Learning to distinguish: Shared perceptual features and discrimination practice tune behavioural pattern separation

Dorottya Bencze[1,2], Ágnes Szőllősi[1,2], and Mihály Racsmány[1,2]

[1] Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary

[2] Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary

Correspondence concerning this article should be addressed to Dorottya Bencze. E-mail: bencze.dorottya@ttk.bme.hu. Postal address: Department of Cognitive Science, Budapest University of Technology and Economics, 1111-Budapest, Hungary, Egry Jozsef utca 1. Telephone: +3614633525.

Word count for main text (excluding the title page, abstract, keywords, references, and figure captions/legends): 10422.

Abstract

Pattern separation is a computational mechanism performed by the hippocampus allowing the reduction of overlap between sensory inputs with similar perceptual features. Our first aim was to develop a new paradigm sensitive to the behavioural consequences of pattern separation (mnemonic discrimination). For this purpose, we constructed morphed face stimuli with parametrically changing levels of similarity. After encoding participants saw studied items and similar lure faces. Perceptual similarity affected false recognition and there was a gradual reduction in discrimination accuracy with the increment of similarity between the stimuli. However, confidence ratings were sensitive to smaller changes (Experiment 1) than the other test type with "old"/"similar"/"new" response options (Experiment 2). Mnemonic discrimination relies strongly on retrieving details of the original stimulus. Therefore, we investigated whether pattern separation can be tuned by retrieval in the form of a discrimination task (Experiment 3). Our findings suggest that repeatedly encountering the stimuli within a two-alternative forced-choice task (in comparison with the repeated presentation of the material) increased both the correct identification and the false recognition of similar stimuli two days after encoding. We conclude that basic computational mechanisms of the hippocampus can be tuned by a task that requires discrimination between studied and new stimuli.

Introduction

Tools for Assessing the Behavioral Consequences of Hippocampal Pattern Separation

One crucial feature of human memory is the ability to encode and store events in the form of distinct, episodic memory representations (Tulving, 1972, 2002). This ability requires the resolution of interference between overlapping representations, and strongly depends on a computational mechanism in the hippocampus, called pattern separation, that allows the reduction of overlap among sensory inputs that share similar features (for reviews, see Hunsaker & Kesner, 2013; Rolls, 2013; Yassa & Stark, 2011). As a consequence of pattern separation, neuronal activities of brain circuits become distinct in response to different but still similar stimuli (Bakker, Kirwan, Miller, & Stark, 2008; Gilbert & Kesner, 2006; Hunsaker & Kesner, 2013). At a behavioral level, as a result of pattern separation, one becomes able to correctly identify and reject a stimulus that is only similar to (but not the same as) a previously perceived item (see Kirwan & Stark, 2007; Stark, Kirwan, & Stark, 2019).

The Mnemonic Similarity Task (MST) is a modified recognition memory task that places strong demands on mnemonic discrimination between similar items, which is the reason why it is a frequently used tool to assess the behavioral consequences of pattern separation in humans (for a recent overview, see Stark et al., 2019). In this task, images are shown to the participants in an incidental learning situation and later memory is tested in the form of a recognition memory task. Following the tradition of recognition memory research, participants are presented with old target items (stimuli encountered at encoding) and completely new foils. As a modification, half of the target stimuli are not presented on the recognition test. Instead, critical lure items are shown that are visually similar images to ones presented at encoding.

In the original version of the MST, participants have three response options: "old", "new", and "similar". The principal idea is that the correct identification of a lure indicates successful pattern separation, as can be measured by the so-called Lure Discrimination Index, i.e., the ratio of "similar" responses given to the lures (minus the ratio of "similar" responses given to the foils, when response bias is controlled for). Since this Lure Discrimination Index seems to be sensitive to hippocampal integrity (rather than a standard recognition hit rate, i.e., the correct identification of a target item), the

MST has become a benchmark test to assess hippocampal dysfunctions in normal and pathological ageing (Stark & Stark, 2017; Stark, Stevenson, Wu, Rutledge, & Stark, 2015; Stark, Yassa, Lacy, & Stark, 2013; Yassa, Lacy, Stark, Albert, Gallagher, & Stark, 2011) as well as in various neurological disorders (Planche, Ruet, Charre-Morin, Deloire, Brochet, & Tourdias, 2017; Poch, Toledano, Garcia-Morales, Prieto, Garcia-Barragan, Aledo-Serrano, Gil-Nagel, & Campo, 2020; Wesnes & Burn, 2013).

The MST has been adopted in many variants, using a diverse assortment of stimuli. The most widespread version of the MST is performed using everyday objects, manipulating the similarity between targets and lures across a wide range of characteristics, such as the color, size, and shape of the objects, just to name a few (Stark et al., 2019). The similarity level of the lures in this version is not determined by changing the physical parameters of the images, but by the rates of false alarm judgements given to the specific lures on a memory test performed by an independent group of participants (Lacy, Yassa, Stark, Muftuler, & Stark, 2011). Nevertheless, there have been studies using the MST with parametric changes in the perceptual features of the stimuli to create different levels of similarity in lures. One of the most common manipulations used in such experiments is changing the degree of rotation of the lures compared to the original stimuli (Motley & Kirwan, 2012). However, restricting the manipulation of similarity to one type of feature is suggested to allow participants to employ specific strategies for identifying lures or to give short verbal labels to the stimuli (e.g., "red car, 45 degrees") (see Stark et al, 2019).

As a further example, for studies designed to investigate behavioral pattern separation faces are a type of stimuli especially suited to create lures. For example, Kirwan and Stark (2007) used photographs of faces, changing their various properties between targets and lures (e.g., expression, clothes, lighting), whereas in another study Chang, Murray, and Yassa (2015) used morphed faces with parametrically changing similarity levels. Faces can be characterized by consistent primary features (the positions of eyes, nose, and mouth are fairly constant), and therefore, it is relatively easy to construct stimuli with the perceptual similarity held under control by morphing faces into each other. Remembering faces is suggested to be a nonverbal activity and most of the time faces can be described only in vague terms (Ellis, Shepherd, & Davies, 1980; Lindsay, Martin, & Webber, 1994). Actually, in certain cases, verbalization of facial images can impair their visual recognition (known as verbal

overshadowing, see e.g., Dodson, Johnson, & Schooler, 1997; Meissner & Brigham, 2001), indicating that verbal labels would not aid the discrimination of similar faces. For this reason, experimental designs using facial stimuli in the MST could reduce the inter-individual variability in verbalization strategies compared to object stimuli.

The Role of Retrieval in Lure Discrimination and in Long-Term Memory Retention

Strongly related to hippocampal pattern separation and its behavioral manifestation (lure discrimination), current models of recognition memory stress the crucial role of memory retrieval in the proper discrimination of highly similar stimuli. These models propose that a so-called recall-to-reject process is acting during the correct rejection of a new item that is similar to an old target item (see Clark & Gronlund, 1996; Rotello & Heit, 1999; Rotello, Macmillan, & Van Tassel, 2000). In other words, when there is an overlap between the features of an old and a new item, one must retrieve the original stimulus to detect differences between them. Accordingly, computational models of the hippocampal formation also suggest that for the correct discrimination between targets and highly similar lures individuals tend to use the recall-to-reject strategy (Norman & O'Reilly, 2003). Many MST studies propose that when giving a "similar" response to a lure, participants might first retrieve the associated target to correctly reject the lure (e.g., Kim & Yassa, 2013; Kirwan & Stark, 2007; Yassa, Lacy, Stark, Albert, Gallagher & Stark, 2011). That is to say, these ideas all emphasize the important role of *memory retrieval* in successful lure discrimination and in pattern separation (see also Kirwan & Stark, 2007). Based on this concept, one could raise the question whether retrieval processes known to affect memory retention are able to change the functioning of pattern separation and its behavioral manifestation.

Another line of research long established that practicing the material via retrieval promotes subsequent memory performance even more than additional study opportunities do (Putnam, Nestojko, & Roediger, 2016; Roediger & Butler, 2011; Roediger & Karpicke, 2006). In addition to its facilitative effect on memory retention, retrieval reduces interference effects when the study material includes stimuli that share similar features (Racsmány & Keresztes, 2015; Szpunar, McDermott, & Roediger, 2008).

5

The majority of studies investigating the beneficial impact of retrieval reported a facilitating effect on memory in tasks measuring recall performance (e.g., Karpicke & Roediger, 2008; Wheeler, Ewers, & Buonanno, 2003). The effects of retrieval on recognition memory performance have been investigated by fewer studies compared to recall, and at first glance, their findings present a somewhat less conclusive picture (see the meta-analytic review of Rowland, 2014). Nevertheless, several studies found that retrieval benefits recognition memory performance (e.g., Chan & McDermott, 2007; Read, 1979; Verde, 2004; but see Lockhart, 1975; McDaniel, Anderson, Derbish, & Morrisette, 2007). Chan and McDermott (2007) proposed that although the effect of retrieval is not always apparent in classic recognition memory performance, it can change the processes underlying recognition decisions. The authors found that source memory performance, performance in an exclusion memory task as well as in a Remember/Know task increased following retrieval practice even when recognition hit rates were unaffected. They concluded that retrieval mainly affects recognition decisions when the task requires the retrieval of specific details associated with the encoding episode, which is not always reflected in standard recognition hit rates. This view has been supported by later studies as well (e.g., Guran, Lehmann-Grube, & Bunzeck, 2019; Verkoeijen, Tabbers, & Verhage, 2011; but see Shaffer & McDermott, 2020).

In sum, based on the empirical findings described above, one can assume that there is a strong interplay between pattern separation, memory retrieval, and memory retention. First, retrieval is known to lead to a reduction in interference effects (Racsmány & Keresztes, 2015; Szpunar et al., 2008). Relatedly, hippocampal pattern separation refers to the process of interference resolution at a computational level. Additionally, retrieval promotes later recognition memory when the task requires the retrieval of specific details of a memory representation (Chan & McDermott, 2007; Guran et al., 2020; Verkoeijen et al., 2011). Accordingly, for successful pattern separation and lure discrimination, the formation of detailed memory representations is needed, as one needs to compare the details of a previously seen item and a visually similar lure (see Kirwan & Stark, 2007). Based on this set of ideas and empirical findings, it can be assumed that pattern separation (as a process that requires access to specific details associated with the target information) can be tuned by retrieval, leading to an increase in lure discrimination performance at a behavioral level.

