



**UNIVERSIDAD
DE GRANADA**



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**

Master's degree in Cognitive Neuroscience and Clinical Neuropsychology
University of Padova

Master's degree in Cognitive and Behavioral Neuroscience
University of Granada

Master's Thesis:

**The Best Cues Evolve with Memory:
a New Challenge for the Encoding Specificity
Hypothesis**

Department of General Psychology (UniPD)
Department of Experimental Psychology (UGR)
Mind, Brain and Behavior Research Center (CIMCYC)

Supervisor:

Juan Linde Domingo

Co-Supervisor:

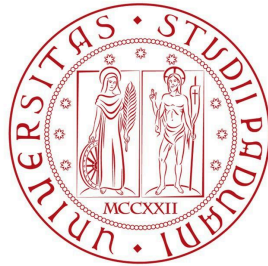
Carlos González García

Student:

Mattia Delmarco

Academic Year:


2023-2024



DECLARATION OF AUTHORSHIP AND ORIGINALITY FOR MASTER'S DISSERTATIONS

Mattia Delmarco, with ID document no.: CA55211FS, a student on the Master's Degree in Cognitive and Behavioural Neuroscience at the Faculty of Psychology, in relation to the Master's Dissertation "The Best Cues Evolve With Memory: a New Challenge for the Encoding Specificity Hypothesis" submitted for defence and evaluation in the 2023/2024 academic year, hereby declares that they attest to the originality of this work, understood in the sense that they have not used sources without duly citing them.

Granada, 28 June 2024


Mattia Delmarco

Contents

Abstract.....	5
Introduction.....	6
Neural Bases of Cued Recall and Memory Trace Evolution.....	7
Experiment 1.....	9
Methods.....	10
Participants.....	10
Apparatus.....	10
Stimuli.....	10
Procedure.....	11
Encoding phase.....	11
Distractor Phase.....	12
Retrieval and Recognition Phase.....	12
Design.....	13
Software.....	14
Pre-registered Analyses Plan.....	14
Additional Analyses.....	14
Results.....	15
Slopes Comparison for RT in the Retrieval Task.....	15
Bootstrapping Analysis on RT.....	16
Post-Hoc Power Analyses on RT.....	17
Comparison Recognised as Identical vs. Actual Identical.....	17
RM ANOVAs on RT.....	17
Slopes Comparison for Accuracy in the Retrieval Task.....	17
Bootstrapping Analysis on Accuracy.....	18
Post-Hoc Power Analyses on Accuracy.....	19
Comparison Recognised as Identical vs. Actual Identical.....	19
RM ANOVAs on Accuracy.....	19
Accuracy in the Recognition Task.....	20
Summary of the Results.....	20
Experiment 2.....	21
Methods.....	21
Participants.....	21
Apparatus and Stimuli.....	22
Procedure.....	22
Design, Software and Analyses.....	22
Results.....	22
Slopes Comparison for RT in the Retrieval Task.....	22
Bootstrapping Analysis on RT.....	22
Post-Hoc Power Analyses on RT.....	23
Comparison Recognised as Identical vs. Actual Identical.....	24
RM ANOVAs on RT.....	24
Slopes Comparison for Accuracy in the Retrieval Task.....	24

	4
Bootstrapping Analysis on Accuracy.....	24
Post-Hoc Power Analyses on Accuracy.....	25
Comparison Recognised as Identical vs. Actual Identical.....	26
RM ANOVAs on Accuracy.....	26
Accuracy in the Recognition Task.....	26
Comparison Between Experiments.....	26
Summary of the Results.....	27
Discussion.....	27
Limitations and Conclusion.....	31
References.....	32

Abstract

The Encoding Specificity Hypothesis, a cornerstone in memory research, posits that the successful recollection of an event is cue dependent and relies on the degree of match between cues present at retrieval and encoding conditions. Here, we challenge this principle by proposing that, since neural memory traces and their psychological experience change over time, the optimal cue should change accordingly, matching the current state of the encoded memory. We examined this hypothesis using a paired-associate paradigm, where participants encoded pairs of objects and recalled one of them upon being presented with the other encoded object (i.e., identical cue), a similar version of it or a different version of it. We tracked the current state of the memory by means of a recognition task on the cues, which could follow (Experiment 1) or precede (Experiment 2) the retrieval of the paired associate. In Experiment 1, participants were faster retrieving a target memory when presented with cues recognised as identical, followed by recognised as similar and then different. This effect was stronger than the one of the actual similarity of the cues. Furthermore, non-encoded cues recognised as identical led to faster response times (RTs) compared to encoded identical cues, contrasting with the Encoding Specificity Hypothesis. No significant difference was observed for the accuracy at retrieval and Experiment 2 failed to replicate the effect on RT, likely due to the early recognition phase reactivating the event memory. We discuss the results as preliminary evidence of a potentially necessary perspective change: from encoding-retrieval match, to cue updating based on memory evolution.

Keywords: Encoding Specificity Hypothesis, cue updating, memory evolution, cue similarity, paired associate paradigm

The Best Cues Evolve with Memory: a New Challenge for the Encoding Specificity Hypothesis

“We remember an event if it has left behind a trace and if something reminds us of it” (Tulving, 1974, p. 74). Memory cues, often conceptualized as specific pieces of information present in the individual’s cognitive environment and linked to the memory of the event (Tulving, 1974), are necessary for successful retrieval (Semon, 1896). Even in the absence of a clear external, physical cue (e.g., the color of the screen; Dulsky, 1935), internal aspects (e.g., pharmacological state, mood; Eich et al., 1975; Eich & Metcalfe, 1989) or contextual ones (Godden & Baddeley, 1975; Reddy & Bellezza, 1983) might play a crucial role (Roediger et al., 2017). Indeed, upon being provided with the right cue, people can retrieve the original event even when memories seem to have faded (Tulving, 1974) or to be affected by clinical conditions (e.g., in Alzheimer disease and Korsakoff disease; Kirk & Berntsen, 2018; Mcdowall, 1979). These findings highlight a fundamental difference between memory availability and memory accessibility (Tulving & Pearlstone, 1966), reinforcing the idea of the cue dependency of memory retrieval (Frankland et al., 2019).

More in general, the successful retrieval of an encoded event highly depends on the type of cue provided, with certain cues being better than others (Roediger et al., 2017). One type of task that has been widely used to assess the efficiency of distinct kinds of cues in eliciting the reinstatement of the memory is the “paired associate learning task” (Calkins, 1896). In this task, one stimulus is presented with or in temporal proximity to another stimulus; at test, only one of the elements of the pair is shown and participants are asked to remember its associate. For instance, if an association between pairs of words has been encoded, the presentation of one of them at test (i.e., cue) would elicit the recall of the associated word (Tulving & Osler, 1968). Thus, by varying the type of cue presented (e.g., presenting colors and words or images and words; Roediger et al., 2017), it is possible to determine the degree to which certain stimuli are more effective than others in triggering the episodic recollection of the entire event (Thomson & Tulving, 1970). For instance, pictures and objects are overall more effective than words as cues to allow the recall of a paired associate (Iodice et al., 2015; Kirk & Berntsen, 2018), while odor cues are generally more related to a stronger emotionality of the memory reinstated (Herz, 1998).

Crucially, the paired associate task evolved into more complex versions in which the cue provided at test was not necessarily one of the stimuli presented during the study phase, but an item thought to have an effect on memory retrieval (Roediger et al., 2017). For instance, Thomson and Tulving (1970) presented pairs of words that were only weakly related at a semantic level; at test, either one of the two words or a word that was strongly associated with the target word was provided as cue. What they observed is that, even though the degree of semantic association between pairs of words generally determines the extent to which participants will remember the paired associate (Low & Roder, 1983), weakly related words can be more effective if they are the only words paired with the to be remembered words at encoding (Thomson & Tulving, 1970).

This effect turned out to be very consistent across most of the paired associate studies (Roediger et al., 2017), forming the basis of the “Encoding Specificity Hypothesis” (Tulving & Thomson, 1973): “retrieval success is dictated by the extent to which the context (or cues) at retrieval matches that which was present during

encoding” (Guskjolen & Cembrowski, 2023, p. 3212). After its initial proposal, this hypothesis turned quickly into a more general principle due to the high number of studies that replicated the effect across different experimental conditions, paradigms and even internal processes: better recall if the mental operation performed at encoding matches the one at retrieval (Morris et al., 1977), if the global environmental context, or one of its relevant features, remains unchanged (Godden & Baddeley, 1975; Geiselman & Glenny, 1977), if the mood is consistent across the experiment phases (Eich & Metcalfe, 1989), or if the same body posture and physiological state are repeated (Dijkstra et al., 2007; Miles & Hardman, 1998).

Nevertheless, several limitations of the Encoding Specificity Hypothesis have also been reported (for a review, see Roediger et al., 2017). For instance, Postman (1975) did not replicate the effect observed by Thomson and Tulving (1970), finding that non-encoded but strongly semantically related words were more effective cues than encoded but weakly related words. This result aligns with a previous finding by Santa and Lamwers (1974), who demonstrated that the effect of non-encoded strongly related cues depends on whether participants are instructed on how to use them to recall the target words. Since Thomson and Tulving (1970) did not provide specific instructions, Santa and Lamwers (1974) concluded that the Encoding Specificity Principle could be more an experimental artifact than a real phenomenon. Furthermore, the encoding specificity findings failed to replicate also when other types of “cues” were investigated: mental operations (Dewhurst & Brandt, 2007; Dewhurst & Knott, 2010), general context (Bjork & Richardson-Klavehn, 2014; Fernandez & Glenberg, 1985) and mood dependency (Bower et al., 1978). Thus, even by taking a more moderate perspective with respect to Santa and Lamwers (1974)’s, the Encoding Specificity Principle cannot explain many of the findings present in the literature and evidence for the effect with specific types of cues (e.g., mood dependency) is sparse and requires additional research (Roediger et al., 2017).

Neural Bases of Cued Recall and Memory Trace Evolution

Further understating of the Encoding Specificity Principle and its possible limitations comes from a significant body of research that has investigated the neural bases of cued recall. This research builds on the idea that if we remember specific events only when they left a trace and a cue can tap into it (Tulving, 1974), then memory retrieval at the neural level can be understood as a cue-induced behavioral expression of the engram (Frankland et al., 2019). The engram, or memory trace, is defined as the physical change that happens in the nervous system whenever we learn something (Guskjolen & Cembrowski, 2023; Josselyn et al., 2015). This change, that occurs through plasticity, involves neurons from multiple and functionally connected brain regions (i.e., “unified engram complex”; Roy et al., 2022), and it is associated with single cells (Guskjolen & Cembrowski, 2023) as well as neural population modifications (DeNardo et al., 2019).

Cues are deemed capable of reinstating the memory of the encoded event by means of a hippocampal mechanism called “pattern completion” (McClelland & Goddard, 1996; Rolls, 2013). After the presentation of a cue (which can be seen as a partial information with respect to the originally encoded event), the hippocampus would compare its neural representation with the stored representations of already encoded events, reinstating the most similar stored representation to the current input cue (thus, “completing” the partial representation of the cue; Knierim & Neuneubel,

2016; Rolls, 2013). A subsequent old/new signal would track the degree of similarity of the neural representation of the cue with the most similar stored representation (Bakker et al., 2008; Chen et al., 2011). If the similarity is sufficiently high, the memory of the associated stored event is quickly (~1500 ms after stimulus onset; Staresina & Wimber, 2019) reinstated (Chen et al., 2011; Knierim & Neuneubel, 2016). This process involves the reactivation of cortical neurons whose synaptic connections with the hippocampus were reinforced during encoding (a concept defined as “hippocampal index”; Teyler & DiScenna, 1986) and it is not limited to the recollection of single instances, but encompasses also the recollection of multi-element events (e.g., recall two objects when a third one, which was associated with the other two, is presented; Horner & Burgess, 2014; Horner et al., 2015).

In accordance with what reported so far, the Encoding Specificity Principle would predict that the maximal reactivation of the engram should be elicited by the cue that mostly matches the overall encoding conditions (Frankland et al., 2019; Jung et al., 2023). Crucially, this view necessarily assumes a complete staticity of the memory trace, since the same exact cue presented at encoding is predicted to always be the one capable of tapping into the memory trace of the encoded event. However, it is known that memory traces change over time, moving from a hippocampal dependent representation to a more distributed neocortical one (Sekeres et al., 2018), with a corresponding change in its psychological value: memories lose their episodic details to become more gist-like (Gilboa & Moscovitch, 2021), following a semanticization trend (Lifanov et al., 2021; Krenz et al., 2023). Moreover, every time a memory is retrieved, a re-encoding process happens (Moscovitch et al., 2016; Shohamy & Wagner, 2008), causing the updating of the memory and its consequent modification based on the contextual information at hand (Schacter et al., 2011) and schemas/prototypes (Kerrén et al., 2024). These findings are reinforced by a growing line of research that reported some encoding-retrieval dissimilarities in the memory trace. For instance, genetic labeling studies with mice highlighted that retrieval memory traces can involve different populations of neurons compared to encoding (DeNardo et al., 2019). This aligns with recent functional magnetic resonance imaging (fMRI) studies that have found differences between encoding and retrieval in the voxels and brain areas involved in the representation of certain items, suggesting that “the encoded represented [*sic*] might be transformed and reinstated across brain regions” (Xiao et al., 2017, p. 2996). Similar results were obtained in studies using representational similarity analysis (RSA) and time-resolved multivariate decoding of electroencephalogram (EEG) data, confirming that the encoding representations of specific items are not necessarily reinstated in the subsequent retrieval phase (Liu et al., 2021) and they are accompanied by a reversal flow of the information (Linde-Domingo et al., 2019). On the one hand, during encoding, above chance decodability of low-level perceptual features precedes the one for more conceptual features. On the other hand, this pattern reverses during retrieval, indicating faster decodability of high level conceptual information compared to low level perceptual one (Linde-Domingo et al., 2019).

