Evidence that neural information flow is reversed between object perception and object reconstruction from memory

Juan Linde-Domingo¹, Matthias S. Treder², Casper Kerrén¹ & Maria Wimber¹

Remembering is a reconstructive process, yet little is known about how the reconstruction of a memory unfolds in time in the human brain. Here, we used reaction times and EEG time-series decoding to test the hypothesis that the information flow is reversed when an event is reconstructed from memory, compared to when the same event is initially being perceived. Across three experiments, we found highly consistent evidence supporting such a reversed stream. When seeing an object, low-level perceptual features were discriminated faster behaviourally, and could be decoded from brain activity earlier, than high-level conceptual features. This pattern reversed during associative memory recall, with reaction times and brain activity patterns now indicating that conceptual information was reconstructed more rapidly than perceptual details. Our findings support a neurobiologically plausible model of human memory, suggesting that memory retrieval is a hierarchical, multi-layered process that prioritises semantically meaningful information over perceptual details.
When Rocky Balboa goes back to his old gym in the movie Rocky V, the boxing ring and the feeling of the dusted gloves in his hands trigger a flood of vivid images from the past. Like in many other movies featuring such mnemonic flashbacks, the main character seems capable of remembering what the room looked like years ago, who was there at the time, and even an emotional conversation with his old friend and coach Mickey. Perceptual details like colours, however, are initially missing in the scene, like in a faded photograph, and only gradually saturate over time. This common way of depicting memories in pop culture nicely illustrates that the memories we bring back to mind are not unitary constructs, and also not veridical copies of past events. Instead, it suggests that remembering is a reconstructive process that prioritises more meaningful components of an event over other, more shallow aspects. We here report three experiments that shed light onto the temporal information flow during memory retrieval. Once a reminder has elicited a stored memory trace, are the different features of this memory reconstructed in a systematic, hierarchical way?

Surprisingly little is known about the time course of memory recall, considering our vast knowledge about the information processing hierarchy during visual perception. Visual object recognition is generally assumed to progress from low-level perceptual features, processed in early visual areas, to increasingly higher levels of integration and abstraction along the inferior temporal cortex. What if a mental representation is re-created from memory, without much external stimulation? Retrieving a scene from Rocky V will elicit semantic knowledge about the film (e.g. the actor being Sylvester Stallone), but also mental images that can include fairly low-level details (e.g. whether the scene was in colour or in grey scale). How the brain manages to bring back each of these features when reconstructing an event from memory remains an open question. The present series of experiments tested our central working hypothesis that the stream of information processing is reversed during memory reconstruction compared with the perception of an external stimulus.

Over the last years, multivariate neuroimaging methods have made it possible to isolate brain activity patterns that carry information about externally presented stimuli, but also about internally generated mnemonic representations. Importantly, it has been shown that parts of the neural trace that an event produces during its initial encoding are reinstated during its later retrieval. Most studies focused on the reactivation of abstract information, including a picture’s category or the task context in which it was encoded. Evidence also exists for the reactivation of low-level perceptual details in early visual areas. Moreover, a growing literature using electrophysiological methods is beginning to shed light onto the timing of such reinstatement, typically demonstrating neural reactivation within the first second after a reminder, and sometimes very rapidly. However, because all existing studies focused on a single feature of a memory representation (e.g., its semantic category), the fundamental question whether memory reconstruction follows a hierarchical information processing cascade, similar to perception, has not been investigated.

We hypothesise that such a processing hierarchy does exist, and that the information flow is reversed during memory retrieval compared with perception. That is, based on the widely accepted idea that memory reconstruction depends on back-projections from the hippocampus to neo-cortex, we expect that those areas that are anatomically closer to the hippocampus (i.e., high-level conceptual processing areas along the inferior temporal cortex) are involved in the reactivation cascade faster than relatively remote areas (i.e., low-level perceptual processing areas). Therefore, we assume that once a reminder has initiated the reactivation of an associated event, higher-level abstract features will be reconstructed before lower-level perceptual features, producing an inverse temporal order of processing compared with perception.

We tested this reverse reconstruction hypothesis in a series of two behavioural and one electroencephalography (EEG) experiment. All studies used a simple associative memory paradigm where participants learn arbitrary associations between word cues and everyday objects, and are later cued with the word to recall the object. In order to test for a processing hierarchy, it was important to independently manipulate the perceptual and conceptual contents of these objects. Therefore, objects varied along two orthogonal dimensions: one perceptual dimension, where the object was either presented as a photograph or a line drawing; and a semantic dimension where the object represents an animate or inanimate entity (Fig. 1a). The two behavioural experiments measure reaction times (RTs) while participants make perceptual or semantic category judgments for objects that are either visually presented on the screen, or reconstructed from memory. The EEG experiment uses a similar associative recall paradigm together with time-series decoding techniques, allowing us to track at which exact moment in time perceptual and semantic components of the same object are reactivated, and to create a temporal map of semantic and perceptual features during perception and memory reconstruction. Our behavioural and electrophysiological findings consistently support the idea that memory reconstruction is not an all-or-none process, but rather progresses from higher-level semantic to lower-level perceptual features.

Results

Behavioural experiments. Our two behavioural experiments used RTs to test our central hypothesis that the information processing hierarchy reverses between the visual perception of an object and its reconstruction from memory. We assumed that the time required to answer a question about low-level perceptual features (photograph vs. drawing) compared to high-level semantic features (animate vs. inanimate) of an object reflects the speed at which these types of information become available in the brain. If so, reaction time patterns should reverse depending on whether the object is visually presented or reconstructed from memory: during perception, RTs should be faster for perceptual compared with semantic questions reflecting a forward processing hierarchy; during retrieval, RTs should be faster for semantic compared with perceptual questions if there is a reversal of that hierarchy.

Both experiments used a 2 × 2 mixed design (Fig. 1b, c), where all participants answered perceptual and semantic questions (factor question type, within-subjects) about the objects. Importantly, one group of participants was visually presented with the objects while answering these questions, whereas the other group recalled the objects from memory (factor task, between-subjects). The main difference between the two experiments was that in Experiment 1, both types of features were probed for each object; and in Experiment 2, objects were presented on background scenes (not of interest for the present purpose; see Methods section).

Overall accuracy in both experiments was near ceiling for the visual reaction time task (Experiment 1: M = 96.88%; SD = 2.40%; Experiment 2: M = 97.19%, SD = 2.99%), and high for the memory reaction time task (Experiment 1: 83.15%; SD = 0.92; Experiment 2: M = 66.23%, SD = 15.35%). Note that Experiment 2 was more difficult because participants had to memorise background scenes in addition to the objects’ semantic and perceptual features. In both experiments, only correct trials were used for all further RT analyses.
mental manipulations of interest. We used single trial RTs as explained by the underlying distribution. They are able to capture variance visual perception and memory reconstruction, we used general-
directly test for a reversal of the reaction time pattern between RTs show the expected perception-to-memory reversal

Fig. 1 Stimuli and design of the behavioural experiments. a Illustration of the orthogonal design of the stimulus set. In all experiments, objects (a total of 128) varied along two dimensions: a perceptual dimension where objects could be presented as a photograph or as a line drawing; and a semantic dimension where objects could belong to the animate or inanimate category. b In the visual reaction time task, participants were prompted on each trial to categorise the upcoming object as fast as possible, either according to its perceptual category (photograph vs. line drawing) or its semantic category (animate vs. inanimate). c During the encoding phase of a memory reaction time task, participants were asked to create word-object associations (a total of eight per block). Reaction times were then measured during the retrieval phase, where subjects were presented with a reminder word, and asked to recall and categorise the associated object according to its perceptual (photograph vs. line drawing) or semantic (animate vs. inanimate) features. Button press symbols indicate at which moment in a trial RTs were collected.

