Characterising the effect of semantic and perceptual similarity in episodic memory

Loris Naspi

Doctor of Philosophy

The University of Edinburgh

2021

Declaration

a) This thesis was composed by the student.

b) The work comprising this thesis is the student’s own.

c) The work has not been submitted for any other degree or professional qualification.

d) The included publications are the student’s own work

Loris Naspi

August 2021

I dedicate this thesis to my mum, who believed in me more than anyone else

Acknowledgments

I first would like to thank my supervisor Dr Alexa Morcom for her guidance throughout my PhD, for showing me how to conduct good science, but also for her endless patience. I have really enjoyed theoretical debates on the functioning of episodic memory and I am grateful to her for allowing me to pursue my ideas. Special thanks go to Dr Paul Hoffman, which was always supportive when I needed help.

I would also like to thank Dr Barry Devereux, Dr Alex Doumas, and Tobias Thejll-Madsen for their assistance with the publication of two manuscripts that are part of the current thesis, as well as Arianna Moccia, my shoulder to lean on when complaining about the world.

I am grateful to my fellow PhD students in the Department of Psychology for making the PhD experience more enjoyable, before, during, and after the COVID-19 pandemic. A warm thank goes to Nikos, who spent part of his time helping me with more complicated aspects of programming and was companion of adventure outside the working environment; Riccardo, which was a good friend since the very first day of my doctorate, and all the other PhD students that I had the pleasure to meet during these 4 years. These include: Andreea, Esperanza, Berengere, Federica, Emma, Carolane, Shereen, and Guillermo.

I would also like to thank:
My girlfriend, Valeria, for standing me with unconditional love.
My whole family, for supporting me throughout my studies.
All my friends outside the academia that made these 4 years special: Jessica, Lucas, Giuseppe, Andrea, Davide, and many others.

A warm thought goes to:
Nonna Anna, Nonno Alfredo, Nonno Ivano, and Zio Romano, who would have been proud of this achievement.

Ciao Zio. Con affetto, “Stecca”

Contents

Lay Summary

Human memory is very efficient. People can generally remember many events with great accuracy and they can also discriminate them from similar episodes to avoid memory errors. False recognition is a specific memory error in which novel information is erroneously remembered as previously been encountered. Although false recognition can occur in healthy young adults, it becomes substantially more frequent in older age. These errors have been attributed to similarity of meaning or appearance and aging is associated with an increased in reliance on semantic processing, thought to be responsible for false recognition due to meaning overlap between similar events. However, perceptual similarity may also have a role. Since items that are similar in meaning also look similar (e.g., “my rented blue mini cooper” versus “another light blue mini countryman parked next to mine”), it is difficult to study their individual contributions to memory. Understanding what causes false recognition is essential for future research into strategies or interventions to prevent these errors, particularly in the context of age-related decline. A fuller understanding of age-related decline in memory performance can also inform basic models of memory. My thesis aimed to disentangle the role of semantic and perceptual similarity in false recognition using two behavioural studies with young and older adults. Here, I used objective measures of semantic and perceptual similarity derived from cognitive models, together with subjective ratings, to assess the influence of these types of similarity on true and false recognition and their age-related differences. This thesis also investigated the neural correlates that are engaged when people initially store memories. It is thought that whether an event’s representation is discriminated from similar memory representations depends critically on how it is encoded. However, the precise encoding mechanisms involved are not yet clear. Thus, I conducted a functional magnetic resonance brain imaging (fMRI) study to reveal whether specific pattern of semantic and perceptual information at encoding could inform later true and false recognition of common objects.

