Feature-specific reactivation and representational geometries in the temporal structure of episodic memory retrieval

Michael Christopher Postzich

Thesis submitted to the University of Glasgow
for the degree of a Doctor of Philosophy (PhD)

School of Psychology and Neuroscience
College of Medical, Veterinary & Life Sciences
University of Glasgow
16.12.2025

Abstract

Episodic memory describes the ability to recall events and situations we have encountered in the past. These memories usually contain multiple elements with lots of conceptual, gist information as well as perceptual details. During perception of visual objects, perceptual and semantic features are represented at different parts along the ventral visual stream hierarchy resulting in a faster accessibility of perceptual feature information compared to semantic features. Previous research indicates that during object retrieval a reversal of this hierarchy is observable with semantic features being accessible before perceptual ones. The endeavour of the current thesis will be to evaluate this reverse stream effect of memory for its generality and robustness, for underlying cognitive components and spatio-temporal representational formats in the brain.

First, in a set of behavioral studies, feature-specific reaction times are evaluated as a measure of mapping the temporal dependencies of features within an episodic long term memory trace. Second, a detailed analysis of these feature-specific reaction time results with the help of cognitive mathematical models of decision making is presented. Third, the spatio-temporal dynamics of feature representations during the retrieval process are accessed by combining MEG, fMRI and Deep Neural Network encodings of naturalistic images using a Representational Similarity-based Fusion approach.

Results indicate that the reverse stream of feature accessibility during memory retrieval is a robust effect that generalizes over different feature dimensions, cues and attentional states. Reaction time modelling suggests that this effect is explained by differences in evidence sampling speeds from memory for perceptual and semantic features. In the fusion analysis of brain imaging data, successful retrievals showed prolonged

activations of high-level areas along the visual stream compared to early visual areas and semantic, gist-like feature compared to low-level feature representations.

Together, these results provide complementary and detailed evidence for reversed hierarchical structure of an episodic long term memory trace. Further implications and future directions are discussed.

Acknowledgements

This thesis is the result of four years of the most intensive encounter with cognitive neuroscience and memory research of my life so far. These years have been filled with exciting projects, many new methods and analysis techniques, some (hopefully) presentable results, good friends, great discussions, a global pandemic, experiencing Brexit in real time and many more unforgettable memories. For all of these I would like to thank a few people.

The biggest thank you goes out to Marion Brickwedde. You have been the constant in my life and without you nothing of this would have existed. Thank you for all the patience, fun, encouragement, scientific discussion, food and endearment that help me through these years. You are the best part of my life!

I would like to thank my supervisor Maria Wimber for giving me a chance to find out a little more about this curious ability of ours to remember who we are and what makes us "us". Your creativity, encouragement and knowledge have guided me through this PhD without impeding me in exploring on my own. You are a great supervisor!

I'd like to thank my parents Barbara and Michael and my siblings Josephine and Martin. Where I am now, I wouldn't be without you.

I really enjoyed my PhD time in Birmingham and Glasgow because of all the wonderful people I met there that helped me through the troughs of PhD life, enhanced the peaks and made the workplace overall a warm and welcoming environment full of curiosity and pints of ideas. I like to thank my lab colleagues Sander, Marije, Catharina, Julia, Marit, Mircea, Luca, Danying, Chloe, Ben, Simon Hanslmayr and Ian Charest.

And not to forget, a big cheers to the Glasgow Crowd: Alex, Bianca, Christoph, Christopher, Jack, Jelena, Katarina, Laura, Ralitsa, Simon, Tarryn. It was a wonderful experience to spend a year with you all in Glasgow!

Declaration

All work in this thesis was carried out by the author. This work has not been submitted for any other degree. No large language models were used in the writing of this manuscript.

Table of Contents

Publications

Poster Presentation

Postzich, C., Linde-Domingo, J., & Wimber, M. Tracking encoding-retrieval transformations using EEG-based representational structures. Poster presented at: Society for Neuroscience (SFN). 2021 November 8-11; Chicago, IL, USA (online)

Postzich, C., Linde-Domingo, J., ter Wal, M., & Wimber, M. Evidence for the neural information flow during memory reconstruction from feature-specific reaction times. Poster presented at: International Conference of Cognitive Neuroscience (ICON). 2022 May 18-22; Helsinki, Finland

Postzich, C., Linde-Domingo, J., Kerren, C., & Wimber, M. Tracking encoding-retrieval transformations using EEG-based representational structures. Poster presented at: Annual Meeting of the Organization of Human Brain Mapping (OHBM). 2022 June 19-23; Glasgow, United Kingdom

Postzich, C., Daube, C., Kay, K., Charest, I. & Wimber, M. Tracking the reconstruction of naturalistic images from memory using similarity based fusion of MEG and fMRI data. Poster presented at: International Conference on Learning and Memory. 2023 April 26-30; Huntington Beach, CA, USA

Abbreviations

EEG	Electroencephalography
MEG	Magnetoencephalography
fMRI	functional Magnetic Resonance Imaging
DNN	Deep Neural Network
ERP	Event-related potential
µV	mikro Volt
µT	mikro Tesla
RSA	Representational Similarity Analysis
RDM	Representational Dissimilarity Matrix
MVPA	Mulitvariate Pattern Analysis
LDA	Linear Discriminant Analysis
SVM	Support Vector Machne
SSM	Sequential Sampling Models
(H)DDM	(hierarchical) Drift Diffusion Model
VVS	Ventral Visual Stream
M	Mean
SD	Standard Deviation
SEM	Standard Error of the Mean
GLMM	Generalized Linear Mixed Model
ANOVA	Analysis of Variance

List of Figures

List of Tables

Chapter 1: General Introduction

The human mind encompasses a wide variety of intricate and astounding abilities from focusing attention selectively to solving difficult tasks in a new and creative fashion. But none of those capacities are as fascinating as the fact that our wake everyday experiences are continuously and effortlessly saved on the fly to be later recalled and reexperienced in detail. A key aspect of this ability is the associative structure of memory that binds together multiple elements of one episode as well as multiple features of one element (Tulving, 1983, 2002). When you remember your breakfast yesterday, the episodic memory will bring back a host of semantic, gist-like experiences (e.g. remembering a cup of coffee because one always drinks coffee in the morning) but also finer perceptual details that are unique to this certain episode (e.g. remembering drinking from a red cup instead of the favourite green cup that was still in the dishwasher).

1.1 Memory systems in cognition and in the brain

Apart from the question how memory is actually stored and represented in the brain, distinctions between different memory systems have been established in cognitive science. A general distinction between memory systems is made based on the temporal extend of memory storage: 1. Sensory memories, defined for the visual domain as iconic and the auditory domain as echoic memories, which are very short-lived activations of sensory areas that vanish quickly if not actively held in mind. 2. The short-term or working memory system, a capacity-limited system that is able to sustain external or internal information over a longer

period of time (up to 30 seconds). 3. A long-term memory store of possibly unlimited capacity. This memory system can be further divided into non-declarative memories (Schacter, 1987; Squire, 1992) that are not consciously accessible like learned skills and habits (e.g. playing an instrument, driving a car) and conditioned responses, and a declarative part (Cohen & Squire, 1980) with its two major components semantic (facts, general knowledge) and episodic memory (events of one's own life). As Tulving (1983) described it in his seminal work, episodic memory is distinguished from semantic memory by it's associative coupling of content to the place and time that it was encountered in by us which he termed the "what", "where" and "when" as defining attributes of this memory system. It is also the only memory system that is focused on the past and enables a mental travel back in time (i.e. into the situation we have once encountered).

The case for multiple memory systems in the brain was first convincingly made, when Canadian psychologist Brenda Milner started to study a now famous patient called HM (B. Milner et al., 1968; Scoville & Milner, 1957). He had suffered from severe epilepsy and opted for an experimental treatment by neurosurgeon William Scoville who removed large extends of HM's hippocampal system. In the weeks after the surgery, HM recovered nearly without any cognitive impairments (Kensinger et al., 2001; Schmolck et al., 2002) except for a complete anterograde amnesia (i.e. he wasn't able to make any new memories). He was still able to remember life events and facts about the world he learned prior to the surgery. He was also able to acquire new procedural memories over 3 days of practice in a mirror drawing task even without being aware of his learning history (B. Milner, 1965). Further research of animal models and patients with brain lesions has shown that both episodic and semantic memory seem to heavily depend on the hippocampus (Squire et al., 2004).

Early computational modelling of long-term memory encoding and retrieval came to the conclusion that at least two storage systems are needed. This was based on the training of neural networks on list learning tasks that showed that while those mathematical networks (a simplification of cerebral neural networks) were perfectly able to recall items from learning a first list, this knowledge disappeared (or better, was overwritten) as soon as a second list was learned afterwards (a process called catastrophic interference). Humans on the other hand were well able to recall items from both list with only some interference. Therefore, it was assumed that long-term memory relied on a fast-learning system (proposed to be the hippocampal system), that learns new material quickly but also forgets old material quickly, and slow-learning system (neo-cortex) where new material is added slowly and effectively stored forever (McClelland et al., 1995).

1.2 How are memory traces represented in the brain?

The idea of a memory trace was first introduced as the term memory engram by Richard Semon (1906) and describes a physical substrate in the brain that is a persistent change, is able to be reactivated by internal and external cues (ecphory), holds content that was encountered during encoding (formation) and is a physical entity independent of memory related processes such as encoding or retrieval (Josselyn et al., 2015).

A possible explanation of how an episodic memory trace could be achieved is Teyler&DiScenna's Index model (Teyler & DiScenna, 1986; Teyler & Rudy, 2007). It posits that during encoding all elements of an episode (sensory details, multimodal information etc.) are represented as distributed patterns in the cortex. These patterns are then condensed along the ventral visual stream entering the hippocampus where all information is bound into an

assembly of index cells. This index can be conceptualized as a librarian who knows where a certain book (i.e. neocortical pattern) is located without having to know what is written in the book (i.e. what this pattern represents). For a retrieval of the full original experience, a fraction of the original pattern (retrieval cue) suffices to activate the hippocampal index which then in turn activates the whole pattern in the cortex, a process termed pattern completion (Marr, 1971). Recent evidence showed the existence of cell assemblies coding for the conjunction of unrelated stimuli forming during paired associate learning (Kolibius et al., 2023).

The idea that episodic and semantic memory first depend on and then over time become independent of the hippocampus is called the systems consolidation theory (Squire, 1992; Squire & Alvarez, 1995; Squire & Zola-Morgan, 1991) and has been challenged by evidence from lesion patient data that actual episodically detailed memory in opposition to semantic memory never seems to become fully hippocampally independent (Winocur & Moscovitch, 2011). Instead, it is proposed that memories that get independent of the hippocampus achieve this by a transformation away from context and perceptual details to a semantic and gist-like representation (Winocur & Moscovitch, 2011). This is achieved through a process called semantization (Heinen et al., 2023) where perceptually detailed memory representations are transformed by strengthening conceptual features instead of perceptual features (Favila et al., 2020; Paller & Wagner, 2002; Xue, 2018).

1.3 How are visual features represented in the brain?

For initial visual perception, it has been well established, that processing of information after the primary visual areas follows two different streams (Goodale & Milner,

1992a; Mishkin et al., 1983a). An occipito-parietal also called “where” stream, implicated in spatio-visual and motor-related processing, and an occipito-temporal processing “what” stream along the inferotemporal cortex, that is mostly concerned with object processing. The transformation of visual information along this ventral visual stream (VVS) follows an anatomical and functional hierarchy (Barlow, 1972; Martin & Barense, 2023; Yamins & DiCarlo, 2016), from a highly detailed and neuronally dense representation in early visual areas that is conjunctively transformed toward a sparse distribution of high-level feature representation which enable object recognition in higher areas like the inferior temporal cortex (Desimone et al., 1984; Gross et al., 1972; Quiroga et al., 2005).

Studies probing the capacity and fidelity of visual long-term memory have found that humans are able to remember a huge amount images of scenes and objects over days with great accuracy (Brady et al., 2008; Konkle et al., 2010b, 2010a). It seems that formed memory traces are able to differentiate learned images from perceptually and semantically close lures. In their review, Brady et al. (2011) argue for a dependence of this ability on stored knowledge, that is a hierarchical organization of features where high-level, conceptual information can be diagnostic for low-level perceptual information. This is supported by studies showing that providing semantic labels to ambiguous shapes improved memory (Koutstaal et al., 2003), object memory interference is based on conceptual overlap between exemplars of a category than perceptual overlap (Konkle et al., 2010a) and improved memory for objects matching pre-existing schema (Van Kesteren et al., 2012). This implies that categories are important to retrieve the whole memory trace with semantic as well as perceptual information belonging to it (Brady et al., 2011).

1.4 How can we track reactivation in reaction times?

If the features bound up by a memory trace are stored in hierarchical fashion, there should be a way to test this in behavior. In a first study to use a feature-specific recall task for object vision, Linde-Domingo et al. (2019) could show such a hierarchy that is behaviorally distinguishable using objects that varied on a perceptual (line drawings or photographs) and semantic dimension (animate or inanimate). Over three experiments they invited participants and split them into two groups. The first group participated in a visual task in which they were presented with objects and had to classify them according to their perceptual ("Is the object a line drawing or a photograph?") and semantic ("Is the object animate or inanimate?") dimensions. In a memory task participant first learned to associate objects with words and were then in a recall phase presented with the words again as a memory cue, had to retrieve the object from memory and classify it like in the visual task. So, the main difference between both task was whether feature-based information came from initial perception or memory recall. It was shown that while in the visual task classification of perceptual features was faster than classification of semantic features, indicative of a forward stream of object processing, in the memory task semantic was faster than perceptual classification, giving first evidence to a reversal of the visual hierarchy of a LTM trace.

This finding has been replicated by Lifanov et al. (2021), where they tested how this difference in feature availability behaved over multiple test intervals up to one day. They showed that the effect actually gets larger and an additional dependency analysis showed that unsuccessful recall of the semantic feature was predictive of unsuccessful recall of the perceptual feature but not vice versa. There is also evidence for the same prioritization of semantic over perceptual detail content in multi-item working memory (Kerrén et al., 2023).

All in all, there seems to be preliminary evidence that reaction time measures of feature-specific classification from memory can track the hierarchical organization of the memory trace. It is still a question how reliable this measure is under experimental boundary conditions (e.g. could this reversal of information be result of attentional biases toward semantic features during learning or of the type of cues that are used in the memory task?). If feature-specific reaction times can be established as a stable and easy measure of feature accessibility, it would be interesting to see whether mathematical cognitive models could give an insight into and reveal which underlying factors are driving the classification process.

1.5 Reaction time modelling

For the modelling of perceptual decision-making reaction time distributions, Sequential Sampling Models (Townsend & Ashby, 1984) have been established that understand the information integration for a decision as an accumulation of evidence over time towards a boundary. A choice for a response option is made the moment the process hits a boundary (in a two-choice setting: choice A or choice B).

A well-established model for linking decision-based reaction time distributions to underlying cognitive information integration has been the Drift Diffusion Model (Ratcliff, 1979; Ratcliff et al., 2016). In its original formulation, a decision process comprises a decision-related evidence accumulation process and a nondecision time component. The accumulation process is modelled as a diffusion process with a constant drift rate δ that starts after stimulus encoding and walks through the decision space towards an upper or lower decision boundary α. The drift rate can be understood as how easy information is accessible (i.e. stimulus quality) and the decision boundary describes a speed accuracy trade-off (i.e.

higher boundaries lead to longer reaction times and less errors and lower boundaries lead to faster reaction times and more errors). The starting point of the decision can also vary and is modelled as a parameter called bias β. A bias toward one response is usually expected if participants have prior knowledge over the likelihood of a certain response being correct. After a boundary is reached by the accumulation process, a motor response is made. Both stimulus encoding time and motor response are lumped together into the nondecision time τ. To allow for differences in the reaction time distributions for correct and erroneous responses, the original DDM includes a trial-to-trial variability parameter for the drift-rate, the starting point bias and the nondecision time.

These additional parameters complicate model estimation and typically require huge amounts of reaction time trials per participants and condition (Boehm et al., 2018). A simplified version of the DDM without trial-to-trial variability parameters has been proposed (Wagenmakers et al., 2007; van Ravenzwaaij et al., 2017; but see Ratcliff, 2008) having the additional property of an analytically tractable likelihood function (Navarro & Fuss, 2009).

1.6 What methods can be used to investigate patterns in the cortex?

Measuring brain activity during an experiment with neuroimaging methods produces high dimensional, multivariate datasets, for example for each participant there will be one matrix with 3 spatial and 1 trial dimensions of voxel activations for fMRI studies or a matrix of a trials, a sensor/electrodes and a time point dimension for M/EEG. Trials are hereby defined as replications of experimental stimuli or conditions. Here we will only refer to both as stimuli. Typical univariate analysis compares the brain's average response to certain

stimuli directly. The drawback of this method when dealing with high-dimensional data is the choice between running a search-light analysis at all voxels or time-points incurring inflation of family-wise error rates or having to predefine and average regions- and time-windows-of-interest.

Therefore, the advent of methods like Multivariate Pattern Analysis (MVPA) and Representational Similarity Analysis (RSA) has been hugely influential in the field of cognitive neuroscience (Cox & Savoy, 2003; Haxby et al., 2001, 2014). Instead of averaging signal within a brain region or over electrodes, these methods compare pattern vectors of brain activity. These pattern vectors can be for example a cluster of voxel activations from a fMRI region or a collection of MEG sensor activation values at 200 ms after a stimulus onset. As can be seen, the big advantage of pattern-based methods is taking information at a single voxel/sensor level into account.

1.7 Multivariate Pattern Analysis (MVPA)

Generally, MVPA tries to quantify how patterns of brain activation in response to two different stimuli differ from each other. For our example, one could take the voxel activations in early visual cortex as a pattern vector and compare pattern vectors for trials in which a house and trials in which a face is presented. To compare vector patterns, MVPA either uses measure based on distance estimation or classifier decoding ability. As pattern vectors of size N can be understood as points in a N-dimensional feature space, the distance between those points can be quantified in different ways among them direct and angular distance measures. Common measures of the direct distances between the points are the Euclidean distance or its multivariate version the Mahalanobis distance. or they are based on the angular distances

of the points compared to the origin like the cosine or correlation distance (Grootswagers et al., 2017; Guggenmos et al., 2018; Walther et al., 2016). Decoding measures on the other hand use statistical classifier techniques to find a decision threshold that best separates two distributions of pattern vectors. The most popular among them in cognitive neuroscience are Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Support Vector Machines (SVM).

In machine learning, cross validation (CV) has become the gold standard in establishing an unbiased classification performance. The idea of cross-validation is to randomly split the data sample (trials) into a larger training set and a smaller test set, fit the model parameters on the training set and test these parameters on the test set. This procedure is repeated several times, each time with a new training-test split and the resulting performance value is estimated as the average performance over all splits. While distance measures are usually estimated directly from the data and unlike cross-validated measures do not require multiple trials, there are reformulations of the Euclidean and Correlation distance using cross-validation (Guggenmos et al., 2018).

1.8 Representational Similarity Analysis (RSA)

Similarly to the MVPA, an RSA compares pattern vectors of the brain response to stimuli. In an RSA approach, each stimulus or condition is compared to each other thereby building up a stimulus-by-stimulus comparison matrix called a representational dissimilarity matrix (RDM). All measures described for MVPA can be used for to build a RDM of all stimulus comparisons. A RDM transforms the original data modality from its own signal

space into a multivariate representational geometry (Diedrichsen & Kriegeskorte, 2017; Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte & Kievit, 2013).

This geometry describes how a certain feature space represents a whole dataset of stimuli. For fMRI studies, the feature space can be chosen in a region-based approach to contain all voxels falling within predefined ROI resulting in one RDM per ROI, or, in a searchlight-based approach, to contain a sphere of voxels that is shifted through the whole brain volume (or over surface map) resulting in one RDM per center voxel of the searchlight. For M/EEG studies, the feature space is usually chosen to be all sensors/electrodes and RDMs are computed with a time window that is slid over the time dimension resulting in one RDM per time point. For Deep Neural Networks (DNN), the feature space can be the hidden activation of network layers in response to an input image or sentence (Kriegeskorte, 2009).

1.9 RSA-based fusion

As mentioned above, RSA transforms signal from a modality specific space (signal change for fMRI, μV or fT for M/EEG) into the same representational geometry space. Because of this, it is possible to correlate RDMs from different data modalities a procedure called second-order correlation or RSA fusion (Cichy & Oliva, 2020). The resulting measure indicates when the representational formats of two data modalities are the most similar. A fusion of fMRI and M/EEG RDMs for example combines the spatial resolution of the fMRI with the temporal resolution of M/EEG and shows at which point in time which regions represent the stimulus pool the more similar than others. This approach has been used to show a hierarchical forward stream in information processing during object perception (Cichy, Pantazis, et al., 2016; Kriegeskorte, Mur, Ruff, et al., 2008).

Furthermore, a fusion approach does allow for correlation of neural RDMs with model RDMs, that capture hierarchical feature representations similar to the sensory areas of the brain (Kriegeskorte, 2009). Potential candidate models could be biologically inspired convolutional Deep Neural Networks (DNN) that have been pretrained on a large amount of visual data and have an astounding object recognition capability rivalling human performance (Krizhevsky et al., 2017). For the visual system, DNN RDMs have been widely used to explain feature-specific activations in fMRI and M/EEG signals (Bone et al., 2020; Cichy, Khosla, et al., 2016; Heinen et al., 2023; Kaniuth & Hebart, 2022).

1.10 Overview over the following Chapters

In the Second Chapter, I will present a host of behavioral reaction time experiments testing the robustness of the reverse stream effect first described in Linde-Domingo et al. (2019) and how well feature-specific reaction times are able to map the hierarchical structure of the memory trace. Three different variations of the original experiment will be reported, exploring the questions in how far the reverse stream effect generalizes to different semantic and perceptual feature dimensions, whether it is dependent on the cue material that is used to associate the objects with, and whether this effect could be simply a result of attentional biases during learning.

In the Third Chapter, a detailed cognitive modelling analysis of the whole reaction time dataset acquired and detailed in the Second Chapter will be presented. Using a Hierarchical Bayesian Prior Structure and state-of-the-art sampling algorithms, a DDM will be fit to the reaction time distribution of over 400 participants in 9 memory and 6 visual tasks,

yielding posterior parameter estimates that will be compared between perceptual and semantic feature classification.

In the Fourth Chapter, the retrieval process for naturalistic images will be tracked by using an RSA-based fusion approach. For this, 29 participants will be recorded in the MEG while they first learn image word pairs and then are presented with the word in order to recall the image and reconstruct it from memory. MEG data will be transformed into a representational geometry format and will be correlated (fused) with the region-specific RDMs from a fMRI dataset of different participants and feature-specifc RDMs from DNNs.

Chapter 2: Feature-specific reaction times as a generalizable approach to measuring the state of a memory trace

2.1 Introduction

Our episodic memory system has the ability to effortlessly bind content elements that appeared together at the same time and in the same place into a memory trace. For example, we might remember in detail the last time we met some friends in a park. The elements contained in a memory, however, are collections of features in themselves. For example, one of the friends might have brought a dog along and we still remember semantic details about it (it was a Golden Retriever) as well as perceptual details (its fur was brown).

Research on memory content has typically focused on the accuracy of the recalled features or stimuli. We believe that the timing of content availability holds crucial information that can be used to test the neural architecture of the memory retrieval process. We show here that reaction times represent a reliable method to assess differences in content accessibility that are robust to variations of features being tested, cues to initiate recall, and attentional demands during learning.

Episodic long-term memory crucially relies on the medial temporal lobe and especially the hippocampus as a system of rapid, one-shot learning (McClelland et al., 1995; Scoville & Milner, 1957). According to the dominant view of episodic memory formation, a memory trace is established with the help of a hippocampal neuron assembly (called an index) that binds together cortical patterns that represent the content of an episode (Teyler & DiScenna, 1986). As such, the episodic memory trace associatively binds together the

constituting content's perceptual and semantic features. The recall of a past episode is thought to require a pattern completion process, where a reminder (e.g., one element) triggers the reconstruction of the remaining elements (Horner et al., 2015; Marr, 1971). As such, it was shown that feature activation across different levels of the visual hierarchy coincided with episodic memory retrieval and were predictive of memory vividness and recall accuracy (Bone et al., 2020). Furthermore, detailed episodic memory was dependent on a reactivation of features in the posterior hippocampus and early visual cortex (Bone & Buchsbaum, 2021).

Regarding the spatial and temporal representation of features during initial perception, visual information processing has been shown to follow a hierarchical structure. Retinal input is transformed along the ventral visual stream from low-level details in early visual areas to high-level semantic categories in later areas (Cichy, Pantazis, et al., 2016; Desimone et al., 1984; Martin et al., 2018). Research on visual long-term memory shows the importance of semantic information over perceptual details (Konkle et al., 2010a; Van Kesteren et al., 2012) suggesting a reuse of the visual hierarchy but in the opposite direction (Brady et al., 2011). If the features of a long-term memory trace are spatially organized in a reversal of hierarchy, it would also suggest a temporal distinction in availability of semantic compared to perceptual features during retrieval.

Linde-Domingo et al. (2019) tested this feature-specific hypothesis with in an associative cued recall task where one group of participants first learned and later retrieved visual objects from memory. To test object retrieval participants had to classify the object on either a semantic or perceptual dimension. Here, classification of semantic features was overall faster than of perceptual features indicating a temporal hierarchy of information accessibility. This reverse stream effect was in another study shown to get larger over multiple testing events especially after one day (Lifanov et al., 2021), revealing an interesting

sensitivity of this feature-specific reaction time difference to track the ‘semanticization’ of the memory trace (Heinen et al., 2023). Although, these results are encouraging, they rely on a cued recall paradigm using action verbs as cues and only one perceptual and semantic dimension. Accordingly, it is unclear whether feature-specific reaction times as a method are a generalizable tool to probe the organization of a long-term memory trace.

Here, we show that reaction time measures of feature-specific decision processes are a robust measure, which can be used to delineate differences in the temporal availability of such stored information. We believe that reaction time measures are so far underutilized in memory research and are capable of providing reliably readout about differences in feature availability that could elucidate the state of a memory trace. For our hypothesis, by analyzing feature-specific reaction times to map a memory trace via the observed reverse processing stream during object retrieval, we would expect to find this effect independently of three different variations of the original experiment:

First, this effect should be generalizable to different perceptual and semantic feature dimensions that objects can vary upon. Therefore, we would predict to find the same forward processing stream indicated by faster reaction times for perceptual features compared to semantic in a visual task and the same reverse processing stream with faster reaction times for semantic compared to perceptual feature classification in a memory task when different feature dimensions are tested.

Second, since the memory task relies on a cue to identify the object and retrieve it from memory, an advantage of semantic features could be explained by the semantic nature of the cue (i.e. words in the original studies). To address this possible confound, we will include two types of cues in our paradigm (i.e. words and scenes) and compare whether the reverse stream effect is dependent on words or also translate to scene cues. Additionally in a

second and third study, we will pair objects with arbitrary locations on the presentation screen and use these locations as a cue. As these locations are semantically meaningless and should not influence perceptual or semantic features of the object associated with them, the emergence of a significant reverse stream would show that feature-specific reaction time are able to robustly track a hierarchical nature of a visual memory trace independent of cue confounds.

Third, memory is known to be influenced by attentional focus during learning especially when shifting the encoding focus towards deeper semantically meaningful features compared to shallow perceptual ones (Craik & Lockhart, 1972). In two additional experiments we will shift the attention during learning of objects towards semantic of perceptual features, to investigate whether the feature-specific reaction time difference is due to attentional biases towards semantic features or not.

2.2 Results

2.2.1 Feature Variation Results – The reversed stream effect generalizes to different features

In the first set of experiments, we asked whether the reversal of the perceptual-to-semantic gradient between perception and memory retrieval holds for different object features, beyond the specific features used in previous work. For Exp. 1, we created a new stimulus pool of 96 objects that belonged to either manmade or natural categories (semantic features) and presented them in either big or small size on screen (perceptual feature). One

group of participants performed a purely visual task where on each trial, they were presented with a semantic (natural or manmade?) or perceptual (small or large?) question first, and then had to classify the following object as fast and accurately as possible. Another group performed a memory task, where they first learned to associate the objects with cue words, and then later recalled the objects from memory when prompted with a cue word. Like in the visual task, a memory trial started with a semantic or perceptual question but was then followed by cue word prompting participants to retrieve the associated object from memory and to classify it according to the question type. During visual classification, we expected faster reaction times for perceptual than semantic questions if objects are processed along a forward visual stream from lower-level perceptual to higher-level semantic analysis. Vice versa, faster reaction times for semantic than perceptual classification in the memory task would be indicative of a reverse processing stream during retrieval (Lifanov et al., 2021; Linde-Domingo et al., 2019).