RTs show the expected perception-to-memory reversal. To directly test for a reversal of the reaction time pattern between visual perception and memory reconstruction, we used generalised linear mixed-effect models (GLMM). GLMMs are ideal for modelling single trial (e.g. RT) data, without assumptions about the underlying distribution. They are able to capture variance explained by fixed and random variables, including the experimental manipulations of interest. We used single trial RTs as target (dependent) variable. Fixed effects were the kind of task (visual vs. memory), question type (perceptual vs. semantic) and the interaction between task and question type. Participant IDs and slopes were included as random factors (including intercept).

Consistent with the reverse reconstruction hypothesis, we found that the interaction between task (visual vs. memory) and question type (perceptual vs. semantic) significantly predicted RTs in Experiment 1 ($F_{1,9220} = 18.027, P < 0.001$) and Experiment 2 ($F_{1,3280} = 10.588, P = 0.001$). To test whether the interaction was produced by differences in the expected direction (perceptual < semantic during encoding, and semantic < perceptual during retrieval), planned comparisons were then performed for the visual and memory task independently, with question type as fixed effect. We found a significant effect of question type in the visual task (Experiment 1: $B = -0.042, t = -3.973, P < 0.001$; Experiment 2: $B = -0.048, t = -2.457, P = 0.014$), where the negative coefficient indicates that the model indeed predicted lower RTs for perceptual compared to semantic questions. A significant effect of question type was also found in the memory task, following the opposite pattern: positive coefficients now indicate significantly faster RTs during semantic than perceptual questions (Experiment 1: $B = 0.156, t = 2.551, P = 0.011$; Experiment 2: $B = 0.165, t = 2.523, P = 0.012$).

For descriptive proposes, Fig. 2 also illustrates the distribution of participant-averaged RTs (Fig. 2a, b). During the visual task, participants on average were faster at answering perceptual (Experiment 1: $M = 795\text{ ms}; SD = 235\text{ ms}$; Experiment 2: $M = 733\text{ ms}; SD = 211\text{ ms}$) than semantic (Experiment 1: $M = 842\text{ ms}, SD = 185\text{ ms}$; Experiment 2: $M = 797\text{ ms}, SD = 235\text{ ms}$) questions. When performing the same task on objects reconstructed from memory, they were now slower responding to perceptual (Experiment 1: $M = 2502\text{ ms}; SD = 561$; Experiment 2: $M = 3348\text{ ms}, SD = 754$) than semantic (Experiment 1: $M = 2334\text{ ms}; SD = 534$; Experiment 2: $M = 3133\text{ ms}, SD = 660\text{ ms}$) questions.

Reaction time analyses thus support our central hypothesis that the speed of information processing for different object features reverses between perception and memory, a pattern replicated between Experiments 1 and 2.

Accuracies support a reversal between perception and memory. Next we investigated if a similar pattern was present in terms of accuracy (Fig. 2d, e). We used a GLMM with a logistic link function and a binary probability distribution for our target variable (accuracy, correct or incorrect on a given single trial). Fixed effects were the type of task (visual vs. memory), question type (perceptual vs. semantic), and the interaction between the two factors. Participant IDs and slopes were selected as random factors, including intercept.

In both experiments, the interaction between task (visual vs. memory) and question type (perceptual vs. semantic) significantly predicted participants’ accuracy (Experiment 1: $F_{1,11260} = 12.215, P < 0.001$; Experiment 2: $F_{1,4124} = 8.383, P = 0.004$). When running planned comparisons separately for the visual and the memory task in Experiment 1, results for the visual task revealed that question type significantly predicted accuracy ($F_{1,2886} = 5.066, P = 0.024; B = -0.420, t = -2.251, P = 0.024$), suggesting that accuracy for perceptual questions ($M = 97.42\%$; SD = 2.68%) was higher compared to semantic questions ($M = 96.33\%$; SD = 1.99%). In the memory task, question type also predicted accuracy ($F_{1,5374} = 5.374, P = 0.001; B = 0.251, t = 3.222, P = 0.001$), with negative coefficients indicating that participants were more likely to give a correct answer in response to semantic ($M = 85.83\%; SD = 7.57\%$) than perceptual ($M = 82.63\%; SD = 8.79\%$) questions, in line with a reversed processing stream. Experiment 2 showed a similar trend in accuracy profiles. GLMM analyses for the visual task indicated that question type
significantly predicted accuracy ($F_{1,2062} = 4.371, P = 0.037$; $B = -0.585, t = -2.091, P = 0.037$), with better performance for perceptual ($M = 97.97\%$; $SD = 2.77\%$) than semantic questions ($M = 96.41\%$; $SD = 3.07\%$). In contrast, for the memory task we found evidence for the prioritisation of higher-level information (semantic accuracy $M = 69.57\%$; $SD = 15.17\%$) over low-level details (perceptual accuracy $M = 62.89\%$; $SD = 15.09\%$). Here, question type again predicted accuracy in the expected direction ($F_{1,2062} = 6.707, P = 0.010$), with more accurate answers to semantic than perceptual questions ($B = 0.319, t = 2.590, P = 0.010$).

Altogether, the findings from our two behavioural experiments support our main hypothesis that during retrieval of a complex visual representation, the temporal order in which perceptual and
semantic features are processed reverses compared with the initial perception. The results suggest that RTs can be used as a proxy to probe neural processing speed, as previously argued. In the next sections, we report the findings from an EEG study that more directly taps into the neural processes that we believe are producing the behavioural pattern.

**EEG experiment.** While the existing literature suggests that RTs tap into neural processing speed, we wanted to obtain a more direct signature of feature activation from brain activity. We therefore applied multivariate pattern analysis to EEG recordings, with the goal to pinpoint when in time, on an individual trial, the perceptual and semantic features of an object could be decoded from brain activity. We expected that perceptual information becomes available before semantic information when an object is visually presented on the screen, and expected the order of these peaks to reverse when the object is recalled from memory. The design closely followed the behavioural experiments, with the important difference that each participant now carried out a visual encoding phase that served to probe visual (forward) processing, and a subsequent recall phase used to probe mnemonic (backward) processing. The trial timing was optimised for obtaining a clean signal during object presentation and recall, rather than for RTs (Fig. 3). We therefore presented the perceptual and semantic questions only during the recall phase, and at the end of each trial, such that the questions would not bias processing towards perceptual or semantic features.

**Accuracy in the EEG study.** In the retrieval phase of the EEG experiment, subjects were again cued with a word and asked to retrieve the associated object. They on average declared to retrieve the object on 93.60% of the trials (SD = 5.89%), with an average reaction time of 3046 ms (SD = 830 ms; minimum = 1369 ms; maximum = 5124 ms). We then asked two questions at the end of each trial, one perceptual and one semantic, which participants answered with an overall mean accuracy of 86.37% (SD = 6.6). Mirroring the behavioural experiments, average hit rates were 87.65% (SD = 6.57%) for semantic questions, and 85.08% (SD = 6.53%) for perceptual questions. A GLMM showed that the fixed factor question type predicted accuracy ($F_{1,1129} = 7.706, P = 0.006$), with perceptual questions showing a significantly lower hit rate than semantic questions ($B = -0.225, t = -2.776, P = 0.006$). Note that EEG participants were instructed to prioritise accuracy over speed, such that no meaningful RT measures could be obtained in this experiment.