Study Objectives

To investigate whether pattern separation can be tuned by repeatedly re-encountering the material in a recognition memory task involving discrimination, we developed an experimental paradigm to examine the behavioral manifestation of pattern separation using lure stimuli with similarity levels manifested in the overlap of physical features between target and lure stimuli, manipulated in a parametric manner. Facial stimuli were chosen because unfamiliar faces are difficult to label verbally (Ellis et al., 1980; Lindsay et al., 1994). Additionally, due to their consistent configuration of primary features lure faces are easy to generate with various levels of perceptual overlap. Parametrically changed faces have been successfully used in one previous MST study to investigate pattern separation. Chang and colleagues (2015) examined a face recognition specific phenomenon, the other race effect, in relation to pattern separation performance. For the aforementioned reasons, we created lure facial stimuli overlapping with the original face by morphing two unique faces.

Our experimental design incorporates several methodological features used by previous MST studies. In designing and testing our paradigm we combined and compared several of these features: different types of encoding and recognition tasks, multiple levels of lure similarity. In the first two of three experiments we used the generated stimuli first in a task widely used in the recognition memory literature to measure discrimination (Egan, 1958) and then in the most popular version of the MST, which is predominantly used in neuroimaging studies (Stark et al., 2019). Consequently, in Experiment 1, we tested lure discrimination in a task where participants made recognition confidence judgements on a scale ranging between "Sure it was new" and "Sure it was old". Then, in Experiment 2, we assessed the behavioral correlates of pattern separation in a task using "old"/"similar"/"new" response options. In brief, the main purpose of the first two experiments was to develop a task that pertains with facial stimuli and is sensitive to lure discrimination performance when memory performance is assessed by recognition confidence ratings (widely used in the experimental psychological literature) and by the ratio of "old", "similar" and "new" decisions (widely used in the neuroimaging literature). Except for this difference in the types of the recognition memory test, all other experimental conditions were the same in Experiment 1 and 2.

In Experiment 3, we used the design developed in the first two experiments to examine whether pattern separation can be tuned by retrieval in the form of a discrimination task. We included a practice phase after initial encoding where participants practiced discriminating the original faces from novel faces in a two-alternative forced-choice recognition memory task. The effect of discrimination tuning on pattern separation was assessed later in a recognition task with "old"/"similar"/"new" response options.

Experiment 1

The aim of the first experiment was to develop a paradigm that allows the investigation of lure discrimination for facial stimuli. We created a set of stimuli where similarity was manipulated parametrically on a continuum of overlapping features between the original (target) and a new face by morphing. To create lure faces with different degrees of overlap to the original ones, we used a morphing continuum in steps of 25%, generating lures that were 25%, 50%, and 75% dissimilar to the original face. Discriminations between previously seen target faces, partially overlapping lures, and completely new foils were measured in a recognition memory task at the final stage.

We aimed to create an experimental design where lure discrimination relies more heavily on the detailed recollection of the original stimulus, requiring the participants to employ the recall-to-reject strategy. Some studies propose that participants tend to use the recall-to-reject strategy for successful lure discrimination in the traditional form of the MST (e.g., Kirwan & Stark, 2007), however, some recent findings suggest that the feeling of familiarity also contributes to lure discrimination performance (Kim & Yassa, 2013; Szőllősi, Bencze, & Racsmány, 2020). Other studies examined lure discrimination in a recognition task involving multiple similar lures presented together with their corresponding target item in the same task (see: Holdstock, Mayes, Roberts, Cezayirli, Isaac, O'Reilly, & Norman, 2002). Holdstock and colleagues (2002) found that lure discrimination in a yes/no recognition task was impaired in a patient with selective hippocampal damage, and it was reported that healthy individuals mainly use recall-to-reject strategy in the same task (Migo, Montaldi, Norman, Quamme, & Mayes, 2009). For this reason, we measured lure discrimination performance in a recognition task where beside each target face (faces encountered in the encoding phase) all versions of its corresponding morphed

lures were presented along with morphed and unmorphed foils. This modification in the experimental design was implemented for the purpose of increasing demand for using the recall-to-reject strategy instead of relying on the sense of familiarity for lure discrimination. Encountering multiple versions of the same item during the recognition task would discourage participants from deciding based on familiarity, since an item could feel familiar from either the encoding phase, or from encountering some morphed version of it in the recognition task.

In recognition memory studies one way to examine recognition memory performance is the analysis of receiver operating characteristics (ROC) (Egan, 1958). The ROC curve is a plot of hits ("old" responses given to the old items) against false alarms (incorrect "old" responses given to the new items). When only "old" and "new" response options are used in a recognition memory task, response criterion can be manipulated experimentally (Hockley, 2011; see also Stretch & Wixted, 1998). Another way is when participants are instructed to make a recognition confidence judgment on a scale. In this latter case, each level reflects a response criterion and no intervention is needed to experimentally manipulate the response criteria (for an overview, see Weidemann & Kahana, 2016). Relatedly, a couple of recent studies also combined the MST with confidence ratings (e.g., Loiotile & Courtney, 2015; Stark et al., 2015; Szőllősi et al., 2020). In these studies, the signal detection framework has been used together with traditional MST scores and/or tasks. The main advantage of the signal detection approach in pattern separation research is that the various memory discrimination scores derived from the ROC analysis yield information about the overlap between the representation of the two measured stimulus types, e.g., the overlap between target and lure representations (see Loiotile & Courtney, 2015). These studies used ROC analysis with the stimulus set of the traditional MST, while in the present study we aimed to expand this line of research by using parametrically changed stimuli from the same semantic category (faces). Due to these multiple reasons, we asked participants to make recognition confidence judgments. Specifically, participants were asked to make a decision confidence judgment on a 6-point scale ranging between "Sure it was new" and "Sure it was old".

An additional matter of design that differs between typical recognition memory paradigms and tasks designed to assess behavioral pattern separation is the instructions in the encoding phase. The behavioral manifestation of putative neural mechanisms enabling pattern separation is often assessed by

recognition memory tasks following incidental encoding (see Stark et al., 2019). However, in both the recognition memory literature and studies investigating the effect of retrieval on later memory, performance is often measured following intentional learning (e.g., Chan & McDermott, 2007; Verde, 2004). There has been a long line of research that investigated the effect of encoding type (incidental/intentional) on recognition performance. The results presented by the studies of this subject are varied: some authors report the advantage of intentional encoding (e.g., Block, 2009; Noldy, Stelmack, & Campbell, 1990), other findings show better memory performance following incidental encoding (e.g., Bernstein, Beig, Siegenthaler, & Grady, 2002; Choe, Kardan, Kotabe, Henderson, & Berman, 2017), and there are studies observing no difference between the two encoding types (e.g., Bird, 1976; Ferrara, Puff, Gioia, & Richards, 1978). However, to the best of our knowledge, there has only been one study examining the impact of encoding type on lure discrimination performance (Stark, Stevenson, Wu, Rutledge & Stark, 2015), and none using a signal detection approach. Stark and colleagues found no difference in lure discrimination between incidental and intentional encoding, although they did not account for different levels of similarity in lures. Therefore, we aimed to expand this line of research by exploring the effects of encoding on lure discrimination in case of varying levels of similarity without strong a priori predictions. In our experiment, half of the participants studied the material in an incidental encoding situation, while the remaining participants were explicitly instructed to study the material.

Materials and Methods

Participants

In all three experiments participants were Hungarian undergraduate students and received course credit for participation. All participants were Caucasian. No participant had a history of psychiatric and neurological disorders and they all had normal or corrected-to-normal vision. All participants gave written informed consent. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology and was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

The sample size of the groups in both Experiment 1 and 2 was based on previous MST studies (Kirwan & Stark, 2007; Loiotile & Courtney, 2015; Motley & Kirwan, 2012; Stark, Stevenson, Wu, Rutledge, & Stark, 2015).

There were 38 participants in Experiment 1 (6 men; age range: 19-26 years). Participants were randomly assigned to either the incidental encoding group ($N = 19$, 4 men; age range: 19-24 years) or to the intentional encoding group ($N = 19$, 2 men; age range: 19-26 years). No participant was excluded from the sample.

Stimuli

For stimuli exemplars, see Figure 1a. In all three experiments we used greyscale images of faces as stimuli. We used Caucasian, young adult male faces since findings of previous studies indicate that participants show a comparable recognition memory performance for male faces regardless of the subjects' gender, while female faces are remembered less by male participants (Lewin & Herlitz, 2002). The faces were presented from a front view with the hair, ears, and shape of the face masked, and without any facial hair or other salient characteristics (e.g., glasses). The stimuli were obtained from freely available online databases (DeBruine & Jones, 2017; Minear & Park, 2004).

Four stimulus types were used in the recognition memory task: (1) targets (original faces that were presented at encoding), (2) lures (morphed faces), (3) morphed foils (new faces that were used during the morphing process but were not presented at encoding), and (4) unmorphed foils (new faces that were not used during the morphing process and were not presented at encoding). Morphed face stimuli were created by using the photo morphing software Abrosoft FantaMorph 5.3.8 (Abrosoft Co.) to generate morphs between two unique faces along a continuum in steps of 25% (25%, 50%, and 75% morphs). We aimed to encompass the whole morphing continuum from 0% to 100% (here, 0% refers to the target items, whereas 100% refers to the foils). Therefore, we used a morphing step of 25%, generating lures that were 25%, 50%, and 75% dissimilar to the original face stimuli that were presented at encoding. We use the terms "25% morph", "50% morph", and "75% morph" to indicate that the lure contains features of a face that is different from the original face in 25%, 50%, and 75% proportions, respectively. Consequently, a 25% morphed lure resembles the original face greatly, a 50% or 75%

morphed lure less so. In this continuum of similarity target stimuli could be placed at the 0% point of the spectrum with morphed foils being the 100% point, sharing no features with the targets. Unmorphed foils are completely outside of this continuum. We used the morphed lures in the recognition task, along with the two types of foil faces. Morphed foils are faces that were not presented previously, but were used in the morphing process, and thus share visual features with some of the morphed lures. Unmorphed foils are faces not used in the morphing process and consequently they do not share features with any of the other types of stimuli.

On the recognition test an equal number of each stimulus type was presented. Even though this design feature resulted in a lower number of targets compared to the traditional MST, we did not describe the rate of stimulus types in the task instructions. Based on previous findings it is unlikely that without making the participants aware of the distribution of the stimulus types this manipulation would cause them to shift their criterion resulting in response bias (see e.g., Estes & Maddox, 1995; Franks & Hicks, 2016, Koop, Criss & Malmberg, 2015; for previous MST studies using a different distribution of stimuli from the traditional design see: Chang et al., 2015; Motley and Kirwan, 2012).

