Supplemental Figures

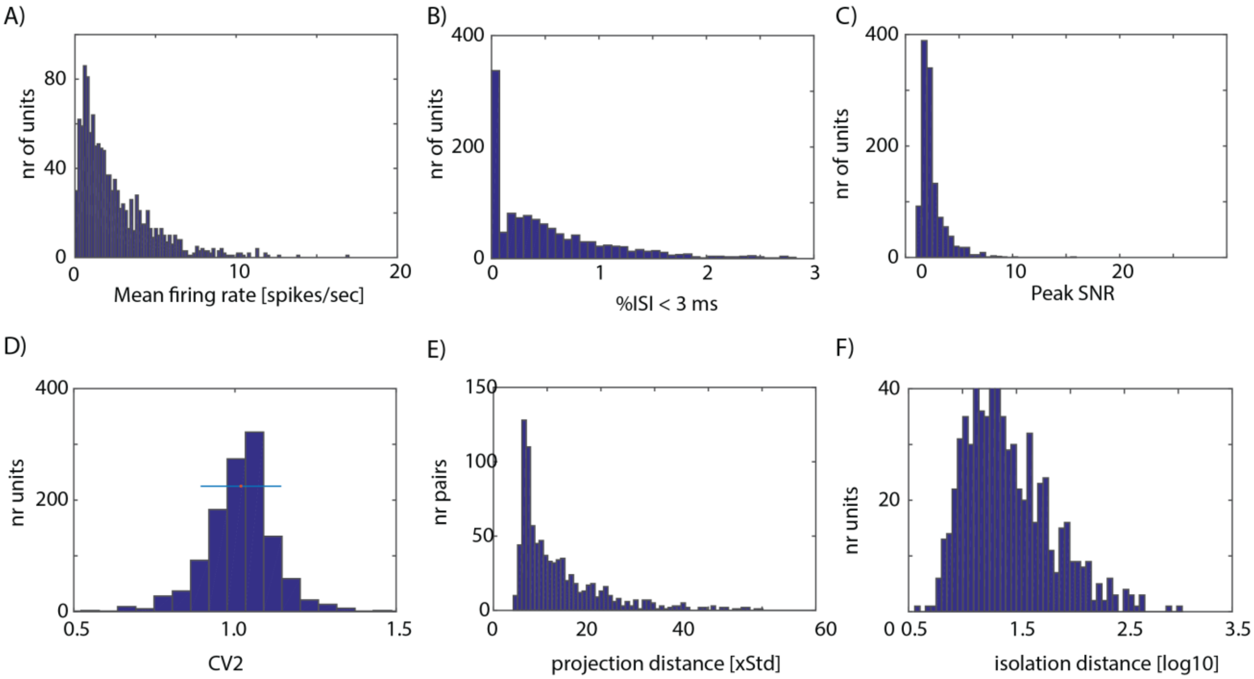

Figure S1: Spike sorting metrics, Related to STAR Methods. A) Mean firing rate of all recorded neurons. B) Histogram of refractory violations (%ISIs<3ms). On average 0.54±0.66% of ISIs were <3ms, indicating well isolated neurons. C) Signal-to-noise ratio (SNR) of the mean waveform of each neuron. Mean SNR was 2.5±1.2. D) Modified coefficient of variation for all recorded neurons. Mean was 1.02±0.12, indicating that neurons were well approximated by a Poisson process. E) Projection test distance [60] between all possible pairs of neurons recorded on the same wire. The average separation was 12±8 (in units of standard deviation), indicating that clusters associated with putative single neurons were well separated from each other. F) Isolation distance [61] for all neurons. Average was 45±76, indicating that clusters associated with putative single neurons were well separated from all other (including the noise) clusters.