**Evidence for a reversal in single-trial classifier fidelity.** To determine the temporal trajectory of feature processing on a single trial level, we carried out a series of time-resolved decoding analyses. Linear discriminant analysis (LDA, see Methods section) was used to classify perceptual (photograph vs. drawing) and semantic (animate vs. inanimate) features of an object based on the EEG topography at a given time point, either during object presentation (encoding) or during object retrieval from memory (cued recall).

Our first aim was to confirm that there was a forward stream during encoding. Two separate classifiers were trained and tested to classify the perceptual (photograph vs. drawing) and the semantic category (animate vs. inanimate) of the to-be-encoded object, respectively, in each trial and time point per participant (see Fig. 3). Decoding was performed in separate time windows from 100 ms before stimulus to 500 ms post-stimulus. Our main interest was to determine the specific moment in each trial at which the perceptual and semantic classifiers showed the highest fidelity (Fig. 3b, c). For the encoding data, we thus identified the absolute $d$ value peak per trial within 500 ms of stimulus onset. This approach allowed us to compare, within each trial, whether the classification peak for perceptual features occurred earlier than the peak for semantic features. Similarly, we used the cued recall time series to find the time points of maximum decidability of perceptual and semantic features during memory retrieval.

**Behavioural analyses showed that an interaction between type of task (i.e. visual or memory) and question type (i.e. perceptual or semantic) significantly predicted accuracy.** In Experiment 1, both types of questions were asked for each object representation. This allowed us to measure the difference in RTs between perceptual and semantic questions ($X$-axis) on a trial-by-trial level ($Y$-axis) during the visual task (left panel) and the memory task (right panel). Curved lines represent an expected normal distribution. The solid horizontal lines indicate the 50% point of the distribution (i.e., half of the trials), and dashed horizontal lines indicate the trial with a value closest to zero, where the perceptual-semantic difference is flipping from positive (pink) to negative (blue). If differences were normally distributed, the solid and dashed lines would be on top of each other. 

**Behavioural RT and accuracy results.** A Box plots representing reaction times in Experiments 1 and 2 for perceptual (blue) and semantic (pink) questions when an object was physically presented on the screen (visual task, left) or cued by a reminder (memory task, right). We found that RTs were significantly predicted by an interaction between question type and kind of task ($P < 0.001$). For illustrative purposes the $Y$-axis in (a) and (b) is logarithmically scaled. In Experiment 1, both types of questions were asked for each object representation. This allowed us to measure the difference in RTs between perceptual and semantic questions ($X$-axis) on a trial-by-trial level ($Y$-axis) during the visual task (left panel) and the memory task (right panel). Curved lines represent an expected normal distribution. The solid horizontal lines indicate the 50% point of the distribution (i.e., half of the trials), and dashed horizontal lines indicate the trial with a value closest to zero, where the perceptual-semantic difference is flipping from positive (pink) to negative (blue). If differences were normally distributed, the solid and dashed lines would be on top of each other. 

Accuracy results in Experiment 1 for perceptual (blue) and semantic questions (pink) when the object was presented on the screen (visual task) or had to be recalled (memory task). Behavioural analyses showed that an interaction between type of task (i.e. visual or memory) and question type (i.e. perceptual or semantic) significantly predicted accuracy. Box plots representing accuracy in Experiment 2 during the visual and memory task, where the significant interaction effect between type of task and question type was replicated. In all box plots, the line in the middle of each box represents the median, and the tops and bottoms of the boxes the 25th and 75th percentiles of the samples, respectively. Whiskers are drawn from the interquartile ranges to the furthest minimum (bottom) and maximum (top) values. Crosses represent outliers.
suggesting that fidelity peaks for perceptual classification occurred earlier than those for semantic classification. This result validates our peak method, and confirms that low-level features are processed before high-level features during visual perception. The results also suggest that an analysis that takes into account the paired difference between the classifier maxima from each single trial is more sensitive than a GLMM that uses the distributions of all single trials (not revealing a robust difference at encoding).

Importantly, following the same procedure, we next analysed the differences between the perceptual and semantic classifier peaks during memory reactivation, to test if the order reversed during retrieval compared with encoding. The single-trial approach ensured that the relative temporal order of perceptual and semantic peaks within a trial would be preserved even if the retrieval process was set off with varying delays across trials. A one-tailed clustered Wilcoxon signed rank test, revealed a significant difference ($T = 34.602, P < .001$) when comparing perceptual with semantic $d$ value peaks (leaning towards the red side in Fig. 4c). Critically, the one-tailed test in this case confirms our central hypothesis that during memory retrieval, semantic information can be classified in brain activity significantly earlier than perceptual information, suggesting that memory recall prioritises semantic over perceptual information.

ERP results are consistent with a reversed processing. In a final step, we sought to corroborate our classifier-based findings by conventional event-related potential (ERP) analyses. If the differences picked up by the LDA classifier were produced by a signal that is relatively stable across trials and participants, these signal differences would also be visible in the average ERP time courses. A comparison of the ERP peaks during encoding and retrieval would then reveal the same perception-to-memory reversal as found in our multivariate analyses.

Firstly, a series of cluster-based permutation tests (see Methods) was performed during object presentation to test for ERP differences between perceptual and semantic categories. A perceptual contrast of the waveforms for photographs and line drawings revealed a significant positive cluster ($P_{corr} = 0.008$) between 136 ms and 232 ms after stimulus onset, with a maximum difference based on the sum of $T$ values at 188 ms, and located over occipital and central electrodes (see Fig. 5a). Contrasting objects from the different semantic categories (animate and inanimate) revealed a later cluster over frontal and occipital electrodes ($P_{corr} = 0.001$) from 237 ms until 357 ms after stimulus presentation, with a maximum difference at 306 ms (see Fig. 5a). The peak semantic ERP difference for encoding thus occurred ~120 ms after the peak perceptual difference, consistent with the existing ERP literature.

Similar contrasts between perceptual and semantic categories were then carried out during retrieval, again aligning trials to the button press. We found a significant perceptual cluster distinguishing the recall of photographs and line drawings over occipital electrodes ($P_{corr} = 0.046$) between 1390 and 1336 ms before participants’ responses, with a maximum difference at 1360 ms prior to response (see Fig. 5b). Comparing ERPs for the different semantic categories, we found a significant cluster
distinguishing the recall of animate from inanimate objects over frontal electrodes \((P_{\text{corr}} = 0.032)\) between 1781 and 1735 ms before object retrieval, with a maximum difference at \(-1770\) ms (see Fig. 5b). Therefore, during memory retrieval, the peak semantic ERP difference occurred \(-1646\) ms before the peak perceptual difference. Note that the timing of these effects is well aligned with the timing of the classifier results (see Fig. 4). Qualitatively, the ERP results thus mirror the results of our multivariate analyses, again supporting the reversal hypothesis.

An additional analysis was carried out to statistically test for an interaction on the ERP level between type of task (encoding vs. retrieval) and representational features (perceptual vs. semantic). In each participant, we identified the time point of the maximum ERP difference in each of our four comparisons of interest (i.e. photographs/drawings during encoding/retrieval; and animate/inanimate objects during encoding/retrieval). A 2 × 2 within-subjects ANOVA revealed a significant interaction between type of task and type of representational feature \((F_{1,42} = 7.798, P = 0.011)\).