Overall, these results suggest that the encoded representations keep transforming, with substantial changes found even within the encoding phase and between an initial short term memory phase and a long term memory one (Liu et al., 2021). Accordingly, since memory retrieval is cue dependent (Tulving, 1974), one

should expect not only an overall change of the representation of the original event (Liu et al., 2021; Winocur & Moscovitch, 2011), but also of the most effective cue that can tap into that (new) representation. In other words, if what is needed to reinstate a memory is a (partial) match between the current input representation (i.e., cue) and the stored representation of the event (Knierim & Neuneubel, 2016), then if the stored representation changes, the cue that can tap into it must change accordingly.

Therefore, we propose a new perspective on the effectiveness of different types of cues in triggering the episodic recollection of an event, predicting that originally encoded cues are not necessarily the best cues (as suggested by the Encoding Specificity Hypothesis; Tulving & Thomson, 1973) when confronted to versions of them that are closer to the *current* state of the memory. Here, we refer to “best cues” as those cues that elicit a higher performance in terms of accuracy (Thomson & Tulving, 1970) and speed (i.e., memory strength and ease of access to the memory; Madigan et al., 2000; Ratcliff, 1978) in the retrieval of the paired associate. Conversely, the “current memory state” is determined by means of a recognition task *on the cues*: we argue that if participants recognised the cue as being the one encoded (i.e., the current state of the memory corresponds to the image shown) this would have led to an overall facilitation in the retrieval of the paired associate, regardless of whether that cue was the actually encoded one.

Experiment 1

In order to test this prediction at a behavioral level, we created a new paired associate task in which participants had to encode pairs of pictures of objects and recall one of them upon being presented with an identical picture of the other object of the pair, a similar version of it or a different version of it. This type of manipulation is not a novelty in the memory field, as some studies reported that the higher the degree of similarity of the cues with respect to the encoded ones, the higher the performance of the participants at retrieval, with identical cues being always the most effective ones, as predicted by the Encoding Specificity Hypothesis (Smith et al., 2013; Yum, 1931). However, these studies did not control for the subjective recognition of the cues. Thus, it is not possible to ascertain whether the performance benefit was driven by the actual similarity of the cues or by the subjective experience of those cues as identical, similar or different with respect to the encoded ones. In Experiment 1, we aimed at determining this by directly comparing the effects of the actual similarity of the cues and their subjective recognition. By considering the trials in which the subjective recognition of the cue does not correspond to its actual similarity with the encoded one (i.e., non-matching trials), we will be able to determine whether the encoding specificity results depend on the characteristics of the cue itself or on the subjective experience of the participants. In particular, we aimed at testing three main hypotheses:

- H1: We hypothesized that the subjective recognition of the cues will have an effect on the response time (RT) for correct retrieval responses and accuracy in the retrieval task. In particular, we predicted increasingly higher RT and lower accuracy moving from cues recognised as identical to cues recognized as different, regardless of the actual similarity of the cues.
- H2: We hypothesized that the effect of the subjective recognition of the cues will be stronger than the one of their actual similarity. In particular, in non-matching

trials, the subjective recognition of the cues, but not their actual similarity, will determine RT (for correct retrieval responses) and accuracy scores.

- H3: We hypothesized that, in non-matching trials, cues recognised as identical will lead to faster RTs (for correct retrieval responses) and overall higher accuracy compared to actual identical cues.

Methods

This study was pre-registered and the pre-registration document can be found at the following link:

<https://docs.google.com/document/d/1N97XTDCQO3znINrtF-MdMD7I26qiP-lkAZEAcBYvOoU/edit?usp=sharing>

Participants

Forty-eight young adults were recruited through the SONA platform (<https://ugr-cimcyc.sona-systems.com/>) of the “Centro de Investigación Mente, Cerebro y Comportamiento” (CIMCYC) of the University of Granada, and were offered course credits as compensation for participating in the study. Following the pre-registered plan, three participants had to be removed from the dataset since their accuracy in either the recognition or the retrieval task was above 1.5 times the interquartile range (IQR) with respect to the 3rd quartile or below 1.5 IQR with respect to the 2nd quartile. Thus, the final sample was composed of 45 participants (35 females), of age between 18 and 37 ($M = 22.6$, $SD = 3.48$). Ethics approval was obtained before data collection. In order to estimate the sample size needed to reach a power above 0.80 in the main analysis planned on RT (i.e., comparison of the slopes of the two factors; see Pre-registered Analyses Plan), we performed Monte Carlo sampling iteratively with a different number of participants (up to 200 possible participants) on an ideal dataset created modeling the distribution of the data observed in a pilot study. For each sample size, we estimated the power as the proportion of simulations where the Bayes Factor for the Bayesian t-test comparing the two conditions under investigation was above 3 or below $\frac{1}{3}$. A sample size of at least 40 was deemed necessary.

Apparatus

Participants were tested on an Intel Core i7-3770 3.40 GHz computer running PsychoPy software (version 2023.2.3; Peirce et al., 2019) and responded to stimuli presented on a 19-inch BenQ LCD monitor (1920 x 1080 pixels) at a viewing distance of about 57 cm.

Stimuli

Throughout the experiment, the screen background was white, while the fixation cross and the displayed texts were black. Experimental material consisted of coloured object images taken from the Object Memorability Image Normed Database Software (O-MINDS; Duncan Lab, 2022) and from the Objects and Similar Lures Database of the Mnemonic Similarity Task (MST, Stark Lab; Kirwan & Stark, 2007). Each image was 128 pixels wide by 128 pixels tall. The stimuli set was composed of a total of 288 object images, equally divisible in four main everyday-life categories: food, clothes, tools and electronic devices (Figure 1). Memory targets and their paired associates (i.e., cues; Figure 2a) were taken from two fixed and distinct sets of 72 images (18 images per object category). However, each of the 72 paired objects had a similar and a different counterpart, defined with respect to their degree of perceptual similarity with the originally presented information. In both cases, the same object (i.e., same identity)

was represented, but similar cues presented some small perceptual differences (i.e., the original cue was rotated or was slightly modified in terms of colors or features; Figure 2b), while different cues presented larger ones (e.g., totally different color with respect to the one of the original cue; Figure 2c).

Figure 1

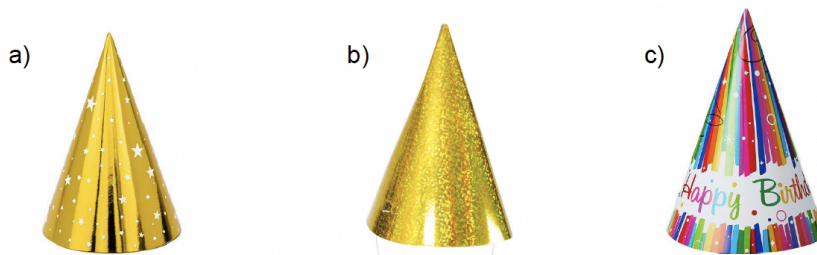
Examples of Stimuli From Each Category



Note. Examples of stimuli from each of the four categories used in the study: a) Food, b) Clothes, c) Tools and d) Electronic Devices. The distinction between tools and electronic devices was based on electricity usage.

Figure 2

Types of Cue



Note. Example of the distinct kinds of cue categories for the same object: a) original cue: a replica of the object presented during the encoding phase; b) similar cue: a slightly modified version of the original cue; c) different cue: a significantly altered version of the original cue.

Procedure

After providing voluntary informed consent, participants were invited to carefully read the instructions on the screen, notifying the experimenter once they had finished reading them. This step allowed the experimenter to ask participants to repeat the instructions, clarifying any doubts before the actual experiment started and ensuring that participants thoroughly understood the various tasks of the experiment. The instructions explained in detail the three main phases of each block, providing examples of what the participants should have done in each of the tasks. Each block was composed by an encoding and a retrieval phase, interleaved by a distractor task (Figure 3). A total of 6 blocks were presented and the three phases were repeated in each block (in succession), but with different stimuli.

Encoding phase. Participants were shown 12 pairs of objects and asked to memorize each couple and the specific perceptual details of each image, for subsequent testing (Figure 3a). In particular, in order to elicit a deeper processing of the visual stimuli (Craik & Lockhart, 1972), participants were instructed to memorize the pair by imaging an interaction between the two objects: for instance, a camera and a hammer could have been memorized by imaging hammering the camera. Each

encoding trial started with a central fixation cross that lasted for 0.5 seconds. At this time point, two object images were presented, one on the right and one on the left side of the screen. Participants could press the spacebar to move to the next pair but, if they did not do that before 8 seconds, the program would have automatically progressed to the next trial. A total of 72 pairs was presented across blocks and each of them was composed of a “target” object and an “original cue” object. Both the 72 target objects (18 instances per category) and the 72 cue objects (18 instances per category) were randomly distributed across blocks, without any repetition. Thus, the object pairs presented differed across participants, since the specific associations between the 72 targets and the 72 original cues were randomized. Also the position (left or right part of the screen) of both cues and targets was randomized across trials, so that no predictable pattern could influence the participants' responses and learning process. Crucially, participants did not know which of the object presented would have been used as target and which one as cue in the subsequent recall phase, an uncertainty designed to prevent any bias in attention allocation during encoding, ensuring that participants memorized both objects in a similar way (Craik & Lockhart, 1972).

Distractor Phase. In this second phase of the block, participants were asked to determine whether a single-digit number was odd or even (Figure 3b). In particular, in each trial, a random single digit between 1 and 9 appeared on the screen for 1.5 seconds. Participants were asked to respond within this time-frame by pressing the right arrow key for even and the left one for odd numbers. In all the 36 trials (~90 seconds, in total), responses were accompanied by a visual feedback: the screen background would have turned red in case of wrong responses and green in case of correct ones. The aim of this phase of the study was to prevent rehearsal by distracting participants with another task and also to introduce a controlled time interval between the encoding and the retrieval phase.

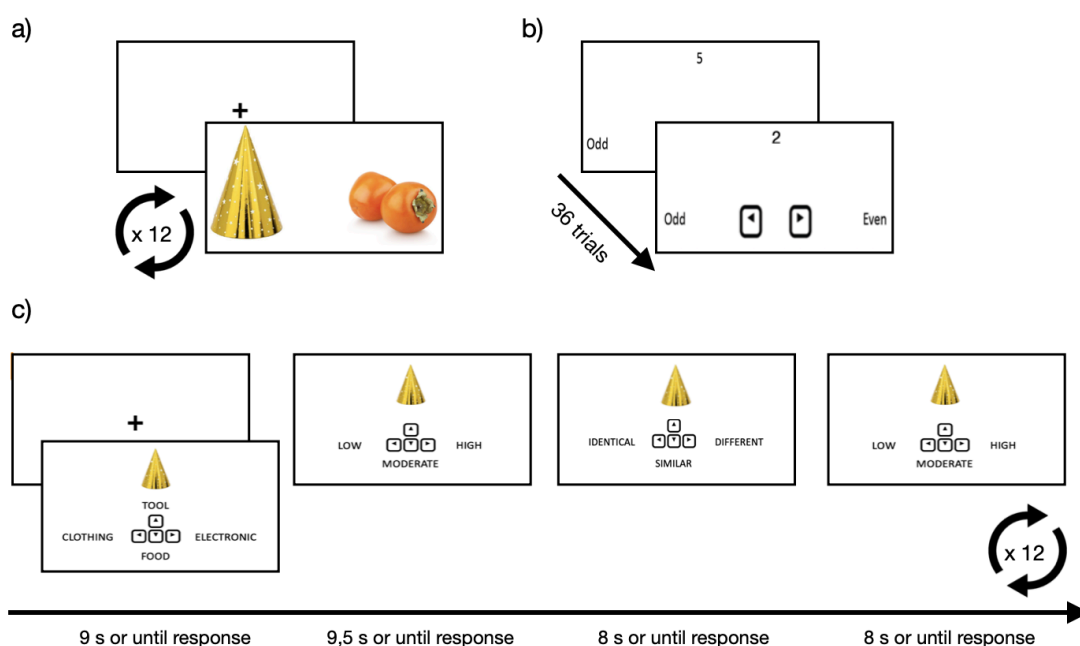
Retrieval and Recognition Phase. This last phase of the block was composed of two different tasks and each of the 12 trials included both of them, in succession (Figure 3c): right after completing the retrieval task for a given pair, participants performed the recognition task for the same cue shown during retrieval.

