioral categories.

Differential latency

We binned spike trains into 1ms bins and computed the cumulative sum. We then averaged the cumulative sums of all individual trials of a neuron that belong to the same condition. To allow averaging of all MS neurons, NS neurons were inverted so that the preferred response of all MS neurons was a firing rate increase. For VS neurons, the best contrast was used as determined by ROC analysis. We then compared, at every point of time, whether the cumulative sums of a group of neurons were different ($p<0.05$, pairwise t-test). We repeated this procedure after randomly scrambling the labels to estimate the null distribution. Corrections for multiple comparisons were performed using a cluster-size correction. The maximal number of consecutively significant data points in the null distribution was used as the minimal cluster size. The first point of time of the first significant cluster was used as the estimate of the differential latency[15]. Note that this method is not sensitive to baseline firing rate differences between neurons because the latency estimate is pairwise for each neuron individually.

Regression analysis

We used the regression model $S(t) = \alpha_0(t) + \alpha_1(t)N + \alpha_2(t)C$ to estimate whether the firing rate S was significantly related to the factors novelty/familiarity (N) or category (C). Both factors were binary (0/1) to make the effect size comparable. We quantified the effect size of each regressor using the effect size metric ω^2, which is better suited for our purposes than more traditional variance explained or p-value metrics[23]. This is because ω^2 is not biased for small numbers of trials and tends towards zero if a factor has no explanatory power[58]. To estimate ω^2 for the factor category regardless of tuning of a neuron, we fit 5 models to each neuron, each contrasting one category with the remaining four. We then averaged the resulting ω^2. Spike counts S(t) were computed for a 500ms window that was moved in steps of 50ms. Here, $\omega_i^2 = \frac{[SS_i - df_i * MSE]}{[SS_{tot} + MSE]}$ where SS_i is the sum of squares of factor I, SS_{tot} the total sum of squares of the model and MSE the mean square error of the model. Models were fit and effect sizes calculated using the effect size toolbox functions mes1way and mes2way[23]. We averaged $\omega^2(t)$ across all neurons (Fig. 6). The null distribution was estimated by randomly scrambling the labels and fitting the same model. This was repeated 1000 times to estimate the 99% confidence interval of the null distribution. Estimates of latency were based on the first time the actual value was located outside of the 99% confidence interval. To estimate potential interactions, we also fit the model $S(t) = \alpha_0(t) + \alpha_1(t)N + \alpha_2(t)C + \alpha_3(t)N * C$ and estimated $\omega^2(t)$ for each main factor and the interaction (Fig. S3).

Population decoding

We pooled all recorded neurons into a pseudo-population. Firing rates were z-scored individually for each. We used a maximal correlation coefficient classifier (MCC) as

implemented in the ndt toolbox[59]. The MCC estimates a mean template $\overline{x_i}$ for each class i and assigns the class $i^* = \underset{i}{\operatorname{argmax}}\, corr(x^*, \overline{x_i})$ for test trial *x**. We used 10-fold cross-validation, i.e. for each iteration 10 trials from each class where chosen randomly from each neuron. 1 trial from each class was used for testing and the remaining 9 for training. All possible train/test splits were tested and this process was repeated 50 times with different subsets of trials, resulting in a total of 500 runs. Spikes were counted in bins of 500ms size and advanced by a stepsize of 50ms. For each point of time, a different classifier was trained. We converted the resulting confusion matrix into mutual information MI $I(S; R)$[60] to estimate the information that the overall population response R provides, in a single trial, about the stimulus S. We estimated the null distribution by repeating above procedure 200 times after randomly scrambling the labels. To estimate the variability of MI across different neurons we repeated above procedure after selecting a group of 200 (all units) or 20 (MS neurons) with replacement from the overall group. We repeated this procedure 50 times, each time estimating the peak MI (Fig. 7e). To estimate whether the same subset of neurons is informative about high-and low confidence trials we trained decoders using all or only high confidence trials, and subsequently tested the decoders with only high or low trials. For decoding of error trials, which are relatively rare, we used larger bin sizes and smaller number of trials (Fig. 7f–h). Thus, we used 6-fold cross-validation (5 training trials, 1 testing), a binsize of 1.5s with stepsize of 50ms and estimated the variability across neurons by randomly sub-selecting with replacement a group of 30 MS neurons. We again used the peak MI of each run and repeated this procedure 500 times (Fig. 7f). For estimating overall readout ability (Fig. 7g–h), we used a single 1.5s long time window starting 200ms after stimulus onset.

Waveform analysis

The trough-to-peak time d[37] is the time between the trough and the point of time of maximal amplitude after the trough of the mean waveform. The mean waveform is the average of all spikes assigned to the cluster. For visualization, all waveforms were normalized to their maximal amplitude and were inverted if their maximum was positive. A spike waveform was considered short if d<0.6ms.

Spike-train variability

Variability was quantified for each neuron using two metrics: the modified coefficient of variation (CV2) and the burst index (BI). The BI is equal to the proportion of ISIs less than 10ms long and the CV2 was used as defined in[61]. The CV2 is insensitive to underlying rate changes and is thus the appropriate metric to use in place of the normal CV[62].

Decision making model

The input to the model is the spiking activity $S_k^{i,j}(t)$ of a NS and FS neuron *i* and *j* in trial *k*. The difference $D(t)_k = S_k^j(t) - S_k^i(t)$ is then integrated over time. Spikes are counted in bins of 250ms, advanced with a step-size of 100ms. Firing rates of neurons were z-scored using the mean and standard deviation of the baseline (1s before stimulus onset). The model has two state variables $EV_{fam}(t)$ and $EV_{nov}(t)$, which accumulate as following:

$\mathrm{EV}_{\mathrm{fam}}(t)=\int_0^t f(D(t))$ and $\mathrm{EV}_{\mathrm{nov}}(t)=\int_0^t f(-D(t))$ where $f(x)$ = max(0, x) is a rectification non-linearity (Fig. 8a). The decision is "familiar" if $EV_{fam}(t) > EV_{nov}(t)$ and "novel" otherwise. Except for Fig. 8i–j, the decision was made 2.5s after stimulus onset. The balance of evidence is $\Delta E(t) = EV_{fam}(t) - EV_{nov}(t)$. We evaluated the model for all possible pairs (n=951) of NS/FS neurons that had at least 3 behaviorally correct trials in each category (TP high/low, FN high/low). For each, we evaluated every possible pair of trials within the same behavioral category. As a control, we randomly scrambled the high and low-confidence labels for each neuron while keeping the trial identity (new/old) labels intact. This abolished the difference in balance of evidence as expected (Fig. 8g). To correlate ΔE and EV with performance, we computed for every cell pair separately the Spearman correlation coefficient between confidence (high or low) with |ΔE| t=2.5s. We evaluated this trial-by-trial correlation for all trials remembered correctly by the subject and the model (excluding errors made by the model) as well as all trials where the subject was incorrect ("errors"). To make this comparison unbiased, we used the same number of high and low confidence trials by subsampling the larger group randomly. To evaluate the decision latency of the model, we terminated the decision when $|\Delta E(t)| > \Delta E_{Th}$. The decision time was equal to the first point of time at which this condition was satisified. ΔE_{Th} was set to 50% of the |ΔE| value reached at 2.5s for every cell pair.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank J. Kaminski, R. Adolphs, C. Anastassiou, U. Maoz, J. Wertheimer and W. Einhaeuser for discussion, Z. Fu for spike sorting, C. Heller for performing some of the surgeries, the staff of the Epilepsy Monitoring Units at Huntington Memorial Hospital and Cedars-Sinai Medical Center for invaluable assistance, particularly J. Schmidt. We thank K. Birch and H. Babu for assistance with patient care and surgery, and L. Philpott and M.-T. Le for neuropsychological testing.

Funding

This work was supported by the Cedars-Sinai Medical Center Department of Neurosurgery (to U.R.), National Institute of Mental Health Conte Center at Caltech (P50 MH094258), and the Gustavus and Louise Pfeiffer Research Foundation (to U.R.).

References

1. Kahana, MJ. Foundations of human memory. New York: Oxford University Press; 2012.
2. Petrusic WM, Baranski JV. Judging confidence influences decision processing in comparative judgments. Psychon Bull Rev. 2003; 10:177–183. [PubMed: 12747505]
3. Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. Science. 2009; 324:759–764. [PubMed: 19423820]
4. Smith JD, Shields WE, Washburn DA. The comparative psychology of uncertainty monitoring and metacognition. Behav Brain Sci. 2003; 26:317–339. discussion 340–373. [PubMed: 14968691]
5. Kepecs A, Mainen ZF. A computational framework for the study of confidence in humans and animals. Philosophical transactions of the Royal Society of London. 2012; 367:1322–1337. [PubMed: 22492750]
6. Pouget A, Dayan P, Zemel RS. Inference and computation with population codes. Annual review of neuroscience. 2003; 26:381–410.

7. Kepecs A, Uchida N, Zariwala HA, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. Nature. 2008; 455:227–231. [PubMed: 18690210]
8. Squire LR, Stark CE, Clark RE. The medial temporal lobe. Annual review of neuroscience. 2004; 27:279–306.
9. Kreiman G, Koch C, Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. Nature neuroscience. 2000; 3:946–953. [PubMed: 10966627]
10. Viskontas IV, Quiroga RQ, Fried I. Human medial temporal lobe neurons respond preferentially to personally relevant images. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:21329–21334. [PubMed: 19955441]
11. Logothetis NK, Sheinberg DL. Visual object recognition. Annual review of neuroscience. 1996; 19:577–621.
12. Rolls ET. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron. 2000; 27:205–218. [PubMed: 10985342]
13. Rutishauser U, Mamelak AN, Schuman EM. Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. Neuron. 2006; 49:805–813. [PubMed: 16543129]
14. Rutishauser U, Schuman EM, Mamelak AN. Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:329–334. [PubMed: 18162554]
15. Xiang JZ, Brown MW. Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology. 1998; 37:657–676. [PubMed: 9705004]
16. Wilson FA, Rolls ET. The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. Exp Brain Res. 1993; 93:367–382. [PubMed: 8519331]
17. Rutishauser U, Ross IB, Mamelak AN, Schuman EM. Human memory strength is predicted by theta-frequency phase-locking of single neurons. Nature. 2010; 464:903–907. [PubMed: 20336071]
18. Green, D.; Swets, J. Signal Detection Theory and Psychophysics. Wiley; 1966.
19. Manns JR, Hopkins RO, Reed JM, Kitchener EG, Squire LR. Recognition memory and the human hippocampus. Neuron. 2003; 37:171–180. [PubMed: 12526782]
20. Rutishauser U, Schuman EM, Mamelak AN. Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. Journal of neuroscience methods. 2006; 154:204–224. [PubMed: 16488479]
21. Viskontas IV, Knowlton BJ, Steinmetz PN, Fried I. Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. Journal of cognitive neuroscience. 2006; 18:1654–1662. [PubMed: 17014370]
22. Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA. A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci. 1996; 13:87–100. [PubMed: 8730992]
23. Hentschke H, Stuttgen MC. Computation of measures of effect size for neuroscience data sets. The European journal of neuroscience. 2011; 34:1887–1894. [PubMed: 22082031]
24. Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychological review. 2006; 113:700–765. [PubMed: 17014301]
25. Vickers, D. Decision processes in visual perception. New York; London: Academic Press; 1979.
26. Fried I, MacDonald KA, Wilson CL. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. Neuron. 1997; 18:753–765. [PubMed: 9182800]
27. Wixted JT, Squire LR, Jang Y, Papesh MH, Goldinger SD, Kuhn JR, Smith KA, Treiman DM, Steinmetz PN. Sparse and distributed coding of episodic memory in neurons of the human hippocampus. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:9621–9626. [PubMed: 24979802]
28. Marr D. Simple memory: a theory for archicortex. Philosophical transactions of the Royal Society of London. 1971; 262:23–81. [PubMed: 4399412]