In the recognition task phase each target face (faces encountered in the encoding phase) had its original version as well as each type of its morphed versions presented along with morphed and unmorphed foils. Presenting each version of the stimuli was a modification added to decrease the probability of participants relying on familiarity when making decisions about lure items. We used two types of foils in the recognition task (morphed and unmorphed foils). Since morphed foils share visual features with some of the morphs, there is a possibility that they could evoke some sense of familiarity; while unmorphed foils do not share any features with the other stimulus types, not evoking the same feeling of familiarity. In all three experiments the faces were presented on a grey background at the size of 8.6° x 8.6° visual angles.

(Figure 1 about here)

Experimental Design and Procedure

12

Experiment 1 consisted of two phases separated by a 5-minute delay: an encoding phase and a recognition memory task. The experimental procedure is illustrated in Figure 1b.

During the encoding phase participants saw 20 images of faces in random order for three consecutive cycles, with all faces being presented in each cycle. There was no delay between the encoding cycles. The faces were presented in the middle of the computer screen for 7000 ms each with a pre-stimulus interval (PSI) of 1000 ms. Participants in the intentional encoding condition were instructed to memorize the faces. Participants in the incidental encoding condition made a pleasant/unpleasant decision of the presented faces (for the previous application of this instruction type, see e.g., Burgess & Gruzelier, 1997) using the mouse of the computer to click on the chosen option. Responses were given with a computer mouse instead of a keyboard in order to avoid the confounding effect of motor learning across encoding conditions. Response options ("Unpleasant", "Rather unpleasant", "Rather pleasant", and "Pleasant") remained written on the bottom of the computer screen for the duration of each encoding cycle. The instructions were presented at the beginning of each cycle, and participants could start the task by pressing the Space bar.

The encoding phase was followed by a 5-minute delay. In the first three minutes of the delay participants were given an arithmetic distractor task when they solved a list of exercises on paper without using a calculator. The aim for the distractor task was to prevent rote rehearsal. Following the arithmetic distractor task, there was a short, 2-minute response practice phase where participants were presented with the labels of the confidence scale (e.g., "It is probably new") and were asked to press the corresponding response button on a standard keyboard of the computer ("X" in this example). The labels were presented in random order during the response practice phase with each label remaining on the screen for 5000 ms (with a PSI of 1000 ms).

In the recognition task, participants saw 124 images of faces in the middle of the computer screen. Twenty images were targets, i.e., exact repetitions of the faces presented during the encoding phase. Sixty images were morphed lure faces, 20 of each type of morph (25%, 50%, 75% morphed). Each face from the encoding phase and each type of its morphed versions were presented during the recognition task. Twenty images were foils that were used in the morphing process (morphed foils) and twenty images were unmorphed foils. The first four images presented in the memory task were fillers

that were not used in the experiment previously, in order to avoid the effects of primacy. Responses given to the fillers were not analyzed.

The faces were presented in a random order (with the exception of the fillers) and each face remained on the screen for 5000 ms preceded by a 1000-ms PSI. Participants were asked to make a recognition confidence judgement on a 6-point scale. The corresponding keys were the "Y", "X", "C", "B", "N", and "M" keys on a QWERTZ keyboard, where Y = "Sure it was new", M = "Sure it was old". The response labels of the confidence scale remained on the screen for the duration of the recognition task. Stimulus presentation was controlled by Psychtoolbox 3.0.12 (Brainard, 1997), a set of MATLAB (version 2014a, The MathWorks, Inc., Natick, Massachusetts, US) and GNU Octave (GNU General Public License; Eaton, Bateman, Hauberg, & Wehbring, 2018) functions.

Data Analysis

To investigate the discrimination of target items from each type of morphed lures and foils, we conducted ROC analyses using the ROC Toolbox (The Regents of the University of California, Oakland, California, US) developed in the MATLAB computing environment (R2018a, The MathWorks, Inc., Natick, Massachusetts, US) by Koen, Barrett, Harlow, and Yonelinas (2017). Separate ROC curves were generated by plotting confidence ratings given to the target faces against false alarm rates for each type of lures (incorrect "old" responses given to either 25%, 50%, or 75% morphed lures) and against foil faces (incorrect "old" responses given to the morphed or unmorphed foils). ROC curves were fitted to the data of each participant individually.

We aimed to assess the accuracy of the discrimination between targets and lures and the discrimination between targets and foils without using any model-specific assumptions. Several measures have spread in the recognition memory literature, however, these indices all depend on assumptions of the underlying distributions. One important advantage of analyzing ROC curves is that it provides an opportunity to calculate a reliable measure of performance (sensitivity) without any model assumption (an index called the Area Under the Curve, *AUC*; Green & Moses, 1966; Macmillan & Creelman, 2005; Pollack & Hsieh, 1969; Pollack & Norman, 1964; Verde, Macmillan, & Rotello, 2006). The AUC has been used in previous studies combining the MST with the signal detection approach as

14

well (Loiotile & Courtney, 2015; Szőllősi et al., 2020). For the purpose of measuring lure discrimination, Loiotile and Courtney (2015) suggest the AUC as a memory discrimination measure in case one would not want to make assumption about the underlying distributions of recognition strength, and reported comparable results using this measure to their analysis on the same data using d_a (a more frequently used discrimination measure of signal detection theory, that assumes normal distribution). Therefore, we used the *AUC* as a measure of discrimination accuracy between targets and morphed lures as well as between targets and foil items.

We analyzed the effect of encoding type (incidental and intentional) and the type of the stimuli that needed to be discriminated from the targets (25%, 50% or 75% morphed lures, morphed foils, and unmorphed foils) on the *AUC* values signifying accuracy of the discrimination. For this purpose, we conducted a 2 x 5 (Encoding x Stimulus type) mixed-design ANOVA with the degrees of freedom adjusted with Greenhouse-Geisser corrections when appropriate.

Results

The *AUC* values are presented in Figure 2. The ANOVA indicated no effect of encoding type on the *AUC* values of the ROC curves. Specifically, there was no significant main effect of Encoding, $F(1, 36) = 1.576$, $p = .217$, $\eta^2_p = 0.042$, and no Encoding x Stimulus type interaction, $F(2.69, 96.76) = 1.261$, $p = .292$, $\eta^2_p = 0.034$. On the other hand, we found a significant main effect of Stimulus type, $F(2.69, 96.76) = 334.080$, $p < .001$, $\eta^2_p = 0.903$.

Further contrast analyses (using Bonferroni corrections) showed a significant difference in the accuracy of all target-morphed lure and target-foil discriminations, with an increased accuracy for item types that were more dissimilar to the targets. The *AUC* was significantly lower for target vs. 25% morph than it was for target vs. 50% morph discrimination, $F(1, 36) = 114.856$, $p < .001$, $\eta^2_p = 0.761$. Target vs. 50% morph discrimination had lower accuracy than target vs. 75% morph discrimination, $F(1, 36) = 145.980$, $p < .001$, $\eta^2_p = 0.802$, and target vs. 75% morph discrimination was less accurate than the discrimination of targets from morphed foils, $F(1, 36) = 42.181$, $p < .001$, $\eta^2_p = 0.539$. Target vs. morphed foil discrimination had a lower *AUC* than target vs. unmorphed foil discrimination, $F(1, 36) = 32.234$, $p < .001$, $\eta^2_p = 0.472$.

In sum, the ability to discriminate target items from morphed lures and foils was not affected by the type of encoding (incidental or intentional). The accuracy of the discrimination decreased when the similarity between targets and lures was higher. The accuracy was the highest when the foil face did not share any feature with any of the presented faces.

(Figure 2 about here)

Experiment 2

The results of Experiment 1 indicated that the paradigm is sensitive to lure discrimination performance for facial stimuli as measured in a task that assessed recognition confidence frequencies. However, in the behavioral pattern separation literature, participants typically have three response options: "old", "similar", and "new" (see e.g., Kirwan & Stark, 2007; Stark et al., 2013). In this case, the ratio of "similar" responses given to the lure items is the measure of lure discrimination performance. Therefore, in our second experiment, participants were not instructed to make confidence judgments (as in Experiment 1); instead, they were asked to make an "old"/"similar"/"new" decision. All other experimental conditions were the same as they were in Experiment 1.

Materials and Methods

Participants

In Experiment 2 there were 41 participants (3 men; age range: 19-26 years). They were randomly assigned to either the incidental encoding group ($N = 20$, 1 man; age range: 19-26 years) or to the intentional encoding group ($N = 21$, 2 men; age range: 19-24 years).

Experimental Design and Procedure

The stimuli and the encoding phases of the experiments were identical in Experiment 1 and 2. The experimental procedure is illustrated in Figure 1b.

In the following recognition memory task participants saw 164 faces (20 targets; 20 of 25%, 50%, and 75% morphs; 20 morphed; 20 unmorphed foils and 4 fillers) in the middle of the computer screen. In contrast to Experiment 1, instead of a confidence judgement, participants were instructed to decide whether they had seen the exact same face before ("old"), or the face was completely new, or the face they saw was just similar to a face they had seen during the encoding phase ("similar"). The response labels (M = "Old", V = "Similar", and Y = "New" on a QWERTZ keyboard) remained on the screen for the duration of the recognition task. The faces were presented in a random order (with the exception of the fillers) and each face remained on the screen for 5000 ms (PSI = 1000 ms). Similar to Experiment 1, the recognition task was preceded by a 5-minute delay, consisting of a 3-minute arithmetic distractor task and a 2-minute response practice phase. During the response practice phase participants saw the labels of the response options (e.g., "Old") and were instructed to press the corresponding response button ("M" in this example). The labels were presented for 5000 ms with a PSI of 1000 ms.

Data analysis

To investigate whether the type of encoding (incidental and intentional) had an effect on recognition performance, we analyzed the recognition score (rate of "old" responses to targets minus the rate of "old" responses to unmorphed foils) of the target items in a two-tailed independent-samples t-test. For morphed lures, we examined the effect of encoding type and the level of similarity to the target face (25%, 50%, and 75% morphed faces) on the lure discrimination scores and false recognition scores of morphed lures. Lure discrimination scores for all lure similarity levels were calculated as the rate of "similar" responses to lures minus the rate of "similar" responses to unmorphed foils to account for the participants' overall tendency of giving "similar" responses (see e.g., Stark et al., 2013; Yassa et al., 2011). False recognition scores were calculated as the rate of "old" responses to lures minus the rate of "old" responses to unmorphed foils. We conducted two 2 x 3 mixed-design ANOVAs followed by a list of paired contrast analyses of the morphing levels for lure discrimination and false recognition scores for the lure items.

Results

For the recognition scores of the target items we found no difference between the incidental (M = 68.0%, SE = 4.0) and intentional (M = 73.3%, SE = 3.5) encoding $t(39)$ = 1.006, p = .321, d = 0.314.