Participants who saw the objects visually rather than reconstructing them from memory were, unsurprisingly, significantly faster overall (main effect of task, $F(1,57) = 132.69$, $p < .001$, $\eta_p^2 = .7$), and the semantic classification was overall faster than the perceptual one (main effect of feature type, $F(1,57) = 7.42$, $p < .01$, $\eta_p^2 = .12$). However, the two main factors of feature type and task were interacting significantly, $F(1,57) = 9.8$, $p < .005$, $\eta_p^2 = .15$, confirming our hypothesis of faster perceptual classification in the visual task, $t(32) = -3.02$, $p < .005$, $d = -0.53$, and faster semantic classification in the memory task, $t(25) = 1.98$, $p < .05$, $d = 0.39$.

The accuracies partly mirrored the reaction time results, with more correct trials in the visual compared to the memory task, $F(1,57) = 50.3$, $p < .001$, $\eta_p^2 = .47$, and overall more correct semantic classification than perceptual, $F(1,57) = 23.85$, $p < .001$, $\eta_p^2 = .29$. A

significant interaction between both factors, $F(1,57) = 7.32$, $p < .01$, $\eta_p^2 = .11$, indicated higher accuracy for semantic than perceptual feature type in the memory task, $t(25) = -3.72$, $p < .005$, $d = -0.73$, and, unexpectedly, as well in the visual task, $t(32) = -2.17$, $p < .05$, $d = -0.38$. So far, the reaction time findings replicate the results of our previous studies and generalise the perception-to-memory flip to new feature variations.

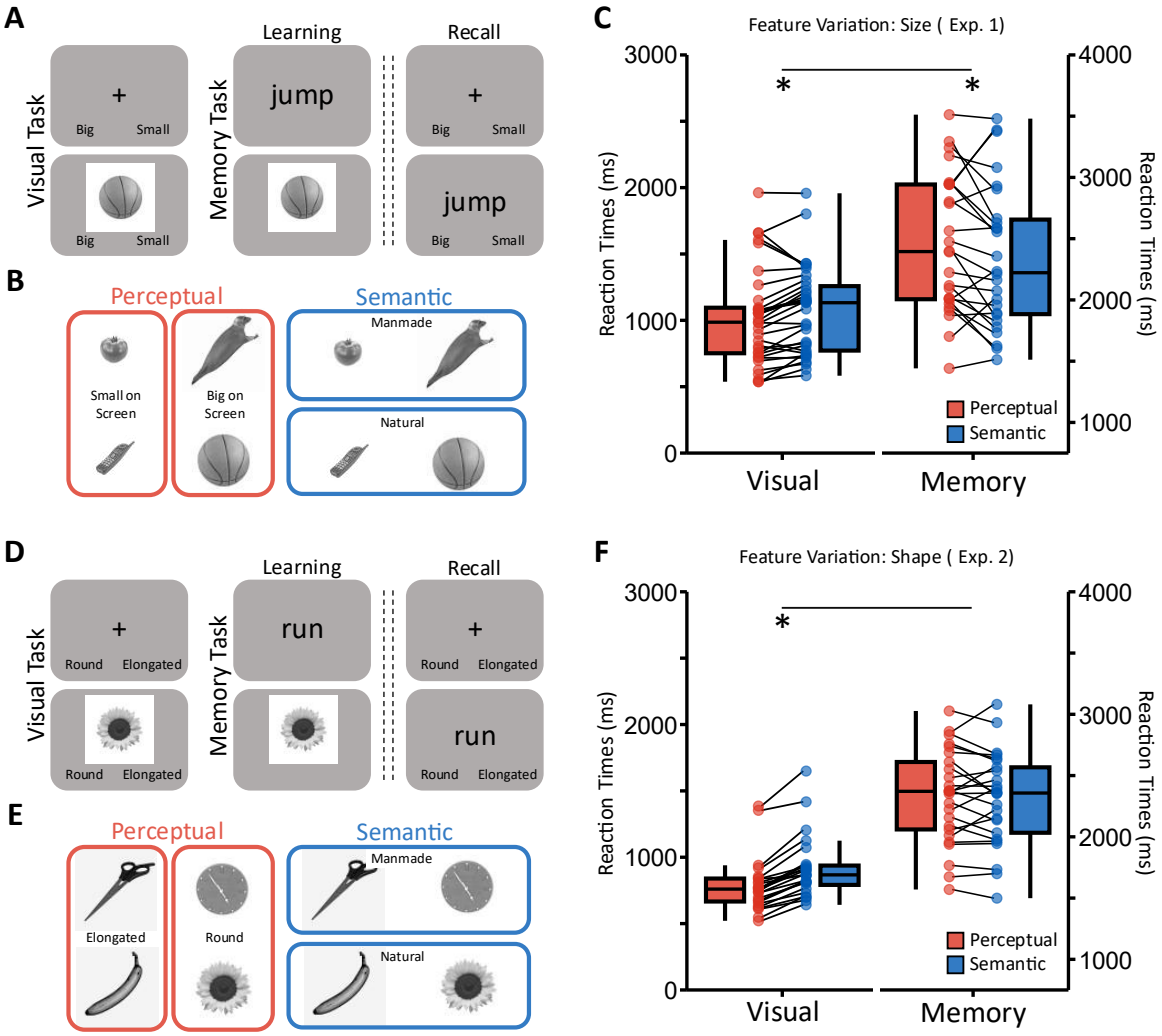

Fig. 1: Paradigm and Results of the Feature Variation Experiments. A: Exemplary depiction of the cued recall paradigm used in Exp. 1. B: Exemplary depicitions of the object images used in Exp. 1 with perceptual and semantic dimensions. C: Boxplots of the reaction time distribution of the task by feature type interaction for Exp. 1. D: Exemplary depiction of the cued recall paradigm used in Exp. 2. E: Exemplary depicitions of the object images used in Exp. 2. F: Boxplots of the reaction time distribution of the task by feature type interaction for Exp. 2. C&F: Dots indicate single participant‘s average reaction times. Bars indicate significant differences of main effects of task or feature type ($p < .05$) and stars indicate significant differences of within tasks ($p < .05$).

In Exp. 2, we varied object shape as a perceptual feature. A stimulus pool of 128 objects was chosen, with manmade or natural objects again serving as the semantic dichotomy, while along the perceptual dimension the objects could now be either elongated or round in appearance. Again, there was a main effect of task such that the visual classification group showed faster reaction times than the memory group, $F(1,47) = 271.31$, $p < .001$, $\eta_p^2 = .85$. but this time there was no significant difference between feature types, $F(1,47) = 0.85$, $p > .05$. There was a significant interaction between task and feature type, $F(1,47) = 26.08$, $p < .001$, $\eta_p^2 = .36$, with post-hoc analyses revealing a significant forward stream in the visual task, $t(23) = -9.97$, $p < .001$, $d = -2.035$, but no significant difference between semantic and perceptual classification in the memory group, $t(24) = 0.73$, $p > .05$.

For the accuracy measures no significant main effect of feature was observed, $F(1,47) = 0.2$, $p > .05$, but the significant main effect of task, $F(1,47) = 35.1$, $p < .001$, $\eta_p^2 = .43$, was qualified by an interaction with feature, $F(1,47) = 7.9$, $p < .01$, $\eta_p^2 = .14$. Post-hoc t-tests within tasks showed again that perceptual features were more often correctly classified than semantic features in the visual task, $t(23) = 3.06$, $p < .01$, $d = 0.62$, and no significant difference occurred when features were classified from memory, $t(24) = -0.9$, $p > .05$.

Taken together, the results of Exp. 1 and 2 show that reaction times reliably track the direction in which object features are processed during visual perception and memory. Exp. 1 replicated our previous findings of forward (perceptual before semantic) processing stream during perception, and a backward (semantic before perceptual) processing stream during retrieval. Importantly, the reversal of feature processing during memory retrieval was absent in Exp. 2, using shape as a perceptual feature. Amongst the feature variations used in this series of studies and previous ones (Lifanov et al., 2021; Linde-Domingo et al., 2019; ter Wal

et al., 2021), the shape feature was in fact the only one that did not produce a semantic-over-perceptual advantage during memory recall. Albeit a post-hoc interpretation, we realised that the shape of an object is bound to its semantics (e.g., knives are elongated while melons are round-shaped), and therefore does not qualify as a purely perceptual feature that can be freely varied across items, like size or colour. As a result, recalling the identity of a memorised object (e.g., that it was a turtle) automatically provides the shape information, explaining why perceptual reaction times matched semantic ones in this version of the task. As an intermediate conclusion, it can thus be summarized that the perception-to-memory reversal in the feature processing hierarchy generalizes to novel feature dimensions as long as those dimensions are clearly separated.

2.2.2 Cue Variation Results – The reversed stream effect is independent of the type of cue used to prompt an object memory

In Exp. 1 and previous work (Lifanov et al., 2021; Linde-Domingo et al., 2019), we find a robust advantage of semantic over perceptual information when an object is recalled from memory. While consistent with our hypothesised reversal in feature processing, an alternative explanation could be that we used verbal retrieval cues in all these studies. Words are inherently semantic in nature, and participants may thus naturally bind the cue words to the objects on the level of their meaning. This may result in a semantic bias, leading to faster retrieval of semantic information faster during memory recall. To test for such a possible bias, in the next experiments we varied the types of cues that were paired with objects.

Instead of verb-object associations, participants in Exp. 3 learned triplets consisting of a scene, a verb and an object, and each associated was later probed once with the scene

cue and once with the verb cue. Feature-specific reaction times for semantic (animate vs inanimate) and perceptual (photo vs drawing) features were measured using the same setup as in Exp. 1 and 2. Having replicated the forward stream during perception several times, we did not include a visual group in this Experiment.

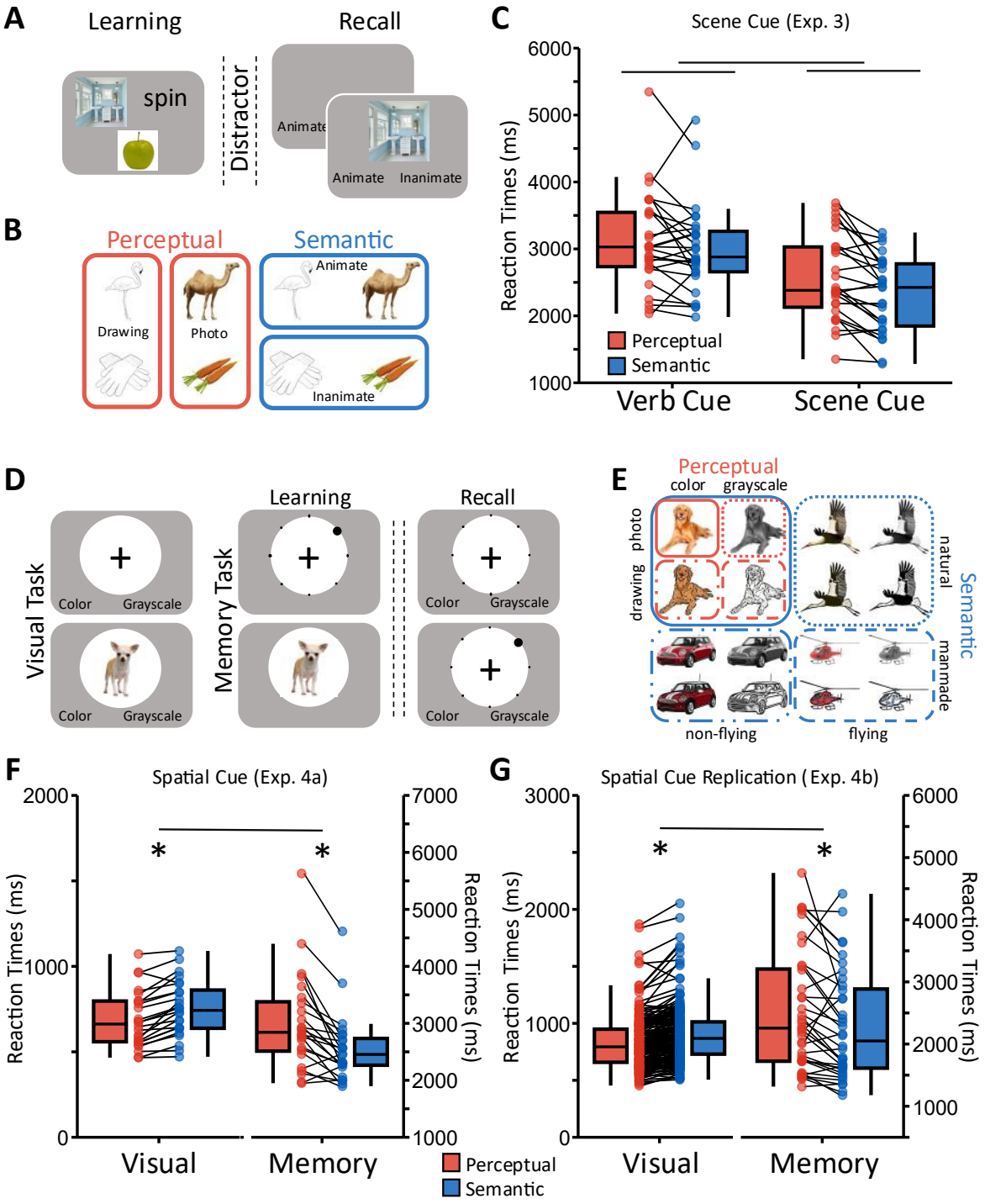

Fig. 2: Paradigm and Results of the Cue Variation Experiments. A: Exemplary depiction of the cued recall paradigm used in Exp. 3. B: Exemplary depicitions of the object images used in Exp. 3 with perceptual and semantic dimensions. C: Boxplots of the reaction time distribution of the cue type by feature type interaction in Exp. 3. Dots indicate single participant‘s average reaction times. Bars indicate significant differences of main effects of task or feature type ($p < .05$). D: Exemplary depiction of the cued recall paradigm used in Exp. 4a and 4b. E: Exemplary depicitions of the object images used in Exp. 4a and 4b with perceptual and semantic dimensions. F: Boxplots of the reaction time distribution of the task by feature type interaction in Exp. 4a. G: Boxplots of the reaction time distribution of the task by feature type interaction in Exp. 4b. F&G: Dots indicate single participant‘s average reaction times. Bars indicate significant differences of main effects of task or feature type ($p < .05$) and stars indicate significant differences of within tasks ($p < .05$).

Overall, we found a significant main effect of cue type, $F(1,24) = 78.6$, $p < .001$, $\eta_p^2 = .77$, where scenes were more efficient cues than words, leading to faster reaction times. More importantly, we found a significant main effect of feature type on reaction times, $F(1,24) = 11.55$, $p < .005$, $\eta_p^2 = .32$, indicating that independent of cue type, semantic object features were accessed faster than perceptual ones $F(1,24) = 1.02$, $p > .05$. Accuracies showed a similar pattern of results, with scene cues eliciting more correct responses than word cues, $F(1,24) = 27.8$, $p < .001$, $\eta_p^2 = .54$, and semantic features showing more correct responses than perceptual features, $F(1,24) = 11.99$, $p < .005$, $\eta_p^2 = .33$. Again no significant interaction between cue type and feature type was observed, $F(1,24) = 0.27$, $p > .05$.

Although these results support the hypothesis that the reverse memory reconstruction stream is cue invariant, even scenes carry some semantic information, and participants may therefore tend to form associations on the level of meaning. To test whether the semantic-over-perceptual advantage persists with meaningless cues, a slightly altered paradigm was adopted in Exp. 4a. In this version of the memory task, participants associated objects from different categories with one of eight locations along a white circle on the screen (see Fig. 2 D). The location cue served as a prompt to then retrieve and classify the associated object according to one of multiple dimensions and levels (exemplar, category, perceptual and

semantic; see Methods and Fig. 2 E). While all of these dimensions were tested in both the visual and the memory groups, for the analysis presented here, we focus on the perceptual dimension of color (coloured vs greyscale) and the semantic dimension of animacy (living vs non-living), in line with previous studies. Since this experiment used a new setup and stimulus pool, we also conducted a visual task in a separate group of participants, who simply classified the objects directly as presented on the screen, without the location circle, according to the feature prompted at the beginning of each trial (Fig. 2 D).

We observed the expected significant interaction between task and feature type, $F(1,46) = 30.78$, $p < .005$, $\eta_p^2 = .4$, due to a significant forward perceptual-semantic difference in the visual task, $t(23) = -6.04$, $p < .001$, $d = -1.23$, and a significant reverse difference in the memory task, $t(23) = 4.75$, $p < .001$, $d = 0.97$. Task, $F(1,46) = 288.02$, $p < .001$, $\eta_p^2 = .86$ and feature type, $F(1,46) = 44.27$, $p < .01$, $\eta_p^2 = .49$, showed significant main effects. All accuracy effects were non-significant (Fs < 0.177).

To corroborate this decision, we replicated this experiment again with a bigger sample size (n = 40) and an adjusted design that included only color as perceptual and animacy as semantic feature dimension.

This replication study (Exp. 4b) found the same significant interaction, $F(1,234) = 96.8$, $p < .001$, $\eta_p^2 = .29$, and main effects task, $F(1,234) = 472.55$, $p < .001$, $\eta_p^2 = .67$, and feature type, $F(1,234) = 69.71$, $p < .001$, $\eta_p^2 = .23$, with the same perceptual-over-semantic advantage in the visual group, $t(199) = -12.91$, $p < .001$, $d = -0.91$, and the reverse pattern in the memory group, $t(35) = 3.51$, $p < .005$, $d = 0.59$. These results were mirrored in the accuracy measure, with a significant interaction, $F(1,234) = 56.7$, $p < .001$, $\eta_p^2 = .20$, and main effects task, $F(1,234) = 38.0$, $p < .001$, $\eta_p^2 = .14$, and feature type, $F(1,234) = 58.2$, $p < .001$, $\eta_p^2 = .20$, and post-hoc tests showing higher accuracy for semantic than perceptual

features in the memory task, $t(35) = -3.75$, $p < .005$, $d = -0.63$, and a tendency toward the opposite pattern in the visual task, $t(199) = 1.93$, $p = .055$, $d = 0.14$.

Taken together, Exp. 3, 4a and 4b provide a clear indication that non-verbal cues, and even semantically meaningless cues like a position on a screen, lead to the same advantage of semantic over perceptual information retrieval. One could thus conclude that the feature processing hierarchy when a visual object is reconstructed from memory is hard-wired and highly robust to feature and cue variations.

2.2.3 Attention Variation Results – The reverse stream effect in memory is independent of attentional demands during learning

Exp. 3 and 4 showed that the semantic feature advantage during memory retrieval generalizes from verbal to visual scene cues, and even location cues. However, it could still be argued that humans have a natural tendency to form new associations on the level of an item's (in our case, an object's) meaning. In Exp. 5, we therefore investigated whether paying attention to meaning vs surface features of the objects during encoding affects the size of the feature processing gap during recall. Participants associated visual scenes with objects, performing either a shallow or deep encoding task. In the shallow task, they were asked to judge whether the object was a photograph of a line drawing. In the deep task, they were asked to classify each object as flying or non-flying. The memory test was performed similar to previous experiments, probing either perceptual (color vs greyscale) or semantic (natural vs manmade) features. No visual group was collected, since reaction times for these features were already available from previous experiments, demonstrating a clear forward stream.

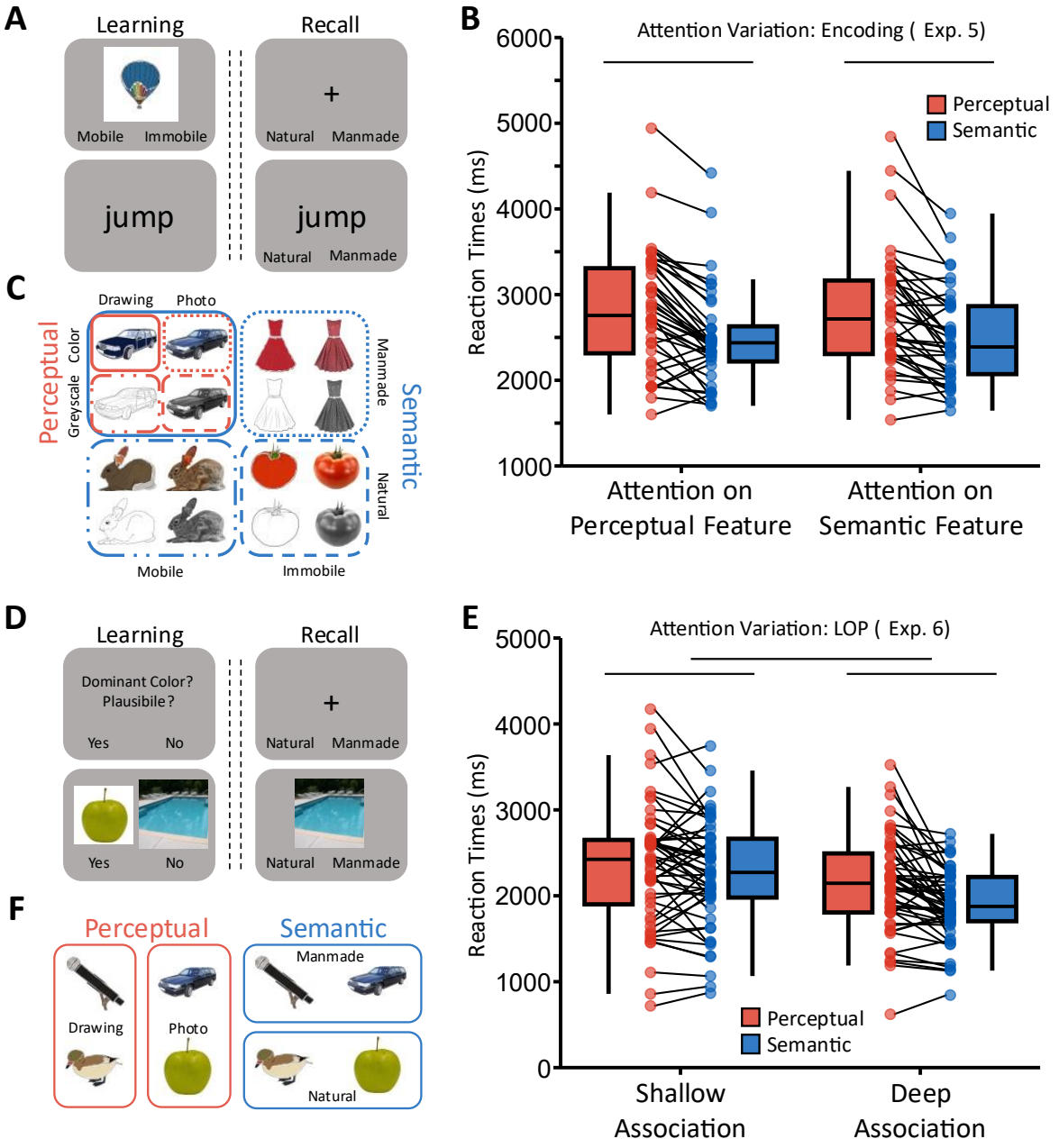

Fig. 3: Paradigm and Results of the Attention Variation Experiments. A: Exemplary depiction of the cued recall paradigm used in Exp. 5. B: Exemplary depicitions of the object images used in Exp. 5 with perceptual and semantic dimensions. C: Boxplots of the reaction time distribution of the encoding focus by feature type interaction in Exp. 5. D: Exemplary depiction of the cued recall paradigm used in Exp. 6. F: Exemplary depicitions of the object images used in Exp. 6 with perceptual and semantic dimensions. E: Boxplots of the reaction time distribution of the encoding focus by feature type interaction in Exp. 6. B&E: Dots indicate single participant‘s average reaction times. Bars indicate significant differences of main effects of encoding focus or feature type ($p < .05$).

We found a significant main effect of feature type, $F(1,37) = 53.82$, $p < .001$, $\eta_p^2 = .59$, replicating the reverse stream effect. Neither the main effect for the encoding focus, $F(1,37) = 0.35$, $p > .05$, nor the interaction, $F(1,37) = 0.0002$, $p > .05$, reached significance. Accuracies pointed in the same direction with only the feature comparison being highly significant, $F(1,37) = 167.9$, $p < .001$, $\eta_p^2 = .82$, (other Fs$(1,37) < 0.82$).

While a non-significant interaction supports our hypothesis that attention during encoding does not modulate the reverse stream effect, this first attention experiment failed to demonstrate a main effect of attention itself, which would be expected based on the large levels-of-processing literature (Craik & Lockhart, 1972). This absence might indicate that the attention manipulation at encoding was not successful. We reasoned that the deep-shallow manipulation may have affected processing of the object, but not the way in which the scene-object association was formed, the latter presumably influencing the way in which the association would later be retrieved.

In Exp. 6 we therefore manipulated the attentional focus with respect to the association that was formed between the visual scene cue and the object during learning. In the learning trials, participants were presented with pairs of scenes and objects, and were instructed to form an association either by focusing on common, shallow perceptual details (i.e. do you find the dominant color of the object in the scene) or on common, deep semantic features (i.e. Is it plausible for the object to appear in this scene). In the recall phase, scenes were then used as cues to recall the object, and we probed perceptual (color vs greyscale) and semantic (natural vs manmade) features on each trial, similar to the previous experiments.

Reaction time results replicated the reverse stream effect with a significant main effect of feature, $F(1,43) = 19.27$, $p < .001$, $\eta_p^2 = .31$. In this experiment, we also found a significant main effect of encoding focus, $F(1,43) = 62.39$, $p < .001$, $\eta_p^2 = .59$, showing that

as expected, deep associations during encoding led to overall faster object accessibility during retrieval than shallow associations. The two main effects were not qualified by an interaction effect, such that the semantic-perceptual gap during retrieval did not vary with the attentional focus during encoding, $F(1,43) = 0.89$, $p > .05$.

A significant interaction between attention and feature type was observed for the accuracies, however, $F(1,43) = 6.26$, $p < .05$, $\eta_p^2 = .13$, and while the shallow encoding condition showed a significant reverse stream effect, $t(43) = -8.4$, $p < .001$, $d = -1.27$, this effect (i.e., the semantic-perceptual gap) was increased in the deep encoding condition, $t(43) = -11.09$, $p < .001$, $d = -1.67$. Both encoding focus and feature type showed a significant main effect on accuracies (Encoding focus: $F(1,43) = 49.38$, $p > .001$, $\eta_p^2 = .53$; Features: $F(1,43) = 136.19$, $p < .001$, $\eta_p^2 = .76$).

2.2.4 A forward and a backward stream can be reliably shown over multiple datasets

In an effort to quantify the overall forward processing stream during visual processing and the reverse stream during memory recall, we combined all of the experiments that we reported above and previously published data using the same paradigm (Orig. Exp. 1 and 2 from Linde-Domingo et al., 2019) and modelled the perceptual to semantic feature reaction time difference with a Generalized Linear Mixed Model (GLMM). These models allow for single trial modelling of hierarchical data (in our case, reaction times from subjects nested in experiments) with non-gaussian error distributions and link functions that are more tailored towards reaction time distributions (Lo & Andrews, 2015). Visual and memory task data was modelled separately. Within experiment and participant, only objects that were correctly

classified on perceptual and semantic features were included, so that we could estimate the average difference between perceptual and semantic feature accessibility on an experiment level. This modelling approach is akin to a meta-analysis with individual participant data availability (Stewart et al., 2012).