29. Macmillan, NA.; Creelman, CD. Detection theory. Mahwah, NJ: Lawrence Associates; 2005.
30. Zhang Y, Meyers EM, Bichot NP, Serre T, Poggio TA, Desimone R. Object decoding with attention in inferior temporal cortex. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:8850–8855. [PubMed: 21555594]
31. Wixted JT. Dual-process theory and signal-detection theory of recognition memory. Psychological review. 2007; 114:152–176. [PubMed: 17227185]
32. Clark SE, Gronlund SD. Global matching models of recognition memory: How the models match the data. Psychon Bull Rev. 1996; 3:37–60. [PubMed: 24214802]
33. Gold JI, Shadlen MN. The neural basis of decision making. Annual review of neuroscience. 2007; 30:535–574.
34. Ratcliff R. A theory of memory retrieval. Psychological review. 1978; 85:59.
35. Hanks TD, Kopec CD, Brunton BW, Duan CA, Erlich JC, Brody CD. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature. 2015; 520:220–223. [PubMed: 25600270]
36. Viskontas IV, Ekstrom AD, Wilson CL, Fried I. Characterizing interneuron and pyramidal cells in the human medial temporal lobe in vivo using extracellular recordings. Hippocampus. 2007; 17:49–57. [PubMed: 17143903]
37. Mitchell JF, Sundberg KA, Reynolds JH. Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron. 2007; 55:131–141. [PubMed: 17610822]
38. Peyrache A, Dehghani N, Eskandar EN, Madsen JR, Anderson WS, Donoghue JA, Hochberg LR, Halgren E, Cash SS, Destexhe A. Spatiotemporal dynamics of neocortical excitation and inhibition during human sleep. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:1731–1736. [PubMed: 22307639]
39. Vigneswaran G, Kraskov A, Lemon RN. Large identified pyramidal cells in macaque motor and premotor cortex exhibit "thin spikes": implications for cell type classification. J Neurosci. 2011; 31:14235–14242. [PubMed: 21976508]
40. Robbins AA, Fox SE, Holmes GL, Scott RC, Barry JM. Short duration waveforms recorded extracellularly from freely moving rats are representative of axonal activity. Frontiers in neural circuits. 2013; 7:181. [PubMed: 24348338]
41. Stuart G, Schiller J, Sakmann B. Action potential initiation and propagation in rat neocortical pyramidal neurons. J Physiol. 1997; 505(Pt 3):617–632. [PubMed: 9457640]
42. Hamann, S. The human amygdala and Memory. In: Whalen, PJ.; Phelps, EA., editors. The Human Amydala. New York: The Guilford Press; 2009. p. 177-203.
43. Weierich MR, Wright CI, Negreira A, Dickerson BC, Barrett LF. Novelty as a dimension in the affective brain. Neuroimage. 2010; 49:2871–2878. [PubMed: 19796697]
44. Metcalfe, J. Metamemory. In: Roediger, HL., editor. Learning and Memory: A Comprehensive Reference. Oxford: Elsevier; 2008. p. 349-362.
45. Metcalfe, J. Evolution of Metacognition. In: Dunlovsky, J.; Bjork, R., editors. Handbook of Metamemory and Memory. New York: Psychology Press; 2008. p. 29-46.
46. Hampton RR. Rhesus monkeys know when they remember. Proc Natl Acad Sci USA. 2001; 98:5359–5362. [PubMed: 11274360]
47. Perry CJ, Barron AB. Honey bees selectively avoid difficult choices. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:19155–19159. [PubMed: 24191024]
48. Foote AL, Crystal JD. Metacognition in the rat. Curr Biol. 2007; 17:551–555. [PubMed: 17346969]
49. Middlebrooks PG, Sommer MA. Metacognition in monkeys during an oculomotor task. Journal of experimental psychology. 2011; 37:325–337. [PubMed: 21171807]
50. Fortin NJ, Wright SP, Eichenbaum H. Recollection-like memory retrieval in rats is dependent on the hippocampus. Nature. 2004; 431:188–191. [PubMed: 15356631]
51. Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]
52. Nelson TO. A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological bulletin. 1984; 95:109–133. [PubMed: 6544431]

53. Ratcliff R, Gronlund SD, Sheu CF. Testing Global Memory Models Using Roc Curves. Psychological review. 1992; 99:518–535. [PubMed: 1502275]

54. Pouzat C, Mazor O, Laurent G. Using noise signature to optimize spike-sorting and to assess neuronal classification quality. Journal of neuroscience methods. 2002; 122:43–57. [PubMed: 12535763]

55. Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsaki G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. Journal of neurophysiology. 2000; 84:401–414. [PubMed: 10899214]

56. Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish AD. Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience. 2005; 131:1–11. [PubMed: 15680687]

57. Diba K, Buzsaki G. Hippocampal Network Dynamics Constrain the Time Lag between Pyramidal Cells across Modified Environments. Journal of Neuroscience. 2008; 28:13448–13456. [PubMed: 19074018]

58. Olejnik S, Algina J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological methods. 2003; 8:434–447. [PubMed: 14664681]

59. Meyers EM. The neural decoding toolbox. Frontiers in neuroinformatics. 2013; 7:8. [PubMed: 23734125]

60. Quian Quiroga R, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. Nature reviews. 2009; 10:173–185.

61. Rutishauser U, Tudusciuc O, Wang S, Mamelak Adam N, Ross Ian B, Adolphs R. Single-Neuron Correlates of Atypical Face Processing in Autism. Neuron. 2013; 80:887–899. [PubMed: 24267649]

62. Holt GR, Softky WR, Koch C, Douglas RJ. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. Journal of neurophysiology. 1996; 75:1806–1814. [PubMed: 8734581]