A final follow-up suggests that these ERP differences are not driven by a specific combination of perceptual and semantic features. For each of the clusters identified in the above ERP analysis, we ran a 2 × 2 within-subjects ANOVA, averaging the signal separately for the four types of sub-categories (animate-photographs, animate-line drawings, inanimate-photographs, inanimate-line drawings, see Supplementary Figure 1). We did not find a significant interaction between semantic and perceptual categories in any cluster during encoding (perceptual cluster: \(F_{1,20} = 2.125, P = 0.160\); semantic cluster: \(F_{1,20} = 0.403, P = 0.533\)), and thus no evidence indicating that our main ERP clusters were produced by a difference in one of the sub-categories that constitute the orthogonal dimension.

Altogether, the ERP results confirm that perceptual aspects are coded in brain activity earlier than semantic aspects during visual processing, but semantic differences dominate the EEG signal earlier than perceptual ones during retrieval.

**Fig. 4** EEG multivariate analysis results. For illustrative purposes, box plots show group peak distributions of \(d\) values for perceptual and semantic categories during encoding (a; Perceptual peaks: \(M = 259\) ms, \(SD = 24\) ms; Semantic peaks: \(M = 267\) ms, \(SD = 43\) ms) and retrieval (b; Perceptual peaks: \(M = -1646\) ms, \(SD = 247\) ms; Semantic peaks: \(M = -1772\) ms, \(SD = 177\) ms) after averaging peaks within participants. All box plot elements represent the same metrics as in Fig. 2. c Measuring classifier fidelity in terms of \(d\) value peaks on a single-trial level allowed us to measure the pairwise time distance between perceptual and semantic peaks during encoding (left panel) and retrieval (right panel). Y-axis represents each individual trial, with trials accumulated across participants. The time distance between classifier peaks (time of perceptual peak minus time of semantic peak on a given trial) is represented on the X-axis. The curved line represents an expected normal distribution. The solid horizontal line indicates the 50% point (half of the trials), and the dashed horizontal line indicates the point where the temporal distance values change sign from perceptual < semantic (blue) to semantic < perceptual (pink).
Fig. 5 Univariate analysis results. a Left panels represent ERP group differences (T values) across time in those electrodes that formed a significant cluster during object presentation, locked to the onset of the stimulus. Top left panel shows the contrast of photographs vs. line drawings, and the bottom left panel differences between animate vs. inanimate objects. Scalp figures next to each contrast illustrate the maximum cluster’s topography, averaged across the significant time-window, with all significant electrodes in a cluster being marked with an asterisk. b Right panels show ERP group differences (T values) over time in those electrodes that are contained in the maximum significant clusters during memory retrieval, time locked to participants’ responses. The top right panel shows the perceptual contrast, and the bottom right panel the semantic contrast. Cluster topographies for each comparison are located next to each panel, and the temporal extent of significant clusters is shaded in colour.

Discussion

How does the neural fingerprint of a memory unfold in time when triggered by a reminder? While it is widely accepted that visual object recognition starts with low-level perceptual followed by high-level abstract processing\(^3,4,6,8\), much less is known about the mnemonic feature processing cascade. Here we demonstrate that the reconstruction of a visual memory does depend on a hierarchical stream too, but this mnemonic stream follows the reverse order relative to visual processing. Across three experiments, we found highly converging evidence from RTs and accuracy (Experiments 1 and 2), multivariate classification analyses, and from univariate ERP analyses (Experiment 3), all indicating that conceptual information is prioritised during retrieval.

In the behavioural studies, participants were significantly faster at detecting low-level perceptual than abstract, conceptual differences during a visual classification task, while the object was presented on the screen. Critically, when probing the features of objects recalled from memory, the reverse effect was found: subjects required significantly less time to correctly retrieve semantic information about the object compared to perceptual details (see Fig. 2a, b). This reversal was corroborated by a significant interaction between the kind of feature (perceptual or semantic) and the kind of task (visual perception or memory recall task). Based on signal-detection models\(^5,8,29\), the RT findings suggest that during memory reconstruction, the decision threshold to identify abstract information of a mnemonic representation is reached before sufficient low-level information is available. The response latency pattern therefore supports our central hypothesis that the temporal order in which features come online is reversed when retrieving a stored representation of an object, relative to its perception. In addition to RTs, the same reversal pattern was present in accuracy profiles in both experiments (see Fig. 2d, e). These findings suggest a prioritisation of abstract semantic information over perceptual details of a mnemonic representation, consistent with hierarchical memory system models\(^30\).

The EEG results fully support the conclusions drawn from the behavioural studies. We used temporally resolved multivariate decoding analyses to observe when in time, during object perception and retrieval, the perceptual and semantic features of an object are maximally decodable from brain activity patterns. These analyses were carried out such that the relative temporal order of the perceptual and semantic classifier peaks could be directly compared in each single trial. When an object was visually presented during encoding, the maximum fidelity in classifying perceptual information (photograph vs. drawing) occurred \(\sim 100\) ms earlier than the maximum for semantic information (animate vs. inanimate) (see Fig. 4a). This finding is consistent with a predominantly feed-forward processing as described previously\(^3-6,8\). Note that perceptual and semantic peaks during visual perception only differed statistically when comparing their relative timing on a single trial level, suggesting that such an analysis is more sensitive to detecting relatively small timing differences in noisy data. When we asked participants to reactivate an object’s representation from memory, semantic peaks were found \(\sim 300\) ms earlier than perceptual peaks (see Fig. 4b). Like in the behavioural experiments, a consistent reversal between perception and memory was supported by a significant interaction between the type of feature that was probed (perceptual or semantic), and the type of task participants were engaged in (encoding or retrieval). Finally, we also found the same reversal pattern in the ERP peaks when comparing the maximum ERP difference between perceptual and semantic object classes. During object perception, the largest perceptual ERP cluster occurred \(\sim 100\) ms before the semantic ERP cluster, whereas during retrieval the perceptual cluster followed the semantic one with a lag of about 400 ms (see Fig. 5). In summary, our results provide robust evidence for our main prediction that semantic features are prioritised over perceptual features during memory recall, in the opposite direction of the well-known forward stream of visual-perceptual processing. Follow-up studies will need to test whether this reversed stream is robust under different conditions, for example in tasks that explicitly vary the encoding demands to emphasise perceptual over semantic aspects of an event. If semantic information is always prioritised, this would suggest a hardwired characteristic of the output pathways from the hippocampus back to neocortex. Alternatively, and maybe more likely, the retrieved representation will to some degree also depend on what Marr\(^22\) called the “internal description” of a stimulus during encoding, including the rememberer’s goals and attentional state.

In our studies, the behavioural data were acquired separately from the EEG data, in a setting that was optimised for measuring...
RTs. Studies simultaneously measuring RTs and neural activity suggest that a meaningful relationship exists between EEG classifier fidelity values and human behaviour. In line with signal detection models, it has been argued that the distance between two or more categories in a neural representational space serves as decision boundary that guides behavioural categorisation. For example, Carlson et al. used fMRI-based activation patterns in late visual brain regions in an object animacy task. They found that the faster the RT on a given trial, the further away in neural space the object was represented relative to the boundary between semantic categories. Similarly, an MEG study showed that the decision values during time points of maximum decodability, derived similar to our EEG decoding peaks, were strongly correlated with RTs for visual categorisation. Both studies thus suggest that during object vision, single-trial decoding measures reflect a distance between categories in neural space that translates into behaviour. Our findings indicate that this brain–behaviour relationship extends to mental object representations during memory reconstruction.