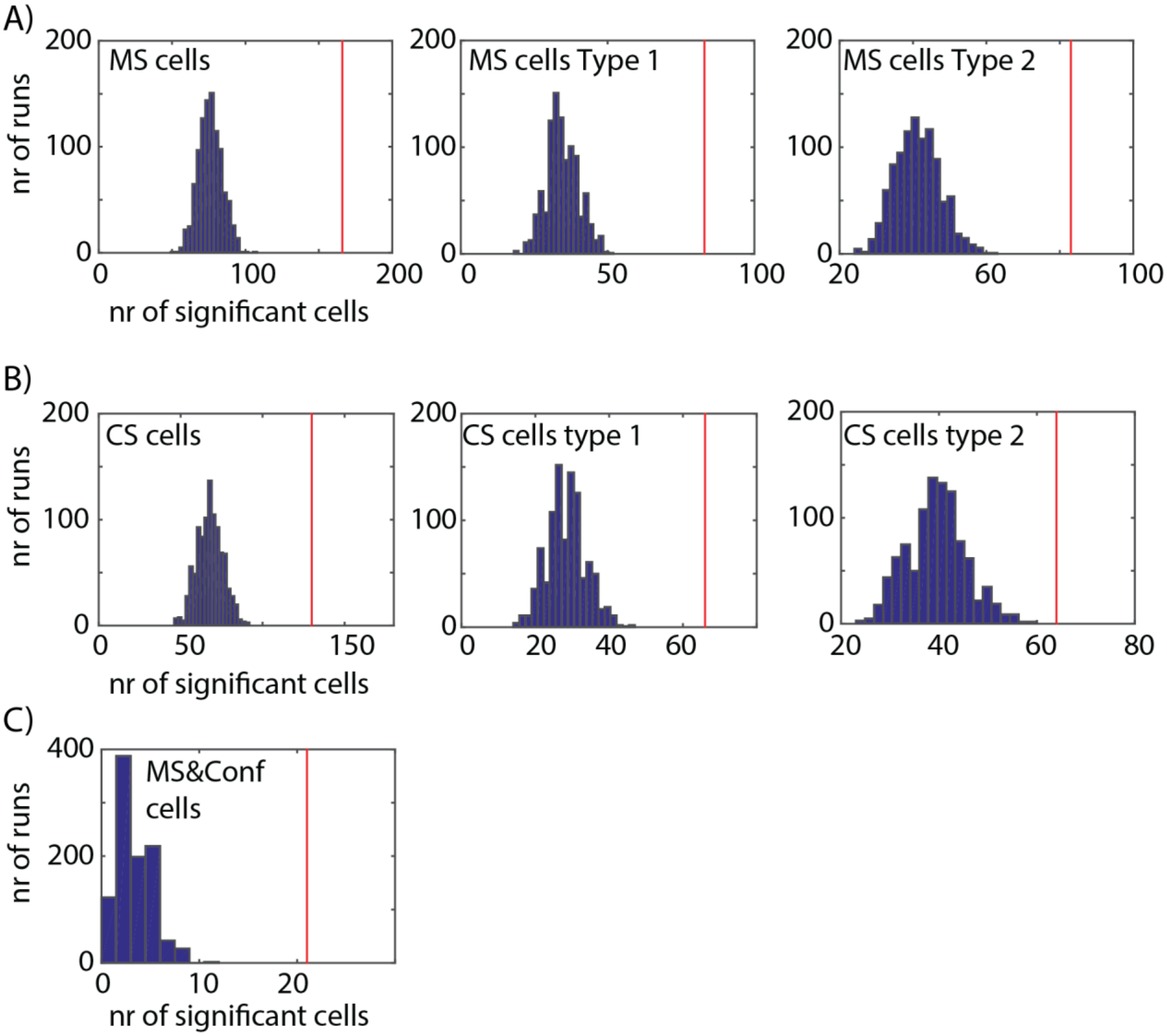

Figure S2: Null distributions of number of cells selected, Related to Figure 3. Shown are, for each selection criteria, the number of selected cells (red line) as well as the null distribution (blue) that is expected by chance. The null distribution was estimated using a bootstrap procedure during which all analysis was identical except the trial identity labels were randomly permuted. (A) MS cells (all), followed by Type 1 and Type 2 separately. (B) CS cells (all), followed by Type 1 and 2 separately. (C) Number of cells that qualified both as MS and CS cells. For none of the conditions did the null distribution exceed the observed value, resulting in p=1/B (where B=1000 bootstrap runs) for all selections.