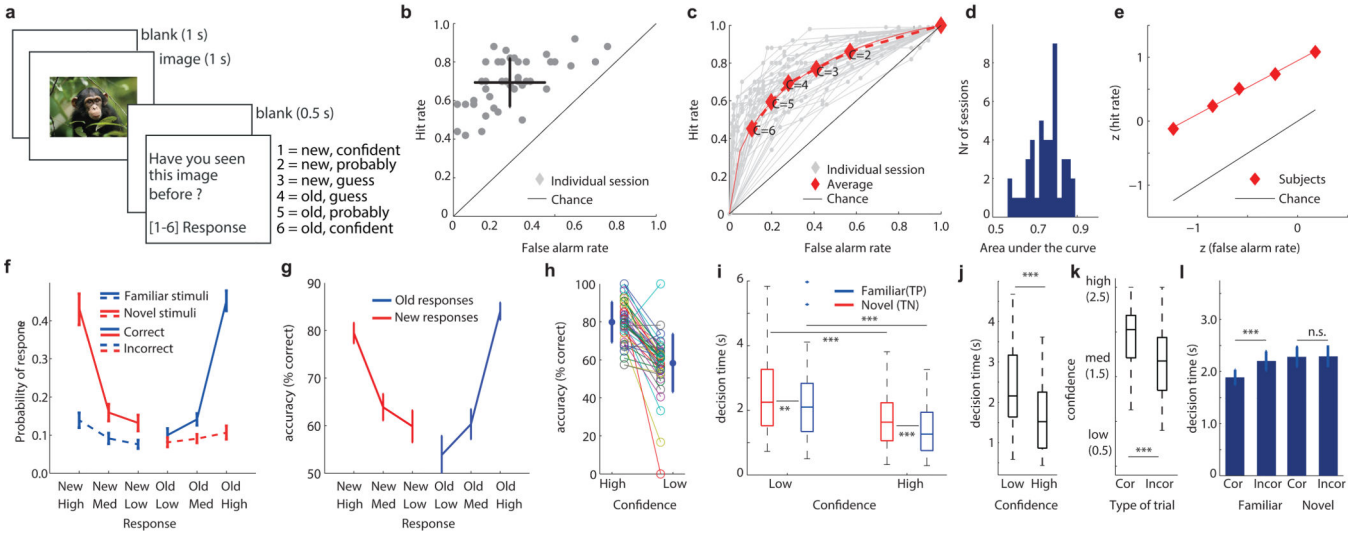

Fig 1. The recognition memory task and behavioral results

(a) Task. (b) Performance as a function of proportion of trials correctly and incorrectly identified. Each point is one session (n=44), black is the mean performance ±s.d.. (c) Behavioral ROC curve for individual sessions (gray) and average (red). Each data point is a different confidence. (d) AUC values of all sessions. (e) z-transform of the average ROC shown in (c). The slope of the red line (least-square fit) is the metric used in the text. (f) Probability of responses, conditional on the ground truth (red or blue). At all levels of confidence, subjects were more likely to be correct than incorrect (straight and dashed lines, respectively). (g) Choice accuracy as a function of confidence, shown separately for new and old reponses. (h) Accuracy was significantly different between high and low confidence trials ($p<1e-10$, paired ttest). Each color is a different session, with average ±s.d. on the left/ right. (i) Decision time was significantly larger (slower) for low compared to high confidence trials (correct trials only; paired Wilcoxon signed rank test, $p<1e-5$ for both novel and familiar stimuli) and significantly larger for novel compared to familiar stimuli for both low and high confidences ($p=0.01$ and $p<1e-4$, respectively). (j) Decision time was significantly slower for low compared to high confidence incorrect trials (paired Wilcoxon singed rank test, $p<1e-6$). (k) Errors were made with less confidence than correct trials ($p<1e-8$, paired Wilcoxon signed rank test). (l) Correct familiar decisions were made faster than incorrect decisions (paired comparison matched for confidence, see methods). (i-k) Boxplots represent quantiles (25%, 75%), line is median and wiskers show range. Outliers are marked. *<=0.05, **<=0.01, ***<=0.001. P-values are uncorrected for multiple comparisons.