How does the reverse reconstruction hypothesis fit with existing knowledge about the neural pathways involved in memory reconstruction? It is generally accepted that during memory formation, information flows from domain-specific sensory modules via perirhinal and entorhinal cortices into the hippocampus. Recent evidence suggests that during visual processing, the coding of perceptual object information is preserved up to relatively late perirhinal processing stages. The hippocampus is considered a domain-general structure whose major role is the associative binding of the various elements that constitute an episode. The hippocampal code later allows a partial cue to trigger the reconstruction of these different elements from memory. This memory process likely depends on back-projections from the hippocampus to neocortical areas, causing the reactivation of memory patterns in (a subset of) the areas that were involved in perceiving the original event. Such reactivation has consistently been reported in higher-order sensory regions related to processing of complex stimuli and task information, but also in relatively early sensory cortex, suggesting that in principle, higher-level and lower-level information can be reconstructed from memory. Recent evidence, however, suggests that the structure of complex naturalistic events (movies) is transformed from perceptual to mnemonic codes during retrieval. This finding is in line with the idea that remembering prioritises higher-order meaningful information over lower-level details.

While the reverse reconstruction hypothesis is neurobiologically plausible and has strong intuitive appeal, direct empirical evidence so far has been lacking. Indirect evidence comes from an fMRI study showing that within the medial temporal lobe, regions involved in visual object and scene processing are also activated when retrieving objects and scenes from memory, but with a delay relative to perception, consistent with a reversed information flow. Intracranial EEG recordings have shown that connectivity between the entorhinal cortex and the hippocampus changes directionality between encoding and retrieval, which could provide the functional basis for cortical reinstatement. Studies in rodents indicate that the hippocampus is in principle capable of replaying the neural code that represents a certain spatial memory in reverse order, in particular when the animal is awake and resting. Finally, work using MEG-based decoding suggests that it is mainly the later visual processing stages that are reactivated during retrieval and mental imagery, consistent with a prioritisation of higher-level information. Our proposal of a reverse processing hierarchy is thus plausible based on functional anatomy and the existing literature, even though it has never been explicitly tested so far.

We regard our reverse reconstruction hypothesis as complementary to existing models that address the nature and timing of different retrieval processes, including the influential dual process model (for a review see ref. 40). Dual process models focus on recognition rather than recall tasks, and on the cognitive processes and operations required to access a stored memory rather than the reactivated features of a memory. Successful recognition presumably can be based on a sense of familiarity, or on the recollection of contextual information from the initial encoding, an influential idea since the introspective analyses of William James. While the original model does not explicitly address the time course of these processes, the EEG literature suggests that familiarity signals occur earlier (~300 ms) than recollection signals (starting from 500 to 600 ms). In contrast, all our experiments probed memory via cued recall, where successful recall strongly depends on the recollection of associative information. Our results suggest that within this recollection process, the semantic “gist” of a memory is accessed before perceptual details. Assuming that familiarity signals reflect a more global and less detailed stage of the retrieval process than recollection signals (an assumption that some find controversial, see ref. 41), the hierarchical progression from an early global semantic signal to more fine-grained recollection might thus be a fundamental principle of retrieval that is shared between recall and recognition memory.

Interesting parallels also exist between our findings and visual learning phenomena like the Eureka effect. The general idea that perception is shaped by stored representations has been proposed over a century ago by von Helmholtz. A wealth of findings support the idea that previous exposures to a stimulus can exert a strong top-down influence on subsequent perception (for a review see ref. 50). Reminiscent of our present findings, Ahissar and Hochstein suggest that such visual learning is a top-down process that progresses from high-level to low-level visual areas. Specifically, they argue that improvements in visual discrimination (e.g. identifying a tilted line among distractors) are guided by high-level information (e.g. “the gist of the scene”) during earlier stages of learning, and increasingly by low-level information (e.g. line orientations or colours) at later stages. If abstract information is reactivated more easily during earlier stages of visual learning, it will influence performance more than detailed information. Even though speculative, the reverse reconstruction framework might thus have explanatory value for findings in related fields.

How our brain brings back to mind past events, and enriches our mental life with vivid images or sounds or scents beyond the current external stimulation, is still a fascinating and poorly understood phenomenon. Our results suggest that memories, once triggered by a reminder, unfold in a systematic and hierarchical way, and that the mnemonic processing hierarchy is reversed with respect to the major visual processing hierarchy. We hope that these findings can inspire more dynamic frameworks of memory retrieval that explicitly acknowledge the reconstructive nature of the process, rather than simply conceptualising memories as reactivated snapshots of past events. Such models will help us understand the heuristics and systematic biases that are inherent in our memories and memory-guided behaviours.

**Methods**

**Participants.** A total of 49 volunteers (39 female; mean age 20.02 ± 1.53 years) took part in behavioural Experiment 1. Twenty-six of them (19 female; mean age 20.62 ± 1.62 years) participated in the memory reaction time task. Five out of these 26 participants were not included in the final analysis due to poor memory performance (c′′ general accuracy) compared with the rest of the group (c′′ = 6.65, P < .01). Another group of 23 participants (20 female; mean age 19.35 ± 1.11 years) volunteered to participate in the visual reaction time task. In a second behavioural
experiment (Experiment 2), 48 participants were recruited (42 female; mean age 19.25 ± 9.1 years). Twenty-four of them performed the memory reaction time task and another 24 took part in the visual reaction time task. For the electrophysiological experiment we recruited a total of 24 volunteers (20 female; mean age 21.91 ± 4.68 years). The first three subjects we recorded performed a slightly different task during retrieval blocks (i.e., they were not asked to mentally visualise the object for 3 s, and they had to answer only one of the perceptual and semantic questions per trial), and were therefore not included in any of the retrieval analyses. Since our paradigm was designed to test for a new effect, we did not have priors regarding the expected effect size. Behavioural piloting of the memory task showed a significant difference in RTs in a sample of $n = 14$. We therefore felt confident that our sample size, in our larger samples of $n = 24$ per group in each of the two behavioural experiments and the EEG experiment.

All participants reported being native or highly fluent English speakers, having normal (20/20) or corrected-to-normal vision, normal colour vision, and no history of neurological disorders. We received written informed consent from all participants before the beginning of the experiment. They were naïve as to the goals of the experiments, but were debriefed at the end. Participants were compensated for their time, receiving course credits or £6 per hour for participation in the electrophysiological experiment. The University of Birmingham’s Science, Technology, Engineering and Mathematics Ethical Review Committee approved all experiments.

Procedure for Experiment 1—Visual reaction time task. Before the start of the experiment, participants were given oral instructions and completed a training block of four trials to become familiar with the task. The main perceptual task consisted of four blocks of 32 trials each (Fig. 1b). All trials started with a jittered fixation cross (500–1500 ms) that was followed by a question screen. On each trial, the question could either be a perceptual question asking the participant to decide as quickly as possible whether the upcoming object is shown as a colour photograph or as a line drawing; or a semantic question asking whether the upcoming object represents an animate or inanimate object. Two possible response options were displayed at the two opposite sides of the screen (right or left). The options for “animate” and “photograph” were always located on the right side to keep the response mapping easy. The question screen was displayed for 3 s, and an object was then added at the centre of the screen. In Experiment 2, this object was overlaid onto a background that filled large parts of the screen. Participants were asked to categorise the object in line with the question as fast as they could, as soon as the object appeared on the screen, by pressing the left or right arrow on the keyboard. To assess whether participants were faster at making perceptual compared to semantic decisions.