Lure discrimination scores of all lure items are illustrated in Figure 3a. The two encoding conditions did not differ in the lure discrimination scores of morphed lure faces. Specifically, neither the main effect of Encoding, $F(1, 39)$ = 0.002, p = .963, η^2_p < 0.001, nor the Encoding x Similarity interaction, $F(1.97, 76.93)$ = 0.247, p = .778, η^2_p = 0.006, was significant. However, morphed faces showed a significant main effect of Similarity in lure discrimination, $F(1.97, 76.93)$ = 25.888, p < .001, η^2_p = 0.399, regardless of the encoding type. Further contrast analysis (using Bonferroni correction to account for multiple comparisons) revealed a significantly lower lure discrimination score for 25% morphs compared to 50% morphs, $F(1, 39)$ = 45.759, p < .001, η^2_p = 0.540. The two weaker morphs (50% and 75%) did not differ significantly in regards of lure discrimination performance, $F(1, 39)$ = 2.411, p = 0.386, η^2_p = 0.058.

False recognition scores of the lure items are illustrated in Figure 3b. False recognition scores of morphed lures showed no significant main effect of Encoding, $F(1, 39)$ = 0.696, p = .409, η^2_p = 0.018, and no Encoding x Similarity interaction, $F(1.69, 66.06)$ = 0.198, p = 0.784, η^2_p = 0.005. However, there was a significant main effect of Similarity, $F(1.69, 66.06)$ = 212.442, p < .001, η^2_p = 0.845. Further contrast analysis showed a significant difference between both 25% and 50%, $F(1,39)$ = 170.384, p < .001, η^2_p = 0.814, as well as between 50% and 75% morphed faces, $F(1,39)$ = 67.596, p < .001, η^2_p = 0.634.

To summarize, the intentional or incidental nature of encoding did not affect the ability to correctly recognize target faces or to correctly label a morphed face as "similar". There was no effect of encoding type on the false recognition of morphed lures either. More importantly, the degree of similarity of the morphed lures to the original face affected the responses given to the lures: strong (25%) morphs were more likely to be labelled as "old" and less likely as "similar" than weaker (50%) morphs, and between weaker morphs, 50% morphs were more likely to labelled as "old" than 75% morphs. In sum, while lure discrimination performance did not show a difference between weaker morphs, "old" response rates given to the lures differed between each morph condition. This pattern of findings

suggests that the lure discrimination index and the false recognition of lures are not the two sides of the same coin.

(Figure 3 about here)

Experiment 3

The results of the first two experiments indicate that the task is sensitive to the discrimination of lure faces when participants' task is to make recognition confidence decisions (Experiment 1) and also when there are three response options (Experiment 2). In Experiment 3, we aimed at investigating whether lure discrimination performance is sensitive to practice, i.e., when participants had a chance to practice discrimination between studied faces and faces they have not encountered before. Since in Experiment 1 and 2, the encoding instruction (incidental/intentional) did not affect memory performance (either sensitivity or discrimination), in Experiment 3, there was only one group of subjects who studied the material in an intentional encoding situation. On the final recognition memory test, there were three response options, just as in the second experiment. Mnemonic discrimination is typically assessed immediately after the presentation of the study material (e.g., Segal, Stark, Kattan, Stark, & Yassa, 2012; Stark, Yassa, Lacy, & Stark, 2013). Similarly, we did not use a long retention interval after the encoding phase, just a short delay of five minutes in Experiment 1 and 2. In contrast, the focus of Experiment 3 was to investigate the effect of retrieval practice on lure discrimination. Previous studies have shown that the beneficial effect of retrieval practice on memory retention can only be detected days or even weeks after learning (Thompson, Wenger, & Bartling, 1978; Wheeler, Ewers, & Buonanno, 2003). Therefore, as a novelty, we used a longer retention interval of two days between learning and the final memory test in Experiment 3.

Materials and Methods

Participants

There were 41 participants in Experiment 3 (6 men, age range: 20-28 years). No participant was excluded from the sample. A sample size equal to the combined two groups of Experiment 2 was chosen because in the previous two experiments the effect of lure similarity was calculated using the whole sample, and, importantly, it resulted in a sample size similar to previous studies' investigating the effect of retrieval practice on recognition (see e.g., Guran, Lehmann-Grube & Bunzeck, 2019; Verkoeijen, Tabbers & Verhage, 2011).

Experimental Design and Procedure

Experiment 3 included some crucial modifications compared to Experiment 1 and 2. Experiment 3 consisted of three phases: and encoding phase, a practice phase, and a final recognition memory task. The experimental procedure is illustrated in Figure 4.

In the encoding phase, participants were presented with 16 images of faces in the middle of the computer screen for three consecutive cycles. We used a smaller stimulus set size of 16 faces compared to the 20 faces in Experiment 1 and 2 to avoid fatigue effect. In each cycle, all faces were presented in random order for 5000 ms preceded by a 1000-ms PSI. Before each encoding cycle participants were instructed to memorize the faces, and they proceeded to the encoding cycle by pressing the Space bar. There was no delay between the encoding cycles.

The encoding phase was followed by a 5-minute arithmetic distractor task where participants solved a list of exercises on paper, without using a calculator. The aim of the distractor task was to eliminate the effect of rote rehearsal on memory.

The following practice phase consisted of five cycles. We included a practice phase with multiple memory tests after initial encoding in Experiment 3, because previous studies have shown that repeated retrieval practice is more beneficial in terms of long-term memory retention and reduces interference effects more effectively than one single retrieval attempt does (Racsmány, Szőllősi & Bencze 2018; Racsmány & Keresztes, 2015; Szpunar, McDermott, & Roediger, 2008; Wheeler & Roediger, 1992). Faces were randomly assigned into a discrimination or a control practice condition (8-8 faces, respectively). Each practice cycle included a discrimination practice and a control block in a randomly determined order within each cycle. In the control blocks participants were presented with

each of the 8 faces in random order for 5000 ms (PSI = 1000 ms). Before each control block, participants were instructed to memorize the faces, and they could proceed to the task by pressing the Space bar.

The discrimination practice blocks consisted of a two-alternative forced-choice recognition task where in each trial participants saw a pair of faces (the practiced face and a new face that was not encountered at encoding) presented next to each other. Before each discrimination practice block participants were instructed to decide which of the two faces they had seen during the encoding task using the keyboard (C = "The left face is old", M = "The right face is old"). The face pairs were presented in a random order and remained on the screen for 5000 ms (PSI = 1000 ms). The stimuli remained on the screen even after an answer was presented, to avoid a discrepancy of the presentation time between the items within and between the conditions. There was no delay between the practice cycles.

Following a two-day retention period, participants' memory for all faces was tested in a recognition memory task, while participants were presented with 100 faces in the middle of the computer screen. Sixteen images were targets, i.e., exact repetitions of the faces learned during the first session of the experiment. There were 16 of each type of morphed lures (25%, 50%, and 75% morphed). Each face had its original version as well as each type of its morphed versions presented in the recognition task. Sixteen images were morphed foils used in the morphing process and 16 images were unmorphed foils. The first four images were fillers. The faces were presented in a random order (with the exception of the fillers) and each face remained on the screen for 5000 ms (PSI = 1000 ms). Similar to Experiment 2, participants made an "old"/"similar"/"new" decision for each face using the keyboard. The response labels (M = "Old", V = "Similar", and Y = "New") remained on the screen for the duration of the recognition task.

(Figure 4 about here)

Data analysis

We investigated the effect of practice strategy on long-term recognition memory performance by conducting two-tailed paired-sample t-tests to compare the recognition scores (rate of "old" responses to targets minus the rate of "old" responses to unmorphed foils) on the final recognition task between

the discrimination practice and control conditions. Additionally, we analyzed the effect of practice strategies on the lure discrimination scores (rate of "similar" responses to lures minus the rate of "similar" responses to unmorphed foils) and false recognition scores (rate of "old" responses to lures minus the rate of "old" responses to unmorphed foils) of different similarity levels of morphed lures on the final recognition task. We conducted three 2 x 3 repeated measures ANOVAs followed by contrast analyses for the lure discrimination and false recognition scores of lure items.

Results

In case of the recognition score of the target items we found no difference between the faces practiced in the discrimination task ($M = 86.4\%$, $SE = 2.3$) and control target faces ($M = 84.0\%$, $SE = 2.6$), $t(40) = 0.840$, $p = .406$, $d = 0.155$.

Regarding lure discrimination performance for morphed stimuli (see Figure 5a), we found a significant interaction between Practice strategy and Similarity, $F(1.97, 78.82) = 5.175$, $p = .008$, $\eta^2_p = 0.115$. Contrast analysis comparing practice strategies on all similarity levels (25%, 50%, 75% morphs) using Bonferroni correction showed that while in case of strong (25%) morphs there was no difference between the practice strategies, $F(1, 40) = 0.276$, $p = 1.000$, $\eta^2_p = 0.007$), discrimination practice resulted in a higher lure discrimination score for both weaker morphs (50% morphs: $F(1,40) = 6.816$, $p = .038$, $\eta^2_p = 0.146$; 75% morphs: $F(1,40) = 18.904$, $p < .001$, $\eta^2_p = 0.321$).

False recognition scores of morphed lures (see Figure 5b) did not show a significant Practice strategy x Similarity interaction, $F(1.94, 77.76) = 1.143$, $p = .323$, $\eta^2_p = 0.028$, however, the significant main effect of Practice strategy, $F(1, 40) = 5.296$, $p < .027$, $\eta^2_p = 0.117$, indicated a general pattern of higher false recognition score for items practiced in the discrimination task. Although, it is difficult to draw conclusions on the effect of practice on the 75% morph due to the extremely low false recognition rate for this stimulus type (discrimination practice: $M = 3.8\%$, $SE = 1.2$, control practice: $M = 2.9\%$, $SE = 1.4$). There was a significant main effect of Similarity as well, $F(1.96, 78.21) = 548.005$, $p < .001$, $\eta^2_p = 0.932$.

To summarize, testing the material by a two-alternative forced-choice recognition task affected long-term discrimination of morphed lure faces that were similar to the practiced items. While correct

recognition of the target items was not affected by the practice strategy, discrimination practice resulted in a facilitated discrimination performance for morphed lures that were less closely similar to the target faces (50% and 75% morphs). Discrimination practice also seemed to have a general effect of increasing false recognition for lures.

(Figure 5 about here)

Discussion

Experiment 1 & 2: A New Tool to Assess Mnemonic Discrimination

Mnemonic discrimination between stimuli that share similar perceptual features is suggested to be the behavioral consequence of hippocampal pattern separation (Kirwan & Stark, 2007; Yassa & Stark, 2011). The aim of the first two experiments was to develop a paradigm that pertains with facial stimuli and is sensitive to mnemonic discrimination performance. For this purpose, we used photographs of faces as stimuli with lure stimuli created by morphing unique faces into each other.