Abstract

This thesis investigated the contribution of semantic and perceptual similarity to mnemonic discrimination, defined as the ability to recognize previously encountered events and to discriminate them from similar ones to avoid false recognition. When mnemonic discriminations fails, false recognition of similar events is typically attributed to shared memory representations (gist) typically characterized in terms of meaning. In two experiments with young adults, I investigated the effects of multiple semantic and perceptual relations in a mnemonic discrimination task involving pictures of objects. I hypothesised that semantic meaning could be processed at least at two different levels: at the concept level, semantic similarity captures features shared by different concepts. I derived this continuous measure from the Conceptual Structure Account, a feature-based model of semantic memory. At the item level, rated semantic similarity between a depicted exemplar and people's internal representation of that concept captures the degree to which an item activates its individual representation. For each item, I also measured indexes of visual form and colour similarity. In all the experiments of the current thesis, the study phase was followed by a recognition memory test including studied items, similar lures, and novel items. In Experiment 1 and 2 of Chapter 2 participants studied single and multiple exemplars for each basic-level concept, respectively. Participants were less likely to recognize studied items with high concept confusability, but also to falsely recognize these lures. Thus, greater emphasis on coarse processing of shared features relative to fine-grained processing of individual concepts weakened the basic-level semantic gist shared between studied items and lures. In contrast, false recognition was higher for those lures judged more similar to people's internal representation of that concept, suggesting reactivation of the studied concept at test by a different exemplar. False recognition was also more frequent for more visually confusable lures. These results suggest that multiple levels of semantic and perceptual similarity contribute to mnemonic discrimination. In Chapter 3, to understand whether initial processing of semantic and perceptual information at encoding is responsible for later true and false recognition, I ran an fMRI study that combined representational similarity analysis (RSA) with the subsequent memory paradigm. To index semantic and perceptual processing that I found relevant in Chapter 2, I created two semantic models to index coarse-grained taxonomic categories and specific object features, respectively, and two perceptual models defining visual properties, like line orientation and colour attributes. Participants encoded images of objects during fMRI scanning. Both perceptual and semantic models predicted later true memory. The strength of the neural patterns corresponding to low-level visual

representations in the early visual and inferotemporal cortex was stronger for items successfully recognized versus forgotten. Similarly, greater alignment of neural patterns with object-specific semantic representations in the inferotemporal cortex also predicted true recognition. However, the strength of the neural patterns reflecting nonspecific taxonomic information was stronger for items later forgotten than remembered in ventral anterior temporal lobe, left inferior frontal gyrus and, preliminary, in the left perirhinal cortex. In contrast, inefficient visual processing in posterior regions was associated with false recognition of similar lures. The results suggest that fine-grained semantic as well as visual analysis contributes to accurate later recognition, while processing visual image detail is critical for avoiding false recognition. To conclude, in Chapter 4 I investigated whether the same semantic and perceptual variables that I found to be significant predictors of young adults' ability to recognize previously studied items, as well as misrecognize similar lures, have a greater impact on older adults' performance on the same task. Thus, I focused on age-related differences. Contrary to our predictions, Experiment 1 did not show any age-related differences in true and false recognition, or semantic and perceptual modulations on recognition memory. However, Experiment 2 revealed an increase in true and false recognition for more confusable concepts in the older relative to young group when multiple exemplars of each basic-level concept were presented at study. I suggested this effect to be due to a reduction of processing semantic relations across concepts. This result is consistent with gist-based theories of aging that assume that semantic gist between studied items and lures can promote true and false recognition. Altogether, the studies described in the current thesis provide evidence that multiple semantic and perceptual relations can contribute to true and false recognition of objects and that these effects are in part mediated by operation that occur during the encoding phase.