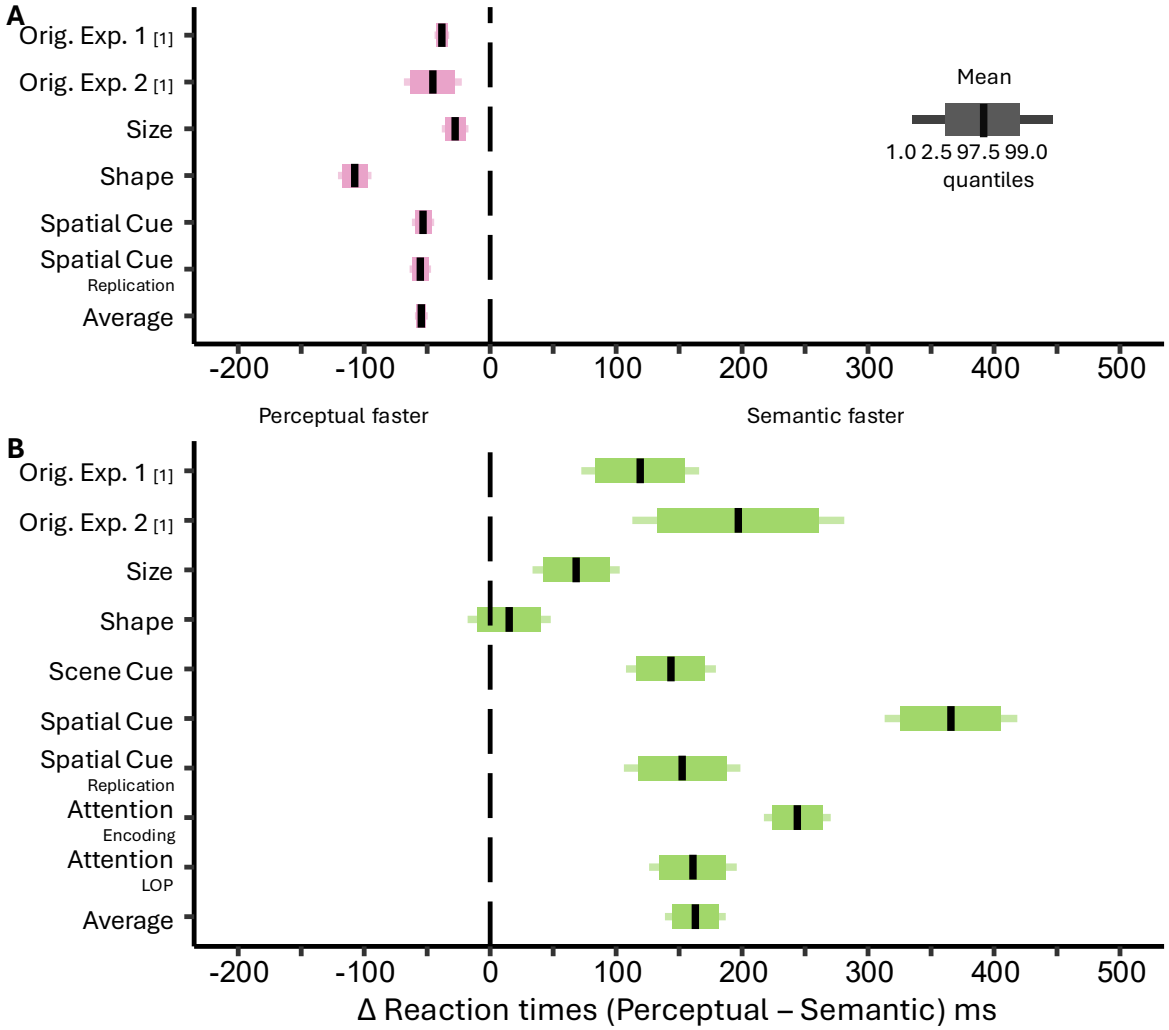

Fig. 4: GLMM Reaction Time Difference between Perceptual and Semantic Feature Type of all Visual and Memory Tasks. A: Difference estimates between perceptual and semantic features for all visual tasks and an average estimate over all tasks. B: Difference estimates between perceptual and semantic features for all memory tasks and an average estimate over all tasks. [1] from Linde-Domingo et al. (2019)

As can be seen in Fig. 4, all visual tasks reliably show a negative difference smaller than zero indicative of a forward processing stream (perceptual to semantic) with an average

of 60.81 (SE = 3.20) ms. Almost all memory tasks show a positive difference larger than zero indicating a reverse processing stream (semantic to perceptual) with an average of 180.68 (SE = 19.58) ms. As mentioned above, the only experiment that failed to show this positive difference was the shape variation (Exp. 2). As discussed in the respective results section, there is an obvious explanation for this absence, namely that shape is part of an object's semantics, which we did not consider when designing the experiment.

2.3 Discussion

The aim of this series of experiments was to establish feature-specific reaction times as a tool to probe the time course of feature reinstatement during memory retrieval. In multiple studies, we measured the accessibility of perceptual and semantic object features via reaction times, comparing the speed of feature access during visual classification with classification from memory. A meta-analysis-like GLMM model over the current and previously published results confirms that feature-specific reaction times reliably show a forward stream (perceptual before semantic) in visual classification of objects, and a reverse stream (semantic before perceptual) when the object is recalled from memory. In the new experiments presented here, we show that this reverse stream is present independent of the exact perceptual and semantic features used, as it generalises to novel feature dimensions, as long as the perceptual feature is not bound to the semantics of the item, as discussed further below. We also showed that this reversal of feature accessibility during retrieval is invariant to cues of varying levels of meaningfulness (i.e. words, scenes or even locations on screen) and that feature-specific reaction times delineate perceptual and semantic feature accessibility during retrieval even when attention during learning is shifted explicitly toward

one of those features. The results thus suggest that associative cued recall of a visual image progresses along a semantic-to-perceptual gradient. This gradient is robust to many manipulations and appears almost like a structural characteristic of the human memory system.

From a brain's perspective, the hippocampus is most closely connected to late stages of the visual processing hierarchy (Felleman & Van Essen, 1991; Suzuki & Amaral, 1994), and it thus makes sense that it would associatively bind and later access visual memories on the level of highly abstract, integrated representations. Interestingly, we failed to find the reverse stream effect when we varied shape as a perceptual feature. Since the shape of an object is inherent to its identity (e.g., a banana cannot be round, a turtle cannot be stick shaped), shape is clearly not a purely perceptual feature. In line with this (posthoc) interpretation, imaging work mapping the visual and conceptual features of objects revealed that visual features that are part of an object's semantic feature space are processed, and integrated with conceptual features, late in the ventral visual processing hierarchy (Martin et al., 2018; Martin & Barense, 2023).

Introducing screen size as a perceptual and naturalness as a semantic dimension in our first experiment corroborates findings from previous studies (Lifanov et al., 2021; Linde-Domingo et al., 2019). Object size is another feature that can or cannot be part of an object's semantics, depending on whether it is confounded with real-world size. In imaging work, the real-world size of an object was shown to activate different areas along a lateral to medial axis of the occipito-temporal cortex independent of their presentation size, while presentation size activates both early visual and higher visual areas (Troiani et al., 2014). This suggests that real-world size is an inherent part of an object's identity and presentation size is a variable perceptual feature. In our Exp. 1, objects were held constant with regards to their real-world

size and varied only in presentation size on screen. Compared to the semantic-type natural/manmade classification, this perceptual dimension produced a reliable forward stream in reaction times during visual processing, and a backward flip during memory recall. When confounded with real-world size, size would behave like a conceptual object feature, and we in fact found evidence for this in previous unpublished work (Linde-Domingo, 2019). Together with imaging work, the first two studies thus suggest that the type of information that is most readily accessible (by the hippocampus) during cued recall is core semantics of an episode's elements. Surface perceptual information that can randomly vary in real life is less readily accessible, possibly related to the fact that this information resides in brain areas further removed from the hippocampus (Suzuki & Amaral, 1994).

This conceptual over perceptual advantage is robust to cue variations, as shown in Exp. 3 and 4a/b. In all these experiments, the extent of the perceptual-conceptual gap did not vary with the meaningfulness of the cue used. Similarly, directing attention during learning towards perceptual or semantic features in Exp. 5 and 6 did not affect the reaction time signatures of the reverse memory stream, even though these manipulations have been long been known to influence what and how well we remember (Craik & Lockhart, 1972). Typically, deeper compared to shallow encoding or attention towards meaning or semantic features rather than to perceptual features lead to better memory performance (Baddeley, 1986; Loaiza et al., 2011). For example, in a recent study, comic images varying in artistic style of drawing (perceptual feature) and thematic content (semantic feature) were incidentally learned with a focus on either the perceptual or the semantic feature and then later tested in a recognition paradigm (Vijayarajah et al., 2023). Attention to thematic content over artist style improved memory performance and led to more detailed memories. It is all

the more surprising that biasing attention towards encoding the perceptual features did not decrease the semantic over perceptual feature advantage during retrieval.

Though not the focus of the present experiments, our experiment contrasting scene cues to word cues showed faster reaction times and higher memory performance for scene than word cues. There are a few possible explanations for this advantage. It has been argued that scenes, and spatial information more generally, play an integral role for episodic memory (Maguire & Mullally, 2013; Robin, 2018) by creating a scaffold that structures memory retrieval. Autobiographical memory research showed that spatial cues compared to event cues lead to faster and more episodically detailed memory retrieval (Sheldon & Chu, 2017). Robin & Olsen (2019) showed that scenes as cues lead to higher memory performance than objects or faces. In our studies, it is impossible to say whether the scene advantage during recall is due to faster perceptual processing of the scene than word cues, or due to scenes acting as more efficient retrieval cues. Peripheral vision from the retina is mostly relayed by magnocellular neurons that are known to be faster than the more focally coding parvocellular neurons (de Haan & Cowey, 2011; Livingstone & Hubel, 1987). It has been shown that scene processing follows a stream among lateral areas between the ventral and parietal stream (Kravitz et al., 2011) and intracranial studies in humans found that scenes information can reach a high-level visual area like the parahippocampal place area within 100 ms after presentation (Bastin et al., 2013) an area that has been implicated in scene recognition (Henriksson et al., 2015). It would therefore be likely that scenic information reaches the hippocampus and initiates a pattern completion process before language information like a meaning of a word. Interestingly, scene words as cues do not yield higher accuracies in recall compared to object words as cues (Horner & Burgess, 2013), which corroborates the notion that the effects found here are not driven by higher-order semantic processing.

In five out of the seven experiments presented here accuracies either showed no effect or mirrored the forward and reverse stream results found in reaction times. In the visual task of the size variation (Exp. 1), higher accuracies are consistently observed for semantic than for perceptual classification while reaction times still show the opposite pattern. Although a puzzling finding, it might be due to different speed-accuracy trade-offs where participants are willing to make more errors in the perceptual classification to be faster than in the semantic classification. Contrary to our reaction time results, accuracy shows a significant interaction between encoding focus and feature type in Exp. 6. Even though higher recall accuracy can be observed for semantic over perceptual features in both shallow and deep association, deep association seems to bolster the difference between perceptual and semantic features. So, while the attentional focus during encoding does not influence the temporal availability of semantic and perceptual information in the memory trace, a deep association strengthens the accuracy of semantic information more than detailed perceptual information. Further studies are warranted to figure out whether this relationship is synergistic (deep encoding enhancing both semantic and perceptual information accuracy) or competitive (deep encoding benefits semantic information storage at the expense of perceptual details).

Reaction time measures are notoriously underused in memory experiments. As our combined results using a meta-analytic approach show, reaction time measures of feature-specific memory decisions can elucidate the time course of the retrieval process. The results presented here show how reliably these tasks can map a forward stream of information processing during visual classification and a reverse stream of information processing during classification from memory. Memory research focused on the retrieval process mainly measures the accuracy of recognition and recall (Wixted, 2009; Yonelinas, 2002). A notable exception are results from recognition memory paradigms showing faster reaction times for

recollection (remember) decisions than for familiarity (know) decision (Dewhurst et al., 2006; Gimbel & Brewer, 2011; Rotello & Zeng, 2008). This pattern of results displayed the opposite direction of what some theoretical accounts argued for and is therefore an example of the fruitfulness reaction time measures hold for memory research.

From here, there are a few avenues for further enquiries. Since we have a whole host of reaction time data from many participants, it would be interesting to fit established reaction time models like Drift Diffusion Models (Ratcliff & Childers, 2015) or Linear Ballistic Accumulator Models (Brown & Heathcote, 2008) to it and check which parameters are mostly affected by the feature-specific differences. These models allow to independently estimate putative cognitive decision processes like the speed of evidence accumulation, speed-accuracy trade-offs, biases in response distributions and stimulus encoding times. Especially the Drift Diffusion Model (DDM) has been applied to memory processes like recognition and cued recall (Aschenbrenner et al., 2016; Ratcliff et al., 2011), showing that evidence accumulation for example could be an indicator of accessibility of information from memory. We would predict that our semantic-to-perceptual effect maps onto the evidence accumulation parameter of DDMs with higher drift rates for semantic than perceptual information accumulation.

Similar to the vast knowledge that exists on the hierarchical structure of object recognition that goes from low-level to mid-level features to high-level visual and finally semantic information (Groen et al., 2017) it is still a question whether a typical retrieval process follows a full reverse stream or a partial reverse processing stream that starts with a semantic gist and then adds low- and mid-level features only when necessary. Our studies were focused on a binary distinction between low-level perceptual and high-level semantic

features, but the inclusion of mid-level feature decisions would be a possibility, maybe enabling testing hierarchical processing models (Balaban et al., 2020).

Nonetheless, the studies presented here generalize our understanding of the temporal accessibility of visual features during retrieval and thereby makes a case for the utility of feature-specific reaction time measures to track the state of LTM traces.

2.4 Methods

2.4.1 General

Participants & Procedure: Experiments 2, 3 as well as the memory tasks of experiment 4a and 4b were administered in person. Participants were recruited the University of Birmingham. Participants were led into a room and seated in front of a computer. They were then presented with instructions followed by a shortened version of the experiment as practice and then completed the experiment. After a debriefing about the purpose of the experiment, participants were dismissed. For experiments 2 and 3 the tasks were programmed and administered with PsychToolbox (Brainard, 1997). For experiment 4a and 4b, tasks were programmed and administered with PsychoPy (Peirce et al., 2019).

Experiments 5 and 6 as well as the memory task for Experiments 1 were administered online. Participants were recruited from the School of Psychology & Neuroscience's Subject Pool at the University of Glasgow and received a personalized link that started the experiment. They were then first presented with instructions, followed by a shortened version of the experiment as practice and then completed the experiment. They were debriefed at the end and contacted again by the experimenter for their compensation. All online experiments

were programmed in PsychoPy/PsychoJS (Peirce et al., 2019) and were hosted on Pavlovia (Open Science Tools, Nottingham, UK).

The visual tasks of experiments 1, 4a and 4b were administered online to external participant samples recruited over Prolific (www.prolific.co). The procedure remained the same as for the other online studies mentioned above.

2.4.2 Experiment 1: Feature Variation (Size)

Participants: For the memory task a total of 32 participants (21 female, 11 male; M_{Age} = 25.9, SD_{Age} = 5.2) were recruited and received course credit or a payment of £10. For the visual task 40 participants (29 female, 11 male; M_{Age} = 30.7, SD_{Age} = 5.9) were recruited and received a payment of £4.

Material & Design: The stimulus pool for this task consisted of 128 colored images depicting everyday objects cut-out in front of a white background. All objects were sampled from 8 distinct categories with 16 objects per category. Half of those categories were manmade (bathroom, kitchen, office, tools) and the other half were natural (fruits, land animals, sea and air animals, vegetables). Since the real-world size can vary substantially and has been shown to influence the visual processing of objects (Konkle & Oliva, 2012), all objects were chosen to minimize variation in real-world size by a heuristic criterion (i.e., real-world size had to be between a human thumb and a medium-size dog). Half of the object images were displayed as big and half as small on screen (.6 and .3 in height units, respectively). For an example of the stimulus pool see Fig. 1 B.

Task: The memory task was repeated over 16 blocks, with each block consisting of a learning phase, a short distractor task and a recall phase (see Fig 1 A). In the learning phase,

each trial started with a jittered fixation cross (0.5 – 1.5 s), followed by an action verb as a cue (1 s). Afterwards a second jittered fixation cross (0.5 – 1.5 s) appeared before the object was presented. Participants were instructed to form an association between the action word and the image, and to indicate when they had successfully formed the association with a button press. Either the button press, or a maximum duration of 10 s ended the trial. Each pair was presented and learned once, with a total 8 pairs of verb-object associations learned per block. Following the learning phase, participants engaged in a short distractor task where they categorized odd or even numbers using the left or right arrow button, respectively. This task lasted for 30 seconds and participants were instructed to classify as many numbers as accurately as possible. Feedback of performance (i.e., number of trials and percentage of correct classification) was given after the task. In the recall phase, each trial started with a jittered fixation cross (0.5 – 1.5 s) that was followed by two response options at the bottom of the screen that lasted for 2 s. These response options were either asking for perceptual features (i.e., 'big' vs 'small') or semantic features (i.e. 'manmade' vs 'natural'). Which option was presented on the left or the right was counterbalanced across subjects. Then participants were cued with the verb (0.5 s) and asked to recall the object presented with this verb and classify the object using the left or right arrow key. This classification period lasted until button press or a maximum of 10 s. The ITI was 0.5 s. Each object was recalled once with a perceptual and once with a semantic feature probe, with the order being counterbalanced across objects, resulting in 16 recall trials per block. The presentation side of the response options was counterbalanced across subjects.

In the visual task, each trial started with a jittered fixation cross (0.5 – 1.5 s) followed by the two response options (asking for perceptual or semantic features) that appeared at the bottom of the screen with one option on the left and one on the right. After 2 seconds, the

object appeared, and participants had to categorize it according to the options as fast and accurately as possible (see Fig 1 A). In a block, 32 objects were categorized twice, once with a perceptual and once with a semantic feature probe (with the order being counterbalanced across objects), resulting in 64 trials per block. The task was repeated over 8 blocks.

2.4.3 Experiment 2 – Feature Variation (Shape)

Participants: 25 participants (17 female, 8 male; M_{Age} = 20.6, SD_{Age} = 2.4) were recruited for a memory task and 24 participants (23 female, 1 male; M_{Age} = 18.7, SD_{Age} = 0.6) were recruited for the visual task. All participants received either course credit or a payment of £6/hour upon finishing their task.

Material & Design: A stimulus pool of 128 everyday objects was created taken from the BOSS database (Brodeur et al., 2014) in such a way that half of the objects were natural, and half were manmade objects. Within these distinct semantic groups objects were chosen such that half of them had a round shape and half had a stick shape, thereby orthogonalizing the semantic dimension of naturalness and the perceptual dimension of shape (see Fig 1 E).

Task: Both the memory and visual task were identical to “Size variation” except for the following changes: the response options were displayed for 3 seconds before a cue (memory task) or object (visual task) appeared; and the perceptual feature response options were “rounded” and “elongated” instead of “small” and “big” (see Fig 1 D).

2.4.4 Experiment 3 – Cue Variation (Scenes)

Participants: This experiment only had a memory group, for which 27 participants (21 female, 6 male; M_{Age} = 20.0, SD_{Age} = 2.2) were recruited and received course credit.

Material & Design: Materials consisted of 128 everyday objects with 64 being animate (birds, insects, mammals, marine animals) and 64 inanimate (fruits, vegetables, electronics, clothes). Each object existed once as a photograph and once as a line drawing, resulting in a stimulus pool of 256 images (for more detailed information see Linde-Domingo et al., 2019; see Fig. 2 B). For each participant, 128 images were selected by randomly choosing 8 photographs and 8 line drawings from each category. As cue material a total of 128 scene images were collected from the SUN Database (Xiao et al., 2010) and 128 action verbs were taken from Linde-Domingo et al. (2019). Out of the scene images, 64 were indoor (depicting homes, shopping, gyms, workplaces) and 64 outdoor (depicting industry, fields, mountains, urban settings). 128 Triplets were formed out of one object, one scene and one action verb (see Fig. 2 A). While the scene and the verb were always presented on the top (which of them presented on the left and the right was counterbalanced between triplets), the object was always presented on the bottom of the screen in the middle. Triplets were balanced such that each object category was paired with each scene category twice.

Task: Since the object pool and the perceptual/semantic dimensions were the exact same ones used in Linde-Domingo et al. (2019), visual classification data was already available, and we only conducted a memory group in this experiment. In the learning phase, each trial started with a jittered fixation cross (0.5 – 1.5 s), followed by a triplet of an action verb, a scene, and an object. Participants had to form a story or mental image including all elements of the triplet and indicate with a button press when they had successfully formed such an association. Maximal presentation duration of the triplet was 10 s. The distractor task and the retrieval task were identical to “Size variation” with two exceptions. First, each object

was cued once with the scene and once with the word (order of cue type was counterbalanced across objects) and the object was classified both times either on the perceptual or on the semantic dimension. Second, the response options were displayed for 3 seconds (see Fig. 2 A).

2.4.5 Experiment 4a – Cue Variation (Spatial)

Participants: 24 participants (20 female, 4 male; M_{Age} = 18.9, SD_{Age} = 0.7) were recruited for the memory task and 24 participants (all female; M_{Age} = 19.6, SD_{Age} = 0.8) for the visual task. They either received a course credit or a payment of £10.

Material & Design: The stimulus pool was considerably different from the previous experiments and consisted of 16 object images with 4 exemplars from each of 4 different categories (dogs, birds, vehicles and aircrafts), orthogonalizing the two semantic dimensions of animacy (living or non-living) and aeromobility (flying or non-flying). Each image was adapted 4 times to create the perceptual dimensions of visual detail (photo or drawing) and color (colored or greyscale), resulting in the 4 possible feature combinations: images were depicted either as photographs or as line drawings and these were either colored or greyscaled (see Fig. 2 E). In addition to the two perceptual and two semantic dimensions, in this experiment we also probed participants' memory on an exemplar level. For the exemplar dimension, the response options were the correct option (e.g. 'chihuahua') and a lure option for the same category (e.g. 'labrador').

Task: For all tasks the background screen was grey with a white circle at the center of the screen. The ITI varied randomly between 0.5 and 1.5 seconds for both the memory and

the visual tasks. Both tasks were repeated over 2 blocks. Participants were able to take self-paced breaks after every 16 trials within a block and between blocks.

The memory task started with a familiarization phase to get participants acquainted to the objects' categorizations. At the start of each trial in this familiarisation phase, a fixation cross appeared in the middle of the screen and two response options appeared at the bottom of the screen. After 2 seconds, the object appeared in the middle of the circle until participant's response. After the response, the object was replaced by a fixation cross for 1 second. During this time the correct name was highlighted in green to give participants feedback on their performance. Each object was tested once on each of the 5 dimension (40 trials per block). In the learning phase, participants had to associate each object with one of eight points spaced equidistantly on the circumference of the white circle. Each location was uniquely associated with one object. Each trial started with a fixation cross, and one of the eight points being enlarged for 1 to 2 seconds followed by the object appearing in the middle of the screen. The object disappeared, starting the next trial, when participants pressed a button to indicate that they had formed an association. Learning trials were occasionally interspersed with test trials, where the object appeared on screen and participants indicated whether they could remember the location on the circle or not. After the button press, they were presented with the eight dots and had to navigate an enlarged grey dot (appearing at a random location) to the associated location with clockwise and/or counterclockwise movements. If successful, the correct location was highlighted in green for 1 second to give participants feedback (see Fig. 2 D). Each object-location association was learned 10 times and tested 3 times (104 trials per block). In the retrieval phase, each trial started with a fixation cross (2 s) and the response options from one dimension on the bottom of the screen with an additional third 'forgotten' option underneath. Next, the eight dots appeared with one

of them enlarged indicating which object participants had to remember and classify as fast and as accurately as possible. They then pressed one of the two choice buttons or the "forgotten" button, upon which the next trial started.

The visual task was identical to the familiarization phase of the memory task without the feedback screen (see Fig. 2 D). Each object was classified 5 times on each of the 5 dimensions, resulting in a total of 200 trials per block.

2.4.6 Experiment 4b - Cue Variation (Spatial Replication)

Participants: For the memory task a total of 38 participants (27 female, 8 male, 3 no information; M_{Age} = 23.4, SD_{Age} = 8.6) were recruited and received course credit or a payment of £10. For the visual task 200 participants (130 female, 70 male; M_{Age} = 33.7, SD_{Age} = 8.5) were recruited and received a payment of £3.49.

Material & Design: The same stimulus pool as in Experiment 4 was used with the exception that all line drawings were omitted. Objects varied on 4 dimensions in this version of the task: exemplar (e.g. "Chihuahua" vs "Labrador"), semantic sub-category (birds vs dogs, vehicles vs aircrafts), perceptual (color vs greyscale) and semantic supra-category (living vs non-living). For the exemplar level, lures were chosen with equal probability from all four sub-categories, and for the sub-category dimension, lures were chosen with equal probability from both supra-categorical semantic dimensions. Note that for the purpose of the present study, to compare with the previous experiments, only the perceptual and supra-categorical semantic probe trials were analysed.

Task: Both memory and visual tasks were identical to Exp. 4a with the following exceptions.

In the familiarization phase, each object was probed once on all 4 feature dimensions (32 trials per block). In the learning phase, each association was learned 8 times and tested twice (80 trials per block). In the retrieval phase, each association was tested twice on all 4 feature dimensions (64 trials per block).

In the visual task, each object was probed 4 times on all 4 dimensions (128 trials).

2.4.7 Experiment 5 – Attention Variation (Feature Focus)

Participants: 40 participants (30 female, 8 male, 2 no information; M_{Age} = 23.8, SD_{Age} = 3.6) were recruited and received either course credit or a payment of £7.

Material & Design: We created a new database of 96 everyday objects from 8 different categories (12 exemplars per category) orthogonalizing the semantic dimensions of mobility and naturalness (mobile-natural: mammals and birds; mobile-manmade: air and ground vehicles; immobile-natural: fruits and vegetables; immobile-manmade: clothes and electronics). Each image was adapted 4 times, creating the two perceptual dimensions of visual detail and color (photograph-color: the original image; photograph-greyscale: greyscale transformation of the image; line drawing-color: line drawing of the image with added color patches matching the image's original colors; line drawing-greyscale: line drawing of the image with color). This two-by-two variation of stimulus features allowed us to use one perceptual (photo vs drawing) and one semantic (mobile vs immobile) dimension as an encoding manipulation, varying what features participants attend to during learning. The other two dimensions were then used to measure feature-specific reaction times at retrieval, probing the remaining perceptual (color vs greyscale) and semantic (natural vs manmade) dimensions (see Fig. 3 A).

Task: During the learning phase, each trial started with a jittered fixation cross (0.5 – 1.5 s), followed by response options for 1.5 seconds. Then the object appeared and had to be classified according to the options. Objects were categorized either according to their mobility (semantic focus) or according to their visual detail (perceptual focus). After the button press or a maximum of 10 seconds, a second fixation cross appeared, followed by the action verb. Participants had to associate the verb and the object presented beforehand and indicate when they had formed an association with a button press that concluded the trial (maximum duration 10 s). The distractor task and the retrieval task were identical to "Size variation" except that the response options were displayed for 1.5 seconds. As mentioned above, to keep the retrieval task similar to previous experiments and to not probe the same features during encoding and retrieval, the perceptual dimension used for the retrieval task was visual detail and the semantic dimension was naturalness.

2.4.8 Experiment 6 – Attention Variation (Association Focus)

Participants: Fifty-two participants (41 female, 9 male, 2 non-binary; M_{Age} = 19.7, SD_{Age} = 3.8) were recruited and received course credit.

Material & Design: For this study, object stimuli were taken from to Experiment 5 and scene stimuli were taken from Experiment 3. For the perceptual dimension only colored photographs and drawings were used, removing the greyscale photographs and drawings.

Task: Like in Experiment 6, we intended to shift participants' focus during learning to either surface perceptual features or meaningful semantic features. This was done in a blocked fashion, with each block containing a sequence of 8 scene-object pairs. Before the start of a block, participants were instructed to compare the object and the scene according

to a dominant color match (perceptual focus) or plausibility (semantic focus). During the learning phase, the response options ('yes' and 'no') and the association instruction ('plausible' or 'dominant color') stayed on the lower and upper part of the screen through the whole phase, respectively. Each trial started with a jittered fixation cross (0.5 – 1.5 s), followed by an object on the left and a scene on the right (see Fig. 3 D). Participants gave their response with a button press (maximum duration 10 s). The distractor task and the retrieval task were identical to "Size variation" except that the response options were displayed for 1.5 seconds, before the verb was then presented as a cue to recall and classify the object according to either the perceptual (photo vs drawing) or semantic (natural vs manmade) dimension.