Retrieval Task. After a fixation interval of 0.5 seconds, a text appeared at the bottom of the screen, indicating which arrow key corresponded to which object category (this correspondence remained the same throughout the experiment and across participants). Starting from 2 seconds post fixation onset, participants were shown, for a maximum of 9 seconds, one of the objects presented at encoding and asked to recall the category of the associated target object (Figure 1), which was not presented on the screen. Crucially, the object presented (i.e., cue) could either be the same object shown during the encoding phase (“original cue”), a similar version of it (“similar cue”) or a different version of it (“different cue”; Figure 2), and participants knew about this manipulation, as it was explained in the initial instructions. Through a process of controlled randomisation, 6 objects per category (24 in total) were selected from the pool of the 72 original cues to be presented during this retrieval phase. The remaining 48 original cues were not used in this phase; instead, they were replaced by their similar or different counterparts: 24 objects (6 per category) per type. However, even though an equal number of original, similar and different cues were presented across blocks, their number within blocks was not kept constant, but randomized, so as to ensure that participants couldn't predict the sequence of cue types. Furthermore, the

order of cue presentation was also randomized, so that it did not correspond to the order in which participants saw the stimuli at encoding. This manipulation allowed to control for possible sequence memory effects (Kahana, 1996). Immediately after choosing the category of the paired object, participants were asked to express their confidence in the choice made. The object presented remained on the screen, while a new text indicated which arrow key to press, depending on the level of confidence: either high (right arrow key), moderate (down arrow key) or low (left arrow key). Again, either the key press or a 9.5 seconds time limit determined the end of this second task.

Recognition Task. Directly after the confidence rating, participants kept seeing the same object in the center of the screen and a new instruction asked them to determine whether the object they were seeing was the same object presented at encoding (left arrow key), a different object (right arrow key) or a similar object (down arrow key). Each trial lasted up to 8 seconds, although pressing one of the arrow keys would terminate it prematurely. Afterwards, participants were again questioned about the confidence in their choice, following the same procedure as in the previous task.

Figure 3
Sequence of Events Within Block



Note. a) Each block started with an encoding phase, in which 12 pairs of objects were presented, interleaved by a fixation cross. b) After the 12 encoding trials, the odd/even distractor task was presented, for a total of 36 trials. c) Finally, in the retrieval and recognition phase, participants were first asked to retrieve the category of the target stimulus associated to the cue presented; then, to rate their degree of confidence about the decision; subsequently, to recognise whether the cue presented was identical, similar or different to the one presented at encoding; lastly, to provide their degree of confidence with regards to this last choice. This sequence was repeated 12 times per block.

Design

This research employed a within subject experimental design. In both recognition and retrieval tasks, participants were assigned to three within-subjects

conditions based on the type of cue presented (“Cue Condition”: identical vs. similar vs. different); further three within-subjects conditions for the retrieval task were defined a posteriori depending on participants’ responses in the recognition task (“Subjective Recognition”: recognised as identical vs. recognised as similar vs. recognised as different). For this second independent variable, the number of trials per condition was determined a posteriori depending on participants’ responses.

For both tasks, the main dependent variables were accuracy (defined as number of correct trials on total number of trials) and RT. For the RT analyses, we considered only correct trials since they ensure that the RT measures reflect successful memory retrieval.

Software

The analyses were conducted using PyCharm (Professional Version 2024.1, JetBrains), RStudio (Version 2024.04.1) and G*Power (Version 3.1). PyCharm was used for writing the Python (Version 3.12.2) code for data cleaning, main statistical analyses and plots. The Python libraries used were: NumPy (2.0.0rc2), Pandas (2.2.2), Matplotlib (3.9.0), SciPy (1.14.0rc2), pingouin(0.5.4), statsmodels (0.14.2). Rstudio was used to run an a priori Bayesian Power Analysis for Experiment 1, using the libraries tidyverse_2.0.0 and BayesFactor_0.9.12-4.7. Finally, we used G*Power to perform the post-hoc Power Analyses of Experiment 1 and the a priori and post-hoc Power Analyses of Experiment 2.

Pre-registered Analyses Plan

In order to assess whether the two factors had an effect on the RT at retrieval and that the effect of Subjective Recognition was stronger than the one of Cue Condition, we computed the mean RT (only for correct retrieval trials) at retrieval per level of Subjective Recognition and Cue Condition, separately, per each participant. Afterwards, we fitted two linear regression models per participant, one considering Cue Condition as a unique predictor and the other one with Subjective Recognition as a unique predictor. We then extracted the slopes from each model and determined whether the slopes of Subjective Recognition and Cue Condition were significantly different from 0, by means of two Wilcoxon Signed Rank Tests. Furthermore, a one-tailed Wilcoxon Signed Rank Test was conducted to determine whether the slopes of Subjective Recognition were steeper than the ones of Cue Condition. The same type of analyses was repeated considering the accuracy at retrieval as the unique dependent variable. Finally, we conducted again the above reported analyses in the subsample of data of incorrect recognition responses (i.e., non-matching trials), so as to verify whether the effect of Cue Condition could depend on the overlap with the Subjective Recognition factor.

Additional Analyses

We repeated the one-tailed Wilcoxon Signed Rank Tests to compare the slopes of Cue Condition and Subjective Recognition on RT and accuracy, after equating the number of trials per each level of the main factors and per participant. Through a bootstrapping procedure, we first determined the minimum number of trials per participant by selecting the level with the lowest number of trials, within factor. Afterwards, we randomly sampled the same number of trials from the other two levels and averaged the RT/accuracy per level and per participant. Then, we fitted two linear regression models per factor and per participant, extracted the slopes and compared them by means of a one-tailed Wilcoxon Signed Rank Test. This procedure was

repeated 1000 times, randomizing every time the trials selected for the two levels with the higher number of trials. Through the distribution of Wilcoxon Test statistics and p-values obtained, we could determine the stability of the result across multiple resampled datasets.

Additionally, we performed post-hoc power analyses of the Wilcoxon Signed Rank Tests comparing the slopes of Subjective Recognition and Cue Condition on the RT and accuracy at retrieval (including also the tests on non-matching trials).

Furthermore, we conducted two one-tailed Wilcoxon Signed Rank Tests to determine whether cues recognised as identical led to faster RT and higher accuracy compared to actual identical cues.

Moreover, we performed four separate repeated measures (RM) ANOVAs with Cue Condition or Subjective Recognition as main within-subjects factors and RT (only for correct retrieval trials) or accuracy as a unique dependent variable. These analyses were needed to assess whether there were significant differences between the levels of the two main factors and determine if the expected gradient-like pattern was present (Smith et al., 2013; Yum, 1931). In case of violation of the sphericity assumption, we applied a Greenhouse Geisser correction (Blanca et al., 2023a), while RM ANOVA was considered robust enough against the normality assumption (Blanca et al., 2023b). However, parallel non parametric Friedman's tests were conducted to assure consistency in the results (we reported the results of the Friedman's tests only in case of inconsistencies). Post-hoc analyses have been performed only in case of significant Fs and corrected for multiple comparisons by means of a Bonferroni correction. Following the same rationale as for the analyses on the slopes, the same four RM ANOVAs have been also performed in the subset of non-matching trials, determining whether potential significant results of the previous four ANOVAs could depend on the overlap between factors.

Finally, we also assessed whether the accuracy at recognition varied significantly across the levels of Cue Condition by means of a RM ANOVA.

Results

Slopes Comparison for RT in the Retrieval Task

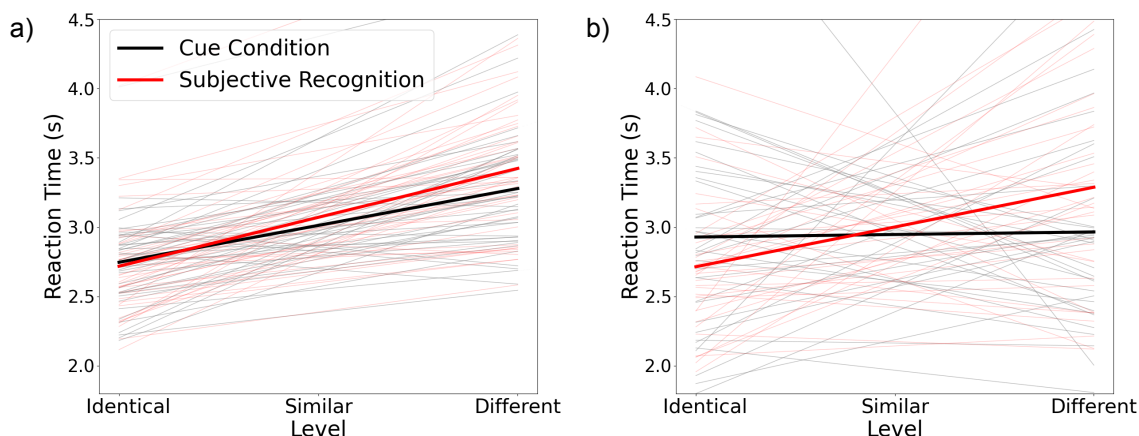
On average, participants needed 3.00 seconds (SD = 0.33) to respond correctly in the retrieval task. To assess whether there was an effect of Subjective Recognition and Cue Condition on the RT for correct retrieval trials, we conducted two Wilcoxon Signed-Rank Tests on the slopes extracted from the linear regression models. Both the slopes of Subjective Recognition (M=0.35, SD= 0.24; W= 7.0, $p < .001$, $r_b = 0.99$) and Cue Condition (M= 0.27, SD= 0.22; W= 34.0, $p < .001$, $r_b = 0.93$) differed significantly from 0. Furthermore, another one-tailed Wilcoxon Signed-Rank Test indicated that the slopes of Subjective Recognition were significantly steeper than the ones of Cue Condition (W= 747.0, $p = .004$, $r_b = 0.44$; Figure 4a).

Considering the high rate of correct responses at recognition (M= 0.62, SD= 0.08), we hypothesized that the effect of Cue Condition might depend on the overlap between the two factors. Indeed, when considering non-matching trials, two Wilcoxon Signed-Rank Tests indicated that the slopes of Subjective Recognition (M= 0.29, SD= 0.53) were significantly different from 0 (W= 153.0, $p = .004$, $r_b = 0.54$), but not the ones of Cue Condition (M= 0.02, SD= 0.57; W= 313.0, $p = .762$, $r_b = 0.06$). Furthermore, a

one-tailed Wilcoxon Test confirmed that the slopes of Subjective Recognition were greater than the ones of Cue Condition ($W= 455.0$, $p= .028$, $r_b= 0.37$; Figure 4b).

Figure 4

RT Regression Lines: Cue Condition vs. Subjective Recognition



Note. Regression lines obtained from the data of Experiment 1 and indicating: a) an overall stronger effect of Subjective Recognition on the RT for correct retrieval trials, compared to Cue Condition; b) the unique effect of Subjective Recognition on the RT for correct retrieval responses when considering non-matching trials. In the figure, light red and gray lines represent the regression lines per participant for Subjective Recognition and Cue Condition, respectively. The two main lines correspond to their average. The initial tick of the y axis, which was 1.8 seconds, was removed for clarity.

Bootstrapping Analysis on RT

Since the number of trials for each level of the two main factors varied depending on NaN values and the participants' responses in the recognition task (Table 1), we evaluated the robustness of the results by bootstrapping to the minimum number of trials per participant. We performed again the one-tailed Wilcoxon Test on the slopes of Cue Condition and Subjective Recognition, but iteratively, for 1000 times. The Wilcoxon Test was significant ($p < .05$) for 829 of the 1000 iterations (82.9%), indicating a moderate to high robustness of the results obtained.

Table 1

Number of Trials per Factor Level - RT

Level	Cue Condition	Subjective Recognition
1	930	1110
2	896	870
3	829	675

Note. The column level represents either 1 = "identical", 2 = "similar", 3 = "different", for Cue Condition; or 1 = "recognised as identical", 2 = "recognised as similar", 3 = "recognised as different", for Subjective Recognition

Post-Hoc Power Analyses on RT

A post-hoc power analysis indicated that the achieved power for the Wilcoxon Signed Rank Test comparing the slopes of Cue Condition and Subjective Recognition was 0.80, confirming that the study was well powered to capture the medium effect size we obtained ($r_b = 0.44$; Goss-Sampson, 2018).

With what concerns the one-tailed Wilcoxon Test on the subsample of data (i.e., non-matching trials), a post-hoc power analysis indicated that we achieved a power of 0.63. Thus, the study was slightly underpowered. Indeed, in order to detect our medium effect size ($r_b = 0.37$; Goss-Sampson, 2018) with a power of 0.80 and considering a significance criterion of $\alpha = 0.05$, we would have needed 57 participants.

Comparison Recognised as Identical vs. Actual Identical

Next, we tested whether there was a significant difference between cues recognised as identical and actual identical cues, when the two conditions didn't match. A one-tailed Wilcoxon Test confirmed that cues recognised as identical ($M = 2.73$, $SD = 0.53$) led to lower RTs for correct retrieval trials compared to actual identical cues ($M = 3.02$, $SD = 0.53$; $W = 301.0$, $p = .030$, $r_b = -0.33$).