In Experiment 1, we compared the discriminations between different stimulus types (targets vs. lures as well as targets vs. foils) by analyzing recognition confidence ratings. According to our results, the task is sensitive to lure discrimination performance. More specifically, there was a gradual reduction in discrimination accuracy with the increment of perceptual similarity between targets and lures. Based on these findings, we can conclude that this test type (when participants are required to make recognition confidence decisions) is sensitive to relatively small changes in overlap between the perceptual features of the presented images (a target face and its lure differing by 25% vs. 50% vs. 75%). Discrimination accuracy was the highest for target-foil discrimination, especially when the foil face did not share features with any of the target stimuli.

Two other methodological aspects of the experimental design should be discussed in details. First, we analyzed lure discrimination performance using the *AUC* (Area Under the Curve) value of a ROC analysis as a measure of accuracy. Together with other authors, we suggest to use this index as a reliable measure of discrimination accuracy, because its value is not affected by response bias and does

23

not depend on model-specific assumptions (Green & Moses, 1966; Macmillan & Creelman, 2005; Pollack & Hsieh, 1969; Pollack & Norman, 1964; Verde et al., 2006). Second, half of the participants were instructed to memorize the stimulus set, whereas the remaining participants encountered with the encoding phase in an incidental learning situation. In our experiment, lure discrimination performance and the discrimination between targets and foils were not affected by the instruction presented before encoding. This result is in line with previous findings showing no difference between incidental and intentional encoding when the incidental task requires deeper processing in the form of a semantic judgement of the stimuli, or the stimuli are relevant to the encoding task (e.g., Bird, 1976; Ferrara, Puff, Gioia & Richards, 1978; Varakin, Frye, & Mayfield, 2012), since in our experiment participants made pleasantness judgements in the incidental encoding condition.

In Experiment 2, participants were not required to give confidence ratings. Instead, they performed a recognition memory test with "old", "similar", and "new" as response options. Lure discrimination performance was measured by the ratio of "similar" responses given to the lures (Lure Discrimination Index, LDI, as it is often called in the neuroimaging literature; see Stark et al., 2019). According to our findings, this test type was sensitive to lure discrimination performance as well. When there was a relatively high overlap between the features of targets and lures, the LDI was lower than it was for weaker morphs. However, there was no difference in lure discrimination performance between weaker morphs (when there was a 50% or a 75% difference in the features of the target face and its lure). These latter findings together with the results of the first experiment indicate that lure discrimination accuracy calculated on the basis of confidence ratings was sensitive to smaller changes in perceptual similarity as compared to the LDI.

A similar threshold in lure discrimination performance was observed in a study investigating a face recognition specific phenomenon, the other race effect, in relation to pattern separation performance (Chang et al., 2015). The authors used morphed lure faces that were dissimilar from the original stimuli in 30%, 40%, 50% and 60% in a recognition task with "old"/"new" decisions (without confidence judgement) to measure the discrimination performance for faces of the same or different ethnicities from the participants'. They found that the correct rejection of morphed lures increased with the lures' dissimilarity to the original face. In case where the faces were the same ethnicity as the participant, the

lure discrimination performance was more tuned, showing a sharp increase in correct rejection between 40% and 50% morphed faces, while lure faces of other races showed a more gradual increase in discrimination performance.

It should be highlighted that there are crucial differences between the present experiments and the study of Chang et al. Chang and colleagues used synthetic faces as stimuli, the encoding was incidental, there were only two response options on the recognition test ("old" and "new"), and lure discrimination performance was analyzed on the basis of the rejection of the lure faces. In contrast, stimuli were photographs of real faces in our experiment for the purpose to create a more realistic experimental situation, we assessed memory performance following incidental and intentional encoding to contrast the possible effects of encoding types, and there was an additional "similar" response option in the test phase of our experiment. Furthermore, Chang and colleagues used lures that were 30%, 40%, 50%, and 60% dissimilar to the original faces, whereas we used a wider continuum ranged between 25% and 75% morphs. In the recognition test phase Chang and colleagues presented either the target face or one version of its morphs, while we presented all versions of the same item in the recognition task order to decease the use of familiarity as a strategy to make lure discrimination decisions. Finally, while this previous study used no delay between encoding and retrieval, we applied a shorter delay of five minutes in Experiment 1 and 2, and we used a longer retention interval of two days in Experiment 3 to examine the long-term effect of discrimination practice on memory retention.

In addition to lure discrimination performance, we analyzed the false recognition of lures as well. Small changes in similarity affected the false recognition of lure faces in Experiment 2, as measured by the ratio of "old" responses given to the lures. Specifically, there was a gradual increase in false alarm rates for the lures with the increment of perceptual similarity between targets and lures (25% vs. 50% vs. 75% difference). In fact, "old" response rate for the lure items is often used as a measure of pattern completion in the neuroimaging literature (see Stark et al., 2019). Pattern completion refers to the process when memories turn into accessible in response to partial/degraded cues (Hunsaker & Kesner, 2013; Marr, 1971; Rolls, 2013). This mechanism is associated with generalization that allows the extraction of regularities from discrete but similar experiences (Keresztes, Ngo, Lindenberger, Werkle-Bergner, & Newcombe, 2018; Shomamy & Wagner, 2008). As it was described in detail

previously, memory retrieval plays a key role in pattern separation, since one must retrieve a previously perceived stimulus when a lure item is presented to detect differences between them (Clark & Gronlund, 1996; Norman & O'Reilly, 2003; Rotello & Heit, 1999; Rotello et al., 2000). Accordingly, pattern separation and pattern completion are suggested to intimately interact with each other and it has been demonstrated that both are associated with specific subregions in the hippocampus. While the dentate gyrus and the CA3 subregion of the hippocampus are involved in pattern separation, pattern completion mostly depends on the intact functioning of the CA1 area (see Yassa & Stark, 2011). In sum, we can conclude, that the paradigm developed and used in Experiment 2 is sensitive to the behavioral correlates of both pattern separation and pattern completion. Crucially, this test type with three response options was more sensitive to the false recognition of lure items as compared to lure discrimination performance as measured by the ratio of "similar" responses given to the lures.

Our results regarding false recognition show a similar pattern observed in the false memory studies where perceptual relatedness was systematically manipulated using morphed faces (Jeye, McCarthy, & Slotnick, 2020; Turney & Dennis, 2017). Turney and Dennis (2017) found an increase of false alarms with higher levels of morph similarity, concluding that false memories at least partially result from the amount of perceptual overlap between the original items and lures. This pattern of false recognition is consistent with the Fuzzy Trace Theory (Reyna & Brainerd, 1995). In case of high perceptual similarity between lure and target, lures are more likely to serve as retrieval cues for episodic detail-rich verbatim traces, while in case of low perceptual similarity, lures are better retrieval cues for gist traces. Successful lure discrimination would require a decision based on verbatim trace, but even in case of a verbatim response, a higher overlap of perceptual details between target and lure is more likely to result in false recognition. Jeye and colleagues (2020) reported a similar overall increase of incorrect "old" responses with the overlap between targets and lures, with the exception of the most similar 20% morph that showed a decrease in false alarm. Maybe this difference from our results could be the ascribed to the differences in similarity levels, or some unique property of the computer-generated faces used in the study.

Finally, it should be highlighted that the encoding type (incidental/intentional) did not affect performance in Experiment 2. This pattern of findings is similar to the results of Experiment 1.

Specifically, the encoding instruction did not affect the ability to recognize target faces (for similar results, see e.g., Bower & Karlin, 1974), lure discrimination performance, and there was no impact of encoding type on the false recognition of lures either. The rationale for using two encoding instruction types is that in experimental psychology, recognition memory is often tested following intentional learning, while tasks assessing mnemonic discrimination are usually used with incidental encoding instructions (e.g., Lacy et al., 2011; Segal, Stark, Kattan, Stark, & Yassa, 2012; Stark et al., 2013). Based on the results of our first two experiments, we conclude that our task can be used to investigate the behavioral consequences of pattern separation and completion with different encoding instructions (incidental and intentional) and with different test types (confidence ratings and recognition task with "old"/"similar"/"new" options).

In sum, we developed a new paradigm to assess the behavioral manifestation of two basic computational mechanisms of the hippocampus. Mnemonic similarity is often assessed by modified recognition memory tasks using photographs of everyday objects as stimuli (Stark et al., 2013; Stark et al., 2015; Stark & Stark, 2017). One advantage of this task type is that a variety of features is used for mnemonic discrimination (color, size, orientation, etc.). Consequently, the task becomes robust to individual strategies and practice effects (see Stark et al., 2019). On the other hand, using images of various objects as stimuli also has disadvantages. For example, in this version, target and lure items are semantically related, while the foil stimuli typically belong to other semantic categories. Consequently, targets and foils can be discriminated on a semantic basis, whereas targets and lures can only be distinguished on a perceptual basis. In contrast, we decided to use facial stimuli due to multiple reasons. First, it is difficult to give verbal labels for unfamiliar faces (Diamond & Carey, 1986; Ellis et al., 1980; Lindsay et al., 1994), therefore, it can reduce the effect of inter-individual differences in verbalization strategies. Relatedly, the discrimination of unfamiliar faces cannot be feasible on a semantic basis. Finally, faces have consistent primary features since the positions of eyes, nose, and mouth are constant. Thus it is relatively easy to construct a stimulus set with the perceptual similarity between targets and lures held under control by morphing faces into each other.

There have been other studies in the neuroimaging literature that used facial stimuli to investigate behavioral pattern separation. For example, Kirwan and Stark (2007) used color photographs

of faces, however, there were several possible differences between lure images, such as lighting, hairstyle, gaze direction, etc. In another study Chang and colleagues (2015) used morphed faces to investigate the effect of a phenomenon specific to face recognition on pattern separation performance. As described before, their findings are in line with the results of our second experiment, further pointing to the conclusion that morphed facial stimuli could be an effective tool to assess pattern separation performance even outside of the context of studies pertaining to face recognition.

Experiment 3: The Effect of Discrimination Practice on Memory Performance

The second aim of the present study was to examine whether pattern separation and its behavioral manifestation can be altered by a recognition memory task that involves discrimination practice. To investigate whether repetition itself affects memory performance, we used a control condition where participants were presented with facial stimuli but were not required to discriminate between them.