2.4.9 Reaction Times and Accuracies

Reaction times were trimmed with a lower cutoff at 150 ms for all visual and 500 ms for all memory tasks. The upper cut-off was defined on a participant specific basis as each participant's overall mean plus three standard deviations. Reaction time analysis only include correct responses. For the accuracy analysis, mean correct responses as a percentage are used.

2.4.10 ANOVA Analysis

For experiments 1, 2, 4a and 4b, reaction times and accuracies were analyzed using a 2 (Task: Visual vs Memory) by 2 (Feature Type: Perceptual vs Semantic) mixed ANOVA. For experiment 3, reaction times and accuracies were analyzed using a 2 (Cue Type: Word vs Scene) by 2 (Feature Type: Perceptual vs Semantic) repeated-measures ANOVA. For

experiment 5 and 6, reaction times and accuracies were analyzed using a 2 (Association Focus: Dominant Color vs Plausibility) by 2 (Feature: Perceptual vs Semantic) repeated-measures ANOVA. Effect sizes for all significant main effects and interactions are reported as partial eta squared.

2.4.11 GLMM Analysis

To quantify the forward and backstream effects over experiments, we combined the data from the experiments described above as well as two experiments described in Linde-Domingo et al. (2019) in a generalized linear mixed model approach that is akin to a meta-analysis (Stewart et al., 2012). Reaction times from correct trials were modelled as a function of Experiment and Feature as fixed effects, with treatment contrasts defined on both factors. To model the hierarchical dependencies, participants were nested in experiment varied as random intercepts (random slopes were not included for model stability). The error was modelled as a gamma distribution with an inverse link function (Lo & Andrews, 2015).

Chapter 3: Modelling the perceptual and cognitive components of the forward and reverse stream

3.1 Introduction

In the previous chapter, it was shown that reaction time measures are sensitive to the accessibility of feature information during retrieval and might thereby be an efficient possibility to tap into the temporal dynamics of a memory trace reactivation. A whole host of behavioral studies was presented that conclusively showed an information processing stream in a forward direction (from perceptual details to semantic category) when participants classified objects directly in a visual task and a reversal of information processing (semantic category before perceptual details) when participants had to retrieve object information from memory (memory task). Our reaction time measures in these experiments were analyzed according to common techniques for the field with ANOVAs for condition comparisons based on participant's average reaction times or a Generalized Linear Mixed Model (GLMM) based on single trial reaction times that allow for specific model error distributions. With an abundance of data, it would be interesting to see, whether more elaborate cognitive models of reaction time distributions can give insight into how the observed reaction time differences between conditions arise.

A popular sequential sampling model for reaction times has been the Drift Diffusion Model (DDM; Ratcliff, 1979; Ratcliff et al., 2016). Here, a decision is modeled as a random walk with a constant drift rate through a decision space enclosed by two decision boundaries. This process begins after an initial stimulus encoding time at a starting point between the two boundaries and ends when it finally hits the upper or lower boundary indicating a decision

has been made. Afterwards a motor output is generated. The initial stimulus encoding and motor output times are combined into a nondecision time parameter. The drift rate indicates how fast information is accumulated or integrated to reach a decision and varies with quality of a presented stimulus and has been shown to correlate with domain-general performance measures such as IQ and working memory capacity (McKoon & Ratcliff, 2012; Schmiedek et al., 2007). The width of the decision boundary on the other hand influences how fast a decision is made independent of how well the person performs and can therefore be seen as a speed accuracy trade off parameter (Lerche & Voss, 2017, 2019; Nunez et al., 2024; Voss et al., 2004). The starting point parameter models prior knowledge about the decision options (e.g. knowing that a red circle is twice as likely to appear on screen as a green triangle).

An example for the usefulness and explanatory power of the DDM can be taken from von Krause et al. (2022) who analyzed the reaction time distributions from 1.2 million people with an age range from 20 to 60 years. Typically, they observed age-related slowing in mean response times which in the past has been taken as an indication for cognitive decline in older participants. But when fitting a DDM parameters to different cohorts they found that while mental processing speed (drift rate) did not vary significantly by age and the decision boundary was wider for old compared to young people. As has been observed before the age-related slowing is not due to cognitive decline but to a more cautious response criterion in older compared to younger participants (i.e. younger participants are more willing to trade accuracy for speed).

Although DDMs are usually fit to reaction time distributions from perceptual decision-making tasks, the original studies for the DDM focused on recognition memory (Ratcliff, 1979), this model has been also applied to recognition- and recall-related decision-making (Arnold et al., 2015; Aschenbrenner et al., 2016; McKoon & Ratcliff, 2012; Ratcliff

et al., 2004). As can be seen from the previous chapter, reaction times from memory tend to be longer with the bulk of reaction times between 2 and 4 seconds. Although our reaction time count per condition is above 100 trials, usually models are fit to reaction time data with 200 to 500 trials per condition. It has been shown though that even the full DDM models can be reliably estimated for longer reaction times and lower trial count (Lerche et al., 2017; Lerche & Voss, 2019). Comparisons between different estimation techniques showed that a hierarchical Bayesian approach showed the best performance for data with lower trial count (Lerche et al., 2017).

To investigate the feature-specific accessibility differences in reaction times that we found in the previous chapter, hierarchical Bayesian Drift Diffusion Models are fit to both a memory dataset (Memory Model) and a visual dataset (Visual Model). These experiments included the two original behavioral experiments from Linde-Domingo et al. (2019) that first showed a reverse stream of information flow during retrieval with faster reaction times for semantic compared to perceptual feature classification. The other experiments included in model investigate the effect of different perceptual and semantic object features (Shape and Size experiments), different cues associated with the objects (Scene Cue, Spatial Cue and Spatial Cue Replication experiments), and different attentional demands during the learning phase (Attention Encoding and Attention LOP experiments). A detailed overview over the different memory and visual experiments can be found in Tab. 1. Experiment- and condition-specific effects of a 3-parameter model (i.e. decision boundary, drift rate and nondecision time) will be estimated as hierarchical priors. In this way we can check whether a certain parameter can reliably describe the differences between perceptual and semantic feature classification that we find in the reaction time averages.

Based on prior studies of memory recognition and cued recall (Arnold et al., 2015; Aschenbrenner et al., 2016; Ratcliff & McKoon, 2008; Spaniol et al., 2006), we would expect to find consistent drift rate differences between both conditions in all memory experiments (except for the Shape experiment, since this experiment failed to find the reverse stream effect, cf. Chapter 2) with higher drift rates for semantic than for perceptual feature classification. We don't expect to find consistent nondecision time and decision boundary differences (again maybe with exception of the Shape study). For the visual model, we would expect to find a consistent opposite effect in drift rates, with higher drift rates for perceptual than for semantic feature classification. Consistent differences between decision boundaries and nondecision times are not expected.

3.2 Results

3.2.1 Posterior Sampling and Parameter Estimation

Single trial reaction time and accuracy data were modelled using a wiener first passage of time likelihood with a decision boundary alpha, a drift rate delta, a nondecision component tau and a starting point bias beta on a single trial level. These single trial level parameters were modelled by a hierarchical prior structure of participant-, experiment-, and condition-specific effects (see Fig. 5 A&B). After a warm-up period, for each model we obtained 7500 samples across 5 chains from the posterior distribution with STAN's No-U-Turn-Sampler.

As can be seen in Fig. 5 C, when we compare reaction time quantiles between our observed data (i.e. the memory and visual reaction times we used to fit our model) and predicted data simulated from the posterior parameter estimates of our model, a general overlap between both data distributions indicates that the hierarchical Bayesian DDM Memory and Visual Models capture the data fairly well.

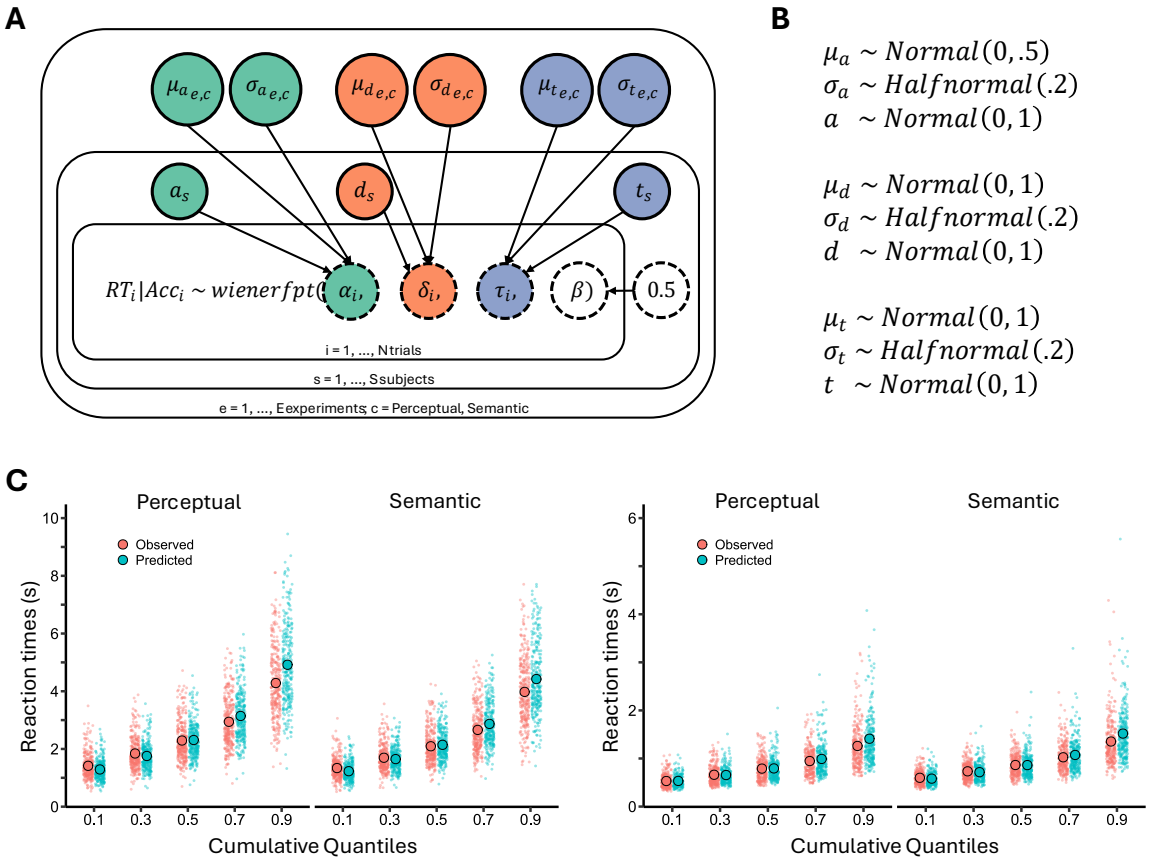

Fig. 5: **A**: Graphical illustration of the hierarchical prior structure used in this study. Parameters, priors and hyper-priors belonging to each of the three DDM parameters are color-coded. Solid line circles represent random variables, dashed lined circles represent deterministic variables. **B**: Overview over prior distributions with parameter values. Normal distributions are parametrized with mean and standard deviation, half-normal distributions with just a standard deviation and a mean of 0. **C**: A comparisons between observed reaction time data and posterior predictive distributions of the model. For both observed and predicted data, small dots are a participant 's average reaction time within quantiles of the reaction time distribution indicated by the x axis. Big dots represent the mean over all participant 's averages. Comparison plots are based on observed data and model data from all memory tasks (left) or all visual tasks (right).

3.2.2 Memory Model Condition Differences

When comparing the three DDM parameters between conditions from a model fit on our memory data, a consistent distinction is only found for the drift rate. Over almost all experiments a significantly higher drift rate for semantic than for perceptual classification from memory is observed (prob($\delta < 0$) = 0 for Exp.1, Exp.2, Size, Scene Cue, Spatial Cue Replication, Attention Encoding, Attention LOP). The only two exceptions from this pattern are the Shape and the Spatial Cue experiment with no significant differences in drift rate. Interestingly, for the Spatial Cue experiment a significant difference in the decision boundary is observed, prob($\alpha > 0$) = 0.005, indicating a higher decision threshold for perceptual than for semantic feature classification. A significant difference in decision boundaries was also

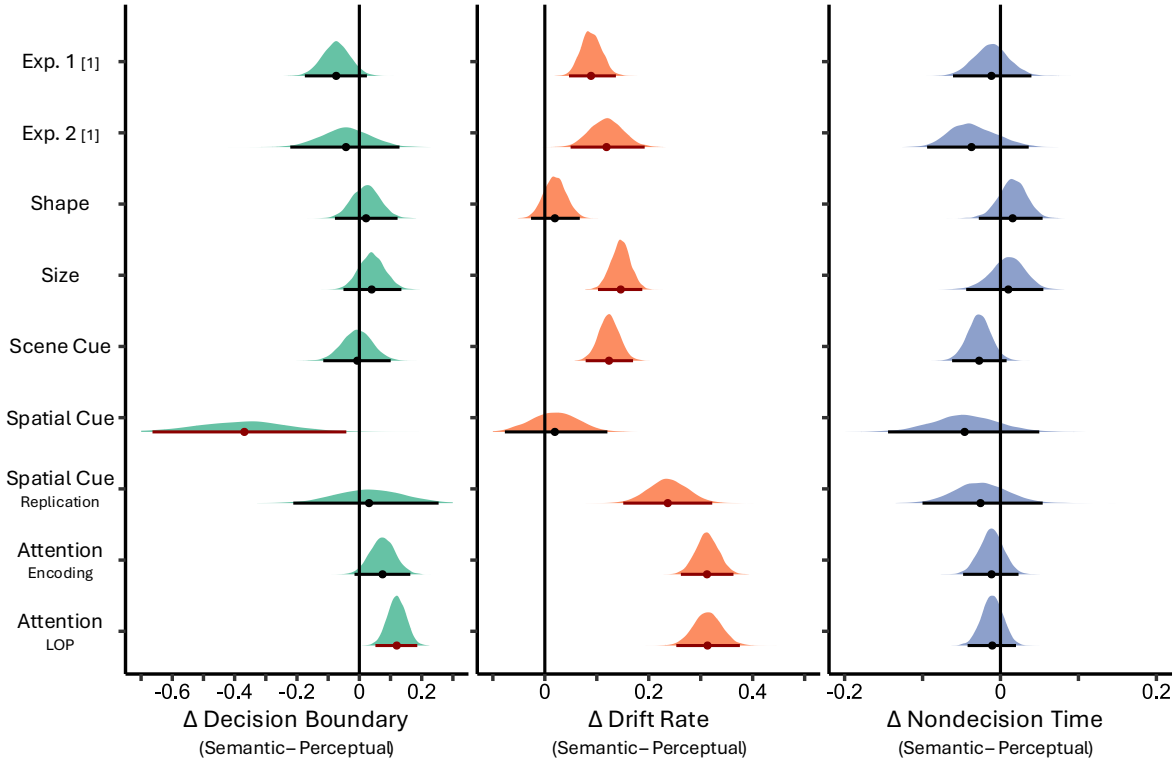

Fig. 6: DDM parameter results for all memory experiments. Parameters are displayed as differences between semantic and perceptual conditions. Left: decision boundary. Middle: drift rate. Right: nondecision time. Points indicate the mean of a parameter estimate distribution, lines indicate the 97.5 percent most likely parameter estimates. Lines are plotted as dark red if the value 0 falls outside of 97.5 of the most likely parameter values. [1] Linde-Domingo et al. (2019)

found for the second Attention Experiment, prob($\alpha < 0$) = 0, but in the opposite direction with a higher decision boundary for semantic than perceptual classification. Over all experiments, no significant differences between conditions found for the nondecision time parameter (see Fig. 6). As predicted, consistent differences in parameters between conditions were only observed for the drift rate but not for the other two parameters.

3.2.3 Visual Model Condition Differences

Other than in the memory experiments and contrary to our hypotheses, in the visual experiments a consistent significant condition difference was found for the nondecision time parameter (prob($\tau < 0$) < 0.02 for all experiments), indicating larger nondecision times for

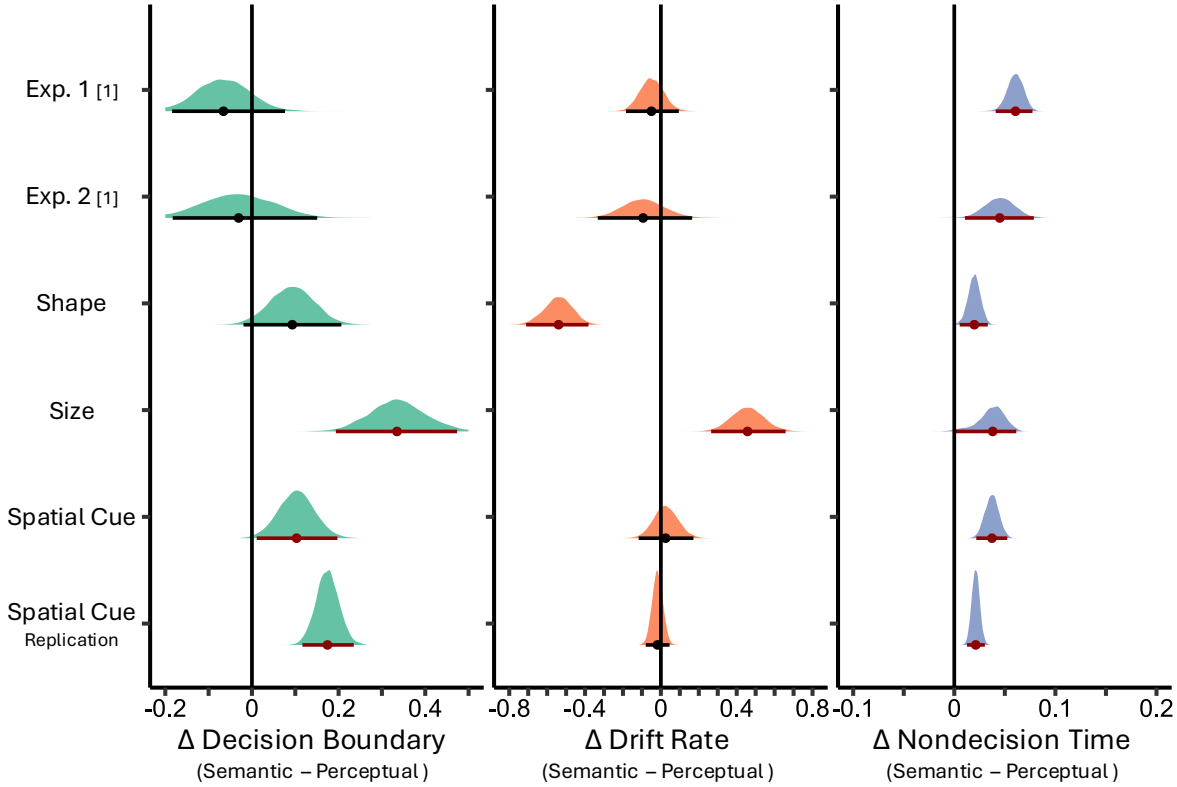

Fig. 7: DDM parameter results for all visual experiments. Parameters are displayed as differences between semantic and perceptual conditions. Left: decision boundary. Middle: drift rate. Right: nondecision time. Points indicate the mean of a parameter estimate distribution, lines indicate the 97.5 percent most likely parameter estimates. Lines are plotted as dark red if the value 0 falls outside of 97.5 of the most likely parameter values. [1] Linde-Domingo et al. (2019).

semantic than for perceptual feature classification. For the decision boundary significant differences were observed for the Size, prob($\alpha < 0$) = 0, the Spatial Cue, prob($\alpha < 0$) = 0.005, and Spatial Cue Replication experiments, prob($\alpha < 0$) = 0, all showing higher decision boundaries when semantic than when perceptual features were classified directly. Additionally, the Shape experiment showed a significantly higher drift rate for perceptual than semantic feature classification, prob($\delta > 0$) = 0, while the Size experiment showed a significant drift rate difference in the opposite direction, prob($\delta < 0$) = 0 (see Fig. 7).

3.3 Discussion

In this study, we applied a well-established mathematical cognitive model to an enormous data set of memory and visual classification studies. By estimating underlying cognitive processes, we aimed to explain the reaction time differences between perceptual to semantic feature classification that switch dependent on whether these classifications are done visually or from memory.

As predicted, for our Memory Model significant differences between the drift rate were found consistently over most experiments. These differences also had the predicted direction with higher drift rates for semantic compared to perceptual feature classification. These results corroborate the claim that semantic feature information is more easily accessible than perceptual features information during the retrieval process, as the drift rate is usually interpreted as a marker for availability of external or internal information (Ratcliff, 1979). Interestingly, a significant difference was not found in the Shape and the Spatial Cue experiment. The former result was expected since in this experiment did not show a significant difference between perceptual and semantic feature classification. As discussed

before, this might be due to an confound between perceptual variation and the semantic object identity (e.g. being displayed as a line drawing or a photograph is not inherent to the identity of a turtle, but their round shape is). Instead of a significant difference in drift rates, we found a significant difference in decision boundaries that explained our reaction time effect in the Spatial Cue experiment. This unexpected result might be due to the nature of the paradigm that was used in this particular experiment, as the task was so difficult (recalling objects from spatial position cues) that participants were enabled to take very long recall reaction times (up to 36 seconds). For our analysis we restricted range of reaction time values from 0.5 to 10 seconds (in line with all other experiments), but it is possible that the non-speeded nature of the task changed participants response criterion to trade reaction speed for accuracy.

For the Visual Model, against our prediction a consistent difference was found on the nondecision time parameter and not, as we assumed, on the drift rate, with every experiment showing higher nondecision times semantic than perceptual feature classification. On a second glance, this pattern might not be totally unexpected, as the nondecision time parameter is partly made up of a stimulus encoding period that has to pass before the decision begins. It could be that this parameter indexes the forward stream of information processing that we hypothesize to map with our reaction time paradigm. If the motor response remains equal for both semantic and perceptual classification, the stimulus encoding time should be different, because early perceptual features should be processed earlier and be available more quickly that high level semantic information. By this logic, these results would indicate a "hard-wiredness" of the visual processing stream that can be mapped by reaction time differences. This interpretation comes with the caveat that nondecision times as mentioned above do not only include stimulus encoding times but also motor response times that could

vary between conditions and a valid estimation of the nondecision time component might be reduced in a simplified DDM (Ratcliff, 2008).

The decision boundary difference for our second Attention experiment curiously showing lower decision boundaries for perceptual than for semantic feature classification might indicate a dissociation between the reverse retrieval stream and a levels-of-processing based attention effect.

In DDMs the full distribution of correct and error responses is taken into account in one model where the estimated parameter values correspond to separable patterns of condition difference in reaction times and accuracy. For example, while faster reaction times and higher response accuracy in one condition compared to another should result in different drift rates between those conditions, faster reaction times and lower accuracies in one condition (indicating a speed-accuracy trade-off) should result in different decision boundaries for those conditions. Since almost all memory task showed a pattern of faster reaction times and higher accuracy for semantic features compared to perceptual features. Generally, the DDM has the ability to uncover latent effects in the data that might not be apparent in mean reaction time and accuracies alone (Lerche & Voss, 2020).

It is difficult to fully disentangle the influence that different accuracy distributions between perceptual and semantic feature retrieval could have on the drift parameter findings, although in most experiments average accuracy comparisons between conditions mirror those of the average reaction time differences (i.e., higher accuracies for semantic than for perceptual feature retrieval). Interestingly, in the experiments where this pattern breaks, we observe either no condition difference in drift parameters like in the Spatial Cue experiment or there are both a drift parameter as well as decision boundary differences in the expected direction like in the Attention LOP experiment. It must be kept

in mind though that high accuracies in a condition (few error trials) make it more difficult to reliably estimate the parameters of a DDM including the drift rate (Lüken et al., 2025).

A necessary step for validating the parameter of mathematical cognitive models like DDM and a potential avenue for further analysis could be to link our results to neuroimaging signal (Nunez et al., 2024). This link can be established in different ways as several studies have shown. These possibilities range from taking a neural signal of interest and see how predictive it is for the variation of parameter values over conditions or participants (Bolam et al., 2022; Cavanagh et al., 2011; Ratcliff et al., 2009) to multivariate methods like probing electrophysiological signal for ramping (accumulation) behavior via Canonical Correlation Analysis (van Vugt et al., 2016). In our case, it could be especially interesting to take feature-specific classification output of M/EEG signal like the distance values of objects to a decision hyperplane at certain points in time during perception or memory retrieval and link those DDM parameter estimates (Bolam et al., 2022; Carlson et al., 2014; Philiastides et al., 2006; Ratcliff et al., 2009).

An important question is how the parameters of the DDM might relate to areas and processes in the brain. Multiple cortical and subcortical regions show correlations with parameters such as the prefrontal cortex (Wittkuhn et al., 2018) and thalamus (Turner et al., 2015) with the drift rate and the pre-supplementary motor area (Berkay et al., 2018) and the subthalamic nucleus (Frank & O'Reilly, 2006) with the decision threshold. As a key hub for episodic and long-term memory, the hippocampus has been implicated in memory-based decision-making (Barron et al., 2013; Wimmer & Shohamy, 2012), although linking DDM parameters to intracranial recordings of hippocampal activity failed to show evidence accumulation behavior (van Vugt et al., 2016). When looking at connectivity between brain areas, the hippocampus and the ventro-medial Prefrontal

Cortex appear to be crucial for decision-making from memory (Gluth et al., 2015). Medial frontal cortex neurons representing choices are phase-locked to medial temporal lobe activity when decision must be informed by memory (Minxha et al., 2020).

Shadlen and Shohamy (2016) proposed a model of decision making where information from either high-level visual areas for perceptual decisions or the medial temporal system for memory-based decisions are funneled through the striatum via a "thalamo-cortical" pipe towards parietal and prefrontal areas that represent a decision variable being constantly updated. For the memory case, the potential role of the hippocampus is acknowledged but a detailed interaction of hippocampal-cortical interactions is missing, leaving room for different roles that a pattern completion process of memory retrieval (Norman & O'Reilly, 2003) could have on the decision updating and the different parameters of the DDM. It could be the case that the retrieval itself is represented by the non-decision time as it delays reaction times overall and unspecifically (Kraemer & Gluth, 2023). On the other hand, the retrieval process itself might influence the decision related parameters (especially the drift rate) as it might need multiple updates of the pattern completion in the case of specific sensory information compared to easily targetable semantic or gist information.

By relating the outcome of this mechanistic model to brain data, other important questions have to be addressed, for example where in brain and how a decision variable is established. Converging evidence from monkey data proposes that competing perceptual information is integrated over time in the lateral intraparietal area (LIP) neurons until a decision boundary is hit and eye movement is initiated in a similar fashion as described by sequential sampling models (Shadlen & Newsome, 2001). Generally, it is assumed that association areas specific to any kind of effector work as a decision variable integration

information from perception or memory (Shadlen & Shohamy, 2016). Based on these mechanistic model, parietal activity during memory retrieval has been proposed to resemble a Mnemonic accumulator of information (Wagner et al., 2005). While it is unlikely to be the only role of the parietal cortex in memory, an EEG study (van Vugt et al., 2019) comparing perceptual and memory related decisions in face discrimination task showed that a positive parietal activity after 400 ms correlated significantly with a drift-rate parameter estimated from reaction times and satisfied multiple criteria of an accumulator (O'Connell et al., 2012).

3.4 Methods

3.4.1 Reaction Time Data

Reaction time data were taken from the seven experiments presented in Postzich (2024, PhD Thesis, Chapter 2) and two presented in Linde-Domingo et al. (2019). A detailed description of the participants, paradigms and procedures can be found there. For the convenience of the reader, the key information necessary for our analysis are laid out below.

While each of the nine experiments included a memory task, six experiments additionally included a visual task. All memory and visual tasks used non overlapping samples of participants. Reaction time data were analyzed separately for memory and visual tasks. Out of the experiments from Postzich (2024, PhD Thesis, Chapter 2) in the first, fifth, sixth, and seventh both memory and visual tasks were administered online.