RM ANOVAs on RT

We performed four independent RM ANOVAs to assess whether there was a significant difference between the levels of each factor. The first RM ANOVA revealed that the effect of Subjective Recognition on the RT at retrieval was statistically significant, $F(2, 88) = 59.49$, $p < .001$, $\eta^2 = 0.30$. Post-hoc comparisons highlighted significant differences between the levels of Subjective Recognition: cues recognised as identical elicited a lower RT ($M = 2.75$, $SD = 0.33$) compared to cues recognised as similar ($M = 3.00$, $SD = 0.47$; $t(44) = -4.42$, $p < .001$) or different ($M = 3.46$, $SD = 0.52$; $t(44) = -9.97$, $p < .001$). Furthermore, the RT at retrieval for cues recognised as similar was significantly lower than the one for cues recognised as different ($t(44) = -6.64$, $p < .001$), confirming the expected gradient-like pattern. The same RM ANOVA was repeated considering Cue Condition as a unique factor and was found to be significant, $F(2, 88) = 46.52$, $p < .001$, $\eta^2 = 0.25$. Post-hoc comparisons highlighted a different pattern of results compared to Subjective Recognition: identical cues elicited a lower RT ($M = 2.80$, $SD = 0.34$) with respect to different cues ($M = 3.34$, $SD = 0.47$; $t(44) = -8.28$, $p < .001$), but not compared to similar cues ($M = 2.90$, $SD = 0.39$; $t(44) = -1.70$, $p = .289$); furthermore, the RT at retrieval for similar cues was significantly lower than the one for different cues ($t(44) = -7.62$, $p < .001$).

Following the same rationale as for the analysis on the slopes, we repeated the RM ANOVAs considering only non-matching trials. The RM ANOVA with Subjective Recognition as a unique predictor was significant ($F(2, 76) = 6.85$, $p = .004$, $\eta^2 = 0.09$). Post-hoc comparisons highlighted a unique significant difference between cues perceived as identical ($M = 2.77$, $SD = 0.54$) and cues perceived as different ($M = 3.30$, $SD = 0.96$; $t(38) = -3.30$, $p = .006$). On the other hand, a RM ANOVA with Cue Condition as a unique predictor was non-significant ($F(2, 82) = 1.97$, $p = .153$, $\eta^2 = 0.03$), thus confirming our prediction.

Slopes Comparison for Accuracy in the Retrieval Task

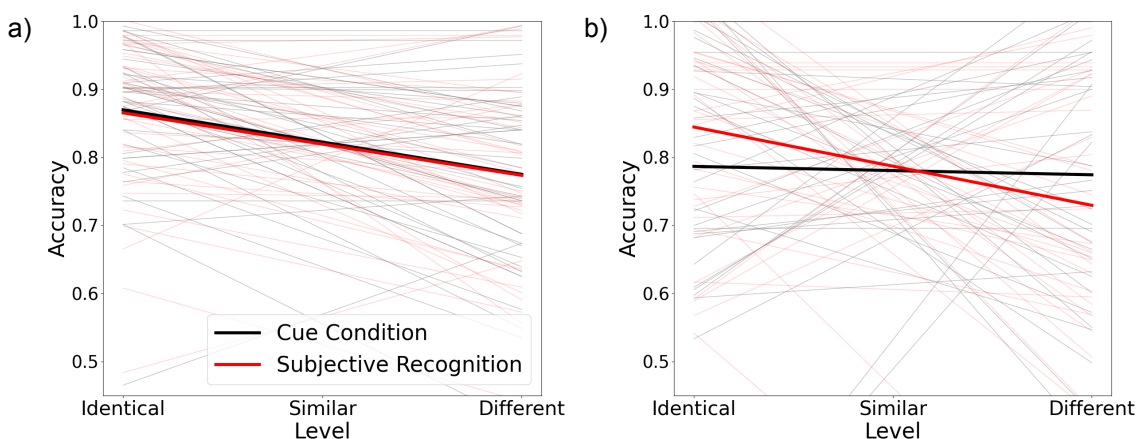
On average, participants had a high accuracy in the retrieval task ($M = 0.82$, $SD = 0.10$). To assess whether there was an effect of Subjective Recognition and Cue Condition on the accuracy in the retrieval task, we conducted two Wilcoxon Signed-Rank Tests on the slopes extracted from the linear regression models. Both the

slopes of Subjective Recognition ($M = -0.05$, $SD = 0.06$; $W = 147.0$, $p < .001$, $r_b = -0.70$) and Cue Condition ($M = -0.05$, $SD = 0.07$; $W = 123.0$, $p < .001$, $r_b = -0.71$) differed significantly from 0. However, a one-tailed Wilcoxon Test indicated that the slopes of Subjective Recognition were not significantly steeper than the ones of Cue Condition ($W = 458.0$, $p = .335$, $r_b = -0.07$; Figure 5a), contrasting with our hypotheses.

As for the RT analyses, we repeated the Wilcoxon Tests considering only non-matching trials. These tests confirmed our hypotheses, indicating that the slopes of Subjective Recognition ($M = -0.06$, $SD = 0.13$) were significantly different from 0 ($W = 141.5$, $p = .013$, $r_b = -0.50$), but not the ones of Cue Condition ($M = -0.01$, $SD = 0.16$; $W = 281.5$, $p = .791$, $r_b = -0.05$). Furthermore, a one-tailed Wilcoxon Test highlighted that the slopes of Subjective Recognition were steeper than those of Cue Condition ($W = 226.5$, $p = .048$, $r_b = -0.32$; Figure 5b).

Figure 5

Accuracy Regression Lines: Cue Condition vs. Subjective Recognition



Note. Regression lines obtained from the data of Experiment 1 and indicating: a) the absence of a significant difference between the effects of Subjective Recognition and Cue Condition on the accuracy at retrieval; b) the unique effect of Subjective Recognition on the accuracy at retrieval when considering non-matching trials. In the figure, light red and gray lines represent the regression lines per participant for Subjective Recognition and Cue Condition, respectively. The two main lines correspond to their average. In the y axis, accuracy represents the rate of correct responses on the total number of responses per level. The initial tick, which was 0.45, was removed for clarity.

Bootstrapping Analysis on Accuracy

Since the number of trials for each level of the two main factors varied depending on NaN values and the participants' responses in the recognition task (Table 2), we evaluated the robustness of the results by bootstrapping to the minimum number of trials, as for the RT analysis. The Wilcoxon Test was significant ($p < .05$) for only 10 of the 1000 iterations (1.0%). This result confirms that the slopes of Subjective Recognition were not steeper than the ones of Cue Condition and that the result obtained didn't depend on the difference in the number of trials between levels of each factor.

Table 2
Number of Trials per Factor Level - Accuracy

Level	Cue Condition	Subjective Recognition
1	1076	1298
2	1076	1068
3	1077	863

Note. The column “Level” represents either 1 = “identical”, 2 = “similar”, 3 = “different”, for Cue Condition; or 1 = “recognised as identical”, 2 = “recognised as similar”, 3 = “recognised as different”, for Subjective Recognition.

Post-Hoc Power Analyses on Accuracy

A post-hoc power analysis indicated that the achieved power for the Wilcoxon Signed Rank Test comparing the slopes of Cue Condition and Subjective Recognition on accuracy was 0.07. Thus, the study was highly underpowered. However, the very small (i.e., irrelevant; Goss-Sampson, 2018) effect size observed ($r_b = -0.07$) suggests that it is highly unlikely that this difference between the two factors could be detected, even increasing the sample size. Indeed, the detection of such a small effect when considering a significance criterion of $\alpha = 0.05$ and power = 0.80 would have required an impractically large sample size of 13385 participants.

Regarding the one-tailed Wilcoxon Test on non-matching trials, another post-hoc power analysis indicated that the achieved power was 0.45. In order to detect our medium effect size ($r_b = -0.32$; Goss-Sampson, 2018) with a power of 0.80 and considering a significance criterion of $\alpha = 0.05$, we would have needed at least 106 participants.

Comparison Recognised as Identical vs. Actual Identical

A one-tailed Wilcoxon Test, assessing whether cues recognised as identical ($M = 0.84$, $SD = 0.16$) led to a higher accuracy compared to actual identical cues ($M = 0.78$, $SD = 0.26$) in non-matching trials, was non-significant ($W = 485.5$, $p = .157$, $r_b = 0.18$). Since our predictions were not confirmed, we ran another one-tailed Wilcoxon Test to assess if the opposite pattern could be true: actual identical cues didn’t lead to higher accuracy at retrieval compared to recognised as identical ones ($W = 334.5$, $p = 0.847$, $r_b = -0.18$)

RM ANOVAs on Accuracy

We performed four independent RM ANOVAs to assess whether there was a significant difference between the levels of each factor. The first RM ANOVA revealed that the effect of Subjective Recognition on the accuracy at retrieval was statistically significant, $F(2, 88) = 11.23$, $p < .001$, $\eta^2 = 0.09$. Post-hoc comparisons highlighted significant differences between the levels of Subjective Recognition: the rate of correct retrieval trials was higher for cues recognised as identical ($M = 0.87$, $SD = 0.11$) compared to cues recognised as similar ($M = 0.82$, $SD = 0.12$; $t(44) = 3.17$, $p = .008$) or as different ($M = 0.77$, $SD = 0.14$; $t(44) = 4.82$, $p < .001$). However, the accuracy for cues recognised as similar did not differ from the one of those recognised as different ($t(44) = 1.83$, $p = .221$). Thus, the expected gradient was not supported by the analyses. The same RM ANOVA was repeated considering Cue Condition as a unique factor and was found significant, $F(2, 88) = 12.96$, $p < .001$, $\eta^2 = 0.10$. Post-hoc comparisons highlighted a different pattern of results compared to Subjective Recognition: the accuracy for

identical cues ($M= 0.86$, $SD= 0.10$) differed significantly from the one for different cues ($M= 0.77$, $SD= 0.14$; $t(44)= 4.88$, $p< .001$), but not from the one for similar cues ($M= 0.83$, $SD= 0.12$; $t(44)= 1.79$, $p= .240$); furthermore, similar cues elicited a higher accuracy at retrieval when compared to different cues ($t(44)= 3.20$, $p= .008$).

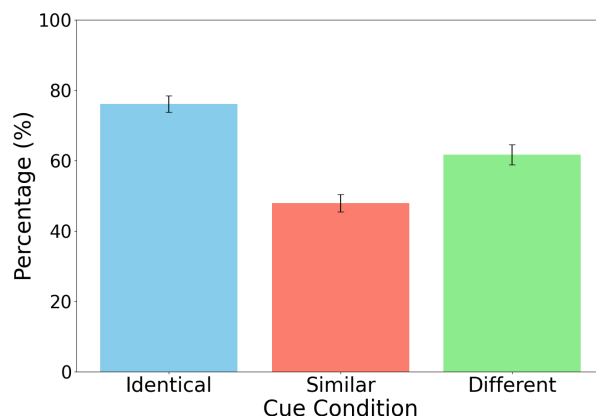
Following the same rationale as for the analyses on RT, we repeated the analyses considering only non-matching trials. Two separate RM ANOVAs revealed that the accuracy varied significantly across the levels of Subjective Recognition ($F(2,80)= 4.03$, $p= .032$, $\eta^2= 0.05$), but not across the levels of Cue Condition ($F(2,86)= 1.70$, $p= .197$, $\eta^2= 0.02$). Post-hoc comparisons highlighted a unique significant difference between cues recognised as identical ($M= 0.84$, $SD= 0.15$) and cues recognised as different ($M= 0.74$, $SD= 0.26$; $t(40)= 2.59$, $p= .040$), while the difference between cues recognised as identical and cues recognised as similar ($M= 0.78$, $SD= 0.16$) approached significance ($t(40)= 2.49$, $p= .051$). However, even though a parallel non parametric Friedman test was significant ($F(1.95, 78.05) = 4.07$, $p= .022$), the non parametric post-hoc comparisons indicated the absence of significant differences across levels of Subjective Recognition.

Accuracy in the Recognition Task

Overall, participants performed well in the recognition task, with an average accuracy of 0.62 ($SD= 0.08$). A RM ANOVA highlighted that the performance at this task (i.e., accuracy) varied depending on Cue Condition ($F(2,88)= 27.03$, $p< .001$, $\eta^2= 0.32$; Figure 6). Subsequent post-hoc tests confirmed that identical cues ($M= 0.76$, $SD= 0.16$) elicited a higher accuracy when compared to similar ones ($M= 0.48$, $SD= 0.16$; $t(44)= 7.57$, $p<.001$) and different ones ($M= 0.61$, $SD= 0.19$; $t(44)= 3.64$, $p= .002$). Furthermore, different cues elicited a higher accuracy than similar cues ($t(44)=3.57$, $p=.003$).

Figure 6

Percentage of Correct Recognition Responses Depending on Cue Condition



Note. The error bars represent the standard error of the mean (SEM) percentage accuracy in the recognition task for each level of Cue Condition.

Summary of the Results

Overall, the results of Experiment 1 partially confirmed our main hypotheses. For what concerns the RT, all three hypotheses were supported by the data. In particular, the results showed an increasingly higher RT moving from cues recognised

as identical to cues recognised as different, confirming the predicted gradient (Smith et al., 2013; Yum, 1931). Furthermore, this effect was stronger than the one of Cue Condition, which disappeared when considering non-matching trials. In this subset, cues recognised as identical elicited a faster RT when compared to actual identical cues, confirming our predictions. On the other hand, the analyses on accuracy revealed only partial evidence for a higher effect of Subjective Recognition in non-matching trials, whereas all the other expected results were not confirmed.