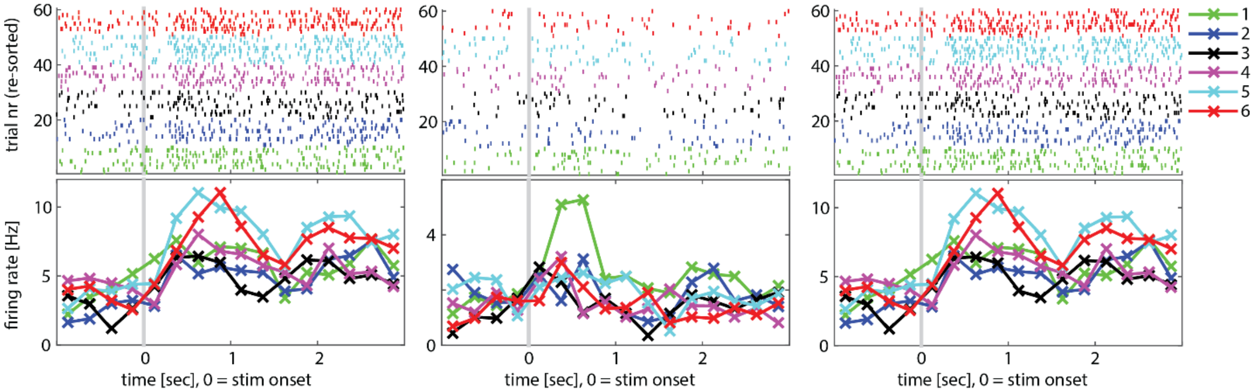

Figure S3: Example number-selective cells during the text control task, Related to Figure 5. Shown are three number-selective units (p<0.05, 1x5 ANOVA). t=0 is stimulus onset (display of the number). Trials were shown in random order but are shown re-sorted for display purposes.

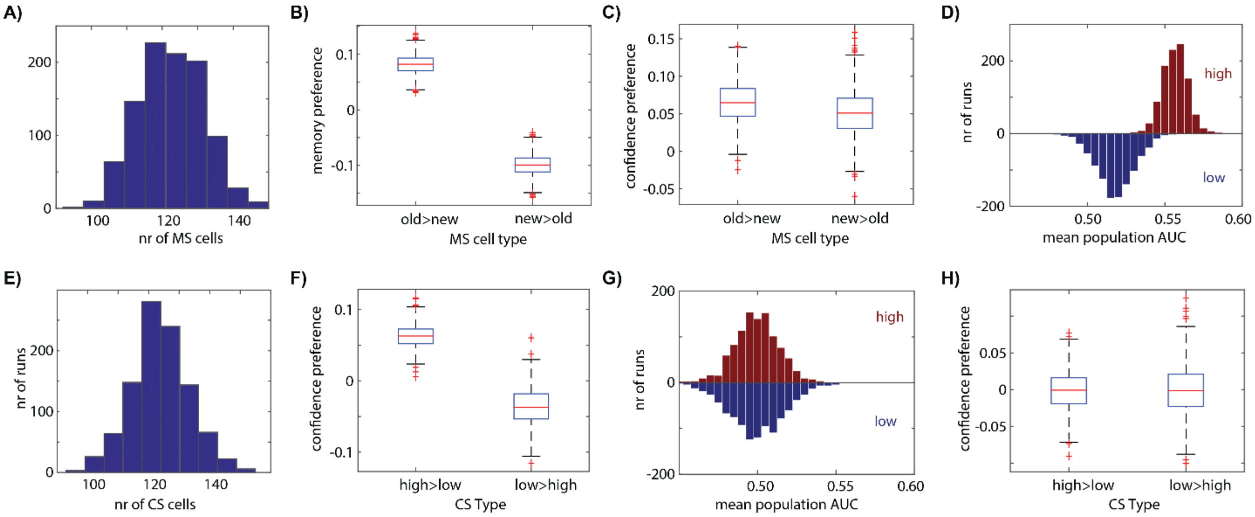

Figure S4: Evaluation of effect of confidence on MS and CS cells using independent selection (training) and evaluation (testing) sets, Related to Figure 3. All results reported are across 1000 bootstrap runs, for each of which a different random subset of 50% of the trials was chosen for selection. All metrics shown in this figure are calculated based on the trials not used for selection. (A) Number of MS cells selected. (B) Average memory preference of the selected MS cells, confirming robustness of tuning of MS cells. (C) Average confidence preference index for MS cells. (D) Average comparison of AUC for high vs. low confidence trials. (E) Number of CS cells selected. (F) Average confidence preference index, confirming robustness of tuning of CS cells. (G-H) Chance control after random scrambling of labels. (G) MS cells show no difference due to confidence (paired t-test p=0.84, AUC 0.50±0.01 vs. 0.50±0.02). (H) CS cells show no difference due to confidence (t-test vs. 0, p>0.05 for both).