A total of 288 participants (M_{Age} = 21.5, SD_{Age} = 4.9) participated in nine different memory tasks. For the six visual tasks, 335 participants (209 female, 113 male, 0 non-binary;

M_{Age} = 29.6, SD_{Age} = 9.4) were tested. Detailed information about the memory and visual tasks can be found in Table 1.

All memory tasks included 3 phases and were repeated over blocks. In the learning phase, cues (e.g. action verbs) and object images were presented together and had to be associated. Object images depicted everyday objects from different categories (e.g. animals, electronics, cars etc.) that varied on perceptual and semantic dimensions. After the learning phase, a short distractor task was administered to clear visual short-term memory. While the learning phase could vary between experiments, the cued recall phase always followed the same schema. Participants were presented with a classification option (e.g. 'animate' vs 'inanimate') indicating on which dimension to classify the object. Then the cue appeared, and participants had to retrieve the object and classified it as fast and accurately as possible on a feature dimension (i.e. "Was the object animate or inanimate?"). The reaction times of interest were recorded between the cue onset and the classification decision. In the memory task from Linde-Domingo et al. (2019) a third response option was given as forgotten. In our current analysis those responses were coded as incorrect. To ensure that the model had enough correct and incorrect reaction times to fit a distribution reliably, we excluded 44 participants who had either less than 55% correct responses or less than 3% incorrect responses. Reaction times below 500 ms and above 10000 ms were excluded.

All visual tasks consisted of only a classification phase that was nearly identical to the cued recall phase of the memory task. Instead of the cue, participants presented directly with the object and had to classify it. Reaction times of interest were recorded between the object onset and the classification decision. Again, for estimation stability, 99 participants with less than 55% correct responses or less than 2% incorrect responses were excluded. Reaction times below 150 ms were discarded.

3.4.2 Hierarchical Bayesian Drift Diffusion Models

Both the visual containing seven experiments and memory dataset containing nine experiments were modelled separately on a single trial level and included reaction times and accuracy for each trial. Reaction times and accuracies were modelled by wiener-first-passage-in-time (wiener-fpt) likelihood with a decision boundary α, a drift rate δ, a nondecision component τ and a starting point bias β on a single trial level (Navarro & Fuss, 2009). The starting point bias β was always set to .5 which is a common choice for modelling accuracy data since one would not expect there to be a systematic bias toward correct or incorrect responses before target onset. All other parameter values varied deterministically as a function of hyperparameters such that,

$$\alpha_i = \exp(\mu_a + \sigma_a * a_s)$$

$$\delta_i = \mu_d + \sigma_d * d_s$$

$$\tau_i = \Phi(\mu_t + \sigma_t * t_s)$$

where a_s, d_s, t_s are varying at a subject-level. These single subject effects are shifted by location parameters μ_a, μ_d, μ_t, resp., and scaled by scaling parameters $\sigma_a, \sigma_d, \sigma_t$, resp. Both location and scale parameters vary between experiments and conditions and can be interpreted as the experiment and condition-wise group mean and group standard deviation. For a graphic illustration of the hierarchical model used here refer to Fig. 5 A and for an overview over the prior distributions for all random variables see Fig. 5 B. Following the suggestions of Ahn et al. (2017), we used an exponential link function for α_i to ensure positive values, and a standard normal cumulative distribution link function for τ_i transforming input values into a range between 0 and 1. The output values of the link function

were additionally scaled to be strictly between a minimum value of 0.1 (since a nondecision time of 0 seconds would be impossible) and the participant's minimum reaction time in seconds. Modelling $\alpha_i, \delta_i, \tau_i$ as a function of a subject-specific effect that is shifted and scaled by hyperparameter values is called a *noncentered parameterization* and is often helpful in drawing out high correlations between group-level parameters and thereby stabilizing model convergence and estimation performance (Ahn et al., 2017).

3.4.3 MCMC sampling and condition difference of parameters

All models were implemented in the probabilistic modelling language STAN (Carpenter et al., 2017) and sampled using the Python's cmdstan interface with the Hamiltonian Monte-Carlo Algorithm and No-U-Turn Sampler. For each model, custom initialization values were passed to five independent sampling chains running in parallel on a high-performance cluster. Each chain ran 1700 iterations with 200 warming up samples. The maximum tree depth was kept at 10 and adaptation delta was set to 0.8. Additional diagnostics from the chains showed no signs for problems in sampling: Divergences were rare and only observed during the warm-up period and Bayesian fraction of missing information (Betancourt, 2016) values were sufficiently high for the memory (range: 0.7 – 0.76) and the visual model (range: 0.65 – 0.79). Summary statistics indicated that most parameters achieved $\hat{R}$ values below 1.01 (95.1% for the memory data, 93.7% for the visual data) (Gelman & Rubin, 1992) and sufficient effective sample size.

Experiments	N (f, m, nb, NA)	M Age (SD)	Cue	Perceptual	Semantic
Exp. 1 [1]	26 (19, 7, 0, 0)	19.0 (0.8)	Verbs	Drawing / Photograph	Animate / Inanimate
Exp. 2 [1]	24 (22, 2, 0, 0)	19.5 (0.9)	Verbs	Drawing / Photograph	Animate / Inanimate
Size	32 (21, 11, 0, 0)	25.9 (5.2)	Verbs	Big / Small	Manmade / Natural
Shape	25 (17, 8, 0, 0)	20.6 (2.4)	Verbs	Round / Elongated	Manmade / Natural
Scene Cue	27 (21, 6, 0, 0)	20.0 (2.2)	Scenes / Verbs	Drawing / Photograph	Animate / Inanimate
Spatial Cue	24 (20, 4, 0, 0)	18.9 (0.7)	Spatial Position	Color / Greyscale	Living / Non-Living
Spatial Cue Replication	38 (27, 8, 0, 3)	23.4 (8.6)	Spatial Position	Color / Greyscale	Living / Non-Living
Attention Encoding	40 (30, 8, 0, 2)	23.8 (3.6)	Verbs	Color / Greyscale	Manmade / Natural
Attention LOP	52 (41, 9, 2, 0)	19.7 (3.8)	Scenes	Drawing / Photograph	Manmade / Natural
Memory N_{Total}	**288** (excl 44)	**21.5** (4.9)			
Exp. 1 [1]	23 (20, 3, 0, 0)	19.3 (1.1)	–	Drawing / Photograph	Animate / Inanimate
Exp. 2 [1]	24 (20, 4, 0, 0)	19.0 (0.9)	–	Drawing / Photograph	Animate / Inanimate
Size	40 (29, 11, 0, 0)	30.7 (5.9)	–	Big / Small	Manmade / Natural
Shape	24 (23, 1, 0, 0)	18.7 (0.6)	–	Round / Elongated	Manmade / Natural
Spatial Cue	24 (0, 24, 0, 0)	19.6 (0.8)	–	Color / Greyscale	Living / Non-Living
Spatial Cue Replication	200 (130, 70, 0, 0)	33.7 (8.5)	–	Color / Greyscale	Living / Non-Living
Visual N_{Total}	**335** (excl 99)	**29.6** (9.4)			

Tab. 1: Overview over the studies used for the Drift Diffusion Modelling. Abbreviations: LOP = Levels of Processing, f = female; m = male; nb = non-binary; NA = not available; M = Mean; SD = Standard Deviation; excl = excluded participants. The Cue column lists the material used as cues in the recall task to remember the object. The last two columns list the perceptual and semantic dimensions that the objects used each experiment varied on. [1] Experiments 1 and 2 taken from Linde-Domingo et al (2019).

Posterior parameter distributions were based on the 7500 samples excluding the warm-up period. Our analysis focused on the posterior distributions of μ_a (decision boundary), μ_d (drift rate), and μ_t (nondecision time) that all varied between experiments and conditions. Within each experiment, marginal parameter distributions of the perceptual feature classification were subtracted from the semantic feature classification. A condition

difference was deemed significant, if 97.5 percent of the difference distribution were either larger or small than zero.

3.4.4 Determining model validity via Posterior Predictive Check

For the Posterior Predictive Check (Gelman et al., 2013), we constrained the wiener-fpt likelihood function on the average parameter estimates from the posterior distribution and simulated reaction time datasets with 1000 responses for each participant and each condition. Only correct responses were included in the analysis. For both simulated (predicted) and observed datasets, reaction times were binned into five cumulative quantiles (0-10 %, 11-30 %, 31-50 %, 51-70 %, 71-90 %), averaged within quantile and then plotted in comparison between conditions for both the memory data model and the visual data model (see Fig. 5 C).

Chapter 4: Tracking the reconstruction of naturalistic images from memory using similarity-based fusion of MEG and fMRI data

4.1 Introduction

The act of retrieving information from memory can sometimes be automatic and near instantaneous, for example when a photograph or an odor reminds us of a nice vacation we spent at the beach, and other times it can be effortful and hard like trying to remember a colleague's name while they are approaching at a conference. According to an influential model, episodic recall starts with a part of the original pattern that was experienced serving as a cue that activates a compressed index ensemble of neurons in the hippocampus which will then in turn reactivate the whole pattern of information in the cortex – a process termed pattern completion (McClelland et al., 1995; O'Reilly et al., 2014; Teyler & DiScenna, 1986). The episode is represented as a distributed pattern of neural activity that can entail everything from semantic, gist-like representations to specific perceptual details (Danker & Anderson, 2010; Rissman & Wagner, 2012). While a lot is known about the time course of the visual processing pathways during initial perception (Goodale & Milner, 1992b; Kravitz et al., 2011; Mishkin et al., 1983b) and how it transforms information from low-level perceptual to high-level semantic representations (Groen et al., 2017), the time course of feature reactivation during retrieval is less clear. In this chapter, we will use multivariate combination of fMRI and MEG and Deep Neural Networks to track the feature-specific reactivation of visual images during the time course of retrieval.

General retrieval processes like recollection and familiarity have been well described and delineated by a host of behavioral and event-related potential studies (Rugg & Curran, 2007; Yonelinas, 2002). Especially, the ERP studies point towards a positive parieto-temporal component indexing recollection-based retrieval (Rugg et al., 1998) that starts after 500 ms. Accordingly, memory studies in epileptic patients with intracranial electrodes show that information about a cue reach the hippocampus at about 200 to 300 ms after which a pattern completion process could be activated to retrieve content associated with the cue (Staresina et al., 2016, 2019; Staresina & Wimber, 2019).

The advent of multivariate pattern classification methods made it possible to investigate content or feature reactivation during the retrieval process in neural signal albeit mostly in fMRI (Johnson et al., 2008, 2009; Polyn et al., 2005). The use of pretrained Deep Neural Networks (DNN) as encoding models has further helped to capture specific fMRI reactivation of low-level visual and high-level semantic features during retrieval all throughout the ventral visual stream (Bone et al., 2020).

A recent study looked into the temporal dynamics of feature-based retrieval process. Linde-Domingo et al. (2019) showed participants everyday objects that varied among a perceptual (line drawing vs photographs) and a semantic (animate vs inanimate) dimension while they measured EEG activity. A classifier analysis of the signal revealed that the perceptual dimension could be successfully classified earlier than the semantic dimension when participants saw the images on-screen, in accordance with a forward sweep of information. But when participants had to retrieve the object from memory, classification peaked earlier for semantic features than for perceptual features, indicating a reversal of the hierarchy during retrieval. This effect has been replicated and generalized over multiple control conditions in reaction time studies (cf. Chapter 1) and has also shown a consolidation-

based enhancement over days of repeated retrieval tests (Lifanov et al., 2021). Compared with perceptual processing, retrieval-based processing thus consistently prioritizes abstract-semantic over detailed-perceptual information.

Here we want to take a closer look at the spatio-temporal dynamics of the feature-based retrieval process by fusing different data modalities using a data-driven, multivariate tool to investigate the processing of stimuli and their features called Representational Similarity Analysis (RSA). The idea is to transform the neural signal acquired by M/EEG or fMRI into a representational geometry by comparing neural activation patterns in response to all stimulus combinations (Kriegeskorte et al., 2008). The resulting Representational Dissimilarity Matrix (RDM) indicates how all stimuli relate to each other on a distance basis and makes it possible to directly compare how different data modalities (MEG/EEG, fMRI, DNNs) represent stimuli (Kriegeskorte & Kievit, 2013). Thereby RDM-based correlation (fusion) analysis has the ability to combine the spatial resolution of fMRI with the temporal precision of MEG and the feature-based focus of DNNs, which is why it has been extensively applied to object processing along the visual stream (Arbuckle et al., 2019; Cichy et al., 2017; Cichy, Pantazis, et al., 2016). Combing MEG and fMRI geometries was successfully used to envision the spatiotemporal dynamics of the ventral visual processing path (Cichy, Pantazis, et al., 2016). Another advantage of a fusion approach is the possibility to combine measurements from different participant samples (Kriegeskorte, 2009), for example Cichy et al. (2014) fused MEG data from humans to intracranial single unit recordings along the ventral temporal cortex in monkeys.

For this research endeavor, we will use images from the Natural Scenes Database (NSD), a special database where each image comes with multiple 7T fMRI whole brain recordings from 8 different participants (Allen et al., 2022). In our present study, we let a

new sample of participants learn NSD images together with words and then cue them later in a retrieval phase with those words for participants to reconstruct the associated image from memory, while MEG is measured. By using the same set of images, we are able to transform both the NSD fMRI data, our MEG data, and Deep Neural Network activations in response to those images into representational geometries allowing us to fuse (correlate) the geometries arising in the different data modalities. Such a multimodal fusion can reveal when in time and where in the brain (or in what layers of a network) the representational structure shows the highest match.

Taken together, we expect to see a forward flow of activation during the initial perception of the image, such that when correlating the representational geometries of different brain regions (fMRI) with the time resolved representational geometries of the MEG, early sensory areas (e.g. early visual cortex) should show significant correlation peaks first, followed by significant peaks from late areas (e.g. higher visual cortex, temporal cortex) later in time. Importantly, during the cued recall of the same images, we expect the opposite pattern of results, with late fMRI areas peaking before early areas, indicating a reverse information flow of image reconstruction from memory.

Similarly, when correlating the feature-specific representational geometries of pretrained deep neural network layers, we expect to find that low visual layer correlations with the MEG peak first during encoding followed by higher visual and semantic layer correlation peaks. Corresponding to the MEG-fMRI fusion predictions, we expect to find a reverse pattern of information flow during retrieval, with semantic and higher visual feature layers showing an earlier correlational peak than low visual feature layers.

As each image is retrieved multiple times, we will in a second step investigate how potential coactivation effects during retrieval change over the course of multiple retrieval

repetitions. While according to a recent theoretical account, multiple retrieval repetitions could strengthen an invariant semantic core of a memory (Antony et al., 2017), this effect might not occur within a session of retrieval repetitions but rather unfolds over longer time periods (Lifanov et al., 2021). Generally, we expect that reaction times (i.e. when in time participants indicate a reconstruction) will decrease over retrieval repetitions while accuracy (i.e. participants indicating a successful reconstruction from memory) should increase.

4.2 Results

In the learning phase of the MEG study, participants were presented with action verbs and naturalistic images from the Natural Scenes Database (Allen et al., 2022) and had to form a mental image or short story associating each pair. This association was, in a subsequent retrieval phase, re-elicited by cueing participants with the action verb and asking them to reconstruct the image mentally onto the screen while we measured their brain activity. Participants indicated the moment they had the image back in mind by pressing a 'Remember' button or a 'Forgotten' button if they couldn't remember. Participants indicated an overall high remembrance rate with an average of 88.1 % (SD = 32.4, see Fig. 8 B). However, since this is a subjective measure, each verb-image association was also tested once objectively during the retrieval phase. This was done by presenting either the correct image or a lure (both were 80% masked to make identification more difficult) to see if participants recognized the correct image. Like before, participants were able to distinguish between correct images and lures very well with an average accuracy of 88.5 % (SD = 32.0, see Fig. 8 B). When correlating both the subjective remembrance rate and image recognition accuracy a highly significant relationship, $r = .88$, $p < .001$, indicated that participants who

remembered the image more often were also better at recognizing the correct image, generally validating their self-report to remember the correct image (see Fig. 8 C).

While participants saw each verb-image pair only once, they retrieved each association six times during the cued recalled phase. This makes it possible to track changes over trials in the accessibility of the memory trace. Behavioral measures of the subjective remember button press showed that participant's average reaction times got faster over retrieval repetitions (one-way repeated-measures ANOVA: $F(5, 140) = 51.44$, $p < .001$, $\eta_p^2 = .65$, $\varepsilon_{GG} = .33$, see Fig. 8 A) and also more accurate ($F(5, 140) = 24.44$, $p < .001$, $\eta_p^2 = .47$, $\varepsilon_{GG} = .53$, see Fig. 8 A). More detailed post-hoc comparisons additionally showed that while the biggest decrease in retrieval time happens between the first and second repetition (Reaction Times: $t(28) = 10.09$, $p < .001$, $d = 1.87$; Accuracy: $t(28) = -4.14$, $p < .01$, $d = -0.77$), there are still significant improvements between the second to fifth (Reaction Times: $t(28) = 3.46$, $p < .05$, $d = 0.64$; Accuracy: $t(28) = -4.41$, $p < .01$, $d = -0.82$) second to sixth (Reaction Times: $t(28) = 3.89$, $p < .01$, $d = 0.72$; Accuracy: $t(28) = -4.32$, $p < .01$, $d = -0.82$). As predicted, multiple retrievals strengthen the accessibility of memory content in associative recall.

4.2.1 MEG-fMRI ROI Fusion

To track the flow of information within the visual system during initial perception of images and then later the recall from memory, we transformed the MEG signal into a representational dissimilarity matrix (RDM) format creating a multivariate measure of how the sensor activity geometrically represents all images at any point in time. The time windows of interest are the initial perception of the image during the learning phase (i.e. when participants are presented with the image during encoding), and the retrieval of the image during the memory phase (i.e. when participants see the action verb as a cue and have to

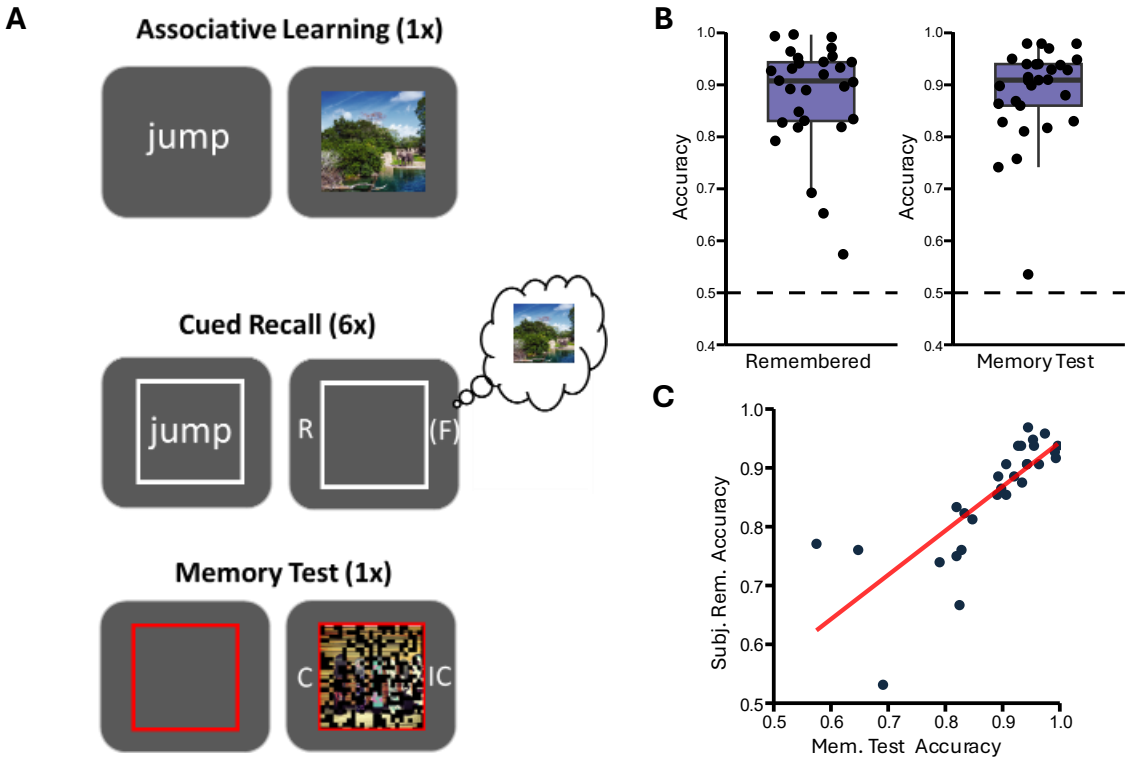

Fig. 8: A: Schematic of the paradigm used in our MEG sample. Participants associatively learned 96 images together with action verbs. Later they were cued with the word and had to try to reconstruct the image 6 times. Memory for each association was tested once per image in a lure recognition test. B: Performance of MEG participants in cued recall (left) and the memory test (right). Boxplots represent the distribution of participant's average accuracies and single dots indicate participants average accuracy. C: Correlation between participants average cued recall accuracy and memory test accuracy. The red line represents a least-squares fit to the data.

reconstruct the image). A separate sample comprised of eight different participants viewed

the same 100 images we used in our MEG study while their brain volumes were scanned in an 7T fMRI scanner (Allen et al., 2022). From this data, we captured the voxel activity of certain predefined ROIs (early visual cortex, mid-lateral, lateral, see Fig. 9 A, for ROI definition see Methods) and transformed these activations into the same RDM format as a multivariate depiction of how each ROI groups the images. In the next step, the MEG RDMs from each time point were fused (i.e., correlated) with the fMRI RDMs, resulting in a coactivation time course that signifies when in time the MEG representations are more similar to a certain ROI representation (see Fig. 9 A). To test the directionality of information processing we used a procedure developed by Michelmann et al. (2016) that transforms each correlation timeseries into a cumulative sum timeseries and then fits an ordinal regression model of ROIs (e.g. early visual first, mid-lateral second, lateral third) at each timepoint in a trial. The resulting timeseries of beta coefficients will show negative deflections if there is a forward processing stream (i.e., forward accumulation of information along those regions), and positive deflections during a reversal of the information flow (see Fig. 9 B).

During the encoding time window (i.e., image onset) cluster activations of the different ROIs show a cascading onset pattern (see Fig. 10 B). Early visual cortex shows significant coactivation with the MEG RDMs first (onset: 70 ms, peak: 120 ms) followed by the mid-lateral ROI (onset: 80 ms, peak: 120 ms) and then later by the lateral ROI (onset: 230 ms, peak: 300 ms). This pattern indicates a clear forward stream of information flow along the visual hierarchy from early sensory areas to high level visual areas represented by

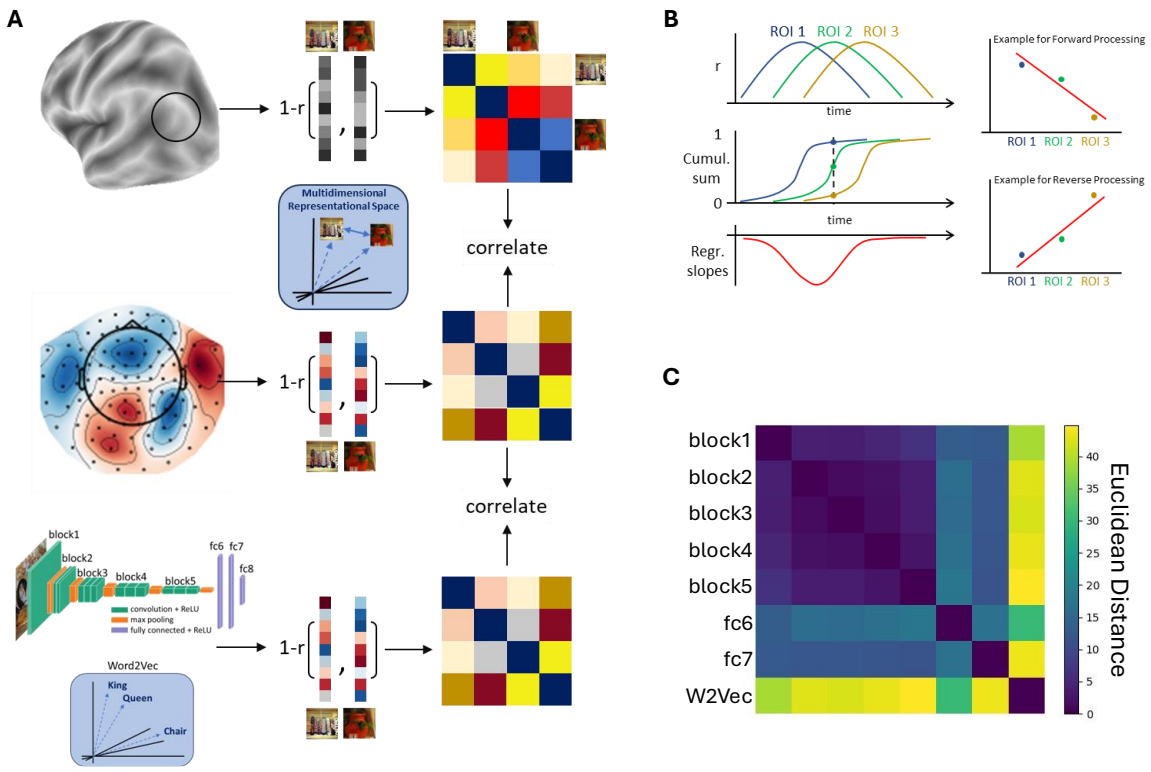

Fig. 9: A: Schematic of the RSA-based fusion approach. Upper row: Region-of-interest voxel activation are used to build one fMRI RDM per ROI. Middel row: Sensor activations at a timepoint t are used to build one MEG RDM per timepoint. Lower row: Hidden layer activations in response to an image or an image description are used to one DNN RDM per layer. MEG and fMRI RDMs are the correlated (fused) to get spatially informed representational time series. MEG and DNN RDMs are correlated (fused) to get feature informed representational time series. B: Schematic of the ordinal regression-based test of directionality. Correlation time series per ROI (or layer) are transformed into a cumulative sum over time and at each timepoint the order of ROIs or (layers) is regressed onto the cumulative sum values. Negative beta coefficients indicate a forward processing while positive beta coefficients indicate reverse processing. C: Distance between the layer representations of the two models (VGG16: block1-5, fc6-7; and Word2Vec).

the lateral ROI. The formal test of directionality showed a significant negative deflection between 110 and 410 ms, indicating a mainly forward stream of information progression.

During the retrieval time window (see Fig. 10 C) cluster activations of all ROIs start after about 500 ms, with the lateral ROI starting first (onset: 640 ms, peak: 1100 ms) followed by the mid-lateral ROI (onset: 700 ms, peak: 810 ms) and then the early visual ROI (onset: 700 ms, peak: 840 ms). Additionally, the high visual area ROI shows prolonged activity (up to 1300 ms) other than both the mid-lateral (up to 850 ms) and early visual ROIs (up to 870

ms). Although there seems to be a trend towards a reverse processing stream from high visual areas being activated first followed by low-level areas during the initial reconstruction process from memory, the directionality measure did not identify a significant deflection of any kind.

4.2.2 MEG Neural Network Fusion

The MEG fMRI fusion approach has the advantage of obtaining region-specific similarity with high temporal precision, showing information flow across the visual system during encoding and retrieval. To get a more detailed view of what types of features are processed at a given timepoint, we use an encoding model approach for our stimulus pool of naturalistic images by using the layer activations from visual (VGG16) and semantic (Word2vec) deep neural network models. As successor to the famous Alexnet (Krizhevsky et al., 2017), VGG16 is also built upon a biologically inspired architecture (Simonyan & Zisserman, 2015) which uses convolutional layers to extract image features and enable classification of objects and scenic content. Images are fed into the model, and activations from three hidden layers (block2 as low visual, block5 as mid visual and dense2 as high visual, see Fig. 10 D) are recorded as new features vectors. Semantic feature representations of the images are obtained by embedding short verbal image descriptions in a Word2vec model (Mikolov et al., 2013). This neural network has been pretrained on a large corpus and transforms input words into a lower dimensional semantic embedding space (e.g. it groups words that are semantically related closer together than semantically unrelated words). For each image or image description, hidden layer activations of VGG16 or Word2Vec, respectively, are used as feature vectors. By comparing the feature vectors for all pairwise

image combinations, layer-specific RDMs are computed and then fused with the MEG RDMs in a very similar way to the MEG-to-fMRI fusion reported above (see Fig. 9 A).