Crucially, none of the results was in line with the Encoding Specificity Hypothesis, since the effect of Cue Condition on accuracy and RT was either comparable to or lower than the effect of Subjective Recognition and the effect of similar and identical cues did not differ significantly.

However, given that in this experiment the retrieval phase preceded the recognition phase, we could not exclude that the RT/accuracy at retrieval determined the subsequent response in the recognition task (i.e., the more accurate or faster at retrieval, the more likely to respond “identical” in the recognition task). Therefore, we designed a follow up experiment in which we inverted the two main phases of the task to account for this potential alternative explanation.

Experiment 2

In Experiment 2 we aimed at further understanding the relationship between the subjective recognition of the cue and the retrieval of its paired associate, by reversing the sequence of events: we presented the recognition phase, followed by the retrieval one. In this new setup, where the serial dependency has been removed, replicating the results of the previous experiment would allow us to rule out the possibility that Subjective Recognition depends on the performance at retrieval. However, besides testing the same hypotheses as in Experiment 1, we also assessed whether there were relevant differences between the two experiments in terms of overall RT and accuracy at retrieval. Indeed, since participants could potentially retrieve the paired associate during the recognition phase, we expected an overall facilitation (e.g., Tambini et al., 2017) in both RT and accuracy in Experiment 2 compared to Experiment 1

Methods

Participants

Forty-five young adults were recruited through the SONA platform of the University of Granada and were offered five euros as compensation for participating in the study. Two participants had to be removed from the dataset since their accuracy in either the recognition or the retrieval task was above 1.5 times the interquartile range (IQR) with respect to the 3rd quartile or below 1.5 IQR with respect to the 2nd quartile. Thus, the final sample was composed of 43 participants (35 females), of age between 18 and 33 ($M = 22.14$, $SD = 3.42$). Ethics approval was obtained before data collection. An a priori power analysis using G*Power was performed to determine the sample size. We used the data of Experiment 1, specifically from the one-tailed Wilcoxon Signed Rank Test on the RT slopes, whose effect size was $r_b = 0.44$. The estimated sample size to detect such an effect, with 0.80 power and a significance criterion $\alpha = .05$, was 46. Due to time constraints, we couldn't reach the estimated sample size. Thus, this experiment was slightly underpowered.

Apparatus and Stimuli

The same apparatus and stimuli from Experiment 1 were used in Experiment 2.

Procedure

The encoding and distractor phases were the same as in Experiment 1. However, the order of the recognition and retrieval tasks was inverted. Furthermore, all the 12 recognition trials of a block were presented one after the other and were followed by all the 12 retrieval trials. The order of the trials in the recognition and retrieval phases was different and they both differed from the order of the encoding trials. We decided to separate the recognition and retrieval phases because we expected that seeing the image of the cue could have reinstated the memory of the paired object already in the recognition phase. Thus, separating the two phases was needed to avoid an excessive facilitation, which could have confounded the results.

Design, Software and Analyses

We used the same design and software for our analyses as in Experiment 1. Additionally, we performed further analyses to compare the performance in the two experiments. In particular, we conducted two Mann-Whitney U-Tests to assess whether there was a significant difference between the two experiments in the overall RT (for correct retrieval trials) and accuracy at retrieval. Furthermore, we repeated the same Mann-Whitney U-Test considering the overall accuracy in the recognition task.

Results

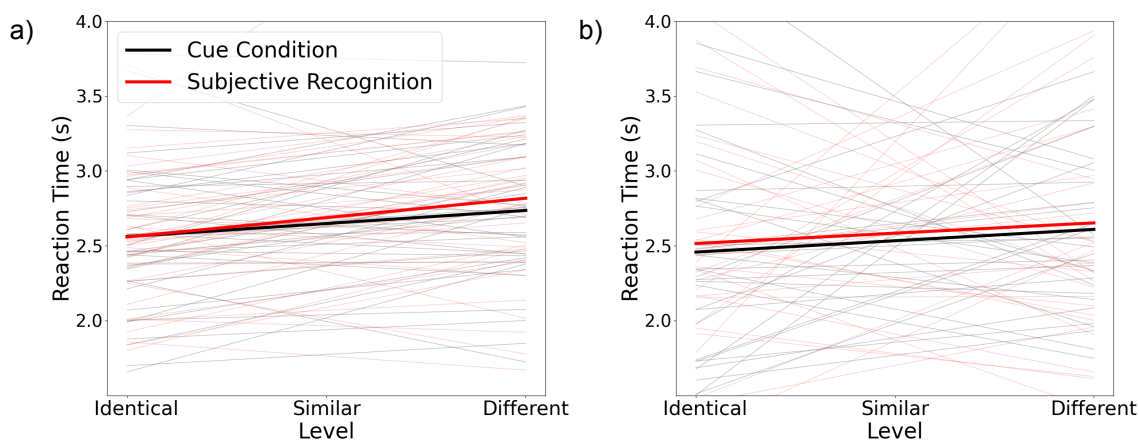
Slopes Comparison for RT in the Retrieval Task

On average, participants needed 2.65 seconds ($SD = 0.41$) to respond correctly in the retrieval task. To assess whether there was an effect of Subjective Recognition and Cue Condition on the RT for correct retrieval trials, we conducted two Wilcoxon Signed-Rank Tests on the slopes extracted from the linear regression models. Both the slopes of Subjective Recognition ($M = 0.13$, $SD = 0.23$; $W = 177.0$, $p < .001$, $r_b = 0.63$) and Cue Condition ($M = 0.09$, $SD = 0.18$; $W = 235.0$, $p = .003$, $r_b = 0.50$) differed significantly from 0. However, a one-tailed Wilcoxon Signed-Rank Test indicated that the slopes of Subjective Recognition were not significantly steeper than the ones of Cue Condition ($W = 557.0$, $p = .158$, $r_b = 0.18$), contrasting with our predictions (Figure 7a).

Furthermore, when repeating the analyses on non-matching trials, two Wilcoxon Signed-Rank Tests checking whether the slopes differed from 0 remained non-significant (for Cue Condition: $W = 223.0$, $p = .313$, $r_b = 0.20$; for Subjective Recognition: $W = 253.0$, $p = .634$, $r_b = 0.10$). Indeed, a one-tailed Wilcoxon Test confirmed that the slopes of Subjective Recognition weren't steeper than the ones of Cue Condition in this subset of data ($W = 267.0$, $p = .597$, $r_b = -0.05$; Figure 7b).

Bootstrapping Analysis on RT

Since the number of trials for each level of the two main factors varied depending on NaN values and the participants' response in the recognition task (Table 3), we evaluated the robustness of the results by bootstrapping to the minimum number of trials, following the same iteration procedure as in Experiment 1. The Wilcoxon Test comparing the slopes of Cue Condition and Subjective Recognition was significant ($p < .05$) for only 116 of the 1000 (11,6%) iterations. These results indicate that not having found an effect in the main analysis didn't depend on the different number of trials per level of the factors.

Figure 7*RT Regression Lines: Cue Condition vs. Subjective Recognition*

Note. Regression lines obtained from the data of Experiment 2, indicating: a) the absence of a significant difference between the effects of Subjective Recognition and Cue Condition on the RT for correct retrieval trials; b) that both the slopes of Cue Condition and Subjective Recognition didn't differ significantly from 0, when considering non-matching trials. In the figure, light red and gray lines represent the regression lines per participant for Subjective Recognition and Cue Condition, respectively. The two main lines correspond to their average. The initial tick of the y axis, which was 1.5 seconds, was removed for clarity.

Table 3*Number of Trials per Factor Level - RT*

Level	Cue Condition	Subjective Recognition
1	791	980
2	780	762
3	736	565

Note. The column "Level" represents either 1 = "identical", 2 = "similar", 3 = "different", for Cue Condition; or 1 = "recognised as identical", 2 = "recognised as similar", 3 = "recognised as different", for Subjective Recognition

Post-Hoc Power Analyses on RT

A post-hoc power analysis indicated that the achieved power for the one-tailed Wilcoxon Signed Rank Test comparing the slopes of Cue Condition and Subjective Recognition on RT was 0.36, suggesting that the study was underpowered. Indeed, the initial sample size was already too small to reliably detect an effect based on the data of Experiment 1. Moreover, an a priori power analysis confirmed that we would have needed 158 participants to reliably detect the small effect size ($r_b = 0.18$; Goss-Sampson, 2018) observed in Experiment 2, considering a significance criterion of $\alpha = 0.05$ and power = 0.80.

With what concerns the one-tailed Wilcoxon Test on non-matching trials, a post-hoc power analysis indicated that the achieved power was only 0.06. However, the effect size observed ($r_b = -0.05$) was very small (i.e., irrelevant; Goss-Sampson, 2018), suggesting that it is highly unlikely that this difference could be detected, even increasing the sample size. Indeed, the detection of such a small effect, when

considering a significance criterion of $\alpha = 0.05$ and power = 0.80, would have required an impractically large sample size of 62123 participants.

Comparison Recognised as Identical vs. Actual Identical

A one-tailed Wilcoxon Test, assessing whether cues recognised as identical ($M = 2.49$, $SD = 0.65$) led to faster RTs compared to actual identical cues ($M = 2.42$, $SD = 0.74$) in non-matching trials, was non-significant ($W = 461.0$, $p = .754$, $r_b = 0.12$), contrasting with our hypothesis.

RM ANOVAs on RT

We assessed whether there were differences between the levels of each factor by means of four independent RM ANOVAs. The first RM ANOVA revealed that the effect of Subjective Recognition on the RT at retrieval was statistically significant, $F(2, 84) = 7.61$, $p = .002$, $\eta^2 = 0.04$. Post-hoc comparisons highlighted a unique significant difference between cues recognised as identical ($M = 2.57$, $SD = 0.45$) and cues recognised as different ($M = 2.82$, $SD = 0.63$; $t(42) = -3.68$, $p = .002$). Thus, the expected gradient was not supported by the analyses. The same RM ANOVA was repeated considering Cue Condition as a unique factor and was found significant, $F(2, 84) = 6.59$, $p = .003$, $\eta^2 = 0.03$. Post-hoc comparisons highlighted a different pattern of results compared to Subjective Recognition: identical cues elicited a lower RT ($M = 2.59$, $SD = 0.50$) compared to different cues ($M = 2.76$, $SD = 0.49$; $t(42) = -3.15$, $p = .009$), but not compared to similar cues ($M = 2.59$, $SD = 0.40$; $t(42) = -0.10$, $p = 1.0$); furthermore, the RT at retrieval for similar cues was significantly lower than the one for different cues ($t(42) = -2.88$, $p = .019$).

Following the same rationale as for the analysis on the slopes, we repeated the RM ANOVAs considering only non-matching trials. However, both RM ANOVAs were found non-significant (for Cue Condition: $F(2,76) = 1.36$, $p = .262$, $\eta^2 = 0.02$; for Subjective Recognition: $F(2,68) = 0.17$, $p = .779$, $\eta^2 = 0.00$).

Slopes Comparison for Accuracy in the Retrieval Task

On average, participants reached a high accuracy in the retrieval task ($M = 0.75$, $SD = 0.13$). To assess whether there was an effect of Subjective Recognition and Cue Condition on the accuracy at retrieval, we conducted two Wilcoxon Signed-Rank Tests on the slopes extracted from the linear regression models. Both the slopes of Subjective Recognition ($M = -0.03$, $SD = 0.07$; $W = 213.0$, $p = .003$, $r_b = -0.53$) and Cue Condition ($M = -0.03$, $SD = 0.07$; $W = 180.0$, $p = .016$, $r_b = -0.46$) differed significantly from 0. However, a one-tailed Wilcoxon Test indicated that the slopes of Subjective Recognition were not significantly steeper than the ones of Cue Condition ($W = 405.0$, $p = .283$, $r_b = -0.10$), contrasting with our hypotheses (Figure 8a).

As for the RT analyses, we repeated the Wilcoxon Signed-rank tests considering only non-matching trials. Nevertheless, both tests assessing whether the slopes differed from 0 didn't reach significance (for Subjective Recognition: $M = -0.03$, $SD = 0.16$; $W = 293.50$, $p = .540$, $r_b = -0.12$; for Cue Condition: $M = -0.01$, $SD = 0.16$; $W = 237.50$, $p = .448$, $r_b = -0.15$), and the slopes of Subjective Recognition weren't steeper than those of Cue Condition ($W = 380.0$, $p = .558$, $r_b = 0.03$; Figure 8b).