All pictures were presented until the participant made a response but for a maximum of 10 s, after which the next trial started. Feedback about participants’ performance was presented at the end of each experimental block. There were 256 trials overall, with each object being presented twice across the experiment, once together with a perceptual and once with a semantic question. Repetitions of the same object were separated by a minimum distance of two intervening trials. In each block, we asked the semantic question first for half of the objects, and the perceptual question first for the other half. The analyses only included trials with correct responses, and excluded all trials with an RT that exceeded the average over subjects by ±2.5 standard deviations (SDs).

Procedure for Experiment 1—Memory reaction time task. The memory version was kept very similar to the visual reaction time task, but we now measured RTs for objects that were reconstructed from memory rather than being presented on the screen, and we thus had to introduce a learning phase first. At the beginning of the trial, the question screen was displayed for 4 s, and the two options to answer were shown on the screen, and we thus had to introduce a learning phase first. At the beginning of the trial, the question screen was displayed for 4 s, and the two options to answer were shown on the screen.
phase, a background scene (1600 × 900 pixels) that was shown on the screen behind each object (286 × 286 pixels), and participants were asked to remember the word–object association. In the third and last object association was tested only once, together with either a perceptual question about the object, a semantic question about the object, or a contextual question regarding the background scene (indoor or outdoor). Therefore, one-third of the objects were tested with a semantic question, one-third with a perceptual question, and one-third with a contextual question. Again, context was not further taken into account in the present analyses.

Procedure for Experiment 3—EEG. Following the EEG set-up, instructions were given to participants and two blocks of practice were completed. The task procedure of the EEG experiment was similar to the memory task in Experiments 1 and 2 except for the retrieval phase (Fig. 3a). Each block started with a learning phase where participants created associations between overall eight action verbs and objects. After a 4-s distractor task, participants were memory for these associations was tested in a cued recall test. In total, the experiment was composed of 16 blocks of eight associations each.

Each trial of the retrieval test started with a jittered fixation cross (500–1500 ms), followed by the presentation of one of the action verbs presented during the learning phase as a reminder. Participants were asked to visualise the object associated with this action verb as vividly and in as much detail as possible while the cue was on the screen. To capture the moment of retrieval, participants were asked to press the up–arrow key as soon as they had the object back in mind; or the down-arrow if they could not remember the object. This reminder was presented on the screen for a minimum of 2 s. A response during this instruction time was included in the analysis (maximum 7.4). Immediately afterwards, a blank square with the same size as the original image was displayed for 3 s. During this time, participants were asked to “mentally visualise the originally associated object on the blank square space”. After a short interval where only the fixation cross was presented (500–1500 ms), a question screen was displayed for 10 s or until the participant responded. A semantic, a linguistic or a contextual question for each of the eight associations was randomly selected, and responses were scored online.

Data collection (behavioural and EEG). Behavioural response recording and stimulus presentation were performed using Psychophysics Toolbox Version 355 running under MATLAB 2014b (MathWorks). For response inputs we used a computer keyboard where directional arrows were selected as response buttons.

EEG was acquired using a Brain Vance amplifier with 128 Ag/AgCl active electrodes. Through a second computer the signal was recorded at a 1024 Hz sampling rate by means of the ActiView recording software (BioSemi, Amsterdam, the Netherlands). For all three experiments it was not possible for the experimenters to be blind to the conditions during data collection and analysis.

GLMM analyses. Generalised linear mixed models (GLMMs) were used to test our alternative hypotheses for accuracy (all experiments), RTs (Experiments 1 and 2), and the relative timing of EEG classifier task (RT) and perceptual classifier task in Experiment 3. We chose GLMMs instead of more commonly used GLM-based models (i.e., ANOVAs or t-tests) because they make fewer assumptions about the distribution of the data, are better suited to model RT-like data43 including our d-value peaks, and can accurately model proportional data that are bounded between 0 and 1 (like memory accuracy). Our conditions of interest were modelled as fixed effects in the GLMM. Unless otherwise mentioned, these were the type of task (visual perception vs. memory retrieval) and the type of feature probed (perceptual vs. semantic). Our central reverse processing hypothesis was tested by an interaction contrast between the factors type of task and task type. Two further planned comparisons were then conducted to test if an interaction was driven by effects in the expected direction (e.g., RTs perceptual < semantic during visual perception, and semantic < perceptual during memory retrieval). For all analyses, participant ID (including intercept) was modelled as a random factor. Wherever possible, we also included slope as a random factor because GLMMs that do not take into account this factor tend to overestimate effects (that is, they are overly liberal48). In all cases, we used a compound symmetry structure based on theoretical assumptions and AIC and BIC values.

Clustered Wilcoxon signed rank test. To compare the pairwise differences between perceptual and semantic d value peaks in each encoding or retrieval trial and test whether the median of these peaks will differ, we used a clustered Wilcoxon signed rank test that clustered the data per participant, using permutation repetitions (2000 permutations). This analysis was run using the R package “clusrank”59.

EEG pre-processing. EEG data was pre-processed using the Fieldtrip toolbox (version from 3 August, 2017) for MATLAB52. Data recorded during the associative learning (encoding) phase was epoched into trials starting 500 ms before stimulus onset and lasting until 1500 ms after stimulus offset. The resulting signal was baseline corrected based on pre-stimulus signal (~500 ms to onset). Retrival epochs contained segments from 4000 ms before until 500 ms post-response. Since these epochs were baseline corrected, we will still contain task-related (i.e., object specific) information, we baseline-corrected the signal based on the whole trial. Both datasets were filtered using a low-pass filter at 100 Hz and a high-pass filter at 0.1 Hz. To reduce line noise at 50 Hz we hand-stop filtered the signal between 48 and 52 Hz. The signal was then visually inspected and all epochs that contained prominent artefacts were discarded. As a result, a minimum of 92 and a maximum of 124 trials remained per participant for the encoding phase, and a range between 80 and 120 trials per subject remained for retrieval. Independent component analysis was then used to remove eye-blink and horizontal eye movement artefacts; this was followed by an interpolation of noisy channels. Finally, all data was referenced to a common-average-reference (CAR).

Time-resolved multivariate decoding. First, to further increase the signal to noise ratio for multivariate decoding, we smoothed our pre-processed EEG time courses using a Gaussian kernel with a full-width at half-maximum of 24 ms. Time-resolved decoding via LDA using shrinkage regularisation52 was then carried out using custom-written code in MATLAB 2014b (MathWorks). Two independent classifiers were trained for each and given time window and each trial (see Fig. 3b) one to classify the perceptual category (photograph or line drawing) and one to classify the semantic category (animate or inanimate). In both decoding analyses, we used undersampling after artefact rejection (i.e. for the category with more trials we randomly selected the same number of trials as available in the smallest category). The pre-processed raw amplitude data of the 128 EEG channels at a given time window for the different conditions for the classifier. LDA classification was performed separately for each participant and time point using a leave-one-out cross-validation approach. This procedure resulted in a decision value (d value) for each trial and time point, where the sign indicates in which category the observation had been classified (e.g., − for photographs and + for line drawings). As a result, a minimum of 92 and a maximum of 124 trials remained per participant for the encoding phase, and a range between 80 and 120 trials per subject remained for retrieval. In order to use the resulting d values for further analysis, the sign of the d values in one category was inverted, resulting in d values that always reflected correct classification if they had a positive value, and increasingly confident classification with increasingly higher values.