To ensure that the visual and the semantic model capture different features of the images used in this study, we compared the RDMs from the last layer of each block of the VGG16 and the layer activations of the Word2Vec model (see Fig. 9 C)

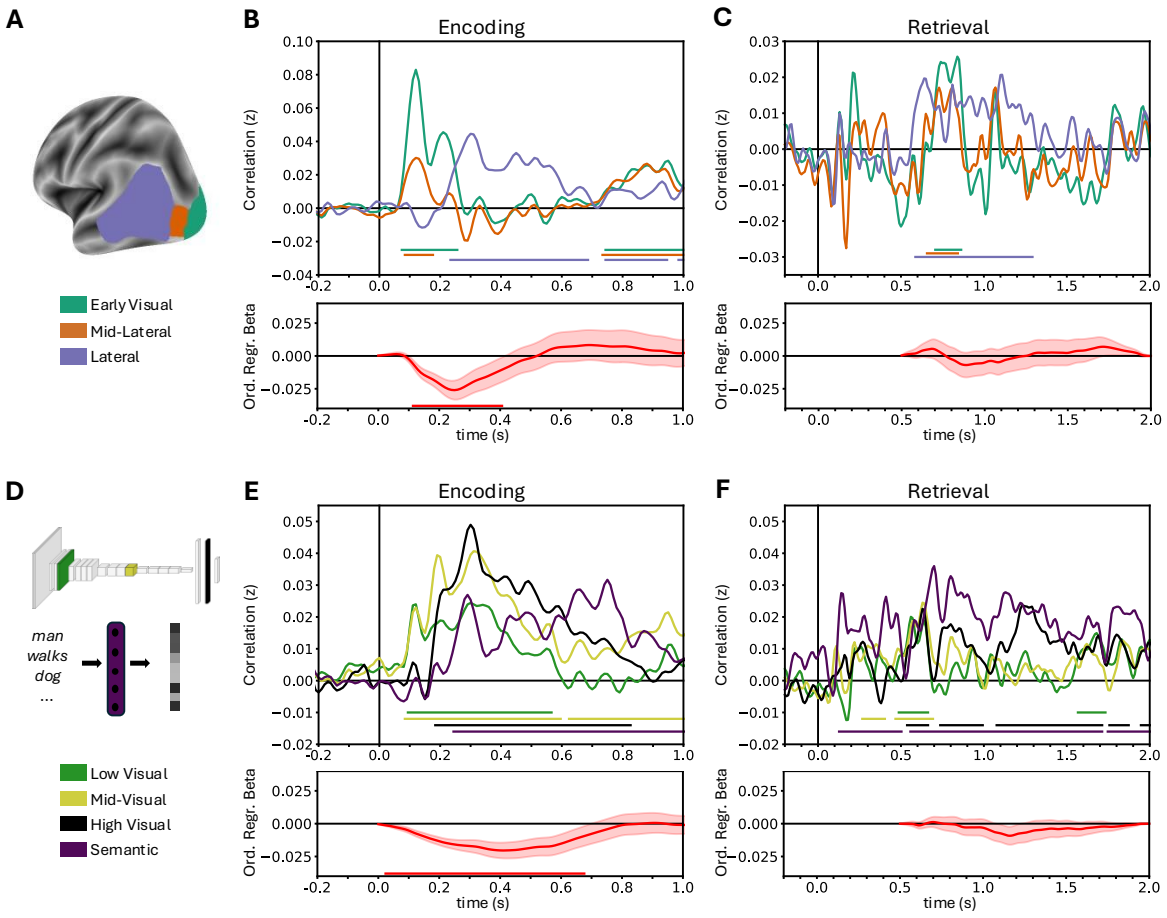

Fig. 10: A: Surface plot of masked areas used as fMRI ROIs. A predefined lateral visual processing stream was used from the NSD. B &C: Upper row: Average correlation timeseries between MEG RDMs and fMRI ROI RDMs. Lower row: Beta coefficients from ROI ordinal regression. Shaded area indicates SEM based on participants. D: Schematic of the visual (VGG16) and semantic DNN model with color-coded layers that were used for RDMs. E&F: Upper row: Average correlation timeseries between MEG RDMs and DNN layer RDMs. Lower row: Beta coefficients from layer ordinal regression. Shaded area indicates SEM based on participants. For B,C,D,F: Horizontal bars indicate significant clusters at $p < .05$ (corrected).

The same cascading pattern as before is observed during the encoding time window, with the low and mid-level features peaking first (onset: 90 ms, peak: 300 ms; onset: 80 ms, peak: 310 ms respectively) followed by a high-level visual layer peak (onset: 180 ms, peak: 300 ms) and then the semantic layer peak (onset: 240 ms, peak: 750 ms). Again, on a feature-specific level a clear forward flow of information from low level, early sensory to high-level, semantic regions is observed. Like in the MEG-to-fMRI fusion, the formal directionality test confirmed this forward processing stream with a significant negative deflection between 20 and 680 ms. This deflection seems to be even more pronounced than in the MEG-to-fMRI fusion.

For the retrieval phase, the results mirror the MEG-to-fMRI fusion results. Except for the semantic model, significant activation clusters (i.e., correlation peaks) are only observed after 500 ms, with an early onset of high visual and semantic layer information that is prolonged over the entire trial (550 to 1720 ms, peak: 700 ms; 530 to 1720 ms, peak: 1230 ms, respectively). Similarity with low and mid-level layers peaked right after 500 ms but only for a short period (onset: 490 ms, peak: 560 ms; onset: 460 ms, peak: 630 ms, respectively). Again, the test of directionality did not reveal any significant deflections.

4.2.3 Fusion analysis split over recall repetitions

To investigate whether the above-mentioned change in behavioral measures over recall repetitions, coincided which changes in the representational geometries emerging during retrieval, we focused our next fusion analysis on objects that were correctly classified in the memory test and that were subjectively recalled correctly at least 4 times. The recall

trials were then split into the first two recall trials and the second two to four recall trials. The fusion analysis remained the same as above.

For the MEG-to-fMRI fusion, the first group of recall trials seems to be dominated by a representational match with the lateral ROI (onset: 530 ms, peak: 610 ms) with no

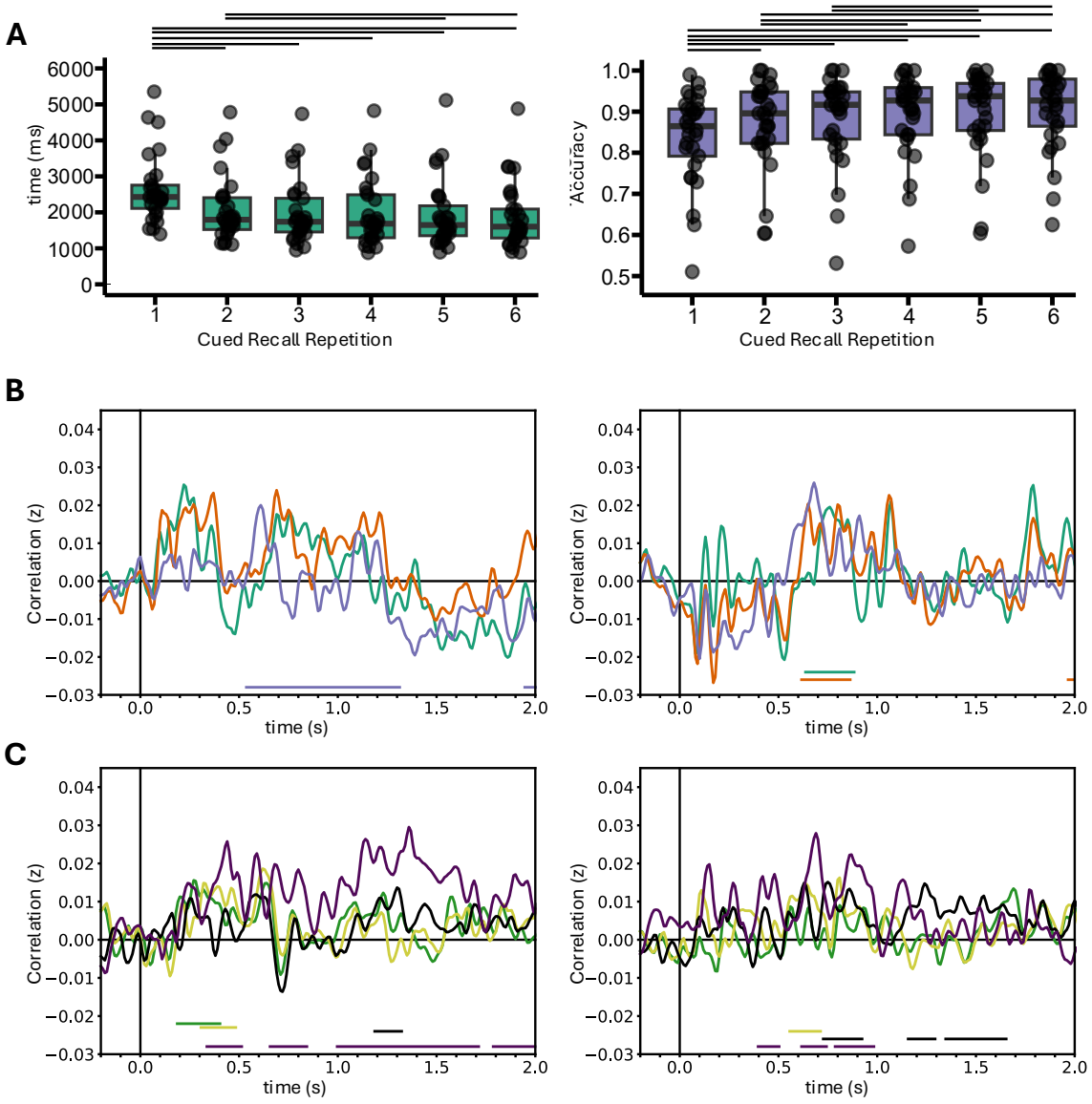

Fig. 11: A: Behavioral results for each of the six cued recall repetitions. Left: Boxplots of participant's average reaction time. Right: Boxplots of participant's average accuracies. Single dots indicate single participants. Lines on the top indicate significant differences between repetitions ($p < .05$, corrected). B: Average correlation timeseries between MEG RDMs and fMRI ROI RDMs split for the first two cued recalls (left) and the second two to four cued recalls (right). C: Average correlation timeseries between MEG RDMs and DNN layer RDMs split for the first two cued recalls (left) and the second two to four cued recalls (right).

significant clusters for the mid-lateral and early visual ROI. The opposite pattern is then found in the second group of recall trials where both mid-lateral (onset: 610 ms, peak: 650 ms) and early visual ROIs (onset: 630 ms, peak: 760 ms), but not the lateral ROI, show significant clusters of correlation. The significant clusters found here overlap almost completely with the clusters found in our fusion analysis including all recall trials.

The MEG-to-DNN fusion shows a similar but more intricate difference between the early and late recall repetitions. For the first two recall trials, there is a dominant correlation with the semantic layer expressed in a late big cluster (onset: 990 ms, peak: 1360 ms). While the low and mid-visual layers show some early activations before 500 ms, for the high visual layer only a small significant cluster is found (onset: 1180 ms, peak: 1300 ms). In the second group of recall trials, the semantic layer shows a smaller activation than before with significant clusters mostly between 390 and 990 ms (peak: 690 ms). Now the high visual layer has a more pronounced pattern of activations with significant correlation clusters between 720 and 1660 ms (peak: 750 ms). While a significant cluster is found for the mid-visual layer (onset: 550 ms, peak: 640 ms), none is found for the low visual layer.

4.3 Discussion

The time course of visual perception has been studied intensively and a clear hierarchical transformation from low-level perceptual to high-level semantic information has been well established (Cichy et al., 2014, 2017; Goodale & Milner, 1992b; Groen et al., 2017; Martin & Barense, 2023; A. D. Milner & Goodale, 2008; Mishkin et al., 1983b). In contrast, evidence for the dynamics of information reactivation during the retrieval process is still sparse (Linde-Domingo et al., 2019; Michelmann et al., 2019). Here we used a multivariate,

data-driven approach of fusing temporally-, spatially- and feature-resolved data modalities to investigate the time course of retrieval.

When testing our main prediction of a reversal of information flow between perception and memory recall, we did encounter a pronounced activation of lateral areas among the visual processing stream in a recollection-based timeframe of ~500 ms after cue onset (Staresina & Wimber, 2019). During the initial perception of the images, we saw the highest and most consistent representational overlap from this lateral ROI stream, in line with previous findings that scene related processing follows a more lateral path situated between the ventral object-oriented and parietal movement-oriented stream (Kravitz et al., 2011). This lateral ROI also includes part of the angular gyrus that is implicated in episodic memory retrieval (Berryhill, 2012; Thakral et al., 2017). Similarly, regarding the MEG-to-DNN fusion, we found pronounced semantic and high-visual layer reactivation during retrieval. Although a test for a reverse flow of information did not show significant positive deflections after 500 ms, the results show that high-level visual and gist-like representations are most reliably reactivated during the retrieval process (Robin & Moscovitch, 2017).

As expected, we observed a significant forward stream during initial perception that was expressed in a forward cascading activation of fMRI ROIs as well as among a hierarchy of DNN layers from low to high visual and semantic features. The results of the fMRI ROI and the DNN layer fusion are well in line with previous fusion results, including studies where participants view objects or more complex scenes (Cichy et al., 2014, 2017; Cichy, Pantazis, et al., 2016).

Interestingly, when taking a closer look into the development of the retrieval time course over multiple recall repetitions, we encountered unexpected results for both types of fusion analyses. For the MEG-to-fMRI fusion, lateral activation was mostly found for early

recall attempts and mid-lateral and early visual activation dominated later recall attempts, indicative of the retrieval becoming increasingly detail-focused over time. A similar but more nuanced switch in feature activation was found in the MEG-to-DNN fusion when comparing recall repetitions, going from more semantic activation in early recall attempts and then shifting towards more visual activation in later attempts. These results are surprising in light of previous suggestions that retrieval can act as a fast route to memory consolidation (Antony et al., 2017), stabilising and in this process also 'semanticising' memories through repeated reactivations. Since this idea was first proposed, empirical findings have lent support to the idea that repeated recall can stabilize memories, but not (as originally thought) in a fast manner, with the 'semanticizing' effects of retrieval often only showing after a longer delay of several days (Ferreira et al., 2019; Lifanov et al., 2021). Repetitions within one session, as conducted in the present study, have even been shown to lead to a gradually faster accessibility of perceptual details (Lifanov et al., 2021), consistent with the present fusion results that suggest a progression, over trials, from reactivation from mainly high-level visuo-semantic to lower-level visual features.

Curiously, descriptive differences in the correlation time series are observed between splits in a time window between cue onset and 500 ms (Fig. 11B), a timeframe in which. In this timeframe processing of the cue is still ongoing. During the first repetitions correlations tend to be greater than during later repetitions, although not significantly so. Since this effect is mainly driven by a change in the temporal representational geometry of the MEG, it might depict a more effortful retrieval process during the first repetitions which would correspond to the behavioral findings of reduction of reaction times and increase of accuracy over repetitions.

In our current analysis, predefined fMRI ROI masks were used from the NSD database that were determined and created alongside the sampling of the database (Allen et al., 2022) and were derived from regions defined in the Wang Atlas (Wang et al., 2015). An early visual cortex ROI included V1-3, a smaller mid-lateral ROI included LO1 and LO2 and a lateral ROI that sits more anterior to the mid-lateral ROI and is enframed by the edges of the superior temporal sulcus and the angular gyrus. Importantly, this last ROI is several times bigger than the mid-lateral ROI.

There is a possibility the differing size of the ROIs, especially the mid-lateral compared to the lateral and early visual ROI have an impact on our results. While larger ROIs more likely encompass heterogenous regions with diverse functional connectivity and activation patterns (Wei et al., 2022), small ROIs might be functionally too specific to find significant covariation with MEG signal (Geerligs et al., 2016). It might be that the distinction between the early and the mid-lateral ROI is too small which could explain why the time courses of both ROIs are very similar. Taking these caveats into account, further analysis would profit from more equally sized ROIs or a searchlight approach allowing for a more fine-grained and even-handed depiction of the representations shared between fMRI and MEG.

A question that is currently strongly debated in the memory field is whether memory retrieval relies on the reactivation of the original patterns established during encoding, or rather a somewhat transformed version of the perceived experience (Favila et al., 2020). Early fMRI studies, including those using multivariate methods, found evidence for an overlap between areas activated during encoding and areas activated during retrieval (Danker & Anderson, 2010), with the latter typically being smaller (Wheeler et al., 2000) and restricted to later sensory processing stages (Danker & Anderson, 2010). Additionally,

memory retrieval was shown to be supported by parietal activity in ERP studies (late positive complex most pronounced in temporo-parietal regions) and in fMRI studies (Johnson et al., 2009; Mecklinger et al., 2016). Early studies assumed that parietal activations reflect memory operations on the retrieved content, rather than content representations in themselves. In recent years, however, it became clear that retrieved content is decodable from parietal activations, and in fact is relatively more decodable during retrieval than perception, suggesting a transformation of information between encoding and retrieval (Favila et al., 2020; Long & Kuhl, 2021). Importantly, the current study uses fMRI patterns from a perceptual exposure condition where participants are viewing the images on screen, and these patterns are used for comparison with the encoding/perception and retrieval MEG patterns. The study might thus not be sensitive to pattern transformations occurring between encoding and retrieval. A comparison with fMRI signal from participants doing the task presented here with the same images could elucidate not only mere reactivation of visual processing patterns, but also pattern transformation. Such an fMRI dataset has actually been collected by our group, but analysis was beyond the time frame of my PhD. Furthermore, a comparison between both MEG retrieval to fMRI encoding and MEG retrieval to fMRI retrieval patterns could elucidate the temporal dynamics between both forms of memory representation, for example by delineating at which time points are different regions of the brain more engaged in pattern reactivation (higher fMRI encoding to MEG retrieval correlations) versus pattern completion (higher fMRI retrieval to MEG retrieval correlations).

Our study used two different samples for the MEG and fMRI which is not uncommon for a representational fusion approach (Cichy et al., 2014; Kriegeskorte, 2009). It has to be kept in mind that this obscures idiosyncrasies in memory encoding and retrieval that could be modelled if both MEG and fMRI samples would use the same participants and focused

only on coactivations that are consistent over participants. Retinotopic maps in early visual areas are preserved across subjects (Benson et al., 2012). Studies comparing fMRI activations between participants who recall the same memory show a high overlap in high-level areas (Chen et al., 2017). However, there might still be a possibility that idiosyncrasies in reactivation could affect perceptual details more strongly than a semantic gist.

In our current analysis, we are restricting representational geometries in the fMRI to certain ROIs. In a next step, we will extend this analysis with a searchlight approach where RDM matrices are computed from a shifting searchlight of voxels throughout the entire brain. This renders a whole brain volume of RDMs enabling us to create a time-resolved movie of coactivations between MEG and fMRI representational geometries. In the past, this approach has been used in visual perception studies to show a forward stream of activation (Cichy, Pantazis, et al., 2016). A more spatially resolved method like this could be fruitful to show evidence of memory reactivation in a reverse or even more complicated fashion, without the need to focus on predefined ROIs.

A potential avenue for further analysis could be the inclusion of eye tracking data that is available for most participants. Eye movements have in the past been shown to covary with memory retrieval (Hannula & Ranganath, 2009). Johansson et al. (2022) showed that gaze patterns during encoding reoccurred during retrieval of the image. The degree to which gaze patterns were reactivated directly or in a rule based transformed fashion could predict the fidelity of memory retrieval. With these results in mind, it would be interesting to see whether the degree of overlap of scan paths between encoding and cued recall could be harnessed as an implicit measure of retrieval quality to sort our trials and sharpen further analysis steps, or even as a covariate that explains the overlap in neural activations at distinct time points (Linde-Domingo & Spitzer, 2023).

4.4 Methods

4.4.1 MEG Participants & Procedure

Twenty-nine participants (16 female; MAge = 25.1, SDAge = 3.9) were recruited from the University of Glasgow Subject Pool to participate in the MEG study. Eligibility criteria included normal or close-to-normal vision, proficient English knowledge and absence of psychological conditions (e.g. ADHD, epilepsy). Before the in-person appointment in the MEG, participants completed an internet version of the training task to get familiar with the task. Upon arrival at the MEG facilities participants signed a consent form and were then prepared for MEG by attaching head position coils to the head. After being seated in the MEG in a magnetically shielded, participants read the instructions for the main task and finished the training task. Participants completed all 8 blocks of the experiment, with typically a larger break between the first and second four blocks. After completion of the task, participants were debriefed and received 30 Pounds in cash as a compensation for their participation. This experiment followed all ethics guidelines and was approved by the CCNi Research Board, Glasgow.

4.4.2 MEG Material & Design

The stimulus pool for this study consisted of 100 naturalistic images taken from the NSD (see "NSD Participants & Material"). 96 images were used for the main task, while 4 images (the same for all participants) were used for the training task.

The main task used here consisted of three phases: An associative learning phase, a distractor number task and a cued recall phase. In the learning phase, each trial started with a jittered fixation cross (0.5 – 1.5 s), followed by an action verb as a cue (1 s). Afterwards a second jittered fixation cross (0.5 – 1.5 s) appeared before the image was presented. Participants were instructed to form an association between the action word and the image and indicate the association with a button press. The association learning was thereby semi self-paced with either the button press or a maximum duration of 10 s ending the trial. Each verb-image pair was presented and learned once, with a total 12 pairs of verb-image associations were learned per block. Following the learning phase, participants engaged in a short distractor number task where they categorized odd or even numbers using the left or right index finger button, resp. This task lasted for 30 s and participants were instructed to classify as many numbers as accurately as possible. Feedback of performance (i.e. number of trials and percentage of correct classification) was given after the task. During the whole recall phase, a square outline of the same size as the images was continuously displayed in the background to aid memory reconstruction. Each retrieval trial started with a jittered fixation cross (0.5 – 1.5 s) that was followed by an action verb. Participants were instructed to retrieve the image that was paired with the action verb and reconstruct it mentally within the square frame to the best of their ability. The action verb disappeared after 0.5 s and two response options appeared outside of the square frame. An "R" for Remembered always appeared on the right and an "(F)" for Forgotten always on the left side of the frame. Participants were told to indicate whether they could reconstruct the image or not by pressing

either their right index button for “R” or their left index button for “F”. After either button press or a maximum duration of 10 s, response options disappeared leaving only the white frame on screen for 3 s. Participants were told to keep their reconstructed image actively in mind during this period. After 3 s the next retrieval trial started, unless there was a random memory test. In case of a memory test, the white frame changed color to red and a fixation cross appeared in the middle of the screen for 1 s. Then a masked image appeared within the frame together with two response options (“C” for Correct and “IC” for Incorrect) on the left and right outside of the frame. Presentation side of response options was counterbalanced between participants. The masked image consisted of either the correct image (i.e. image that was presented with the action verb) or a lure image from the same block. Images were overlaid by a black square with randomly placed transparent tiles blocking of 80 % of the image to challenge recognition performance. Participants were instructed to respond as quickly and accurately as possible using the left and right index button. After either button press or a maximum duration of 6 s the next retrieval trial started. Each of the 12 verb-image associations was retrieved six times (72 trials) and tested one time (12 trials). The memory test appeared randomly after one of the six retrieval trials. This task repeated over 8 blocks testing overall 96 verb-image pairs. The training task was a shortened version of the main task described above and consisted of 4 separate verb-image pairs in the learning phase and 2 retrieval trials and 1 memory test trial per pair.

4.4.3 MEG Data Acquisition

Continuous MEG signals were recorded from 306 channels (204 planar gradiometers, 102 magnetometers; Elekta Neuromag TRIUX, Elekta, Stockholm) at a sampling rate of

1,000 Hz and filtered above 330 Hz. In 28 participants an additional two electrodes were placed above and below the left eye and two electrodes were placed next to the left and right eye to measure vertically and horizontal, resp, eye movement. In the MEG scanner eye activity (gaze patterns, pupil dilation) was additionally monitored with an EyeLink Tracking System (SR Research, Ottawa, Canada) in all participants but was only stable in 25 of 29 participants. Visual stimuli were projected onto a screen (~1.15 m from participant's eyes) in the magnetically shielded room using a PROPixx projector (VPixx Technologies, Saint-Bruno, Canada) with a 1440 Hz refresh rate. All images were presented at the center of the screen in their original size of 425 x 425 pixels (visual angle: ~ 4.48°). Participants were given a left- and a right-hand button box for recording manual responses (fMRI optical response pad; LUMItouch, Photon Response Inc., Burnaby, Canada). Due to a software malfunction, in one block of one participant MEG was not recorded during the learning phase.

Individual anatomical MRI scans (T1-weighted; 1 × 1 × 1 mm voxels; TR = 7.4 ms; TE = 3.5 ms; flip angle = 7°, field of view = 256 × 256 × 176 mm) of participants were acquired in the Glasgow CCNi facilities 3T Scanner (Siemens, Germany).

4.4.4 MEG Data Preprocessing

All data preprocessing was done using the MNE toolbox v1.2.2 (Gramfort et al., 2014) and custom-tailored Python code. Spatiotemporal SSS was applied to the raw data via MNE's inbuilt maxfilter function with a duration window of 100 s (acting as an implicit low-pass filter at 0.01 Hz), a correlation value of .9 and head motion correction from continuous HPI recordings.

To reduce the influence of artifactual noise in the signal, independent component analysis (ICA) was applied to the maxfiltered data. To estimate the signal sources and unmixing matrices, data from all MEG channels was concatenated over blocks, Butterworth filtered between 1 and 40 Hz and then downsampled to 100 Hz. Since ICA can be influenced by noisy data trials, the signal was split into fixed length epochs on which a global rejection threshold was estimated using the autoreject package (Jas et al., 2017). Bad data epochs were then dropped according to the package's recommendations (see below). ICA was estimated with MNE's ICA class using the picard algorithm with as many components to capture 98 % of the data's variance. The estimated source time series and topographies of each participant were visually inspected to detect ocular and cardiac components. For all participant ocular and cardiac components were identified and rejected from the data with a median of 5 components per participant (range: 2 – 7 components).

The maxfiltered and ICA cleaned data was low passed filtered to 100 Hz and then split into epochs. The two ERF windows of interest where the encoding (i.e. onset of the image during the learning phase) and the retrieval cue (i.e. onset of the action verb during the retrieval phase) window. Both windows were cut to 0.5 s before and 4 s after event onset and downsampled to 500 Hz. To avoid slow drifts in the signal, a trial masked robust detrending procedure was used instead of a high pass filter which can distort classifier performance (van Driel et al., 2021). In this procedure trials are cut to long epochs including 15 s before and after each event and a robust trend is iteratively estimated while blocking high amplitude outliers. The ERF window of interest around the event marker is not included in the trend estimation (i.e. masked) to not discard valid low frequency signal. We first removed an order 1 (linear) trend before removing an order 30 trend.

Artifacts in the signal of each participant were automatically detected using the "autoreject" package (Jas et al., 2017). This procedure identifies artifact trials by optimizing the fit between the median signal of all trials and the mean signal of non-rejected trials. Whether a trial is rejected or interpolated is controlled by two hyperparameters with optimal parameter values being determined that were identified by a grid search with cross-validation. For the encoding window a total of 71 trials (2.6 %) were rejected with an average of 2.45 rejections per participant (SD = 4.52) and for the retrieval cue window a total of 737 trials (4.4 %) were rejected with an average of 25.41 rejections per participant (SD = 31.1).

4.4.5 NSD Participants & Material

For our study we used image and fMRI data taken from the Natural Scene Database (Allen et al., 2022). In the creation of the NSD eight participant were presented with over 70000 images during more than one year of weekly 7T fMRI scanning sessions. While most images where only seen once by one participant, a certain subset of images called the "special 100" were presented three times to each of the eight fMRI participants. For this study, we chose these 100 naturalistic images to, first, maximize the amount of available fMRI data (24 brain volumes per image) and, second, achieve a heterogenous sample of images maximizing variance in general brain areas of interest (i.e. early visual cortex, inferior temporal cortex). All images (425 x 425 pixels) from NSD originate from the Mircosoft COCO database (Lin et al., 2014) and are depicting natural scenes with multiple objects from 80 different categories (e.g. people, traffic, animals, food, furniture etc.).