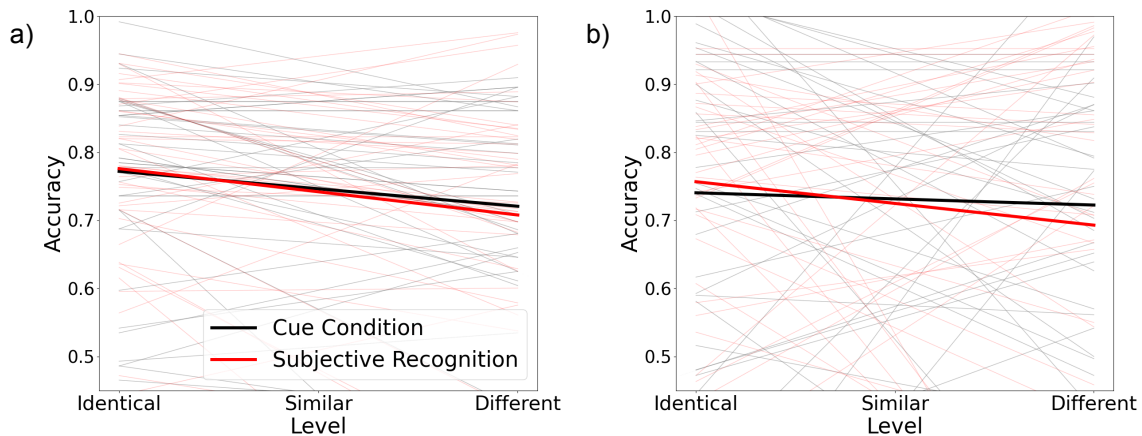
Bootstrapping Analysis on Accuracy

Since the number of trials for each level of the two main factors varied depending on NaN values and the participants' response in the recognition task (Table 4), we evaluated the robustness of the results by bootstrapping to the minimum number of trials, following the same iteration procedure as in Experiment 1. The Wilcoxon Test

was significant ($p < .05$) for 89 of the 1000 iterations (8.9%), confirming the absence of a significant difference between the slopes of Cue Condition and Subjective Recognition.

Figure 8

Accuracy Regression Lines: Cue Condition vs. Subjective Recognition



Note. Regression lines obtained from the data of Experiment 2, indicating: a) the absence of a significant difference between the effects of Subjective Recognition and Cue Condition on the accuracy at retrieval; b) that both the slopes of Cue Condition and Subjective Recognition didn't differ significantly from 0 when considering non-matching trials. In the figure, light red and gray lines represent the regression lines per participant for Subjective Recognition and Cue Condition, respectively. The two main lines correspond to their average. In the y axis, accuracy represents the rate of correct responses on the total number of responses per level. The initial tick, which was 0.45, was removed for clarity.

Table 4

Number of Trials per Factor Level - Accuracy

Level	Cue Condition	Subjective Recognition
1	1032	1284
2	1030	994
3	1029	813

Note. The column "Level" represents either 1 = "identical", 2 = "similar", 3 = "different", for Cue Condition; or 1 = "recognised as identical", 2 = "recognised as similar", 3 = "recognised as different", for Subjective Recognition.

Post-Hoc Power Analyses on Accuracy

A post-hoc power analysis indicated that the achieved power for the one-tailed Wilcoxon Signed Rank Test comparing the slopes of Cue Condition and Subjective Recognition on accuracy was 0.24. Thus, the study was underpowered for detecting the small effect size observed in our study ($r_b = -0.10$; Goss-Sampson; 2018). To detect such an effect with 0.80 power and significance criterion of $\alpha = 0.05$, we would have needed 312 participants.

With what concerns the one-tailed Wilcoxon Test on non-matching trials, a post-hoc power analysis indicated that we achieved a power of 0.15. However, the effect size observed ($r_b = 0.03$) was very small (i.e., irrelevant; Goss-Sampson, 2018),

suggesting that it is highly unlikely that this difference could be detected, even increasing the sample size. Indeed, the detection of such a small effect, when considering a significance criterion of $\alpha = 0.05$ and power = 0.80, would have required an impractically large sample size of 658 participants.

Comparison Recognised as Identical vs. Actual Identical

A one-tailed Wilcoxon Test, assessing whether cues recognised as identical ($M = 0.75$, $SD = 0.18$) led to a higher accuracy compared to actual identical cues ($M = 0.75$, $SD = 0.28$) in non-matching trials, was not significant ($W = 311$, $p = .808$, $r_b = -0.16$), contrasting with our hypothesis.

RM ANOVAs on Accuracy

We performed four independent RM ANOVAs to assess whether there was a significant difference between the levels of each factor. The first RM ANOVA revealed a significant effect of Subjective Recognition on the accuracy at retrieval ($F(2,84) = 7.73$, $p < .001$, $\eta^2 = 0.04$). Post-hoc comparisons highlighted a significant difference between cues recognised as identical ($M = 0.76$, $SD = 0.14$) and cues recognised as different ($M = 0.69$, $SD = 0.20$; $t(42) = 3.10$, $p = .010$) and between cues recognised as similar ($M = 0.77$, $SD = 0.13$) and cues recognised as different ($t(42) = 3.45$, $p = .004$), not providing support for the expected gradient-like pattern. The RM ANOVA for Cue Condition was also significant ($F(2,84) = 3.97$, $p = .024$, $\eta^2 = 0.02$), but none of the post-hoc comparisons survived Bonferroni correction. Furthermore, a parallel non-parametric Friedman's Test was not significant, $F(1.95, 82.05) = 3.11$; $p = .051$.

We repeated the analyses on non-matching trials. Both RM ANOVAs were found non significant (for Subjective Recognition: $F(2,78) = 2.21$, $p = .130$, $\eta^2 = 0.02$; for Cue Condition: $F(2,82) = 0.33$, $p = .717$, $\eta^2 = 0.00$)

Accuracy in the Recognition Task

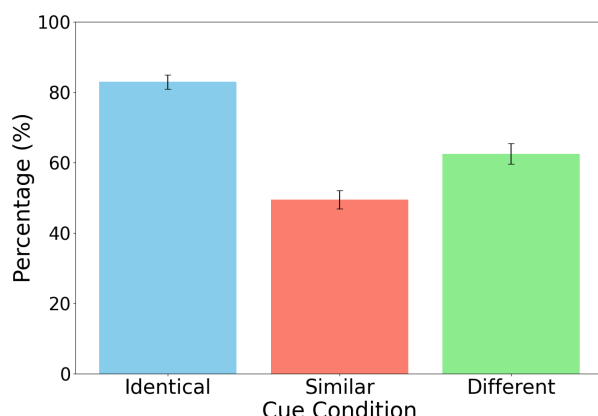
Overall, participants performed well in the recognition task, with an average accuracy of 0.64 ($SD = 0.11$). A RM ANOVA highlighted a significant difference in the task performance, depending on Cue Condition ($F(2, 84) = 45.47$, $p < .001$, $\eta^2 = 0.39$; Figure 9). Post-hoc tests confirmed that identical cues ($M = 0.83$, $SD = 0.13$) elicited a higher accuracy when compared to similar ones ($M = 0.49$, $SD = 0.17$; $t(42) = 10.63$, $p < .001$) and different ones ($M = 0.61$, $SD = 0.21$; $t(42) = 5.65$, $p < .001$). Furthermore, different cues elicited a higher accuracy than similar cues ($t(42) = 3.19$, $p = .008$), replicating the results obtained in Experiment 1. Indeed, a Mann-Whitney U-Test confirmed the absence of a significant difference in the accuracy at recognition between the two experiments ($U = 821.5$, $p = .224$, $r_b = 0.15$).

Comparison Between Experiments

We assessed overall differences in RT and accuracy between the two experiments by means of Mann-Whitney U-Tests. The overall RT at retrieval was significantly higher in Experiment 1 ($M = 3.00$, $SD = 0.32$) compared to Experiment 2 ($M = 2.65$, $SD = 0.41$; $U = 1463$, $p < .001$, $r_{bc} = -0.51$). However, the overall accuracy was higher in Experiment 1 ($M = 0.82$, $SD = 0.10$) compared to Experiment 2 ($M = 0.75$, $SD = 0.13$; $U = 1325.5$, $p = .003$, $r_{bc} = -0.37$).

Figure 9

Percentage of Correct Recognition Responses Depending on Cue Condition



Note. The error bars represent the standard error of the mean (SEM) percentage accuracy in the recognition task for each level of Cue Condition

Summary of the Results

Experiment 2 did not replicate the results of Experiment 1 and none of our hypotheses was supported. With what concerns the RT for correct retrieval trials, the effect of Subjective Recognition did not differ from the one of Cue Condition and none of the two effects showed the expected gradient-like pattern (Smith et al., 2013; Yum, 1931). In non-matching trials, both effects were nullified and cues recognised as identical didn't result in lower RTs at retrieval compared to actual identical cues. The same pattern of results was replicated for the analyses on the accuracy at retrieval.

Finally, the comparison between experiments highlighted an expected RT facilitation in Experiment 2. However, contrary to our predictions, accuracy decreased compared to Experiment 1.

Overall, Experiment 2 was designed with the primary aim of assessing whether the response in the Subjective Recognition task could depend on the RT/accuracy at retrieval. By inverting the retrieval and recognition phases, we still found an effect of Subjective Recognition, both on RT and accuracy. However, if in Experiment 1 all the differences reported were mainly "identical" driven, with this condition being faster in RT and higher in accuracy with respect to the other two, the results of Experiment 2 were mostly "different" driven, with this condition being slower in RT and lower in accuracy. Thus, since the results of Experiment 1 were not replicated, we cannot exclude the possibility that the findings of Experiment 1 could be driven by a dependency of Subjective Recognition on the performance on the previous retrieval task.

Discussion

In this study we aimed at challenging the Encoding Specificity Hypothesis (Tulving & Thomson, 1973) by proposing that the best cue for reinstating the memory of an episodic event is not necessarily the encoded cue, but rather the cue that best matches the current state of the memory. In particular, since neural memory traces (i.e., engrams) and their psychological experience change over time (Liu et al., 2021; Winocur & Moscovitch, 2011) and since cues able to reinstate the whole encoded event are cues whose neural representation is similar to the stored representation of that

event (Knierim & Neunuebel, 2016), then the best cue to reinstate the memory must change in accordance with it.

We examined this hypothesis using a paired-associate paradigm (Calkins, 1896), where participants encoded pairs of objects and, during a test phase, were asked to recall one of them upon being presented with a cue. This cue could be either one of the two originally encoded objects, a similar version of it or a different version of it (Figure 2). Crucially, we tracked the current status of the memory by means of a recognition task on the cues, which could either follow (Experiment 1) or precede (Experiment 2) the paired associate retrieval task. An encoding specificity account would predict that the best cue will always be the original one, since it was the only one encoded, followed by similar and different cues, in this order (Smith et al., 2013; Yum, 1931). On the other hand, from our perspective, this gradient is not the direct reflection of the actual similarity of the cues. Instead, it depends on the subjective recognition of those cues as identical, similar or different, since they would reflect increasingly dissimilar representations with respect to the current status of the memory trace of the encoded event. Indeed, we expected this effect to be stronger than the one of the actual similarity of the cues (i.e., Cue Condition), whose effect was predicted to disappear when considering the trials in which the actual similarity of the cues did not correspond with their subjective recognition (i.e., non-matching trials). Finally, we hypothesized that non-encoded cues recognised as identical would have been more effective than encoded identical cues.

In Experiment 1, our hypotheses were confirmed for the analyses on RT, signaling a gradient-like effect of Subjective Recognition, which was stronger than the one of Cue Condition (Figure 5). Furthermore, the effect of Cue Condition disappeared when considering only non-matching trials (Figure 6) and, in this subset, cues recognised as identical elicited a lower RT compared to actual identical (i.e., encoded) cues. These results suggest that the highest retrieval ease of the memory associated with the cue (Madigan et al., 2000; Ratcliff, 1978) did not necessarily depend on providing the encoded cue, but on providing the version of it that best matched the current state of the memory. Expanding on this result with a neural-based explanation, we propose that it is the degree of similarity between the representation of the visual input (i.e., cue) and the current state of the stored representation of the full event that determines its successful retrieval (Knierim & Neuneubel, 2016). In this sense, since only correct retrieval trials were considered for the RT analyses, the degree of dissimilarity of the representations of cues recognised as similar or different wasn't enough to hinder retrieval, but it was reflected in a higher difficulty in finding the target memory. This aligns with recognition studies (Vieweg et al., 2015; Vieweg et al., 2019) that demonstrated that the time needed to correctly recognise a previously encoded figure varies as a function of stimulus completeness: the more features are removed from that figure (i.e, partial cue), the more time needed to categorize it. From our perspective, cues perceived as more similar would be more "complete", in the sense that they would be closer to the current memory representation of the encoded event.

Additionally, the results support a competition-selection account of memory retrieval (Caplan et al., 2022; Kuhl et al., 2011). Cues recognized as different or similar would elicit more potential associations due to their incomplete match with the current memory representation. This would lead to a higher competition between possible "matching" memories, resulting in more time needed to accurately recall the original

event (Sohn et al., 2003). Indeed, since participants had to try to generate a response regardless of the cue provided, more difficulties were expected in cases in which the association with the original memory was harder to achieve (i.e., when the cue was recognized as similar/different). Future research should test this interpretation by means of representational similarity analysis (RSA) on EEG data, in order to capture the evolution of the memory trace and the timing of its reinstatement depending on the cue provided (see, for instance, Linde-Domingo et al., 2019; Staresina et al., 2016).