Our results showed that following discrimination practice, lure discrimination performance (as measured by the ratio of "similar" responses given to the lure items) was better for weaker morphs (when there was a 50% or a 75% difference in the features of the target face and its lure) as compared to the control condition. Additionally, as a general pattern, discrimination practice increased false recognition scores for the lures (as measured by the rate of "old" responses to lures minus the rate of "old" responses to unmorphed foils), however, the low false recognition rate for 75% morphs makes determining the effect of practice on this lure type difficult. Although we collected only behavioral data, perhaps we can conclude that practice in the form of a recognition memory task that involves discrimination between two stimuli presented simultaneously may alter pattern separation and pattern completion. However, future neuroimaging studies are needed to verify this assumption.

Increased lure discrimination performance indicates that discrimination practice promotes the formation of detailed, unique memory representations. This finding is consistent with the results of previous studies investigated the effect of memory retrieval on subsequent memory performance. These studies showed that retrieval supports interference resolution (Racsmány & Keresztes, 2015; Szpunar et al., 2008), and consequently, the formation of distinct memory representations. Although the beneficial

impact of memory retrieval is not always apparent in case of recognition memory tasks (see the meta-analytic review of Rowland, 2014), here we present results showing that memory performance on a modified recognition task benefits from retrieval practice. Several authors suggest that the crucial factor is the type of the recognition memory task. Specifically, as long as the task requires the retrieval of specific details associated with the encoding situation, the benefits of retrieval practice can be observed in recognition performance (Chan & McDermott, 2007; Read, 1979; Verde, 2004). Our results are in support of this view: even though we found no effect of practice in standard recognition performance for target stimuli, the discrimination performance for lures increased following recognition practice. While successful lure discrimination requires the retrieval of the original item to detect the differences in details between the original and the lure (Clark & Gronlund, 1996; Norman & O'Reilly, 2003; Rotello & Heit, 1999; Rotello et al., 2000), recognition scores of target items in the MST are suggested to rely more heavily on familiarity (Stark et al., 2013). Correspondingly, in the present study lure discrimination proved to be an index of memory performance that was more sensitive to the effects of practice than target hit rate.

In addition to correct lure discrimination, repeated discrimination practice affected the memory performance related to lures by increasing false recognition as well. As it was described previously, lure discrimination serves as a behavioral measure of pattern separation and false alarm for lures serves as an index of a strongly related process, pattern completion (see Stark et al., 2019). Pattern completion plays a key role in re-instating memory representations from incomplete or degraded cues, and in case of explicit retrieval tasks it is suggested to precede pattern separation as a part of a recall-to-reject strategy (Hunsaker & Kesner, 2013). Since recalling the original item is a frequently employed strategy for the discrimination between targets and partially overlapping lures (Kirwan & Stark, 2007; Norman & O'Reilly, 2003), it can be expected that practicing the material with a method that affects pattern separation performance leads to changes in pattern completion as well. Accordingly, as a process that involves retrieval of the original item, pattern completion was facilitated by discrimination practice even when pattern separation, and therefore successful lure discrimination did not follow the reinstatement of the original memory representation, resulting in the false recognition of the lure.

In sum, we found that repeated practice involving a discrimination task increased both long-term pattern separation performance and pattern completion compared to repetition of the material. These findings are in line with the view that retrieval practice only have a beneficial effect on recognition performance if the task requires the retrieval of specific details of the memory representation. Both behavioral pattern separation and pattern completion involve the reinstatement of the original representation, thus they can be simultaneously altered by discrimination practice.

Acknowledgment

Declaration of interest

The authors declare no conflict of interest.

Open Practices Statement

The experiments reported in this article were not formally preregistered.

Data Availability Statement

Datasets related to this article are available at an open source data repository (Open Science Framework,

OSF, https://osf.io/jvme3/).

References

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science, 319*, 1640-1642. https://doi.org/10.1126/science.1152882

Bernstein, L. J., Beig, S., Siegenthaler, A. L., & Grady, C. L. (2002). The effect of encoding strategy on the neural correlates of memory for faces. *Neuropsychologia, 40*, 86-98. https://doi.org/10.1016/S0028-3932(01)00070-7

Bird, J. E. (1976). Effects of intentional and incidental instructions on picture recognition. *Perceptual and Motor Skills, 42*, 555–561. https://doi.org/10.2466/pms.1976.42.2.555

Block, R. A. (2009). Intent to remember briefly presented human faces and other pictorial stimuli enhances recognition memory. *Memory and Cognition, 37*, 667-678. https://doi.org/10.3758/MC.37.5.667

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103*, 751-757. https://doi.org/10.1037/h0037190

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436. https://doi.org/10.1163/156856897X00357

Burgess, A. P., & Gruzelier, J. H. (1997). Localization of word and face recognition memory using topographical EEG. *Psychophysiology, 34*, 7-16. https://doi.org/10.1111/j.1469-8986.1997.tb02410.x

Chan, J. C., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 431-437. https://doi.org/10.1037/0278-7393.33.2.431

Chang, A., Murray, E., & Yassa, M. A. (2015). Mnemonic discrimination of similar face stimuli and a potential mechanism for the "other race" effect. *Behavioral Neuroscience, 129*, 666-672. https://doi.org/10.1037/bne0000090

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review, 3*, 37-60. https://doi.org/10.3758/BF03210740

Choe, K. W., Kardan, O., Kotabe, H. P., Henderson, J. M., & Berman, M. G. (2017). To search or to like: Mapping fixations to differentiate two forms of incidental scene memory. *Journal of Vision*, *17*:8. https://doi.org/10.1167/17.12.8

DeBruine, L. M., & Jones, B. C. (2017). Face research lab London set. Retrieved from http://doi.org/10.6084/m9.figshare.5047666

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. Journal of experimental psychology: general, 115(2), 107. https://doi.org/10.1037/0096-3445.115.2.107

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition, 25*, 129-139. https://doi.org/10.3758/BF03201107

Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2018). GNU Octave version 4.4.1 manual: A high-level interactive language for numerical computations. Retrieved from Network Theory Ltd.

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note AFCRC-TN-58-51). Hearing and Communication Laboratory. Bloomington, IN: Indiana University.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration, 8*, 101-106.

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1075-1095. https://doi.org/10.1037/0278-7393.21.5.1075

Ferrara, R. A., Puff, C. R., Gioia, G. A., & Richards, J. M. (1978). Effects of incidental and intentional learning instructions on the free recall of naturalistic sounds. *Bulletin of the Psychonomic Society*, *11*, 353-355. https://doi.org/10.3758/BF03336852

Franks, B. A., & Hicks, J. L. (2016). The reliability of criterion shifting in recognition memory is task dependent. *Memory & Cognition, 44*, 1215-1227. https://doi.org/10.3758/s13421-016-0633-8

Gilbert, P. E., & Kesner, R. P. (2006). The role of the dorsal CA3 hippocampal subregion in spatial working memory and pattern separation. *Behavioural Brain Research, 169*, 142-149. https://doi.org/10.1016/j.bbr.2006.01.002

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66*, 228-234. https://doi.org/10.1037/h0023645

Guran, C. N. A., Lehmann-Grube, J., & Bunzeck, N. (2019). Retrieval practice improves recollection-based memory over a seven-day period in younger and older adults. *Frontiers in Psychology, 10*, 2997. https://doi.org/10.3389/fpsyg.2019.02997

Hockley, W. E. (2011). Criterion changes: How flexible are recognition decision processes?. In P. A. Higham, & J. P. Leboe (Eds.), *Constructions of remembering and metacognition* (pp. 155-166). London, UK: Palgrave Macmillan.

Holdstock, J. S., Mayes, A. R., Roberts, N., Cezayirli, E., Isaac, C. L., O'reilly, R. C., & Norman, K. A. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans?. *Hippocampus, 12*, 341-351. https://doi.org/10.1002/hipo.10011

Hunsaker, M. R., & Kesner, R. P. (2013). The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neuroscience and Biobehavioral Reviews, 37*, 36-58. https://doi.org/10.1016/j.neubiorev.2012.09.014

Jeye, B. M., McCarthy, C. R., & Slotnick, S. D. (2020). Long-term memory specificity depends on inhibition of related items. *Memory, 28*, 261-269. https://doi.org/10.1080/09658211.2019.1711129

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966-968. https://doi.org/10.1126/science.1152408

Keresztes, A., Ngo, C. T., Lindenberger, U., Werkle-Bergner, M., & Newcombe, N. S. (2018). Hippocampal maturation drives memory from generalization to specificity. *Trends in Cognitive Sciences, 22*, 676-686. https://doi.org/10.1016/j.tics.2018.05.004

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

Kirwan, C. B., & Stark, C. E. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning and Memory, 14*, 625-633. https://doi.org/10.1101/lm.663507

Koen, J. D., Barrett, F. S., Harlow, I. M., & Yonelinas, A. P. (2017). The ROC Toolbox: A toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behavior Research Methods, 49*, 1399-1406. https://doi.org/10.3758/s13428-016-0796-z

Koop, G. J., Criss, A. H., & Malmberg, K. J. (2015). The role of mnemonic processes in pure-target and pure-foil recognition memory. *Psychonomic Bulletin & Review*, *22*, 509-516. https://doi.org/10.3758/s13423-014-0703-5

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning and Memory, 18*, 15-18. https://doi.org/10.1101/lm.1971111

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition – Women's faces make the difference. *Brain and Cognition, 50*, 121-128. https://doi.org/10.1016/S0278-2626(02)00016-7

Lindsay, R. C., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions. *Law and Human Behavior, 18*, 527-541. https://doi.org/10.1007/BF01499172

Lockhart, R. S. (1975). The facilitation of recognition by recall. *Journal of Verbal Learning and Verbal Behavior, 14*, 253-258. https://doi.org/10.1016/S0022-5371(75)80068-5

Loiotile, R. E., & Courtney, S. M. (2015). A signal detection theory analysis of behavioral pattern separation paradigms. *Learning & Memory*, 22(8), 364-369. https://doi.org/10.1101/lm.038141.115

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York, NY: Psychological Press.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society B: Biological Sciences, 262*, 23-81. https://doi.org/10.1098/rstb.1971.0078

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*, 494-513. https://doi.org/10.1080/09541440701326154

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15*, 603-616. https://doi.org/10.1002/acp.728

Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *Quarterly Journal of Experimental Psychology, 62*, 1198-1215. https://doi.org/10.1080/17470210802391599

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, and Computers, 36*, 630-633. https://doi.org/10.3758/BF03206543

Motley, S. E., & Kirwan, C. B. (2012). A parametric investigation of pattern separation processes in the medial temporal lobe. *Journal of Neuroscience, 32*, 13076-13084. https://doi.org/10.1523/JNEUROSCI.5920-11.2012

Noldy, N. E., Stelmack, R. M., & Campbell, K. B. (1990). Event-related potentials and recognition memory for pictures and words: The effects of intentional and incidental learning. *Psychophysiology*, *27*, 417-428. https://doi.org/10.1111/j.1469-8986.1990.tb02337.x

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

Planche, V., Ruet, A., Charré-Morin, J., Deloire, M., Brochet, B., & Tourdias, T. (2017). Pattern separation performance is decreased in patients with early multiple sclerosis. Brain and behavior, 7(8), e00739. https://doi.org/10.1002/brb3.739

Poch, C., Toledano, R., García-Morales, I., Prieto, A., García-Barragán, N., Aledo-Serrano, Á., Gil-Nagel, A., & Campo, P. (2020). Mnemonic discrimination in patients with unilateral mesial temporal lobe epilepsy relates to similarity and number of events stored in memory. *Neurobiology of Learning and Memory, 169*, 107177. https://doi.org/10.1016/j.nlm.2020.107177

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve and of d'e. *Psychological Bulletin, 71*, 161-173. https://doi.org/10.1037/h0026862

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1*, 125-126. https://doi.org/10.3758/BF03342823

Putnam, A. L., Nestojko, J. F., & Roediger, H. L. (2016). Improving student learning: Two strategies to make it stick. In J. C. Horváth, J. M. Lodge, & J. Hattie (Eds.), *From the laboratory to the*

classroom: Translating the science of learning for teachers (pp. 106-133). New York, NY: Routledge.