NSD fMRI Data Acquisition, Preprocessing and ROIs

MRI data was collected on a 7T Siemens Magnetom scanner, using a single-channel-transmit, 32-channel-receive RF coil. For functional scans, Allen et al. (2022) used a gradient-echo EPI sequence (TR = 1,600 ms, TE = 22.0 ms, flip angle 62°, echo spacing 0.66 ms, bandwidth 1,736 Hz per pixel, partial Fourier 7/8, iPAT 2, multi-band slice acceleration factor 3) that covered the whole brain (field-of-view 216 mm (FE) × 216 mm (PE), phase encode direction anterior-to-posterior, matrix size 120 × 120). Each functional volume consisted of overall 84 axial slices, each 1.8mm thick, with 0mm gap, resulting in isotropic 1.8 mm sized voxels.

A general linear model (GLM) including a denoising approach (see Allen et al., 2022) was used to estimate a single beta image per image presentation, approximating the BOLD voxel pattern elicited by this image. For the given image set used in the present study, we thus obtained 100 (images) x 3 (repetitions) x 8 (participants) = 2400 functional volumes.

Predefined region of interest masks were derived from NSD as a subset of "streams ROIs": 1. Early visual cortex ROI that includes V1-3 from Wang et al. (2015), 2. Intermediate lateral ROI that includes LO1 and LO2 from Wang et al. (2015), and 3. High-level lateral ROI that is enframed by the edges of the superior temporal sulcus and the angular gyrus (https://cvnlab.slite.page/p/X_7BBMgghj/ROIs).

4.4.6 Neural Network Data Acquisition, Preprocessing and Layer Information

To capture low to high level visual features of the images, we used a pretrained VGG16 (Simonyan & Zisserman, 2015) from Tensorflow. This convolutional deep neural network (cDNN) follows a similar biologically inspired architecture of the visual cortex like

the previous gold standard AlexNet (Krizhevsky et al., 2017) but improves upon its image classification performance to 93% (top-5 out of 1000 categories). The original VGG16 consists of 13 convolutional and 3 fully connected layers and is trained on the ImageNet database for image classification. Since our dataset is a more naturalistic and complex image set with scenic information and multiple objects, we adapted the original model in the following way: The original three fully connected layers were dropped and replaced by two fully connected layers (2048 neurons) ending in 80 neuron output layer with a sigmoid activation function to enable multi-label classification. Thereby it is possible to transfer-learn the original model to detect multiple objects within an image. We fine-tuned this new model on 20531 training images taken from NSD. These images did not contain the "special 100" used in the MEG and fMRI data and were split into a training and a validation set (90%/10%, resp). Image labels come from MS Coco's 80 categories of objects with an 80x1 vector of 1s and 0s for each image indicating which objects are depicted. As loss function to train the new model we used a sigmoid f1 score (Bénédict et al., 2021). The f1-score function (Powers, 2020) balances precision ("how many retrieved categories are actually relevant?") and recall ("how many of all relevant categories were retrieved?") and is a preferred performance measure in case of multilabel classification (Bénédict et al., 2021; Opitz & Burst, 2019). Instead of freezing layer weights, all layers were trained with different learning rates for the convolutional layers (lr = 1e-6) and the new fully connected layers (lr = 1e-4). The new model was trained for 100 epochs and reached a validation set f1 score ≈ .51. Then, each of the 100 images from our study was input into the model and its layer activations were taken from layer 3 (block1_maxpool), layer 9 (block5_maxpool) and layer 15 (dense2) as a proxy for early, mid- and high visual feature representations.

For a semantic model of image features, we used a pretrained Word2vec model from the nlu package (John Snow Labs). This model embeds input words in a 300-dimensional space that maps out a semantic geometry. This means that words that describe semantically similar concepts (e.g. “king” and “queen”) are grouped closer together and farther away from dissimilar concepts (e.g. “king” and “chair”). For each of our 100 images, we retrieved five short sentence descriptions (e.g. “A man walking a dog in the park”) from MS COCO and input these descriptions into the Word2vec model to get a 300x1 vector as a semantic representation per image.

4.4.7 Representational Similarity based Fusion of Data Modalities

For the MEG data, encoding trials were cropped to -0.5 s and 1.5 s around image onset and retrieval cue trials were cropped to -0.5 s and 4 s around cue onset. Both signals were baseline corrected with an average pretrial time window between 0.5 and 0.2 s. Each of the 102 sensor triplet (2 gradiometers, 1 magnetometer, 306 sensors in total) was spatially combined with the mne’s “as_type()” function into 1 virtual magnetometer. Each encoding trial per participants equals one image. For the general fusion, all retrieval trials of one image were averaged over repetitions resulting in one average trial per image. For the comparison of fusion results over repetitions, only images that were successfully remembered at least 4 times were included in this analysis. Here, all retrieval trials of an image were split into the first two and the remaining two to four recall repetitions resulting in two averaged trials per image. Feature vectors per image were retrieved from the encoding and retrieval trial data in the following manner: A gaussian window (full-width at half maximum ~ 40 ms) was slid over the time dimension with a center at time point t in 10 ms steps. At each time point the

window was gaussian averaged leaving a 102-element vector representing the sensor activations at time point t. Within each time point each image was compared to each other image by computing the correlation distance (1-r) between their sensor activations. The resulting 96x96 array of all comparisons is called a representational dissimilarity matrix (RDM) and describes the representational geometry of the MEG signal at this time point. Accordingly, after looping over time there will be one RDM per time step for both encoding and retrieval cue windows.

For the fMRI data, beta values for the whole brain were normalized and ROI masks were used to extract the voxel activations for each ROI and image as feature vectors within participants. For each participant and ROI, feature vectors of each image were compared to each other image by computing the correlation distance (1-r) resulting in three ROI-specific RDMs (early visual, mid-lateral, lateral).

For the neural networks, within each chosen layer hidden activations of each image were compared to each other image with a correlation distance resulting in 4 layer-specific RDMs (early, mid-, high visual and semantic).

Since all modality specific signals have been transformed into the same space, RDMs can be correlated (so called second order correlation or Fusion) themselves. In the case of MEG-fMRI fusion, each ROI-specific RDM was correlated with all encoding-specific RDMs and retrieval-cue-specific RDMs yielding an ROI encoding and an ROI retrieval-cue time course of second order correlation values. High values at a time point indicate that this ROI represents the images in our study more similar to the sensor activations at this time point. In the case of MEG-DNN fusion, the same procedure was used but instead of ROI-specific RDMs the layer-specific RDMs were used. All RDMs were correlated using Pearson correlation.

4.4.8 Significance Testing

The correlation of participant average accuracies between cued recall and memory test was tested for significance using a one-sample t-test. Reaction time and accuracy data over recall repetitions were analyzed using a one-way repeated measures ANOVA with recall repetitions as a six-level factor. Degrees of Freedom and p-values for the ANOVA were Greenhouse-Geisser corrected if sphericity assumptions were violated. Here ANOVA results are reported with Greenhouse-Geisser's epsilon instead of corrected degrees of freedom. Post-hoc t-tests between repetitions were corrected using a Tukey-HSD criterion.

Encoding and retrieval-cue time courses of second-order correlations were tested using a cluster permutation test. A surrogate dataset of 1000 permutations was created in the following manner: Within one permutation step, the order of each ROI-specific and layer-specific RDM was randomly shuffled and each shuffled matrix was then fused with the encoding and retrieval-cue RDMs (which were not shuffled). Note that the RDMs were only shuffled once and held constant over time to not disturb the autocorrelation of the signal. Each surrogate time series was z-transformed and significant values were defined to exceed a threshold of value alpha (in this case $\alpha value < .05$). The thresholded time series were then screened for clusters of significant values and z values within each cluster were added giving each cluster a z-sum value. For each permutation and time series, the biggest z-sum value was saved, resulting in a vector of the 1000 most extreme surrogate z-sum values. In the real data, significant clusters of activation were determined in the same way as in the surrogate data using the alpha-value and z-sum values for each cluster was computed. These values were then compared against the distribution of most extreme surrogate z-sum values and

were deemed significant if they exceeded a threshold of cluster alpha (in this case $\alpha_{cluster} < .05$). This means that real clusters had to be higher than 5 % of the most extreme clusters from the surrogate (Null) distribution.

Chapter 5: General Discussion

The aim of this thesis was to track the spatiotemporal dynamics of feature-specific retrieval. For episodic memory, an engram or a memory trace is understood as a binding of content that was experienced in a certain situation or event with features ranging from low-level perceptual details (“seeing brown fur”) over gist-like information (“an animal in the park”) up to schemas (“someone walking their dog”) and general abstract knowledge (“a dog is an animal”). Similar to the known hierarchical organization of the visual system, where incoming percepts are transformed from low-level detailed to high-level conceptual information over the brain regions of the ventral visual stream (Desimone et al., 1984; Goodale & Milner, 1992; Gross et al., 1972; Martin & Barense, 2023; Mishkin et al., 1983), converging evidence points towards a similar hierarchical organization of the memory trace with a preference for high-level gist like information over perceptual details (Ahissar et al., 2009; Ahissar & Hochstein, 2004; Brady et al., 2011; Kerrén et al., 2023; Konkle et al., 2010b, 2010a; Lifanov et al., 2021; Linde-Domingo et al., 2019; Van Kesteren et al., 2012).

An episodic long term memory (LTM) trace is posited to bind together cortical patterns representing the elements of an episode in such a manner that encountering a partial pattern (cue) can bring back the whole pattern (other elements and their features of this episode) via a process of pattern completion (Horner et al., 2015; Horner & Burgess, 2013; McClelland et al., 1995; Teyler & DiScenna, 1986; Teyler & Rudy, 2007). While this theory makes little to no assumptions about the spatio-temporal relations and dependencies of these cortical patterns, if these cortical patterns of the elements and their features are represented within the visual hierarchy in a spatially reverse manner as mentioned above, this would imply a temporal difference in the accessibility of features during memory retrieval.

By combining behavioral reaction time studies with computational models, deep neural networks and MEG as well as fMRI data, we were able to delineate in detail, the time-resolved topography of information processing during memory encoding as well as retrieval, expanding on our knowledge of the perceptual forward stream as well as the memory-based reverse stream of information flow.

5.1 Specificity and robustness of the reverse stream effect

Previous studies found initial evidence that this hierarchical structure can be probed with reaction time measures, showing faster accessibility of semantic features compared to perceptual features during retrieval. However, so far this effect has only been shown under very specific conditions, where the association between action verbs and objects were learned (Lifanov et al., 2021; Linde-Domingo et al., 2019). After a break, the words were presented as cues and participants had to recall whether the associated object was either animate versus inanimate (semantic feature) or a line-drawing versus a photograph (perceptual feature). Accordingly, showing the generalizability and robustness of this effect would be a necessary to prove that feature-specific reaction times are able to probe the state of episodic LTM traces.

In a first sequence of studies, the type of semantic and perceptual features participants had to recall were extended. Perceptual object features were presentation size on screen (big vs small; Exp. 1) and the shape of an object (rounded vs elongated, Exp. 2), while semantic object features were always naturalness (manmade vs natural). When objects were directly classified from vision, perceptual classification was faster than semantic classification, in accordance with previous studies (Lifanov et al., 2021; Linde-Domingo et al., 2019) and indicating a forward stream of information processing during perception. As expected, the

opposite reaction time difference was observed in the memory task but reached only significance in Exp 1. Interestingly, we could not replicate this effect when instructing participants to differentiate between different object shapes instead of object sizes, while keeping all other aspects of the experiment identical. As object shape is an identifying feature with semantic content, we could show that the underlying factor causing reaction time differences truly is perceptual versus semantic content.

Previous studies all relied on the same kind of probe to induce recall, an action verb which had been learned alongside the object (Lifanov et al., 2021; Linde-Domingo et al., 2019). To rule out that written action-verbs facilitated semantic-feature recall over perceptual-feature recall, we tested different types of probes. As such, we asked participants to remember scenes alongside an object and an action verb. We then either probed the object via the scene or the action verb. In both cases, semantic differentiation led to faster reaction times than perceptual differentiation. Additionally, in another experiment, we used spatial location of the presentation of an object as recall probe rendering it inherently 'meaningless' on a semantic level. Nonetheless, we again could show that spatial location as probed for recall led to faster reaction times for semantic compared to perceptual features. Our results underline that the reversed hierarchical reactivation of the whole pattern that the memory trace binds together is not influenced by the nature of the partial pattern (cue).

Still, an alternative explanation would be that participants show a bias in learning, by focusing more on semantic compared to perceptual features, which in turn would make them accessible faster when recalling the object. To test this, we manipulated the attentional focus of participants during the learning phase of the experiment by letting them attend either the semantic or the perceptual category of an object in Exp. 6 or by letting participants form their associations between scene cues and objects on a deep or a shallow basis in Exp.7 (Craik &

Lockhart, 1972). In both experiments, an advantage of semantic over perceptual feature accessibility in reaction times could be observed independent of attentional manipulation.

When modelling all tasks mentioned above plus the original tasks presented in Linde-Domingo et al. (2019) in a meta-analytic GLMM, visual feature classification of objects shows a reliable forward stream of information while feature classification from memory shows a reliable reverse stream of information. These results indicate the ability of feature-specific reaction times to access and map temporal differences in the accessibility of features that comprise an episodic LTM trace.

5.2 Modelling the cognitive components of the forward and reverse stream

As robustness of the reverse stream effect could be reliably demonstrated, a next step would be to investigate the underlying cognitive components in both visual encoding (forward stream) and retrieval (reverse stream). We therefore decided to apply the well-established Hierarchical Drift Diffusion Model (HDDM; Ratcliff & McKoon, 2008; Wiecki et al., 2013), which estimates decision-related parameters based on reaction time distributions of each individual person. For example, it can be assumed that a fast integration of perceptual information necessary to reach a conclusion, implied by a steep decision slope, would lead to coherently fast responses. In contrast, a more difficult integration process, would be illustrated by a flat slope indicated by longer reaction times with a higher temporal variance between single responses (Arnold et al., 2015; Lerche et al., 2017; Lerche & Voss, 2017; Nunez et al., 2024; Voss et al., 2004).

The simplified version of the model divides reaction times into 4 different subcomponents of the decision process (van Ravenzwaaij et al., 2017; Wiecki et al., 2013). The starting point describes initial biases towards one of the categories. As our experiment was balanced on a 50% chance level for the categorization task, meaning each category was correct for the same number of trials, we fixed the starting point to have equal probabilities for both categories. The decision-based accumulation rate describes the speed of the decision process, illustrated by how fast decision-relevant information is integrated through the steepness of the decision slope. The decision boundary describes the point at which the decision is reached, illustrating the tradeoff between accuracy and speed of the decision. Lastly, the nondecision time refers to purely perceptual and motor aspects, incorporating the assumed length of the stimulus encoding time prior to the start of the decision process and the motor response after a decision has been made.

To first estimate these components for the recall phase in the reverse stream, we applied the model to all memory datasets presented in Chapter 2 of this thesis. As expected, results showed a clear difference between perceptual and semantic features on the decision-based accumulation rate, implying that semantic information is more readily accessible during retrieval than perceptual information. At the same time, no differences could be found in decision boundaries and nondecision time, which suggests that the reverse stream indeed describes the neuronal basis of memory recall and is not a confound of perceptual processes or individual decision bias.

Furthermore, we analyzed the forward stream by applying the HDDM model (Wiecki et al., 2013) to all six perceptual categorization datasets we collected. However, we only found a significant difference between perceptual and semantic features in the non-decision component, which is difficult to interpret and has not received a lot of attention in previous

literature. However, a possible explanation could pertain to the purely perceptual nature of the early phase of visual recognition and encoding, which is faster for perceptual compared to semantic features, leaving no apparent differences in the cognitive components of the decision process itself. Nonetheless, this subcomponent is not yet well understood and should therefore be interpreted with caution (Verdonck & Tuerlinckx, 2016).

In conclusion, these results provide first indications that differences in encoding semantic and perceptual categories could be based mainly on early perceptual stimulus processing instead of higher-level cognitive aspects, while for the reverse stream, differences could be caused by divergence in the speed of the integration of relevant information, cued by semantic features. Especially, the latter result provides additional evidence that feature-specific reaction times reflect a memory search process.

5.3 Time-resolved brain activity and localization of feature-specific processing differences

In the third chapter we tracked feature processing during retrieval by combining magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI) and Deep Neural Network (DNN) data. The image material originated from the Natural Scene Database (NSD). Here each image has been presented multiple times to eight participants during 7T fMRI scan, which is publicly available for further analysis (Allen et al., 2022).

Additionally, we recruited 29 healthy participants who took part in our MEG experiment. In line with previous studies (Lifanov et al., 2021; Linde-Domingo et al., 2019), they were instructed to learn associations between images (from the NSD) and words. After a short break, they were cued with the word and asked to remember and imagine the

associated image. Through lure recognition trials, we verified that memories were indeed correct, by showing a strongly occluded image coupled to a word and asking participants whether the association was correct or not. Participants had to recall each image six times. During this we recorded MEG to track time-resolved brain activity. A more detailed feature representation of NSD images was achieved by feeding images through biologically inspired visual and semantic Deep Neural Networks (Krizhevsky et al., 2017; Lin et al., 2014; Mikolov et al., 2013; Simonyan & Zisserman, 2015).

The comparison of each combination of image-specific pattern activations, allowed the transformation of each data modality into a representational geometry, which delineates how a brain region (fMRI), a point in time (MEG sensors) or a hidden layer (DNN) represents image specific features.

By correlating MEG and fMRI representational dissimilarity matrixes (RDMs), it is possible to extract information on the time at which, during encoding and retrieval, certain ROIs become active (Cichy, Khosla, et al., 2016; Cichy, Pantazis, et al., 2016; Cichy & Oliva, 2020). MEG-fMRI results revealed a clear forward stream (early visual regions showing peak correlations earlier in time than late lateral regions) during initial perception of images. Accordingly, during retrieval, there was some indication for a primacy of high-level reactivations during retrieval (mainly lateral regions correlations from 500 after cue onset). Specifically, we found long-lasting high-level visual activations (lateral), which were active prior to low-level visual areas (occipital) but continued on even after low-level activations ceased. This is partly in line with previous findings delineating the reverse stream (Lifanov et al., 2024; Linde-Domingo et al., 2019), where high-level features are accessible earlier in time compared to low-level features.

By correlating MEG and DNN RDMs, feature specific reactivations were modulated. Results of this analysis matched well with our prior MEG and fMRI results, showing that high-level semantic features were activated earlier than and outlasted low-level perceptual features. Although the continued activation of high-level areas and features was unexpected, it also reveals interesting details about the retrieval process, possibly indicating that low-level information is tapped into only for a brief period of time, while transformed more abstract high-level information is longer-lasting.

A split of the recall repetitions into the first 2 versus the last 2-4 recalls revealed in both MEG-to-fMRI and MEG-to-DNN fusion that early recalls mainly showed semantic feature and lateral ROI correlations while later recall repetitions showed mid-lateral and early visual ROI correlations and more high-visual features in a time window of 500 ms after cue onset. This is interesting as it shows that multiple recall attempts in short succession seem to strengthen perceptual details instead of a semantization effect that is usually seen after consolidation of memory (Heinen et al., 2023; Lifanov et al., 2021). These results could speak to frequency of recall dependent plastic adaptations, which strengthen the representations of perceptual features within the memory trace. A similar trend could also be observed in feature-specific reaction times, where the difference between perceptual and semantic feature accessibility diminishes over retrieval attempts within one day (Lifanov et al., 2021).

5.4 Implications and future directions

A general question in the field is in how far perceptual details of a memory trace are faithfully stored and reactivated alongside the gist information during the retrieval process or whether only the gist information is saved and perceptual details are reconstructed based

on gist information (Ahissar et al., 2009; Long & Kuhl, 2021; Moscovitch et al., 2016; Schacter et al., 1998). While our data cannot definitively answer this question, it challenges the notion that these are exclusive concepts. The strong semantic recall in initial recalls, could indicate that only gist information is stored and perceptual features are reconstructed along with recall repetitions. Alternatively, it is conceivable that (at least some) perceptual information is faithfully stored, while their retrieval depends on the strength of network weights, dependent on the importance of the episodic memory, e.g. the number of activations or recalls.

Regarding the organization of the long-term memory store, it is still a contested topic whether features are saved and accessible by themselves or whether the storage is object based through which an access of features is possible (Balaban et al., 2020; Brady et al., 2011). Research on forgetting of memories, shows a dependency of perceptual details on gist information, i.e. if the gist or the object identity is lost, low level information also becomes inaccessible (Balaban et al., 2020; Lifanov et al., 2021). In line with a reversal of the visual hierarchy during memory retrieval, object identity (e.g. dog) should be accessible even before an abstract semantic feature (e.g. animate) is available, which should be detectable with feature-specific reaction times and decoding of retrieval-related M/EEG signal.

In this line of research, the focus lied on vision, but it is possible that the reverse stream generalizes to other sensory domain as well. In the auditory domain, a hierarchical processing stream during perception exist similar to the visual domain (Rauschecker & Scott, 2009; Rauschecker & Tian, 2000). A general problem with testing the reverse stream effect in other modalities might be that participants use visual strategies to represent perceptual features during learning. To remedy this, special populations could be tested that are known

for an impoverished or missing visual imagery abilities like in aphantasia (Zeman et al., 2015).

In the future, the method of feature-specific reactions times might be useful for clinical or pedagogical settings where retrieval related components are important as is prevalent, for example, in dyslexia or other learning disorders (Martínez-Briones et al., 2023; Schulz et al., 2008). In fact, in children with developmental learning disorder, it could be shown that while the lexical information store was intact, semantic retrieval from this store was slowed down (Mengisidou et al., 2020). But while children with dyslexia show slowed down semantic retrieval, they display increased accuracy in recognizing incidentally learned objects (i.e. participants were unaware of the subsequent memory task) in a semantic categorization task (Hedenius et al., 2013). Pinpointing the part of the encoding and retrieval process which is affected might be aided by feature-specific reaction time analysis, which could greatly enhance our understanding of these diseases and potentially pave the way for new treatment options. Accordingly, it would be crucial to analyze the detailed state of memory traces from encoding to retrieval in children with developmental learning disorders and dyslexia. Understanding the underlying deficits in detail, could open possibilities for targeted interventions aimed at training perceptual-semantic associations.

5.5 Conclusion

In this thesis, forward stream encoding as well as reverse stream retrieval of episodic memories were analyzed thoroughly regarding their feature representations (low-level perceptual versus high-level semantic) robustness, generalizability, timing, and brain topography. To achieve this, we conducted behavioral reaction time experiments, established

computational models, analyzed combinations of MEG and fMRI data as well as Deep Neural Networks.

We could show that the reverse stream is well generalizable and robust, varying several perceptual and semantic features and cueing modalities. In general, the semantic content of a memory trace is more rapidly available for recall compared to the perceptual content.

Moreover, while the faster accessibility of perceptual compared to semantic information during perception might be due to the necessities of the hierarchical stimulus encoding of the forward stream, our data suggests that the speed of retrieval through the reverse stream relies on the speed at which relevant information can be sampled and integrated from memory.

Lastly, we were able to corroborate our findings with fMRI and deep neural networks, which show that low-level areas and features are active prior to high-level areas and features during encoding. During retrieval, high-level features and areas are active prior to and outlast a brief activation of low-level perceptual areas and features.

We provide strong evidence for a reversed hierarchical processing during memory retrieval in the visual domain, linking the architecture of the brain to the ever-elusive nature of the (young) LTM engram or trace. This could increase the insight into memory encoding and retrieval, but it might be of great value in the research of pathological conditions in the future. As such, while it is known that there are deficits in encoding or retrieval in conditions such as dyslexia, dementia and Schizophrenia, the underlying circuit deficits are unclear. Our framework could therefore provide valuable insight, which might pave the way for the development of new treatment options in the future.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends in Cognitive Sciences, 8(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. In Philosophical Transactions of the Royal Society B: Biological Sciences (Vol. 364, Issue 1515, pp. 285–299). Royal Society. https://doi.org/10.1098/rstb.2008.0253

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. Computational Psychiatry, 1(0), 24. https://doi.org/10.1162/cpsy_a_00002

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. Nature Neuroscience, 25(1), 116–126. https://doi.org/10.1038/s41593-021-00962-x

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a Fast Route to Memory Consolidation. Trends in Cognitive Sciences, 21(8), 573–576. https://doi.org/10.1016/j.tics.2017.05.001

Arbuckle, S. A., Yokoi, A., Pruszynski, J. A., & Diedrichsen, J. (2019). Stability of representational geometry across a wide range of fMRI activity levels. NeuroImage, 186(November 2018), 155–163. https://doi.org/10.1016/j.neuroimage.2018.11.002

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. Psychological Research, 79(5), 882–898. https://doi.org/10.1007/s00426-014-0608-y

Aschenbrenner, A. J., Balota, D. A., Gordon, B. A., Ratcliff, R., & Morris, J. C. (2016). A diffusion model analysis of episodic recognition in preclinical individuals with a family history for Alzheimer's disease: The adult children study. Neuropsychology, 30(2), 225–238. https://doi.org/10.1037/neu0000222

Baddeley, A. D. (1986). Working Memory. Clarendon Press. https://books.google.de/books?id=ZKWbdv__vRMC

Balaban, H., Assaf, D., Arad Meir, M., & Luria, R. (2020). Different features of real-world objects are represented in a dependent manner in long-term memory. Journal of Experimental Psychology: General, 149(7), 1275–1293. https://doi.org/10.1037/xge0000716

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? Perception, 1(4), 371–394. https://doi.org/10.1068/P010371

Barron, H. C., Dolan, R. J., & Behrens, T. E. J. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. Nature Neuroscience, 16(10), 1492–1498. https://doi.org/10.1038/nn.3515

Bastin, J., Committeri, G., Kahane, P., Galati, G., Minotti, L., Lachaux, J. P., & Berthoz, A. (2013). Timing of posterior parahippocampal gyrus activity reveals multiple scene processing stages. Human Brain Mapping, 34(6), 1357–1370. https://doi.org/10.1002/hbm.21515

Benson, N. C., Butt, O. H., Datta, R., Radoeva, P. D., Brainard, D. H., & Aguirre, G. K. (2012). The retinotopic organization of striate cortex is well predicted by surface

topology. Current Biology, 22(21), 2081–2085. https://doi.org/10.1016/j.cub.2012.09.014

Berkay, D., Eser, H. Y., Sack, A. T., Çakmak, Y. Ö., & Balcı, F. (2018). The modulatory role of pre-SMA in speed-accuracy tradeoff: A bi-directional TMS study. Neuropsychologia, 109, 255–261. https://doi.org/10.1016/j.neuropsychologia.2017.12.031

Berryhill, M. E. (2012). Insights from neuropsychology: Pinpointing the role of the posterior parietal cortex in episodic and working memory. In Frontiers in Integrative Neuroscience (Issue JUNE 2012). https://doi.org/10.3389/fnint.2012.00031

Betancourt, M. (2016). Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo. http://arxiv.org/abs/1604.00695

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., Krypotos, A. M., Lerche, V., Logan, G. D., Palmeri, T. J., van Ravenzwaaij, D., Servant, M., Singmann, H., Starns, J. J., Voss, A., Wiecki, T. V., Matzke, D., & Wagenmakers, E. J. (2018). Estimating across-trial variability parameters of the Diffusion Decision Model: Expert advice and recommendations. Journal of Mathematical Psychology, 87, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004

Bolam, J., Boyle, S. C., Ince, R. A. A., & Delis, I. (2022). Neurocomputational mechanisms underlying cross-modal associations and their influence on perceptual decisions. NeuroImage, 247(December 2021), 118841. https://doi.org/10.1016/j.neuroimage.2021.118841

Bone, M. B., Ahmad, F., & Buchsbaum, B. R. (2020). Feature-specific neural reactivation during episodic memory. Nature Communications, 11(1), 1–13. https://doi.org/10.1038/s41467-020-15763-2

Bone, M. B., & Buchsbaum, B. R. (2021). Detailed Episodic Memory Depends on Concurrent Reactivation of Basic Visual Features within the Posterior Hippocampus and Early Visual Cortex. Cerebral Cortex Communications, 2(3). https://doi.org/10.1093/texcom/tgab045