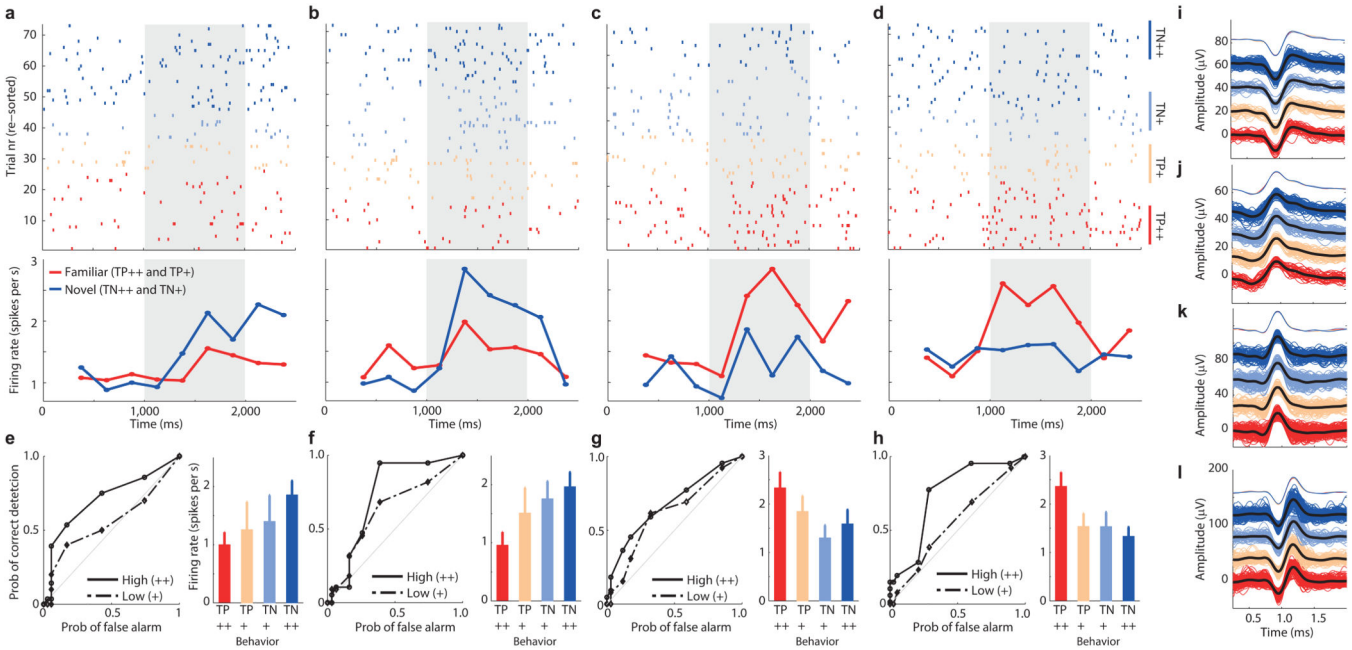

Fig 2. Memory selective (MS) neurons

(a–d) Raster (top) and PSTH (bottom) of four example neurons. (a–b) and (c–d) are NS and FS neurons, respectively. Stimulus onset is at 1000ms (gray). Trials are re-sorted by behavior for display purposes: familiar high confidence (TP++), familiar low confidence (TP+), novel low confidence (TN+), novel high confidence (TN++). Error trials are not shown. In the PSTH, trials are grouped according to TP/TN. (e–h) Single-neuron ROC curves (left) and mean rate (right) for same neurons shown in (a–d). Bar plots show the mean rate in a 1.5s window starting 200ms after stimulus onset. Errorbars are ±s.e. across trials. (i–l) Waveforms of spikes associated with the four different trial types for each neuron, in same order as in (a–d). Top shows mean waveforms superimposed, bottom all individual waveforms associated with the spikes shown in (a–d). Color code is identical to (a–d).

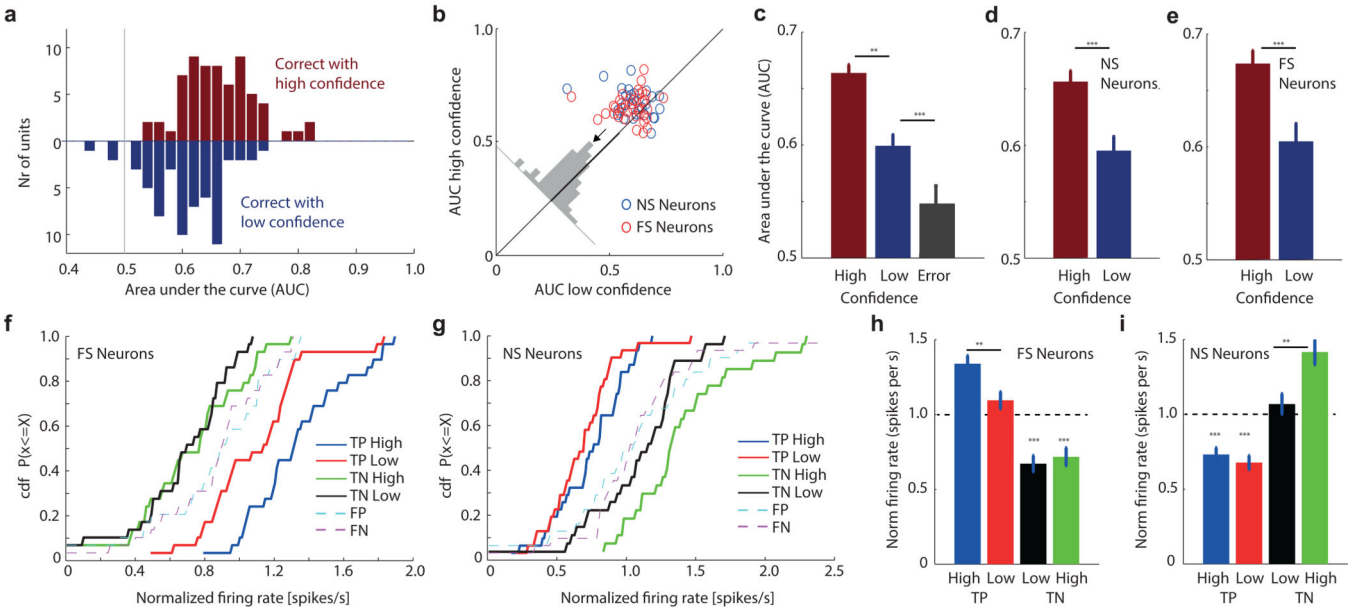

Fig 3. The response of MS neurons is modulated by subjective confidence

(a–e) Single-neuron ROC analysis. (a) AUC of MS neurons, for high (red) and low (blue) confidence, respectively (n=65 units; the two distributions were significantly different, p=0.001). (b) Pairwise comparison of AUC values. For 49/65 units, the AUC was high>low (p<1e-4, sign-test). The average difference was above the diagonal (inset). (c) Average AUC for high, low confidence correct and error trials (FN). FN vs low p=0.0056, high vs low p<1e-5 (pairwise t-test). (d,e) AUC for high confidence trials was significantly larger for both NS (n=29) and FS (n=36) neurons (p=0.0001 and p=0.0003, respectively). (f–i) Comparison of firing rate using baseline normalized responses and grouped by behavior. (f) Activity of FS neurons differentiated high from low confidence familiar trials (n=29, TP high vs. TP low, p=0.0094, ks-test) but not novel trials (n=30, TN high vs. TN low, p=0.74, ks-test). (g) Activity of NS neurons differentiated high from low confidence novel trials (TN high vs. TN low, p=0.03, ks-test) but not high from low familiar trials (TP high vs. TP low, p=0.22, ks-test). (h,i) Mean normalized response across neurons. (h) FS neurons had significantly higher firing rate for TP high compared to TP low trials (paired ttest, p=0.0014). (i) NS neurons had significantly higher firing rate for TN high compared to TN low trials (paired ttest, p=0.0002). *** indicates significant difference from baseline (p<1e-4). Abbreviations: true positive (TP) and negatives (TN) are correctly remembered familiar and novel stimuli. False positives (FP) and false negatives (FN) are wrongly identified novel and familiar stimuli. Errors are ±s.e. across neurons. **<=0.01, ***<=0.001.