Chapter 1: General Introduction

In everyday life it is important to remember previously experienced events but also to discriminate them from highly similar ones to avoid mnemonic discrimination errors. Imagine how many times we rent a car for a new trip. In each occasion, we are encoding a new representation of that specific car. This means that we need to encode a novel representation of its type, shape, and colour in order to remember it. How can we later accurately recognise our car when it is parked in a car park among other semantically and perceptually similar cars? Recognising my car is an example of mnemonic discrimination, which also requires that I can discriminate similar ones as novel. To avoid falsely recognising the novel but similar car, it is important for the representations to be distinguished in memory. Whether discrimination is successful is thought to depend in part on encoding processes, and the nature of the information encoded. According to fuzzy-trace theory, emphasis on semantic *gist* at encoding increases the probability of misrecognizing items overlapping in gist with studied items. Gist is defined as shared semantic representations between sets of events embodied in a memory trace. However, other theories have also emphasised the role of further non-semantic dimensions, including perceptual similarity. Compared with young adults, older adults show reduced ability to discriminate similar events in memory, and this is thought to be associated with age-related changes in encoding processes. Fuzzy trace theory differs with other theories in terms of encoding mechanisms and the type of similarity proposed to drive mnemonic discrimination failures and their age-related increase. In this thesis I sought to establish the type of similarity contributing to false recognition in young and older adults, and to clarify the contribution of encoding-related processes to mnemonic discrimination in young adults. The beginning of this chapter outlines prominent accounts of mnemonic discrimination in healthy young adults, and highlights outstanding questions. Emphasis is placed on fuzzy trace theory, which offer mechanistic explanations of encoding contributions to mnemonic discrimination errors. Then, age-related mnemonic discrimination decline is discussed in relation to the fuzzy trace theory and other relevant accounts. At the end, the thesis aims are outlined.

1.1 Mnemonic discrimination

As many of the events we experience every day are highly similar, substantial overlap between memory traces inevitably occurs (McClelland et al., 1995). *Mnemonic discrimination* is the ability to recognize and differentiate previously encountered events from similar incoming information, and it is an essential component of episodic memory (Klippenstein et al., 2020). Although mnemonic discrimination refers to the ability to differentiate between internal neural representations, the behavioural outcome of this process is the differentiation between external stimuli (Leal & Yassa, 2014; Reagh et al., 2016). In the laboratory, during the study phase, participants are presented with a list of items that usually consists of single pictures or words. Then, in the subsequent test phase, mnemonic discrimination is typically probed by presenting a subset of studied items, semantically and/or perceptually similar lures, and unrelated novel items. Incorrectly identifying lures as “old” produces *mnemonic discrimination errors* or *false recognition* (Stark et al., 2010). Although Young adults can successfully discriminate most studied items and similar lures, false recognition is rather common when words or pictures are used (Koutstaal & Schacter, 1997; Motley & Kirwan, 2012; Seamon et al., 2000; Stahl et al., 2016). Thus, it is essential to understand the factors responsible for mnemonic discrimination errors.

Studies employing images of common objects have typically probed mnemonic discrimination with lure images that represent different exemplars of the same studied basic-level concepts. Intuitively, these lures are both semantically and perceptually similar to studied items. For instance, after studying multiple images (e.g., cat, stool, sheep, ladybug), Koutstaal and Schacter (1997) presented a subset of studied items (e.g., the same cat), together with similar lures (e.g., a different exemplar of cat), and unrelated novel items (e.g., a panther). The authors found that false recognition was more likely for related lures compared with unrelated novel items. These results which were attributed by the authors to semantic and perceptual similarity between studied items and lures, were subsequently replicated in more recent studies, all of which used pictures of common objects (Gutchess & Schacter, 2012; Koustaal et al., 1999; Koutstaal, 2003, 2006). However, a clear-cut separation between semantic and perceptual similarity is difficult with pictorial material since these dimensions often overlap, which makes it difficult to disentangle the relative contribution of each dimension in false recognition (but see Koutstaal et al., 2003 in Chapter 2 for more details).