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. Journal of Vision, 11(5), 4–4. https://doi.org/10.1167/11.5.4

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. Proceedings of the National Academy of Sciences of the United States of America, 105(38), 14325–14329. https://doi.org/10.1073/PNAS.0803390105

Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision, 10(4), 433–436. http://www.ncbi.nlm.nih.gov/pubmed/9176952

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase ii: 930 new normative photos. PLoS ONE, 9(9). https://doi.org/10.1371/journal.pone.0106953

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. Cognitive Psychology, 57(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction Time for Object Categorization Is Predicted by Representational Distance. Journal of Cognitive Neuroscience, 26(1), 132–142. https://doi.org/10.1162/jocn_a_00476

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic

programming language. Journal of Statistical Software, 76(1). https://doi.org/10.18637/jss.v076.i01

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nature Neuroscience, 14(11), 1462–1467. https://doi.org/10.1038/nn.2925

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. Nature Neuroscience, 20(1), 115–125. https://doi.org/10.1038/nn.4450

Cichy, R. M., Khosla, A., Pantazis, D., & Oliva, A. (2017). Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. NeuroImage, 153(November 2015), 346–358. https://doi.org/10.1016/j.neuroimage.2016.03.063

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Scientific Reports, 6(January), 1–13. https://doi.org/10.1038/srep27755

Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. Neuron, 107(5), 772–781. https://doi.org/10.1016/j.neuron.2020.07.001

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. Nature Neuroscience, 17(3), 455–462. https://doi.org/10.1038/nn.3635

Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object

Recognition. Cerebral Cortex, 26(8), 3563–3579. https://doi.org/10.1093/cercor/bhw135

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. Science, 210(4466), 207–210. https://doi.org/10.1126/SCIENCE.7414331

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage, 19(2), 261–270. https://doi.org/10.1016/S1053-8119(03)00049-1

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Danker, J. F., & Anderson, J. R. (2010). The Ghosts of Brain States Past: Remembering Reactivates the Brain Regions Engaged During Encoding. Psychological Bulletin, 136(1), 87–102. https://doi.org/10.1037/a0017937

de Haan, E. H. F., & Cowey, A. (2011). On the usefulness of "what" and "where" pathways in vision. Trends in Cognitive Sciences, 15(10), 460–466. https://doi.org/10.1016/j.tics.2011.08.005

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience, 4(8), 2051–2062. https://doi.org/10.1523/jneurosci.04-08-02051.1984

Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than

knowing. Consciousness and Cognition, 15(1), 147–162. https://doi.org/10.1016/j.concog.2005.05.002

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding,. In PLoS Computational Biology (Vol. 13, Issue 4). https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1005508&type=printable

Favila, S. E., Lee, H., & Kuhl, B. A. (2020). Transforming the Concept of Memory Reactivation. Trends in Neurosciences, 43(12). https://doi.org/10.1016/j.tins.2020.09.006

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex, 1(1), 1–47. http://www.ncbi.nlm.nih.gov/pubmed/1822724

Ferreira, C. S., Charest, I., & Wimber, M. (2019). Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. NeuroImage, 201(June), 115996. https://doi.org/10.1016/j.neuroimage.2019.07.009

Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. Behavioral Neuroscience, 120(3), 497–517. https://doi.org/10.1037/0735-7044.120.3.497

Geerligs, L., Cam-CAN, & Henson, R. N. (2016). Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. NeuroImage, 135, 16–31. https://doi.org/10.1016/j.neuroimage.2016.04.047

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis. Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, 7(4), 409–435. https://doi.org/10.1214/ss/1177011136

Gimbel, S. I., & Brewer, J. B. (2011). Reaction time, memory strength, and fMRI activity during memory retrieval: Hippocampus and default network are differentially responsive during recollection and familiarity judgments. Cognitive Neuroscience, 2(1), 19–26. https://doi.org/10.1080/17588928.2010.513770

Gluth, S., Sommer, T., Rieskamp, J., & Büchel, C. (2015). Effective Connectivity between Hippocampus and Ventromedial Prefrontal Cortex Controls Preferential Choices from Memory. Neuron, 86(4), 1078–1090. https://doi.org/10.1016/j.neuron.2015.04.023

Goodale, M. A., & Milner, A. D. (1992a). Separate visual pathways for perception and action. Trends in Neurosciences, 15(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Goodale, M. A., & Milner, A. D. (1992b). Separate visual pathways for perception and action. Trends in Neurosciences, 15(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1714). https://doi.org/10.1098/rstb.2016.0102

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied

to Time Series Neuroimaging Data. Journal of Cognitive Neuroscience, 29(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. Journal of Neurophysiology, 35(1), 96–111. https://doi.org/10.1152/JN.1972.35.1.96

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. NeuroImage, 173(February), 434–447. https://doi.org/10.1016/j.neuroimage.2018.02.044

Hannula, D. E., & Ranganath, C. (2009). The Eyes Have It: Hippocampal Activity Predicts Expression of Memory in Eye Movements. Neuron, 63(5), 592–599. https://doi.org/10.1016/j.neuron.2009.08.025

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. Annual Review of Neuroscience, 37(1), 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science (New York, N.Y.), 293(5539), 2425–2430. https://doi.org/10.1126/SCIENCE.1063736

Hedenius, M., Ullman, M. T., Alm, P., Jennische, M., & Persson, J. (2013). Enhanced Recognition Memory after Incidental Encoding in Children with Developmental Dyslexia. PLoS ONE, 8(5). https://doi.org/10.1371/JOURNAL.PONE.0063998

Heinen, R., Bierbrauer, A., Wolf, O. T., & Axmacher, N. (2023). Representational formats of human memory traces. Brain Structure and Function, 0123456789. https://doi.org/10.1007/s00429-023-02636-9

Henriksson, L., Khaligh-Razavi, S. M., Kay, K., & Kriegeskorte, N. (2015). Visual representations are dominated by intrinsic fluctuations correlated between areas. NeuroImage, 114, 275–286. https://doi.org/10.1016/j.neuroimage.2015.04.026

Horner, A. J., Bisby, J. A., Bush, D., Lin, W. J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. Nature Communications, 6(May). https://doi.org/10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. Journal of Experimental Psychology: General, 142(4), 1370–1383. https://doi.org/10.1037/a0033626

Johansson, R., Nyström, M., Dewhurst, R., & Johansson, M. (2022). Eye-movement replay supports episodic remembering. Proceedings of the Royal Society B: Biological Sciences, 289(1977), 1–29. https://doi.org/10.1098/rspb.2022.0964

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, Familiarity, and Cortical Reinstatement: A Multivoxel Pattern Analysis. Neuron, 63(5), 697–708. https://doi.org/10.1016/j.neuron.2009.08.011

Johnson, J. D., Minton, B. R., & Rugg, M. D. (2008). Content dependence of the electrophysiological correlates of recollection. NeuroImage, 39(1), 406–416. https://doi.org/10.1016/j.neuroimage.2007.08.050

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. Nature Reviews Neuroscience, 16(9), 521–534. https://doi.org/10.1038/nrn4000

Kaniuth, P., & Hebart, M. N. (2022). Feature-reweighted representational similarity analysis: A method for improving the fit between computational models, brains, and behavior. NeuroImage, 257(November 2021), 119294. https://doi.org/10.1016/j.neuroimage.2022.119294

Kensinger, E. A., Ullman, M. T., & Corkin, S. (2001). Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: Evidence from amnesic patient H.M. Hippocampus, 11(4), 347–360. https://doi.org/10.1002/HIPO.1049

Kerrén, C., Linde-Domingo, J., & Spitzer, B. (2023). Prioritization of semantic over visuo-perceptual aspects in multi-item working memory. BioRxiv. https://doi.org/10.1101/2022.06.29.498168

Kolibius, L. D., Roux, F., Parish, G., Ter Wal, M., Van Der Plas, M., Chelvarajah, R., Sawlani, V., Rollings, D. T., Lang, J. D., Gollwitzer, S., Walther, K., Hopfengärtner, R., Kreiselmeyer, G., Hamer, H., Staresina, B. P., Wimber, M., Bowman, H., & Hanslmayr, S. (2023). Hippocampal neurons code individual episodic memories in humans. Nature Human Behaviour, 7(11), 1968–1979. https://doi.org/10.1038/s41562-023-01706-6

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. Journal of Experimental Psychology. General, 139(3), 558–578. https://doi.org/10.1037/A0019165

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene Memory Is More Detailed Than You Think: The Role of Categories in Visual Long-Term Memory. Psychological Science, 21(11), 1551. https://doi.org/10.1177/0956797610385359

Konkle, T., & Oliva, A. (2012). A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. Neuron, 74(6), 1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036

Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger

adults: testing the semantic categorization account. Journal of Experimental Psychology. Learning, Memory, and Cognition, 29(4), 499–510. https://doi.org/10.1037/0278-7393.29.4.499

Kraemer, P. M., & Gluth, S. (2023). Episodic Memory Retrieval Affects the Onset and Dynamics of Evidence Accumulation during Value-based Decisions. Journal of Cognitive Neuroscience, 35(4), 692–714. https://doi.org/10.1162/jocn_a_01968

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. Nature Reviews Neuroscience, 12(4), 217–230. https://doi.org/10.1038/nrn3008

Kriegeskorte, N. (2009). Relating population-code representations between man, monkey, and computational models. Frontiers in Neuroscience, 3(DEC), 363–373. https://doi.org/10.3389/neuro.01.035.2009

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. Trends in Cognitive Sciences, 17(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2(NOV), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron, 60(6), 1126. https://doi.org/10.1016/J.NEURON.2008.10.043

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386

Lerche, V., & Voss, A. (2017). Retest reliability of the parameters of the Ratcliff diffusion model. Psychological Research, 81(3), 629–652. https://doi.org/10.1007/s00426-016-0770-5

Lerche, V., & Voss, A. (2019). Experimental validation of the diffusion model based on a slow response time paradigm. Psychological Research, 83(6), 1194–1209. https://doi.org/10.1007/s00426-017-0945-8

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. Behavior Research Methods, 49(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2

Lerche, V., & Voss, A. (2020). When accuracy rates and mean response times lead to false conclusions: A simulation study based on the diffusion model. The Quantitative Methods for Psychology, 16(2), 107–119. https://doi.org/10.20982/tqmp.16.2.p107

Lifanov, J., Griffiths, B. J., Linde-Domingo, J., Ferreira, C. S., Wilson, M., Mayhew, S. D., Charest, I., & Wimber, M. (2024). Reconstructing Spatio-Temporal Trajectories of Visual Object Memories in the Human Brain. BioRxiv, 2022.12.15.520591. https://doi.org/10.1101/2022.12.15.520591

Lifanov, J., Linde-Domingo, J., & Wimber, M. (2021). Feature-specific reaction times reveal a semanticisation of memories over time and with repeated remembering. Nature Communications, 12(1), 1–10. https://doi.org/10.1038/s41467-021-23288-5

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland (pp. 740–755). https://doi.org/10.1007/978-3-319-10602-1_48

Linde-Domingo, J. (2019). Retrieval dynamics in episodic memory – from computations to representations [Thesis (Doctorates > Ph.D.), University of Birmingham, UK]. http://etheses.bham.ac.uk/id/eprint/9325

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

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. Nature Communications, 10(1). https://doi.org/10.1038/s41467-018-08080-2

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. Journal of Neuroscience, 7(11), 3416–3468. https://doi.org/10.1523/jneurosci.07-11-03416.1987

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Frontiers in Psychology, 6(August), 1–16. https://doi.org/10.3389/fpsyg.2015.01171

Loaiza, V. M., McCabe, D. P., Youngblood, J. L., Rose, N. S., & Myerson, J. (2011). The Influence of Levels of Processing on Recall From Working Memory and Delayed Recall Tasks. Journal of Experimental Psychology: Learning Memory and Cognition, 37(5), 1258–1263. https://doi.org/10.1037/a0023923

Long, N. M., & Kuhl, B. A. (2021). Cortical Representations of Visual Stimuli Shift Locations with Changes in Memory States. Current Biology, 31(5), 1119-1126.e5. https://doi.org/10.1016/j.cub.2021.01.004

Lüken, M., Heathcote, A., Haaf, J. M., & Matzke, D. (2025). Parameter identifiability in evidence-accumulation models: The effect of error rates on the diffusion decision model and the linear ballistic accumulator. Psychonomic Bulletin & Review, 32(3), 1411–1424. https://doi.org/10.3758/s13423-024-02621-1

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: A manifesto for change. Journal of Experimental Psychology: General, 142(4), 1180–1189. https://doi.org/10.1037/a0033650

Marr, D. (1971). Simple memory: a theory for archicortex. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 262(841), 23–81. https://doi.org/10.1098/RSTB.1971.0078

Martin, C. B., & Barense, M. D. (2023). Perception and Memory in the Ventral Visual Stream and Medial Temporal Lobe. Annual Review of Vision Science, 9(1), 409–434. https://doi.org/10.1146/annurev-vision-120222-014200

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L. Y., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. ELife, 7, 1–29. https://doi.org/10.7554/eLife.31873

Martínez-Briones, B. J., Fernández, T., & Silva-Pereyra, J. (2023). Semantic Priming and Its Link to Verbal Comprehension and Working Memory in Children with Learning Disorders. Brain Sciences 2023, Vol. 13, Page 1022, 13(7), 1022. https://doi.org/10.3390/BRAINSCI13071022

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychological Review, 102(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. Journal of Memory and Language, 66(3), 416–437. https://doi.org/10.1016/j.jml.2011.12.001

Mecklinger, A., Rosburg, T., & Johansson, M. (2016). Reconstructing the past: The late posterior negativity (LPN) in episodic memory studies. Neuroscience and Biobehavioral Reviews, 68, 621–638. https://doi.org/10.1016/j.neubiorev.2016.06.024

Mengisidou, M., Marshall, C. R., & Stavrakaki, S. (2020). Semantic fluency difficulties in developmental dyslexia and developmental language disorder (DLD): poor semantic structure of the lexicon or slower retrieval processes? International Journal of Language & Communication Disorders, 55(2), 200–215. https://doi.org/10.1111/1460-6984.12512

Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. PLoS Biology, 14(8). https://doi.org/10.1371/journal.pbio.1002528

Michelmann, S., Staresina, B. P., Bowman, H., & Hanslmayr, S. (2019). Speed of time-compressed forward replay flexibly changes in human episodic memory. In Nature Human Behaviour (Vol. 3, Issue 2, pp. 143–154). Nature Publishing Group. https://doi.org/10.1038/s41562-018-0491-4

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings, 1–12.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. Neuropsychologia, 46(3), 774–785. https://doi.org/10.1016/j.neuropsychologia.2007.10.005

Milner, B. (1965). Physiologie de l'Hippocampe: Colloque International, No. 107, Editions du Centre National de la Recherche Scientifique, Paris, 1962. 512 pp. 58NF. Neuropsychologia, 3, 273–277. https://api.semanticscholar.org/CorpusID:141235351

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. Neuropsychologia, 6(3), 215–234. https://doi.org/10.1016/0028-3932(68)90021-3

Minxha, J., Adolphs, R., Fusi, S., Mamelak, A. N., & Rutishauser, U. (2020). Flexible recruitment of memory-based choice representations by the human medial frontal cortex. Science, 368(6498), eaba3313. https://doi.org/10.1126/science.aba3313

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983a). Object vision and spatial vision: two cortical pathways. Trends in Neurosciences, 6(C), 414–417. https://doi.org/10.1016/0166-2236(83)90190-X

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983b). Object vision and spatial vision: two cortical pathways. Trends in Neurosciences, 6, 414–417. https://doi.org/10.1016/0166-2236(83)90190-X

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: The hippocampus and neocortex in transformation. Annual Review of Psychology, 67, 105–134. https://doi.org/10.1146/annurev-psych-113011-143733

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. In Journal of Mathematical Psychology (Vol. 53, Issue 4, pp. 222–230). Academic Press Inc. https://doi.org/10.1016/j.jmp.2009.02.003

Nunez, M. D., Fernandez, K., Srinivasan, R., & Vandekerckhove, J. (2024). A tutorial on fitting joint models of M/EEG and behavior to understand cognition. Behavior Research Methods. https://doi.org/10.3758/s13428-023-02331-x

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. Psychological Review, 110(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nature Neuroscience, 15(12), 1729–1735. https://doi.org/10.1038/nn.3248

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. Cognitive Science, 38(6), 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. Trends in Cognitive Sciences, 6(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. Behavior Research Methods, 51(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. Journal of Neuroscience, 26(35), 8965–8975. https://doi.org/10.1523/JNEUROSCI.1655-06.2006

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-Specific Cortical Activity Precedes Retrieval During Memory Search. Science, 310(5756), 1963–1966. https://doi.org/10.1126/SCIENCE.1117645

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature 2005 435:7045, 435(7045), 1102–1107. https://doi.org/10.1038/nature03687

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. Psychological Bulletin, 86(3), 446–461. https://doi.org/10.1037/0033-2909.86.3.446

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? Psychonomic Bulletin and Review, 15(6), 1218–1228. https://doi.org/10.3758/PBR.15.6.1218

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. Decision, 2(4), 237–279. https://doi.org/10.1037/dec0000030

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation, 20(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proceedings of the National Academy of Sciences of the United States of America, 106(16), 6539–6544. https://doi.org/10.1073/pnas.0812589106

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. Trends in Cognitive Sciences, 20(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., College, B. M., & Mckoon, G. (2011). Effects of Aging and IQ on Item and Associative Memory. Journal of Experimental Psychology: General, 140(3), 464–487. https://doi.org/10.1037/aü023810

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. Journal of Memory and Language, 50(4), 408–424. https://doi.org/10.1016/j.jml.2003.11.002

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. Nature Neuroscience, 12(6), 718–724. https://doi.org/10.1038/nn.2331

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proceedings of the National Academy of Sciences of the United States of America, 97(22), 11800–11806. https://doi.org/10.1073/pnas.97.22.11800

Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: Insights from functional brain imaging. Annual Review of Psychology, 63, 101–128. https://doi.org/10.1146/annurev-psych-120710-100344

Robin, J. (2018). Spatial scaffold effects in event memory and imagination. Wiley Interdisciplinary Reviews: Cognitive Science, 9(4), 1–15. https://doi.org/10.1002/wcs.1462

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. Current

Opinion in Behavioral Sciences, 17, 114–123. https://doi.org/10.1016/j.cobeha.2017.07.016

Robin, J., & Olsen, R. K. (2019). Scenes facilitate associative memory and integration. Learning and Memory, 26(7), 252–261. https://doi.org/10.1101/lm.049486.119

Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember—know paradigm. Psychonomic Bulletin & Review, 15(4), 825–832. https://doi.org/10.3758/PBR.15.4.825

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. Trends in Cognitive Sciences, 11(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. Nature, 392(6676), 595–598. https://doi.org/10.1038/33396

Schacter, D. L. (1987). Implicit expressions of memory in organic amnesia: learning of new facts and associations. Human Neurobiology, 6(2), 107–118. http://www.ncbi.nlm.nih.gov/pubmed/3624022

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The Cognitive Neuroscience of Constructive Memory. Annual Review of Psychology, 49(1), 289–318. https://doi.org/10.1146/annurev.psych.49.1.289

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual Differences in Components of Reaction Time Distributions and Their Relations to Working Memory and Intelligence. Journal of Experimental Psychology: General, 136(3), 414–429. https://doi.org/10.1037/0096-3445.136.3.414

Schmolck, H., Kensinger, E. A., Corkin, S., & Squire, L. R. (2002). Semantic knowledge in patient H.M. and other patients with bilateral medial and lateral temporal lobe lesions. Hippocampus, 12(4), 520–533. https://doi.org/10.1002/HIPO.10039

Schulz, E., Maurer, U., van der Mark, S., Bucher, K., Brem, S., Martin, E., & Brandeis, D. (2008). Impaired semantic processing during sentence reading in children with dyslexia: Combined fMRI and ERP evidence. NeuroImage, 41(1), 153–168. https://doi.org/10.1016/J.NEUROIMAGE.2008.02.012

Scoville, W. B., & Milner, B. (1957). Loss Of Recent Memory After Bilateral Hipoocampal Lesions. Journal of Neurology, Neurosurgery & Psychiatry, 20(1), 11–21. https://doi.org/10.1136/jnnp.20.1.11

Semon, R. W. (1906). Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens. Nature, 73(1893), 338–338. https://doi.org/10.1038/073338a0

Shadlen, M. N. N., & Shohamy, D. (2016). Decision Making and Sequential Sampling from Memory. Neuron, 90(5), 927–939. https://doi.org/10.1016/j.neuron.2016.04.036

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. Journal of Neurophysiology, 86(4), 1916–1936. https://doi.org/10.1152/jn.2001.86.4.1916

Sheldon, S., & Chu, S. (2017). What versus where: Investigating how autobiographical memory retrieval differs when accessed with thematic versus spatial information. Quarterly Journal of Experimental Psychology, 70(9), 1909–1921. https://doi.org/10.1080/17470218.2016.1215478

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–14.

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. Journal of Experimental Psychology: Learning Memory and Cognition, 32(1), 101–117. https://doi.org/10.1037/0278-7393.32.1.101

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychological Review, 99(2), 195–231. https://doi.org/10.1037/0033-295X.99.2.195

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. Current Opinion in Neurobiology, 5(2), 169–177. https://doi.org/10.1016/0959-4388(95)80023-9

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. In Annual Review of Neuroscience (Vol. 27, pp. 279–306). https://doi.org/10.1146/annurev.neuro.27.070203.144130

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. Science (New York, N.Y.), 253(5026), 1380–1386. https://doi.org/10.1126/SCIENCE.1896849

Staresina, B. P., Michelmann, S., Bonnefond, M., Jensen, O., Axmacher, N., & Fell, J. (2016). Hippocampal pattern completion is linked to gamma power increases and alpha power decreases during recollection. ELife, 5(AUGUST). https://doi.org/10.7554/eLife.17397.001

Staresina, B. P., Reber, T. P., Niediek, J., Boström, J., Elger, C. E., & Mormann, F. (2019). Recollection in the human hippocampal-entorhinal cell circuitry. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-09558-3

Staresina, B. P., & Wimber, M. (2019). A Neural Chronometry of Memory Recall. Trends in Cognitive Sciences, 23(12), 1071–1085. https://doi.org/10.1016/j.tics.2019.09.011

Stewart, G. B., Altman, D. G., Askie, L. M., Duley, L., Simmonds, M. C., & Stewart, L. A. (2012). Statistical Analysis of Individual Participant Data Meta-Analyses: A Comparison of Methods and Recommendations for Practice. PLoS ONE, 7(10), 1–8. https://doi.org/10.1371/journal.pone.0046042

Suzuki, W. A., & Amaral, D. G. (1994). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. Journal of Neuroscience, 14(3 II), 1856–1877. https://doi.org/10.1523/jneurosci.14-03-01856.1994

ter Wal, M., Linde-Domingo, J., Lifanov, J., Roux, F., Kolibius, L. D., Gollwitzer, S., Lang, J., Hamer, H., Rollings, D., Sawlani, V., Chelvarajah, R., Staresina, B., Hanslmayr, S., & Wimber, M. (2021). Theta rhythmicity governs human behavior and hippocampal signals during memory-dependent tasks. Nature Communications, 12(1), 1–15. https://doi.org/10.1038/s41467-021-27323-3

Teyler, T. J., & DiScenna, P. (1986). The Hippocampal Memory Indexing Theory. Behavioral Neuroscience, 100(2), 147–154. https://doi.org/10.1037/0735-7044.100.2.147

Teyler, T. J., & Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory: Updating the index. Hippocampus, 17(12), 1158–1169. https://doi.org/10.1002/hipo.20350

Thakral, P. P., Madore, K. P., & Schacter, D. L. (2017). A role for the left angular gyrus in episodic simulation and memory. Journal of Neuroscience, 37(34), 8142–8149. https://doi.org/10.1523/jneurosci.1319-17.2017

Townsend, J. T., & Ashby, F. G. (1984). Stochastic Modeling of Elementary Psychological Processes. Cambridge University Press. https://books.google.de/books?id=k8uMswEACAAJ

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. Cerebral Cortex, 24(4), 883–897. https://doi.org/10.1093/cercor/bhs364

Tulving, E. (1983). Elements of Episodic Memory. Oxford University Press.

Tulving, E. (2002). Episodic Memory: From Mind to Brain. Annual Review of Psychology, 53(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Turner, B. M., van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: The neural drift diffusion model. Psychological Review, 122(2), 312–336. https://doi.org/10.1037/a0038894

Van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. Trends in Neurosciences, 35(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. Psychonomic Bulletin and Review, 24(2), 547–556. https://doi.org/10.3758/s13423-016-1081-y

van Vugt, M. K., Beulen, M. A., & Taatgen, N. A. (2016). Is there neural evidence for an evidence accumulation process in memory decisions? Frontiers in Human Neuroscience, 10(MAR2016), 1–13. https://doi.org/10.3389/fnhum.2016.00093

van Vugt, M. K., Beulen, M. A., & Taatgen, N. A. (2019). Relation between centro-parietal positivity and diffusion model parameters in both perceptual and memory-based

decision making. Brain Research, 1715(March), 1–12. https://doi.org/10.1016/j.brainres.2019.03.008

Verdonck, S., & Tuerlinckx, F. (2016). Factoring out nondecision time in choice reaction time data: Theory and implications. Psychological Review, 123(2), 208–218. https://doi.org/10.1037/rev0000019

Vijayarajah, S., McAlister, E., & Schlichting, M. L. (2023). Encoding-phase orientation toward thematic content over perceptual style benefits picture memory. Memory, 31(2), 259–269. https://doi.org/10.1080/09658211.2022.2147954

von Krause, M., Radev, S. T., & Voss, A. (2022). Mental speed is high until age 60 as revealed by analysis of over a million participants. Nature Human Behaviour 2022 6:5, 6(5), 700–708. https://doi.org/10.1038/s41562-021-01282-7

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. Memory & Cognition, 32(7), 1206–1220.

Wagenmakers, E. J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. Psychonomic Bulletin and Review, 14(1), 3–22. https://doi.org/10.3758/BF03194023

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. Trends in Cognitive Sciences, 9(9), 445–453. https://doi.org/10.1016/j.tics.2005.07.001

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. NeuroImage, 137, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. Cerebral Cortex, 25(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Wei, P., Bao, R., & Fan, Y. (2022). Independence of functional connectivity analysis in fMRI research does not rely on whether seeds are exogenous or endogenous. Medicine in Novel Technology and Devices, 15, 100126. https://doi.org/10.1016/j.medntd.2022.100126

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. Proceedings of the National Academy of Sciences of the United States of America, 97(20), 11125–11129. https://doi.org/10.1073/pnas.97.20.11125

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in Python. Frontiers in Neuroinformatics, 7(JULY 2013), 1–10. https://doi.org/10.3389/fninf.2013.00014

Wimmer, G. E., & Shohamy, D. (2012). Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. Science, 338(6104), 270–273. https://doi.org/10.1126/science.1223252

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. In Journal of the International Neuropsychological Society (Vol. 17, Issue 5, pp. 766–780). Cambridge University Press. https://doi.org/10.1017/S1355617711000683

Wittkuhn, L., Eppinger, B., Bartsch, L. M., Thurm, F., Korb, F. M., & Li, S.-C. (2018). Repetitive transcranial magnetic stimulation over dorsolateral prefrontal cortex modulates value-based learning during sequential decision-making. NeuroImage, 167, 384–395. https://doi.org/10.1016/j.neuroimage.2017.11.057

Wixted, J. T. (2009). Remember/Know judgments in cognitive neuroscience: An illustration of the underrepresented point of view. In Learning and Memory (Vol. 16, Issue 7, pp. 406–412). https://doi.org/10.1101/lm.1312809

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970

Xue, G. (2018). The Neural Representations Underlying Human Episodic Memory. In Trends in Cognitive Sciences (Vol. 22, Issue 6, pp. 544–561). Elsevier Ltd. https://doi.org/10.1016/j.tics.2018.03.004

Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nature Neuroscience, 19(3), 356–365. https://doi.org/10.1038/nn.4244

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language, 46(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery – Congenital aphantasia. Cortex, 73, 378–380. https://doi.org/10.1016/J.CORTEX.2015.05.019