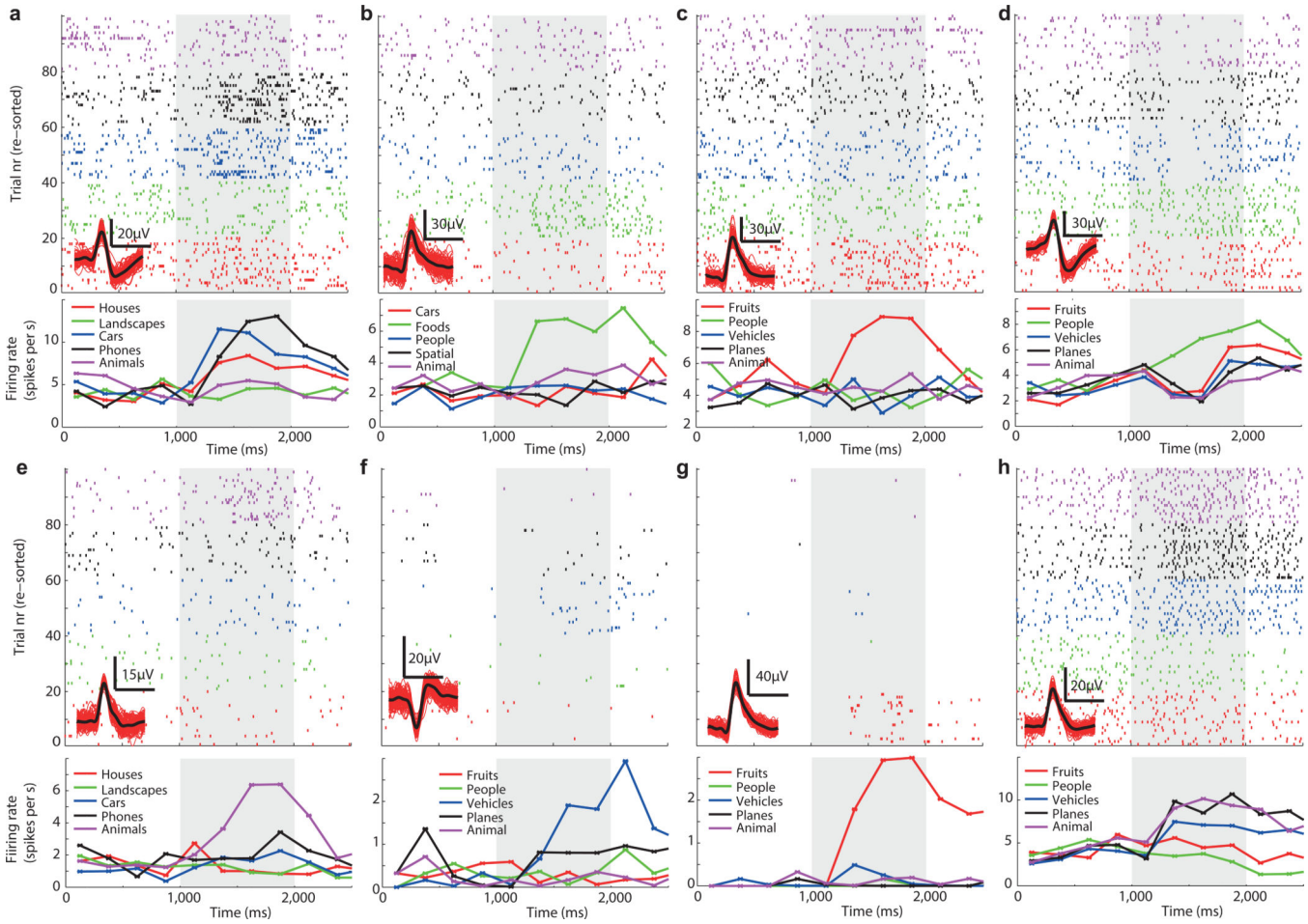

Fig 4. Visually selective (VS) neurons

(a–h). For each, the raster (top) and PSTH (bottom) is shown. Trials are re-sorted for illustration purposes. Visual identity (category) is indicated by color, the legends shows the corresponding label (variable). The inset (bottom left of raster) shows waveforms associated with the neuron shown (red are 100 randomly chosen individual waveforms, black mean waveform, horizontal scalebar is 1ms, vertical as indicated). (a–b,d,f) and (c,e,g–h) are from the hippocampus and amygdala, respectively. All units are from different sessions. Some units respond with a firing increase only to one category (b–c,e–g) whereas others show a mixed response (a,d,h). Stimulus onset was at 1000ms (gray bar). Significance of selection criteria (1×5 ANOVA) was, for A–H, 7e-5, 1e-6, 0.004, 0.003, 5e-9, 0.0004, 3e-12, and 4e-9. PSTH binsize is 250ms.

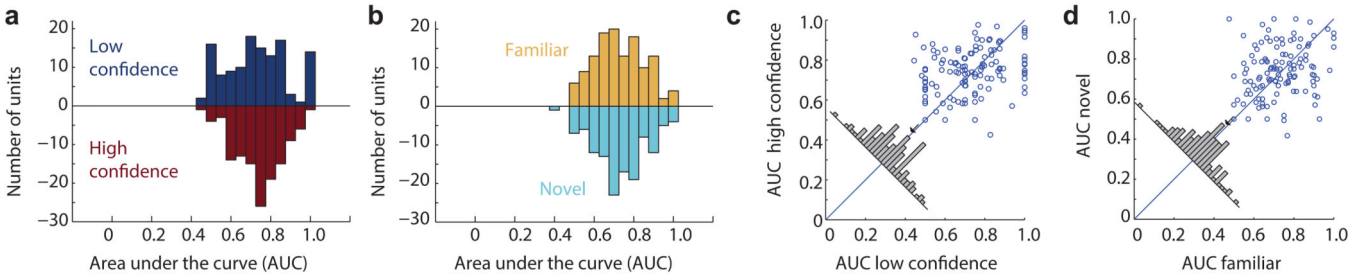

Fig 5. The ability of VS neurons to differentiate visual stimuli is not influenced by confidence judgment or novelty of the stimulus

(a) AUC of VS neurons for low-and high confidence trials (p=0.31, bootstrap test). (b) AUC of VS neurons for novel and familiar trials (p=0.54, bootstrap test). (c) Pairwise comparison of AUC values as a function of confidence (p=0.53, pairwise sign-test). (d) Pairwise comparison of AUC values as a function of familiarity (p=0.41, pairwise sign-test). In (c–d), every data point is one VS neuron (n=128 in total). All pairwise comparisons showed no significant difference. Only correct trials are considered throughout.