With what concerns the accuracy at retrieval, the results indicated the absence of a difference between the effect of Cue Condition and Subjective Recognition, not providing support for either the Encoding Specificity Hypothesis or for our predictions. Indeed, even though the stronger effect of Subjective Recognition in non-matching trials might indicate a tendency towards the predicted pattern, no evidence for a significant difference between the levels of the factor was found in this subset. However, this result is still noteworthy, since it suggests that the effect might depend on the correct recognition trials, indicating the possibility of an interaction between the two factors. Future research should test this hypothesis by fitting generalized linear mixed models (GLMMs) on the accuracy and RT data of the retrieval task, so as to assess whether more complex models can explain better the results.

Another point that is worth discussing is the fact that cues recognized as identical led to more accurate recall responses compared to both recognized as similar and different cues. Indeed, even though this result does not reflect the expected gradient (Smith et al., 2013; Yum, 1931), it can still fit into a representational similarity account. Assuming that the degree of representational similarity varies along a continuum (Knierim & Neuneubel, 2016), recognising a cue as similar could either reflect an almost complete match with the current trace, thus eliciting memory reinstatement, or a high diversity (but not high enough to be classified as different), thus hindering retrieval (Knierim & Neuneubel, 2016; Rolls, 2013). In this sense, it could be that, for most participants, recognising the cue as similar would have reflected a quite different representation with respect to the current “identical” one. Conversely, correct retrieval responses for cues recognised as similar or different might simply correspond to trials in which the degree of representational dissimilarity didn't overcome a threshold for affecting retrieval (Rolls, 2013). Alternatively, the recollection of episodic details of the encoded event might still happen despite differences in the representations (Yassa & Stark, 2011). Indeed, cortical memory reinstatement has been observed also in the presence of hippocampal activity signaling representational dissimilarity during the recognition of similar lures (Lohnas et al., 2018), suggesting the possibility of a complementary role played by the cortex, with respect to the hippocampus. Even though the evidence for this interpretation is still sparse (especially because hippocampal reinstatement might depend on task demand, current goals and type of cue provided; Brunec et al., 2020), future studies should try to determine the nature of these processes by analyzing patterns of neural activity when cues recognised as identical led to mistakes at retrieval and when cues recognised as different led to correct retrieval responses.

On the whole, the results of Experiment 1 confirmed that, at least under our experimental conditions, non-encoded cues can be better than encoded ones in eliciting a fast retrieval of the original event. Additionally, the overall effect of the actual

similarity of the cue might depend on the subjective experience of similarity, challenging the Encoding Specificity Hypothesis.

However, we considered the possibility that the RT/accuracy at retrieval could have driven the subsequent recognition response in Experiment 1. In Experiment 2, we tried to test this alternative explanation by inverting the retrieval and recognition phases. We argued that since the serial dependency was removed, a replication of the results of Experiment 1 would have reinforced our findings. Nevertheless, in this set-up, recognition cannot precede, trial by trial, the retrieval of the object associated to that same cue, because it would potentially elicit the reinstatement of the memory, facilitating the subsequent cued recall. We tried to control for this potential confounder by separating the recognition and retrieval tasks into two distinct blocks, but doing so we lost track of the actual current state of the memory: recognition (i.e., current state of the memory) was assessed in a temporal distance from the subsequent cued recall, not allowing to control for a possible further evolution of the memory trace (Liu et al., 2021). Moreover, even with this set-up, the presentation of the cue could have reinstated the memory of the encoded event, an effect that was expected to facilitate the subsequent cued recall (Tambini et al., 2017). However, this reactivation could have led to incorporating the current perceptual input as part of the modified memory trace (Moscovitch et al., 2016; Olsen et al., 2012; Rasch & Born, 2007), confounding the results. In particular, this process could have strengthened or degraded the memory trace (Bridge & Voss, 2014), depending on whether the contextual information integrated matched the memory of the event or not (De Oliveira Alvares et al., 2013; Scully et al., 2017). In this sense, the recognition phase could have either effectively primed the participants (see, for instance, Bowyer & Humphreys, 1979), facilitating cued recall, or hindered it, leading to subsequent recall difficulties. Indeed, this interpretation would explain why participants in Experiment 2 were overall faster in retrieving the memory of the paired associate, but also committed more mistakes compared to the participants of Experiment 1. Future research should try to determine the nature of this drop in the accuracy and significant decrease in RT observed in Experiment 2, by testing when reactivation (if present) facilitates retrieval and when it impairs it. A first step could be to consider interactions between Cue Condition and Subjective Recognition, so as to clarify which trials led to a failure in episodic recollection and, possibly, find some commonalities among them. Afterwards, implementation of neuroimaging techniques to track memory reactivation (Staresina & Wimber, 2019; Staresina et al., 2016) during the recognition task would be needed to understand if that is the mechanism that led to interference, impacting the overall accuracy.

To conclude, the absence of significant results in Experiment 2 might depend on the paradigm structure. Since we could not exclude a further evolution of the memory trace between recognition and retrieval, the interpretation of the results is limited and does not allow a direct comparison with the results of Experiment 1.

Limitations and Conclusion

To further test the possibility that the best cue evolves with the memory it is associated with, future research should overcome the main limitations of our study. First of all, we argue that the recognition task is problematic to the extent that Cue Condition and Subjective Recognition overlap most of the time. Indeed, the hypothesis testing should be mainly based on the trials in which the two factors do not correspond,

since it is the only way to disentangle the two effects. However, doing so drastically reduces the overall amount of trials, leading to potentially biased results and reduced statistical power. In fact, all the analyses on non-matching trials were underpowered, as indicated by an achieved power of 0.63 for the analyses on RT and of 0.45 for the analyses on accuracy (in Experiment 1). Thus, future research should increase the number of participants up to a minimum of 106, for determining whether an effect on accuracy and RT can be reliably detected. Furthermore, future studies should also consider replacing the categorical recognition task with a continuous scale representing the percentage of subjective perceived similarity. This way it would be easier to test the hypothesis of a linear increase in RT and linear decrease in accuracy as the perceived similarity decreases. However, since Experiment 2 highlighted that recognition cannot precede cued recall without confounding the results, the obligatory sequence of events (i.e., cued recall followed by recognition) doesn't permit to rule out the possibility that Subjective Recognition depends on the accuracy or RT in the cued recall task. Therefore, we believe that future research should consider removing the recognition task and make a priori predictions of which kind of cues will be the most effective for reinstating the memory of the original event. One option could be to first determine the degree of prototypicality of distinct cues. Then, based on the assumption that memory converges towards a prototype over time (Heinen et al., 2023), present cues with varying levels of prototypicality at different time intervals following the encoding event. In this scenario, it could be predicted that the most effective cues, as the time from encoding increases, would be those that are more and more prototypical, since their representation would be more similar to the evolution of the encoded one.

Apart from the paradigm structure, another limitation of our study was the absence of a control on the degree of perceptual dissimilarity between the cues provided. This didn't allow us to determine whether the performance at retrieval could be impacted by an excessive degree of similarity of the similar cues with the original ones. Indeed, we failed to replicate the results obtained by Smith et al. (2013) and Yum (1931), as in our experiments no significant difference was observed between identical cues and similar cues in the accuracy and RT at retrieval. This inconsistency with previous studies could be partially due to the nature of the stimuli used (i.e., video-clips, line drawings; Smith et al., 2013; Yum, 1931). However, it is also possible that similar cues were more similar to identical ones than they were to different ones, leading to identical and similar cues to constitute a unique "perceptual group". In fact, recognition studies highlighted that the more similar a lure is to the encoded target, the more likely it is to categorize it as "old" (Kim & Yassa, 2013); furthermore, studies on cued recall of similar cues (Smith et al., 2013; Yum, 1931) reported an increasingly higher performance at retrieval for increasingly higher levels of cue similarity, which is compatible with the absence of a significant difference between similar and identical cues in our study. Thus, future research should quantify the degree of similarity between types of cues, so as to have a clearer estimate of the impact of Cue Condition, with the possibility of inserting more than just three levels of similarity (in line with the stimuli of the Mnemonic Similarity Task; Stark et al., 2019).

Another limitation of our study concerns the overall high performance of the participants in the retrieval task, which could have impacted the results on accuracy. Future studies should consider increasing the number of encoding trials or the time between encoding and retrieval, so as to induce more errors and quantify relevant

differences in their distribution, depending on the kind of cue provided. Furthermore, the type of index of accuracy used (i.e., hit rate) does not take into account a possible response bias between different types of stimuli (Brady et al., 2022) or the possibility of guessing (Cleary, 2018). Future studies should rely on more appropriate measures of accuracy, integrating the confidence rating into the analyses. One possibility could be to analyze the performance through Receiver Operating Characteristic (ROC) Curves, computed on the hit rate and false alarm rate of each participant at different levels of reported confidence and per type of cue (Brady et al., 2022).

Overall, besides these limitations, the obtained results provide preliminary evidence for a potentially necessary perspective change: from encoding-retrieval match, to cue updating based on memory evolution. If confirmed and supported by neuroimaging evidence, our findings could be crucial for the development of ad-hoc therapies and interventions for cognitive impairments (e.g., Bird & Luszcz, 1993) or, more in general, educational proposals aimed at enhancing learning and memory (e.g., Gallagher, 2017).

References

- Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640–1642. <https://doi.org/10.1126/science.1152882>
- Bird, M., & Luszcz, M. (1993). Enhancing memory performance in Alzheimer's disease: Acquisition assistance and cue effectiveness. *Journal of Clinical and Experimental Neuropsychology*, *15*(6), 921–932. <https://doi.org/10.1080/01688639308402608>
- Bjork, R. A., & Richardson-Klavehn, A. (2014). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes* (1st ed., pp. 313-344). Psychology Press. <https://doi.org/10.4324/9781315807515>
- Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023a). Repeated measures ANOVA and adjusted F-tests when sphericity is violated: which procedure is best?. *Frontiers in Psychology*, *14*. <https://doi.org/10.3389/fpsyg.2023.1192453>
- Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023b). Non-normal data in repeated measures ANOVA: impact on Type I error and power. *PubMed*, *35*(1), 21–29. <https://doi.org/10.7334/psicothema2022.292>
- Bower, G. H., Monteiro, K. P., & Gilligan, S. G. (1978). Emotional mood as a context for learning and recall. *Journal of Verbal Learning and Verbal Behavior*, *17*(5), 573–585. [https://doi.org/10.1016/s0022-5371\(78\)90348-1](https://doi.org/10.1016/s0022-5371(78)90348-1)
- Bowyer, P. A., & Humphreys, M. S. (1979). Effect of a recognition test on a subsequent cued-recall test. *Journal of Experimental Psychology. Human Learning and Memory*, *5*(4), 348–359. <https://doi.org/10.1037/0278-7393.5.4.348>
- Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2022). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*(2), 421–449. <https://doi.org/10.3758/s13423-022-02179-w>

- Bridge, D. J., & Voss, J. L. (2014). Hippocampal Binding of Novel Information with Dominant Memory Traces Can Support Both Memory Stability and Change. *Journal of Neuroscience*, *34*(6), 2203–2213. <https://doi.org/10.1523/jneurosci.3819-13.2014>
- Brunec, I. K., Robin, J., Olsen, R. K., Moscovitch, M., & Barense, M. D. (2020). Integration and differentiation of hippocampal memory traces. *Neuroscience & Biobehavioral Reviews/Neuroscience and Biobehavioral Reviews*, *118*, 196–208. <https://doi.org/10.1016/j.neubiorev.2020.07.024>
- Calkins, M. W. (1896). Association (II.). *Psychological Review*, *3*(1), 32–49. <https://doi.org/10.1037/h0068098>
- Caplan, J. B., Hennies, N., & Sommer, T. (2022). Competition between Associations in Memory. *Journal of Cognitive Neuroscience*, *34*(11), 2144–2167. https://doi.org/10.1162/jocn_a_01900
- Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and CA1 mismatch detection. *Learning & Memory*, *18*(8), 523–528. <https://doi.org/10.1101/lm.2135211>
- Cleary, A. M. (2018). Dependent measures in memory research: From free recall to recognition. In H. Otani & B.L. Schwartz (Eds.), *Handbook of research methods in human memory* (1st ed., pp. 19–35). Routledge. <https://doi.org/10.4324/9780429439957>
- Craik, F. I., & 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](https://doi.org/10.1016/s0022-5371(72)80001-x)
- De Oliveira Alvares, L., Crestani, A., Cassini, L., Haubrich, J., Santana, F., & Quillfeldt, J. (2013). Reactivation enables memory updating, precision-keeping and strengthening: Exploring the possible biological roles of reconsolidation. *Neuroscience*, *244*, 42–48. <https://doi.org/10.1016/j.neuroscience.2013.04.005>
- DeNardo, L. A., Liu, C. D., Allen, W. E., Adams, E. L., Friedmann, D., Fu, L., Guenther, C. J., Tessier-Lavigne, M., & Luo, L. (2019). Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nature Neuroscience*, *22*(3), 460–469. <https://doi.org/10.1038/s41593-018-0318-7>
- Dewhurst, S. A., & Brandt, K. R. (2007). Reinstating effortful encoding operations at test enhances episodic remembering. *Quarterly Journal of Experimental Psychology*, *60*(4), 543–550. <https://doi.org/10.1080/17470210601137086>
- Dewhurst, S. A., & Knott, L. M. (2010). Investigating the encoding—retrieval match in recognition memory: Effects of experimental design, specificity, and retention interval. *Memory & Cognition*, *38*(8), 1101–1109. <https://doi.org/10.3758/mc.38.8.1101>
- Dijkstra, K., Kaschak, M. P., & Zwaan, R. A. (2007). Body posture facilitates retrieval of autobiographical memories. *Cognition*, *102*(1), 139–149. <https://doi.org/10.1016/j.cognition.2005.12.009>
- Dulsky, S. G. (1935). The effect of a change of background on recall and relearning. *Journal of Experimental Psychology*, *18*(6), 725–740. <https://doi.org/10.1037/h0058066>

- Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(3), 443–455. <https://doi.org/10.1037/0278-7393.15.3.443>
- Eich, J. E., Weingartner, H., Stillman, R. C., & Gillin, J. C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Verbal Learning and Verbal Behavior*, 14(4), 408–417. [https://doi.org/10.1016/s0022-5371\(75\)80020-x](https://doi.org/10.1016/s0022-5371(75)80020-x)
- Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, 13(4), 333–345. <https://doi.org/10.3758/bf03202501>
- Frankland, P. W., Josselyn, S. A., & Köhler, S. (2019). The neurobiological foundation of memory retrieval. *Nature Neuroscience*, 22(10), 1576–1585. <https://doi.org/10.1038/s41593-019-0493-1>
- Gallagher, K. M. (2017). Retrieval Cues on tests: A strategy for helping students overcome retrieval failure. *College Teaching*, 65(4), 164–171. <https://doi.org/10.1080/87567555.2017.1304352>
- Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition*, 5(5), 499–504. <https://doi.org/10.3758/bf03197392>
- Gilboa, A., & Moscovitch, M. (2021). No consolidation without representation: Correspondence between neural and psychological representations in recent and remote memory. *Neuron*, 109(14), 2239–2255. <https://doi.org/10.1016/j.neuron.2021.04.025>
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, 66(3), 325–331. <https://doi.org/10.1111/j.2044-8295.1975.tb01468.x>
- Goss-Sampson, M. A. (2018). *Statistical analysis with JASP: A guide for students* (2nd ed.). University of Greenwich.
- Guskjolen, A., & Cembrowski, M. S. (2023). Engram neurons: Encoding, consolidation, retrieval, and forgetting of memory. *Molecular Psychiatry*, 28(8), 3207–3219. <https://doi.org/10.1038/s41380-023-02137-5>
- Heinen, R., Bierbrauer, A., Wolf, O. T., & Axmacher, N. (2023). Representational formats of human memory traces. *Brain Structure & Function*, 229(3), 513–529. <https://doi.org/10.1007/s00429-023-02636-9>
- Herz, R. S. (1998). Are odors the best cues to memory? A Cross-Modal comparison of Associative memory stimuli^a. *Annals of the New York Academy of Sciences*, 855(1), 670–674. <https://doi.org/10.1111/j.1749-6632.1998.tb10643.x>
- Horner, A. J., Bisby, J. A., Bush, D., Lin, W., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, 6(1), Article 7462. <https://doi.org/10.1038/ncomms8462>
- Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *CB/Current Biology*, 24(9), 988–992. <https://doi.org/10.1016/j.cub.2014.03.012>
- Iodice, R., Meilán, J. J. G., & Carro, J. (2015). Improvement of encoding and retrieval in normal and pathological aging with word–picture paradigm. *Aging and Mental Health/Aging & Mental Health*, 19(10), 940–946. <https://doi.org/10.1080/13607863.2014.995590>

- 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>
- Jung, J. H., Wang, Y., Mogle, A. J., Zhang, T., Köhler, S., Frankland, P. W., & Josselyn, S. A. (2023). Examining the engram encoding specificity hypothesis in mice. *Neuron*, *111*(11), 1830-1845.e5. <https://doi.org/10.1016/j.neuron.2023.03.007>
- Kerrén, C., Zhao, Y., & Griffiths, B. (2024). *Metacognitive awareness of memory distortion during recall*. *osf.io*. <https://doi.org/10.31234/osf.io/gzfnh>
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. <https://doi.org/10.3758/bf03197276>
- Kim, J., & Yassa, M. A. (2013). Assessing recollection and familiarity of similar lures in a behavioral pattern separation task. *Hippocampus*, *23*(4), 287–294. <https://doi.org/10.1002/hipo.22087>
- Kirk, M., & Berntsen, D. (2018). A short cut to the past: Cueing via concrete objects improves autobiographical memory retrieval in Alzheimer's disease patients. *Neuropsychologia*, *110*, 113–122. <https://doi.org/10.1016/j.neuropsychologia.2017.06.034>
- Kirwan, C. B., & Stark, C. E. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, *14*(9), 625–633. <https://doi.org/10.1101/lm.663507>
- Knierim, J. J., & Neunuebel, J. P. (2016). Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiology of Learning and Memory*, *129*, 38–49. <https://doi.org/10.1016/j.nlm.2015.10.008>
- Krenz, V., Alink, A., Sommer, T., Rooszendaal, B., & Schwabe, L. (2023). Time-dependent memory transformation in hippocampus and neocortex is semantic in nature. *Nature Communications*, *14*(1), Article 6037. <https://doi.org/10.1038/s41467-023-41648-1>
- Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5903–5908. <https://doi.org/10.1073/pnas.1016939108>
- 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), Article 3177. <https://doi.org/10.1038/s41467-021-23288-5>
- 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), Article 179. <https://doi.org/10.1038/s41467-018-08080-2>
- Liu, J., Zhang, H., Yu, T., Ren, L., Ni, D., Yang, Q., Lu, B., Zhang, L., Axmacher, N., & Xue, G. (2021). Transformative neural representations support long-term episodic memory. *Science Advances*, *7*(41), Article eabg9715. <https://doi.org/10.1126/sciadv.abg9715>
- Lohnas, L. J., Duncan, K., Doyle, W. K., Thesen, T., Devinsky, O., & Davachi, L. (2018). Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. *Proceedings of the National Academy of*

- Sciences of the United States of America*, 115(31), E7418–E7427.
<https://doi.org/10.1073/pnas.1717088115>
- Low, L. A., & Roder, B. J. (1983). Semantic relations between encoding and retrieval in cued recall. *Memory & Cognition*, 11(6), 651–659.
<https://doi.org/10.3758/bf03198290>
- Madigan, S., Neuse, J., & Roeber, U. (2000). Retrieval latency and “at-risk” memories. *Memory & Cognition*, 28(4), 523–528. <https://doi.org/10.3758/bf03201242>
- McClelland, J. L., & Goddard, N. H. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus*, 6(6), 654–665.
[https://doi.org/10.1002/\(SICI\)1098-1063\(1996\)6:6%3C654::AID-HIPO8%3E3.0.CO;2-G](https://doi.org/10.1002/(SICI)1098-1063(1996)6:6%3C654::AID-HIPO8%3E3.0.CO;2-G)
- Mcdowall, J. (1979). Effects of encoding instructions and retrieval cueing on recall in Korsakoff patients. *Memory & Cognition*, 7(3), 232–239.
<https://doi.org/10.3758/bf03197543>
- Miles, C., & Hardman, E. (1998). State-dependent memory produced by aerobic exercise. *Ergonomics*, 41(1), 20–28. <https://doi.org/10.1080/001401398187297>
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519–533. [https://doi.org/10.1016/s0022-5371\(77\)80016-9](https://doi.org/10.1016/s0022-5371(77)80016-9)
- Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic Memory and Beyond: The hippocampus and neocortex in transformation. *Annual Review of Psychology*, 67(1), 105–134.
<https://doi.org/10.1146/annurev-psych-113011-143733>
- Olsen, R. K., Moses, S. N., Riggs, L., & Ryan, J. D. (2012). The hippocampus supports multiple cognitive processes through relational binding and comparison. *Frontiers in Human Neuroscience*, 6, Article 146.
<https://doi.org/10.3389/fnhum.2012.00146>
- 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>
- Postman, L. (1975). Tests of the generality of the principle of encoding specificity. *Memory & Cognition*, 3(6), 663–672. <https://doi.org/10.3758/bf03198232>
- Rasch, B., & Born, J. (2007). Maintaining memories by reactivation. *Current Opinion in Neurobiology*, 17(6), 698–703. <https://doi.org/10.1016/j.conb.2007.11.007>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108.
<https://doi.org/10.1037/0033-295x.85.2.59>
- Reddy, B. G., & Bellezza, F. S. (1983). Encoding specificity in free recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 9(1), 167–174.
<https://doi.org/10.1037/0278-7393.9.1.167>
- Roediger, H. L., Tekin, E., & Uner, O. (2017). Encoding–retrieval interactions. In J.H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 5–26). Oxford: Academic Press.
<http://dx.doi.org/10.1016/B978-0-12-809324-5.21036-5>

- Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, 7, Article 74. <https://doi.org/10.3389/fnsys.2013.00074>
- Roy, D. S., Park, Y., Kim, M. E., Zhang, Y., Ogawa, S. K., DiNapoli, N., Gu, X., Cho, J. H., Choi, H., Kamentsky, L., Martin, J., Mosto, O., Aida, T., Chung, K., & Tonegawa, S. (2022). Brain-wide mapping reveals that engrams for a single memory are distributed across multiple brain regions. *Nature Communications*, 13(1), Article 1799. <https://doi.org/10.1038/s41467-022-29384-4>
- Santa, J. L., & Lamwers, L. L. (1974). Encoding specificity: Fact or artifact. *Journal of Verbal Learning and Verbal Behavior*, 13(4), 412–423. [https://doi.org/10.1016/s0022-5371\(74\)80019-8](https://doi.org/10.1016/s0022-5371(74)80019-8)
- Schacter, D. L., Guerin, S. A., & St Jacques, P. L. (2011). Memory distortion: an adaptive perspective. *Trends in Cognitive Sciences*, 15(10), 467–474. <https://doi.org/10.1016/j.tics.2011.08.004>
- Scully, I. D., Napper, L. E., & Hupbach, A. (2017). Does reactivation trigger episodic memory change? A meta-analysis. *Neurobiology of Learning and Memory*, 142, 99–107. <https://doi.org/10.1016/j.nlm.2016.12.012>
- Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, 680, 39–53. <https://doi.org/10.1016/j.neulet.2018.05.006>
- Semon, R. W. *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens*. Engelmann. <https://doi.org/10.5962/bhl.title.10234>
- Shohamy, D., & Wagner, A. D. (2008). Integrating Memories in the human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron*, 60(2), 378–389. <https://doi.org/10.1016/j.neuron.2008.09.023>
- Smith, S. M., Handy, J. D., Angello, G., & Manzano, I. (2013). Effects of similarity on environmental context cueing. *Memory*, 22(5), 493–508. <https://doi.org/10.1080/09658211.2013.800553>
- Sohn, M., Goode, A., Stenger, V. A., Carter, C. S., & Anderson, J. R. (2003). Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12), 7412–7417. <https://doi.org/10.1073/pnas.0832374100>
- 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, Article e17397. <https://doi.org/10.7554/elife.17397>
- 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>
- Stark, S. M., Kirwan, C. B., & Stark, C. E. (2019). Mnemonic Similarity Task: a tool for assessing hippocampal integrity. *Trends in Cognitive Sciences*, 23(11), 938–951. <https://doi.org/10.1016/j.tics.2019.08.003>
- Tambini, A., Berners-Lee, A., & Davachi, L. (2017). Brief targeted memory reactivation during the awake state enhances memory stability and benefits the weakest memories. *Scientific Reports*, 7(1), Article 15325. <https://doi.org/10.1038/s41598-017-15608-x>

- 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>
- Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86*(2), 255–262. <https://doi.org/10.1037/h0029997>
- Tulving, E. (1974). Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American scientist, 62*(1), 74-82.
- Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77*(4), 593–601. <https://doi.org/10.1037/h0026069>
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 381–391. [https://doi.org/10.1016/s0022-5371\(66\)80048-8](https://doi.org/10.1016/s0022-5371(66)80048-8)
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. <https://doi.org/10.1037/h0020071>
- Vieweg, P., Riemer, M., Berron, D., & Wolbers, T. (2018). Memory Image Completion: Establishing a task to behaviorally assess pattern completion in humans. *Hippocampus, 29*(4), 340–351. <https://doi.org/10.1002/hipo.23030>
- Vieweg, P., Stangl, M., Howard, L. R., & Wolbers, T. (2015). Changes in pattern completion – A key mechanism to explain age-related recognition memory deficits? *Cortex, 64*, 343–351. <https://doi.org/10.1016/j.cortex.2014.12.007>
- Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society, 17*(05), 766–780. <https://doi.org/10.1017/s1355617711000683>
- Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., & Xue, G. (2017). Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. *Journal of Neuroscience, 37*(11), 2986–2998. <https://doi.org/10.1523/jneurosci.2324-16.2017>
- Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34*(10), 515–525. <https://doi.org/10.1016/j.tins.2011.06.006>
- Yum, K. S. (1931). An experimental test of the law of assimilation. *Journal of Experimental Psychology, 14*(1), 68–82. <https://doi.org/10.1037/h0071335>