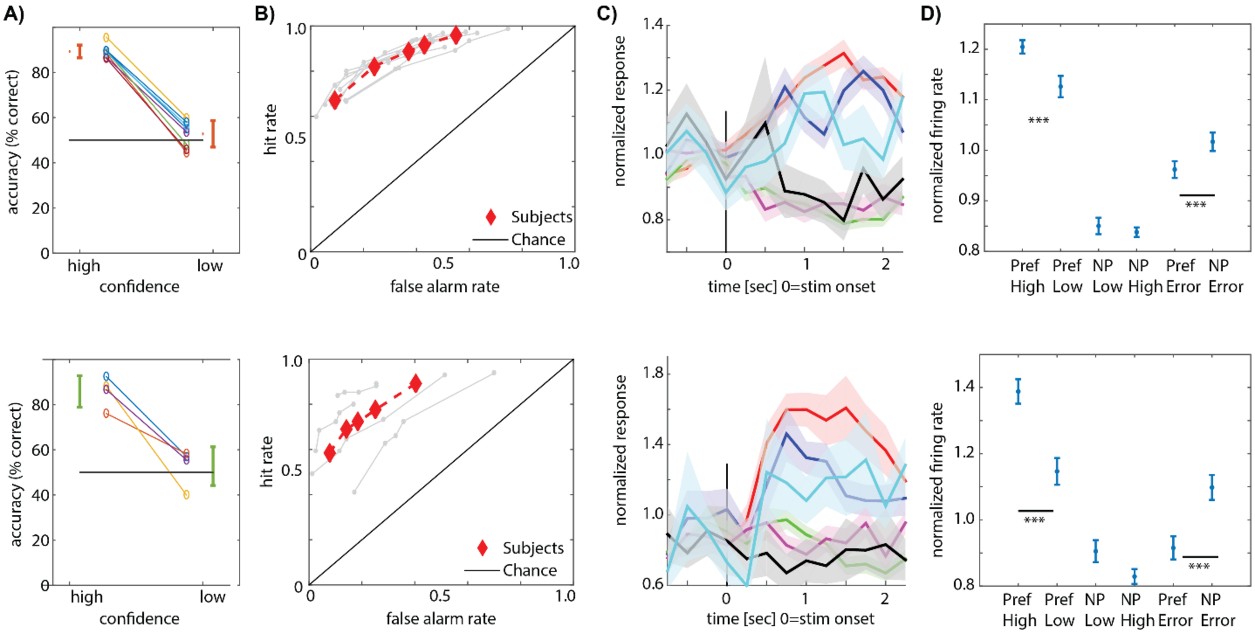

Figure S5: Results shown separately for the two subjects, related to Figure 4. (A,B) Behavior. (A) Accuracy was significantly better for high vs. low confidence (p=2.3e-10 and p=0.00096). (B) Behavioral ROC, with an average AUC of 0.87 and 0.85, respectively. Gray lines are individual sessions, red lines average across all session of each subject. (C,D) Response of MS neurons characterized separately for both subjects. (C) Average PSTH for all MS neurons, with n=135 and n=31 MS neurons, respectively. This number of neurons corresponds to 11.5% and 14.0% of the population of all recorded neurons of each patient. Similarly, CS cells were observed individually in both patients. We found n=107 and n=23 CS cells, respectively (9.1% and 11.0% of the population, respectively). See (D) for statistics. (D) Statistics for MS cell response. Responses were significantly different between high-and low confidence responses for the preferred category (p=0.00092 and p=0.002, respectively). Similarly, responses were significantly different in error trials between the preferred and non-preferred category (p=0.0014 and p=0.00055, respectively), with the response to the non-preferred errors larger and that to preferred errors. The top row is subject NS, the bottom row subject EGS. All numbers in this legend are reported first for NS, followed by EGS.