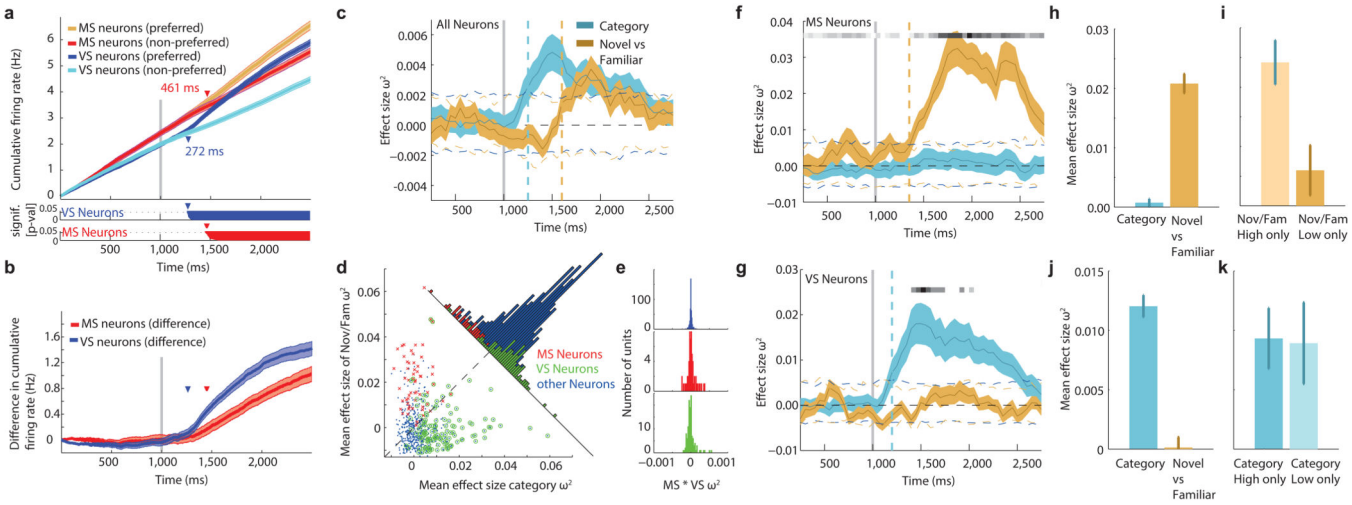

Fig 6. MS and VS neurons signal at different times and only MS neurons are sensitive to confidence

(a–b) Cumulative firing rate for MS and VS neurons. Pairwise comparison (a, bottom; cluster-corrected p-values) between the preferred and non-preferred stimulus reveals differences in timecourse. (b) Pairwise difference for both populations. (c–k) Effect size estimation for populations of neurons based on a regression model. ω^2 is used to estimate effect size. (c) Time course of effect size, averaged across all neurons (N=664) and computed separately for the variable category (blue) and novel/familiar (yellow). Dashed horizontal lines indicate the 99% confidence intervals of the null distribution. Dashed vertical lines indicate first time point significantly above the 99% confidence interval. Stimulus onset is at 1000ms (gray line). (d) Average effect size (1.5s window starting 200ms after stim onset) of category and novel/familiar regressor for each neuron. (e) Product of ω^2 for regressors novel/familiar and category for MS, VS and other neurons. There was no significant correlation. (f) Same metrics as in (c), but for MS neurons only. MS neurons did not distinguish categories. (g) same as in (c), but for VS neurons only. VS neurons did not distinguish novel from familiar stimuli. Black horizontal line in (f–g) indicates proportion of significant units (from white to black) at every point of time, based on the 99% confidence interval. (h) MS neurons have significantly larger effect size for regressor Novel/Familiar compared to category (p=0). (i) Effect size of MS neurons is significantly modulated by confidence (p=0.0049). (j) Average effect size for VS neurons was significantly larger for category information (p=0.0049), and (k) was not sensitive to confidence (p=0.81, right). All p-values are paired t-tests. Binsize is 500ms, stepsize 50ms, error bars and shaded regions represent ±s.e. across neurons.

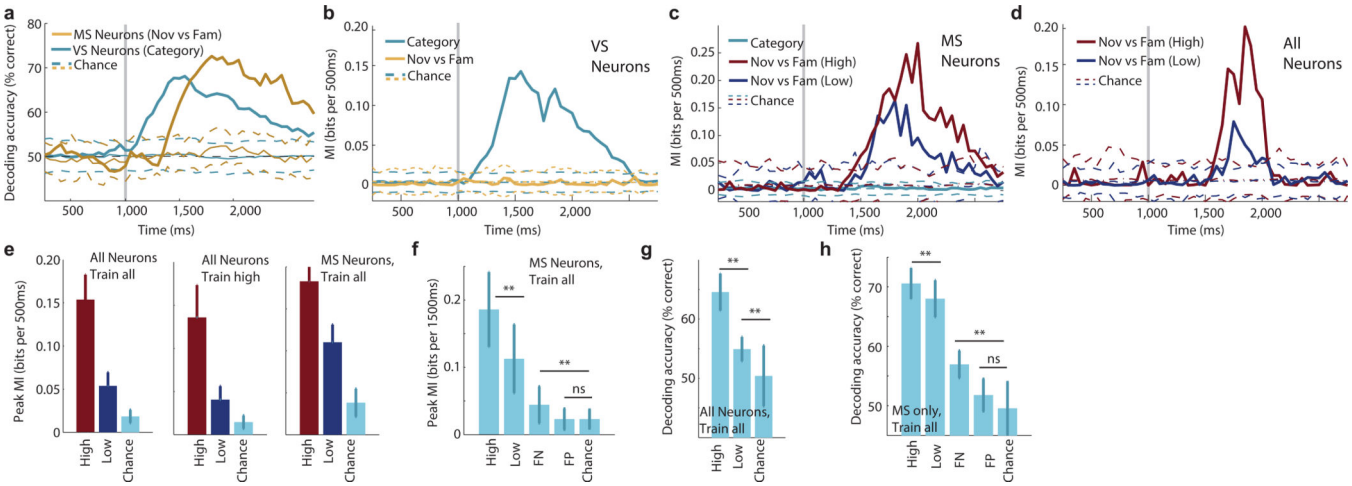

Fig 7. Quantification of population-level information difference due to confidence