Racsmány, M., & Keresztes, A. (2015). Initial retrieval shields against retrieval-induced forgetting. *Frontiers in Psychology, 6*, 657. https://doi.org/10.3389/fpsyg.2015.00657

Racsmány, M., Szőllősi, Á., & Bencze, D. (2018). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 157-166. http://dx.doi.org/10.1037/xlm0000423

Read, J. D. (1979). Rehearsal and recognition of human faces. *American Journal of Psychology, 92*, 71-85. https://doi.org/10.2307/1421480

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. Learning and individual Differences, 7(1), 1-75. https://doi.org/10.1016/1041-6080(95)90031-4

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rolls, E. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience, 7*, 74. https://doi.org/10.3389/fnsys.2013.00074

Rotello, C. M., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject?. *Journal of Memory and Language, 40*, 432-453. https://doi.org/10.1006/jmla.1998.2623

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43*, 67-88. https://doi.org/10.1006/jmla.1999.2701

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*, 1432-1463. http://doi.org/10.1037/a0037559
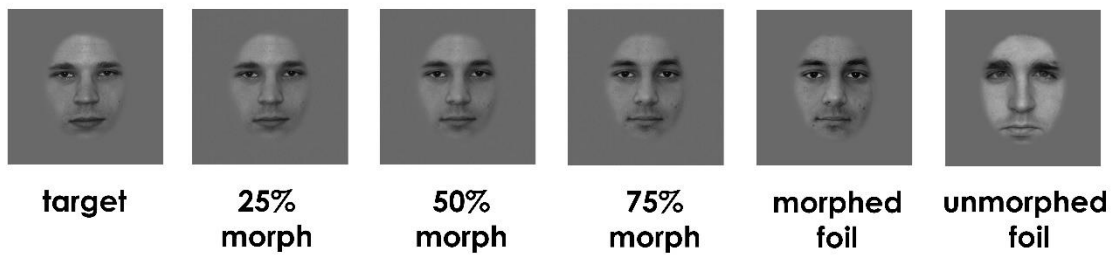
Segal, S. K., Stark, S. M., Kattan, D., Stark, C. E., & Yassa, M. A. (2012). Norepinephrine-mediated emotional arousal facilitates subsequent pattern separation. *Neurobiology of Learning and Memory, 97*, 465-469. https://doi.org/10.1016/j.nlm.2012.03.010

Shaffer, R. A., & McDermott, K. B. (2020). A role for familiarity in supporting the testing effect over time. *Neuropsychologia, 138*, 107298. https://doi.org/10.1016/j.neuropsychologia.2019.107298

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron, 60*, 378-389. https://doi.org/10.1016/j.neuron.2008.09.023

Stark, S. M., Kirwan, C. B., & Stark, C. E. (2019). Mnemonic Similarity Task: A tool for assessing hippocampal integrity. *Trends in Cognitive Sciences, 23*, 938-951. https://doi.org/10.1016/j.tics.2019.08.003

Stark, S. M., & Stark, C. E. (2017). Age-related deficits in the mnemonic similarity task for objects and scenes. *Behavioural Brain Research, 333*, 109-117. https://doi.org/10.1016/j.bbr.2017.06.049

Stark, S. M., Stevenson, R., Wu, C., Rutledge, S., & Stark, C. E. (2015). Stability of age-related deficits in the mnemonic similarity task across task variations. *Behavioral Neuroscience, 129*, 257-268. https://doi.org/10.1037/bne0000055

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia, 51*, 2442-2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1397-1410. https://doi.org/10.1037/0278-7393.24.6.1397

Szőllősi, Á., Bencze, D., & Racsmány, M. (2020). Behavioural pattern separation is strongly associated with familiarity-based decisions. *Memory, 28*, 337-347. https://doi.org/10.1080/09658211.2020.1714055

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1392-1399. https://doi.org/10.1037/a0013082

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210-221. https://doi.org/10.1037/0278-7393.4.3.210

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson (Eds.), *Organization of memory* (pp. 382-402). New York, NY: Academic Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1-25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Turney, I. C., & Dennis, N. A. (2017). Elucidating the neural correlates of related false memories using a systematic measure of perceptual relatedness. *Neuroimage, 146*, 940-950. https://doi.org/10.1016/j.neuroimage.2016.09.005

Varakin, D. A., Frye, K. M., & Mayfield, B. (2012). Intentional memory instructions do not improve visual memory. *International Journal of Brain and Cognitive Sciences, 1*, 18-25. https://doi.org/10.5923/j.ijbcs.20120103.02

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory and Cognition, 32*, 1265-1272. https://doi.org/10.3758/BF03206317

Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of d′, A_z, and A'. *Perception and Psychophysics, 68*, 643-654. https://doi.org/10.3758/BF03208765

Verkoeijen, P. P., Tabbers, H. K., & Verhage, M. L. (2011). Comparing the effects of testing and restudying on recollection in recognition memory. *Experimental Psychology, 58*, 490-498. https://doi.org/10.1027/1618-3169/a000117

Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science, 3*, 150670. https://doi.org/10.1098/rsos.150670

Wesnes, K. A., & Burn, D. J. (2013). Compromised object pattern separation performance in Parkinson's disease suggests dentate gyrus neurogenesis may be compromised in the condition. *Journal of Alzheimer's Disease and Parkinsonism, 3*, 1000131. https://doi.org/10.4172/2161-0460.1000131

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*, 571-580. https://doi.org/10.1080/09658210244000414

Wheeler, M. A., & Roediger III, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*, 240-246. https://doi.org/10.1111/j.1467-9280.1992.tb00036.x

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus, 21*, 968-979. https://doi.org/10.1002/hipo.20808

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34*, 515-525. https://doi.org/10.1016/j.tins.2011.06.006

Figure 1

A

target   25% morph   50% morph   75% morph   morphed foil   unmorphed foil

B

5 min

Encoding 3x

Recognition task

Intentional encoding

Memorize!

Pleasant or Unpleasant?