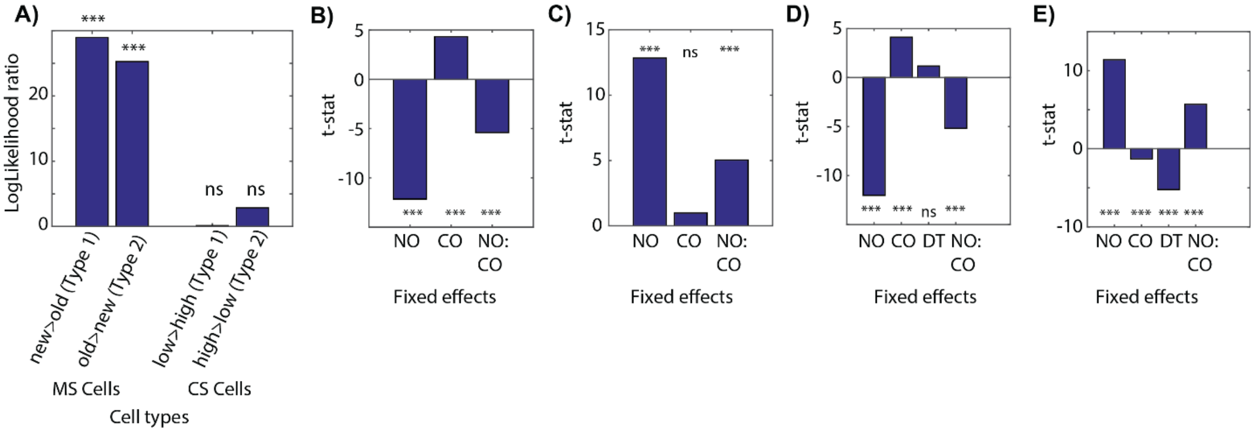

Figure S6: GLM model analysis of population, Related to Figure 3. (A) Model comparison between model 1 (no interactions) and model 2 (Interaction between familiarity and confidence). Model 2 fit the data significantly better for both MS cell types (p=7.3e-8 and p=5.0e-7), but not for either CS cell type (p=0.71 and p=0.09). (B-C) Model fit for all MS cell types 1 (B) and 2 (C). In both cases, there was a significant interaction (p=7.1e-08 and p=4.9e-7, respectively). Note that the sign of the interaction was the same as that for the factor familiarity, confirming that the higher the confidence, the larger the firing rate for the preferred stimulus. (D-E) Model fit for MS cell types 1 (D) and 2 (E) with fixed effect of DT added. Even after accounting for DT differences, the interaction terms remained significant for both MS cell types (p=1.7e-7 and p=1.3e-8, respectively). Note that effects are expressed relative to the baseline condition of "New" and "Low". The sign of each coefficient thus indicates the modulation relative to "New Low". For example, new>old cells have (by definition) higher firing rate to new compared to old trials, thus the effect of familiarity (NO) is negative. Also note that, of course, the main effect of NO is significant by design because of how the cells were selected. However, the result that this panel shows is based on the fixed effects of confidence, DT and their interaction with NO, which are independent of selection. Abbreviations: NO is new/old (factor familiarity). CO is confidence (high or low). NO:CO marks the interaction between these two terms. DT is decision time. *** marks p<0.01.

Supplemental Tables

Table S1, Related to STAR Methods: List of experimental sessions performed. Patients are indicated by initials, followed by the session ID (an internal reference). Each variant (v1-v9) is a unique set of stimuli and was only used once for each subject. Each variant consists of 100 learning trials, but some subjects were only shown 75 learning trials on days on which the patient wasn't comfortable going through all 100. The delay specifies the time that elapsed between the end of the learning block and start of the recognition block. This delay time was on average 36.9±6.1 minutes and varied slightly between sessions depending on the needs of the patient.

Patient (SessionID)	Session nr	Experiment variant	Nr learning (novel) trials	Nr recognition trials	Delay (min)
NS (86)	1	v1	100	200	31
NS (87)	2	v2	75	150	35
NS (88)	3	v3	100	200	39
NS (89)	4	v4	100	200	36
NS (90)	5	v5	100	200	40
EGS (91)	6	v1	75	150	52
EGS (103)	7	v2	75	150	38
EGS (106)	8	v5	75	150	39
EGS (107)	9	v3	75	150	35

NS (108)	10	v7	75	150	29
NS (109)	11	v8	75	150	39
NS (110)	12	v9	75	150	30