Decoding performance was quantified using decoding accuracy and mutual information (MI) between spiking activity and stimulus category and familiarity. (a) Accuracy as a function of time estimated separately for VS and MS neurons while decoding visual category and familiarity, respectively. (b–c) VS neurons (n=128) and MS (n=59) neurons only signal category and novel/familiar information, respectively. (c) Spiking of MS neurons contains more information about familiarity for high-confidence trials. (d) Spiking activity of all recorded neurons (n=606) together contains more information for high-confidence trials. (e) Statistical comparison of MI for high and low confidence trials. A subset of n=200 (all) and n=20 (MS) units was chosen at random from the entire population (bootstrap, 50 runs) and the peak MI was estimated for each run. More information was available for all neurons (left) as well as for MS neurons only (right), and regardless of whether the decoder was trained with all (left) or only high confidence (middle) trials (high vs. low and low vs. chance is $p<0.001$ for all). (f) Decoding of error trials, using a subset of n=30 MS neurons chosen at random from the population. Decoder was trained on all correct trials and separately evaluated on high and low confidence as well as forgotten (FN) and false positive (FP) trials. Performance for FN was above chance ($p=0.003$) but FP was not ($p=0.98$). FN performance was significantly lower than low confidence ($p<1e-5$). (g–h) Quantification of overall readout ability (1.5s window), regardless of time, for all neurons (g) and MS neurons only (h). (e–h) Errorbars are ±s.d. across bootstrap runs. Dashed lines in (a–d) show the mean ±99% confidence interval of the null distribution.

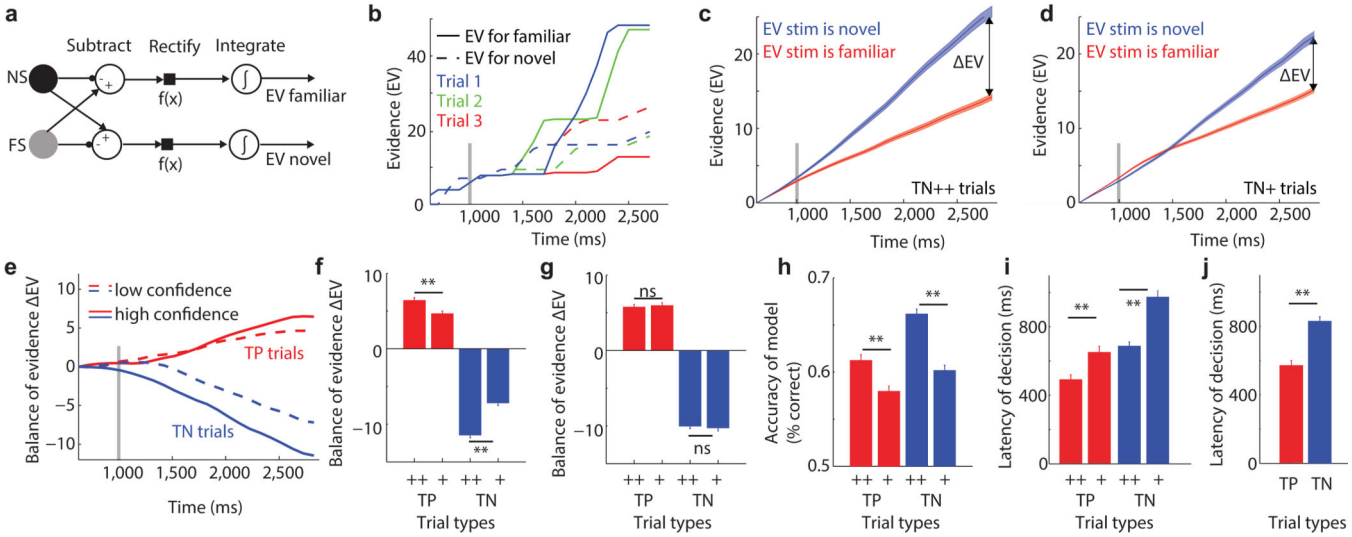

Fig 8. Computational model to decide the familiarity and confidence of a stimulus

(a) Circuit diagram of a race model that integrates the difference of the output of an NS and FS neuron. (b) Model output for three familiar (TP) trials for an example pair of neurons. Decision is made correct for trial 1 and 2, incorrectly for trial 3. (c–d) Model output for all (FS,NS) neuron pairs (n=951) for novel (TN) trials for high (c) and low confidence (d), respectively. Note how the balance of evidence ΔE is larger for high confidence trials. Errorbars are 99% confidence intervals across pairs of neurons. Marked time points are the centers of each bin (binsize 250ms). (e) ΔE as a function of time for all four trial types. Here, $\Delta E = EV_{fam} - EV_{nov}$, making ΔE negative for TN trials. (f) Average ΔE for the last time-point in (e), for all neuron pairs (n=951, errors are ±s.e.). ΔE was significantly larger for high relative to low confidence trials (pairwise t-test, p<1e-6). (g) Control, random reassignment of confidences abolishes the difference while keeping new/old performance intact (p=0.56 and 0.45, respectively). (h) Single-trial model performance for determining the familiarity of a stimulus. Performance was higher for high compared to low confidence trials (pairwise t-test, p<1e-5). (i) Latency to reach a decision, as a function of confidence. High-confidence trials had significantly shorter latency (p<1e-14 and p=0.00022 for TN and TP, respectively; paired t-test across all cell pairs). (j) Familiar (TP) trials were faster than Novel (TN) trials (p<1e-11, paired t-test). All errorbars represent ±s.e.m. across all neuron pairs.