Incidental encoding

7000 ms
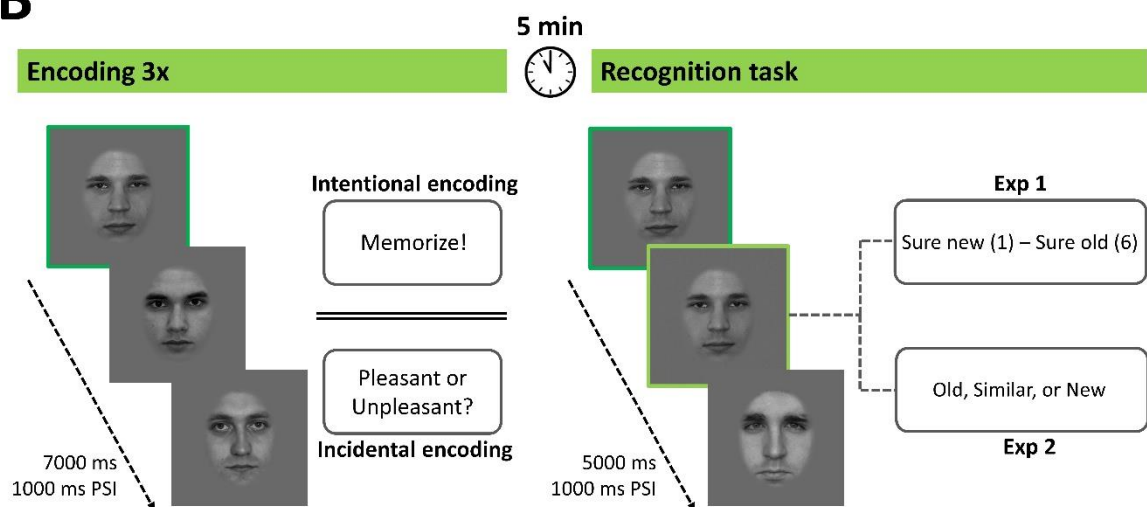1000 ms PSI

5000 ms
1000 ms PSI

Exp 1

Sure new (1) – Sure old (6)
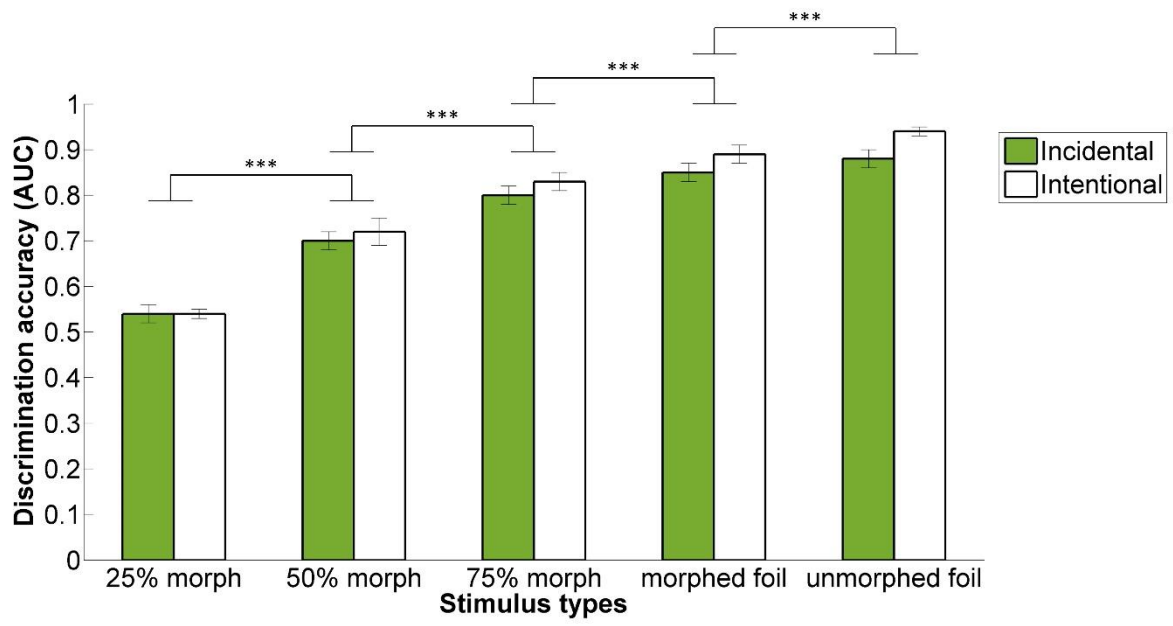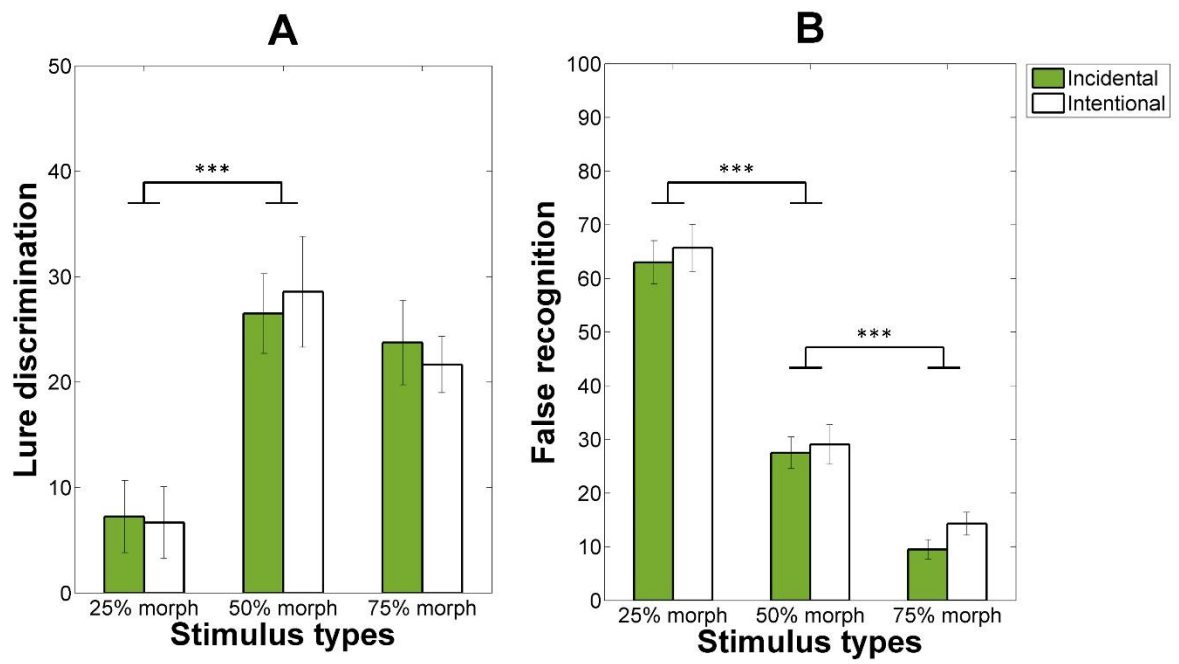
Old, Similar, or New
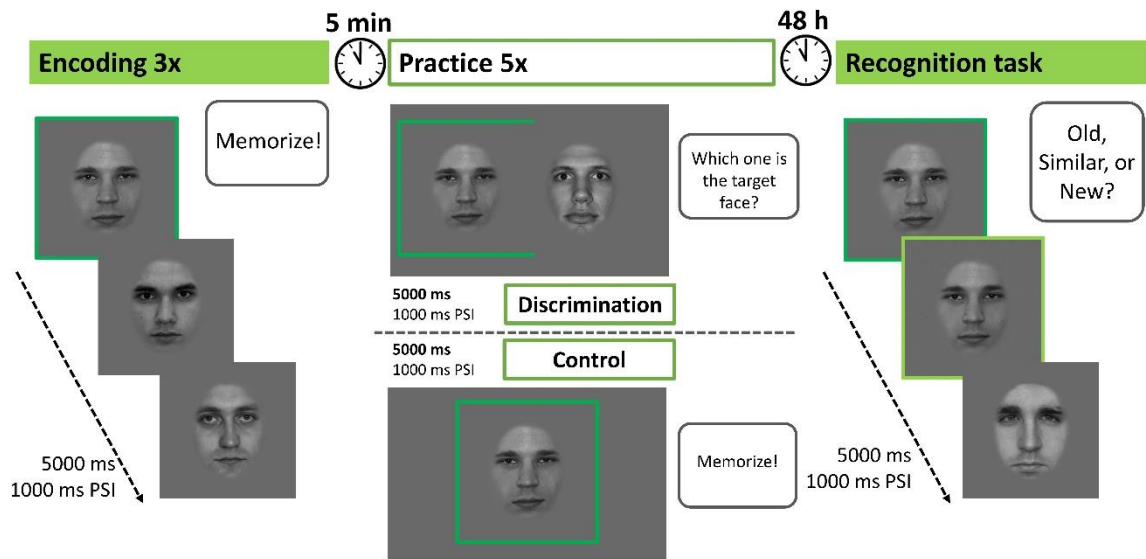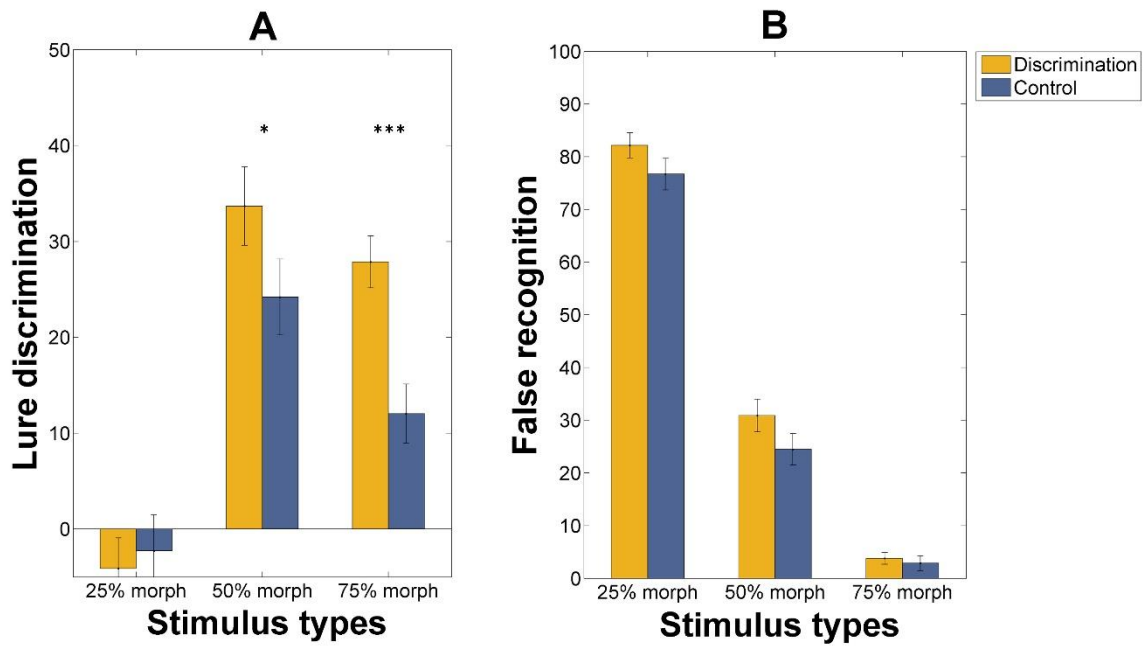
Exp 2

Figure 2

Figure 3

Figure 4

Figure 5

List of figure captions and legends

Figure 1. Stimulus exemplars ('A') and the procedures of Experiment 1 and Experiment 2 ('B').

Note(s). ('A') Targets were exact repetitions of faces presented at encoding; lures were morphed faces that were visually similar to the target faces; morphed foils were new faces that were used during the morphing process but were not presented at encoding; unmorphed foils were new faces that were not used during the morphing process and were not presented at encoding. ('B') Participants were presented with targets, lures, and foils in the recognition test phase of the memory task, and were asked either to make recognition confidence judgments (Experiment 1) or to give "old"/"similar"/"new" decisions (Experiment 2).

PSI = pre-stimulus interval.

Figure 2. Discrimination between targets versus lures (25%, 50%, and 75% morphed faces) as well as targets versus foils (morphed and unmorphed faces) in the two encoding conditions (incidental and intentional), separately (Experiment 1).

Note(s). AUC = Area Under the Curve; error bars represent the standard error of the mean; *** $p < .001$.

Figure 3. Lure discrimination performance and false recognition in the two encoding conditions (incidental and intentional), separately (Experiment 2).

Note(s). ('A') Lure discrimination performance refers to the percentage of "similar" responses to lure (morphed) faces minus the percentage of "similar" responses to unmorphed foils. ('B') False recognition refers to the percentage of "old" responses to lure (morphed) faces minus the percentage of "old" responses to unmorphed foils.

Error bars represent the standard error of the mean. *** $p < .001$.

Figure 4. The experimental design and the procedure of Experiment 3.

Note(s). In Experiment 3 participants were presented with images of faces and were instructed to memorize them. Following this encoding stage, subjects participated in a practice phase where they either performed a forced-choice recognition memory task (discrimination condition) or were asked to memorize the faces (control condition). The final recognition memory test was performed in a different session following a 48-hour delay where participants made "old"/"similar"/"new" decisions. *PSI* = pre-stimulus interval.

Figure 5. Lure discrimination performance and false recognition in the discrimination practice and control conditions (Experiment 3).

Note(s). ('A') Lure discrimination performance refers to the percentage of "similar" responses to lure (morphed) faces minus the percentage of "similar" responses to unmorphed foils. Discrimination practice resulted in a higher rate of correct "similar" responses for weaker morphs (50% and 75% morphs). ('B') False recognition refers to the percentage of "old" responses to lure (morphed) faces minus the percentage of "old" responses to unmorphed foils. Practice increased false recognition for lures regardless of their similarity to the target stimuli.

Error bars represent the standard error of the mean. $* p < .05$; $*** p < .001$.