Our main intention was to identify the specific moment within a given trial at which each of the two classifiers showed the highest fidelity, and to then compare the temporal segments around the two distinctive peaks. We chose positive d value in each trial, separately for the semantic and perceptual classifiers. The time window used for d value peak selection covered 3 s prior to participants’ response and, based on behavioural RTs, only trials with an RT ≥ 3 s were included (rejecting a total of 1439 trials on a group level). For all further analyses we only used peaks with a value exceeding the 95% percentile of the classifier chance distribution (see section on bootstrapping below), such as to minimise the risk of including meaningless noise peaks. The resulting output from this approach allowed us to track and compare the temporal “emergence” of perceptual and semantic classification within each single-trial. When a peak for a given condition did not exceed the 95th percentile threshold, we did not include the trial in further analyses. For encoding trials, including all participants, we excluded 1.77% of the trials based on this restriction. In the case of retrieval trials, all maximum peaks found exceeded the value of the threshold. In addition to this single-trial analysis, we also calculated the average d value peak latency for perceptual and semantic classification in each participant to further verify the existence of these two distinctive peaks. In cases of negative d values, we could observe differences between semantic and perceptual peaks when using this approach, including variance in processing speed across trials, e.g. for more or
Comparisons across time and electrodes. For all four contrasts, we reported the 300 ms to run a series of button press, but we further subdivided this time window into smaller epochs of press. We used the full time window from 3000 ms before until 100 ms after the contrasts during retrieval, in this case aligning all trials to the time of the button semantic (animate vs. inanimate) categories, controlling for multiple comparisons ERPs between the two perceptual (photograph vs. line drawing) and the two semantic categories during encoding, perceptual categories during retrieval, and retrieval. The above procedure resulted in four statistically meaningful clusters of the single-trial classifier peaks.

Generating an empirical null distribution for the classifier. Previous work has shown that the true level of chance performance of a classifier can differ substantially from its theoretical chance level that is usually assumed to be 1/number of categories. A known empirical null distribution of d values would allow us to determine a threshold for considering only those d values peaks as significant whose values are higher than the 95th percentile of this null distribution. We generated such an empirical null distribution of d values by repeating our classifier analysis with randomly shuffled labels a number of times, and combined this with a bootstrapping approach, as detailed in the following.

As a first step, we generated a set of d value outputs that were derived from carrying out the same decoding procedure as for the real data (including the leave-one-out cross-validation), but using category labels that were randomly shuffled at each repetition. This procedure was carried out independently per participant. On each repetition, before starting the time-resolved LDA, all trials were randomly divided into two categories with the constraint that each group contained a similar number of photographs and line drawings, and approximately the same amount of animate and inanimate objects (the difference in trial numbers was smaller than 8%). The output of one such repetition per participant was one d value per trial and time-point, just as in the real analysis. This procedure was conducted 150 times per participant for object perception (encoding) and retrieval, respectively, with a new random trial split and random label assignment on each repetition. For each participant we thus had a total of 151 classification outputs, one using the real labels, and 150 using the randomly shuffled labels.

Second, to estimate our classification chance distribution for the random-epoch data (i.e., trial-averaged) peak analyses, we used the 151 classification outputs from all participants in a bootstrapping procedure. On each of the bootstrapped repetitions, we randomly selected one of the 151 classification outputs (150 from shuffled labels classifiers and one from a real labels classifier) per participant, and calculated the d value group average based on this random selection for each given time point. The real data was included to make our bootstrapping analyses more conservative, since under the null hypothesis, the real classifier output could have been obtained just by chance. This procedure was repeated with replacement 10,000 times. To generate different distributions for the perceptual and semantic classifiers, we ran this bootstrapping approach two times: once where the real labels output from each subject came from the semantic classifier, and once where the real d values came from the perceptual classifier.

Univariate ERP analysis. A series of cluster-based permutation tests (Monte Carlo, 2000 repetitions, clusters with a minimum of two neighbouring channels within a 10 ms time window) was carried out in order to test for differences in ERPs between the two perceptual (photograph vs. line drawing) and the two semantic (animate vs. inanimate) categories, controlling for multiple comparisons across time and electrodes. First, we contrasted ERPs during object presentation in the encoding phase in the time interval from stimulus onset until 500 ms post-stimulus. We then carried out the same type of perceptual and semantic ERP contrasts during retrieval, in this case aligning all trials to the time of the button press. We used the full time window from 3000 ms before until 100 ms after the button press, but we further subdivided this time window into smaller epochs of 300 ms to run a series of T tests, again using cluster statistics to correct for multiple comparisons across time and electrodes. For all four contrasts, we reported the cluster with the lowest p value.

We were mainly interested in the temporal order of the ERP peaks that differentiated between perceptual and semantic classes during encoding and retrieval. The above procedure resulted in four statistically meaningful clusters across subjects: one for each differentiating perceptual categories during encoding, semantic categories during encoding, perceptual categories during retrieval, and semantic categories during retrieval. To statistically test for an interaction in this timing of these clusters, we extracted the time point of the maximum ERP difference for each individual participant, restricted to the electrodes showing an overall cluster effect but over the entire time window for encoding and retrieval. These time points were entered into a 2 x 2 within-subjects ANOVA with the factors type of feature (perceptual or semantic), and type of task (encoding or retrieval), with the only planned comparison in this analysis being the interaction contrast.

Data availability
The data and that support the findings of this study are in [https://doi.org/10.17605/OSF.IO/327EK](https://doi.org/10.17605/OSF.IO/327EK).

Received: 2 March 2018 Accepted: 3 November 2018 Published online: 14 January 2019

References
1. Schacter, D. L. Constructive memory: past and future. Dialog. Clin. Neurosci. 13, 7–18 (2012).
2. Schacter, D. L., Guerin, S. & St Jacques, P. L. Memory distortion: an adaptive perspective. Trends Cogn. Sci. 15, 467–474 (2011).
3. Carlson, T., Tovar, D., Alink, A. & Kriegeskorte, N. Representational dynamics of object vision: the first 1000 ms. J. Vis. 13, 1–19 (2013).
4. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. Nat. Publ. Gr. 17, 455–462 (2014).
5. Clarke, A. & Tyler, L. K. Understanding what we see: how we derive meaning from vision. Trends Cogn. Sci. 19, 677–687 (2015).
6. Lehky, S. R. & Tanaka, K. Neural representation for object recognition in inferotemporal cortex. Curr. Opin. Neurobiol. 37, 23–35 (2016).
7. Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L. & Barense, M. Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. Elife 7, e31873 (2018).
8. Serre, T., Oliva, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. Proc. Natl Acad. Sci. USA 104, 6424–6429 (2007).
9. Chen, J. et al. Shared memories reveal shared structure in neural activity across individuals. Nat. Neurosci. 20, 115–125 (2017).
10. Johnson, J. D., McDuff, S. G. R., Rugg, M. D. & Norman, K. A. Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron 63, 697–708 (2009).
11. Kuhl, B. A., Rissman, J., Chun, M. M. & Wagner, A. D. Fidelity of neural reactivation reveals competition between memories. Proc. Natl Acad. Sci. USA 108, 5903–5908 (2011).
12. Michelmann, S., Bowman, H. & Hanslmayr, S. The temporal signature of memories: identification of a general mechanism for dynamic memory replay in humans. PLoS Biol. 14, 1–27 (2016).
13. Storesina, B. P., Henson, R. N., Kriegeskorte, N. & Alink, A. Episodic reinstatement in the medial temporal lobe. J. Neurosci. 32, 18150–18156 (2012).
14. Wimber, M., Alink, A., Charest, I., Kriegeskorte, N. & Anderson, M. C. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. Nat. Neurosci. 18, 582–589 (2015).
15. Bosch, S. E., Jhee, J. F. M., Fernandez, G. & Dodler, C. F. Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. J. Neurosci. 34, 7493–7500 (2014).
16. Waldhauser, G. T., Braun, V. & Hanslmayr, S. Episodic memory retrieval functionally relies on very rapid reactivation of sensory information. J. Neurosci. 36, 251–260 (2016).
17. Infarpour, A., Fuentemilla, L., Hornner, A. J., Penny, W. & Duzel, E. Replay of very early encoding representations during recollection. J. Neurosci. 34, 242–248 (2014).
18. Sols, I., Dubrow, S., Davachi, L. & Fuentemilla, L. Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. Curr. Biol. 27, 3499–3504.e4 (2017).
19. Staudigl, T. et al. Memory signals from the thalamus: early thalamocortical phase synchronization entrains gamma oscillations during long-term memory retrieval. Neropsychologia 50, 3519–3527 (2012).
20. Wimber, M., Maas, A. S., Staudigl, T., Richardson-Klavehn, A. & Hanslmayr, S. Rapid memory reactionvation revealed by oscillatory entrainment. Curr. Biol. 22, 1482–1486 (2012).
21. Moscovitch, M. The hippocampus as a “Stupid,” domain-specific module: implications for theories of recent and remote memory, and of imagination. Can. J. Exp. Psychol. 62, 62–79 (2008).
22. Marr, D. Simple memory: a theory for architecture. Philos. Trans. R. Soc. Lond. B 262, 23–81 (1971).
23. Kourt-Nelson, Z., Barnes, G., Sejdicinov, D., Dolan, R. & Dayan, P. Temporal structure in associative retrieval. elife 2015, 1–18 (2015).
24. Lo, S. & Andrews, S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Front. Psychol. 6, 1171 (2015).
25. Ritchie, J. B., Tovar, D. A. & Carlson, T. A. Emerging object representations in the visual system predict reaction times for categorization. PLoS Comp. Biol. 11, 1–19 (2015).
26. Jiang, Y., Lee, M. T. & Rosner, B. Wilcoxon rank-based tests for clustered data with R package clusrank (2017).
27. Fabiani, M., Gratton, G., & Federmeier, K. in Handbook of Psychophysiology (eds. Cacioppo, J., Tassinary, L. & Berntson, G.) (Cambridge University Press, Cambridge, 2007).

28. Ashby, F. G. A stochastic version of general recognition theory. J. Math. Psychol. 44, 310–329 (2000).

29. O’Connell, R. G., Dockree, P. M. & Kelly, S. P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. 15, 1729–1735 (2012).

30. Henson, R. N. & Gagnepain, P. Predictive, interactive multiple memory systems. Hippocampus 20, 1315–1326 (2010).

31. Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S. & Ma, J. Reaction time for object categorization is predicted by representational distance. J. Cogn. Neurosci. 26, 132–142 (2014).

32. Eichenbaum, H. Hippocampus: Cognitive processes and neural representations that underlie declarative memory. Neuron 44, 109–120 (2004).

33. Staresina, B. P. & Davachi, L. Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. J. Cogn. Neurosci. 20, 1478–1489 (2008).

34. Squire, L. R., Stark, C. E. L. & Clark, R. E. The medial temporal lobe. Annu. Rev. Neurosci. 27, 279–306 (2004).

35. Davachi, L. Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. 16, 693–700 (2006).

36. Eichenbaum, H., Yonelinas, A. P. & Ranganath, C. The medial temporal lobe and recognition memory. Annu. Rev. Neurosci. 30, 123–152 (2007).

37. Staresina, B. P., Cooper, E. & Henson, R. N. Reversible information flow across the medial temporal lobe: the Hippocampus links cortical modules during memory retrieval. J. Neurosci. 33, 14184–14192 (2013).

38. Fell, J. et al. in 105–114 https://doi.org/10.1073/pk7-319-46687-3_11 (2016).

39. Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nat. Neurosci. 14, 147–153 (2011).

40. Dijkstra, N., Mostert, P., Lange, F. P., de Bosch, S. & van Gerven, M. A. Differential temporal dynamics during visual imagery and perception. eLife 7, 1–16 (2018).

41. Yonelinas, A. P., Aly, M., Wang, W.-C. & Koen, J. D. Recollection and familiarity: examining controversial assumptions and new directions. Hippocampus 20, 1178–1194 (2010).

42. James, W. Principles of Psychology (Henry Holt and Company, New York, 1890).

43. Mecklinger, A. Electrophysiological measures of familiarity memory. Clin. EEG Neurosci. 37, 292–299 (2006).

44. Rugg, M. D. & Curran, T. Event-related potentials and recognition memory. Trends Cogn. Sci. 11, 251–257 (2007).

45. Klimesch, W. et al. Theta synchronization during episodic retrieval: Neural correlates of conscious awareness. Cogn. Brain Res. 12, 33–38 (2001).

46. Bridson, N. C., Fraser, C. S., Herron, J. E. & Wilding, E. L. Electrophysiological correlates of familiarity in recognition memory and exclusion tasks. Brain Res. 1114, 149–160 (2006).

47. Nyhus, E. & Curran, T. Semantic and perceptual effects on recognition memory: evidence from ERP. Brain Res. 1283, 102–114 (2009).

48. Ahissar, M. & Hochstein, S. Task difficulty and the specificity of perceptual learning. Nature 387, 401–406 (1997).

49. Helmholtz, H. Treatise on Physiological optics. English Translation (Optical Society of America, Rochester, 1924–1925). https://doi.org/10.1037/13536-000.

50. Aggelopoulos, N. C. Perceptual inference. Neurosci. Biobehav. Rev. 35, 375–392 (2015).

51. Ahissar, M. & Hochstein, S. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8, 457–464 (2004).

52. Brodeur, M. B., Dionne-Dostie, E., Monreuil, T. & Lepage, M. The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. PLoS ONE 5, e10773 (2010).

53. Brainard, D. H. The psychophysics toolbox. Spat. Vis. 10, 433–436 (1997).

54. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J. Mem. Lang. 68, 255–278 (2013).

55. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1–9 (2011).

56. Lemm, S., Blankertz, B., Dickhaus, T. & Müller, K. R. Introduction to machine learning for brain imaging. Neuroimage 56, 387–399 (2011).

57. Jamalamadi, H., Alizadeh, S., Schönauer, M., Leibold, C. & Gais, S. Classification based hypothesis testing in neuroscience: below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. Hum. Brain Mapp. 37, 1842–1855 (2016).

58. Combrinon, E. & Ferdi, K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J. Neurosci. Methods 250, 126–136 (2015).

59. Konwieczek, A. & Chapelle, O. An analysis of the anti-learning phenomenon for the class symmetric probabilod. Lect. Notes Comput. Sci. (Inc. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.) 3734 LNLI, 78–91 (2005).

60. Stelzer, J. Chen, Y. & Turner, R. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage 65, 69–82 (2013).

Acknowledgements
We thank Alexandria-Andrei Moise, Emma Sutton, Thomas Faherty, Laura De Herde and James Lloyd-Cox for helping with data collection, and Rodika Sokolik for her useful technical support. This work was supported by a European Research Council Starting Grant ERC-2016-STG-715714 awarded to M.W., and a scholarship from the Midlands Integrative Biosciences Training Partnership (MIBTP) awarded to J.L.D.

Author contributions
J.L.D. and M.W. designed the experiments. J.L.D. conducted the experiment. J.L.D., M.S. and C.K. conducted the statistical analysis. All authors contributed to the analysis approach and to data interpretation. J.L.D. and M.W. wrote the first version of the manuscript and all authors contributed in reviewing and editing.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-08080-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.