Other studies reported false recognition for similar lure exemplars of pre-experimentally unfamiliar studied objects due to perceptual similarity (Budson et al., 2001; Koutstaal et al., 1999; Slotnick & Schacter, 2004). Stimuli were taken from categories defined by perceptual similarity. Categories were defined by a prototype figure, and sets of exemplars were created by applying different distortions to the prototype. Participants falsely recognized lure objects from studied categories more often than novel objects from non-studied categories. False recognition could be interpreted as due to similarity between exemplars of perceptually defined categories. According to this line of argumentation, false recognition could not have been due to semantic similarity because the objects and categories were unfamiliar and meaningless. However, the absence of pre-existing semantic information does not exclude the possible contribution of semantic similarity. Specifically, people may either have acquired new semantic knowledge when studying multiple exemplars from abstract categories or see them as meaningful in light of prior knowledge (Pidgeon & Morcom, 2014). If people acquire semantic knowledge during the study phase, then false recognition may have been based on this abstracted semantic similarity rather than on perceptual similarity (Pidgeon & Morcom, 2014). Nonetheless, fuzzy trace theory supports the view that semantic similarity can play a role in false recognition. I consider this theory below.

1.1.1 Fuzzy Trace theory

Fuzzy trace theory posits dual memory mechanisms that act in concert to support true recognition, but that act in opposition with respect to false recognition. This theory suggests that semantic information shared between events is the key to understanding false recognition. Fuzzy trace theory hypothesised that subjects encode separate verbatim and gist traces of studied items when they initially process information. *Verbatim traces* are integrated representations of a studied item's surface content and other item-specific information, which are unique to an item. *Gist traces* are episodic elaborations of semantic information (Brainerd & Reyna, 2002; Reyna & Brainerd, 1995). Fuzzy trace theory assumes that verbatim and gist traces are stored in parallel. A hallmark of fuzzy trace theory's notion of verbatim traces is integration, i.e. that verbatim traces are integrated representations of multiple surface features, so that the retrieval of such traces on memory tests induces vivid mental reinstatement of experiences that accompanied items' earlier presentations. Turning to gist traces, the initial encoding of items' surface features is assumed to initiate a corresponding mechanism of meaning access and extraction, with these different storage mechanisms running in parallel during encoding.

If subjects store separate verbatim and gist traces of individual items, there will be latitude in the types of representations that may be accessed on memory tests, and therefore, the types of representations that are accessed will depend heavily upon the retrieval cues that are provided; verbatim traces are favoured when the surface content of retrieval cues matches that of the studied items, and gist traces are favoured when retrieval cues match studied items in meaning content but not in surface form. Thus, memory performance will depend on the relative availability and accessibility of verbatim and gist traces, which, while influenced also by consolidation and retrieval factors, rely largely on the strength of respective verbatim and gist representations formed at encoding (Brainerd & Reyna, 2002). As long as verbatim traces are still accessible, studied items are better retrieval cues for verbatim than for gist traces on recognition test; thus, successful recognition of studied items is thought to be based predominantly upon the retrieval of verbatim rather than gist traces (Reyna et al., 2002). Similarly, semantically related lures are better retrieval cues for gist than for verbatim traces; thus, mnemonic discrimination errors are based upon the retrieval of gist traces (Reyna & Kiernan, 1994, 1995).

It is important to note that verbatim and gist retrieval are convergent processes in true recognition because they both support mnemonic discrimination of studied items. Fuzzy trace theory assumes that in true recognition retrieval of verbatim traces induces a vivid form of remembering in which subjects consciously re-experience items' occurrence in the study phase, a process called *identity judgment*. However, a studied item may fail to provoke verbatim retrieval and may instead provoke gist retrieval, which usually induces a more general form of remembering based on meaning overlap, a process called *similarity judgment*. With similar lures, however, verbatim and gist retrieval are assumed to be opponent processes, with gist retrieval supporting false recognition errors and verbatim retrieval suppressing them. Gist retrieval that results in misrecognition is again called *similarity judgment.* Verbatim retrieval, on the other hand, works against false recognition and it takes advantage of the vivid mental states that are provoked by verbatim traces. False-but-gist-consistent items may sometimes produce the retrieval of verbatim traces of previously studied items, which can generate mismatches at the level of verbatim detail (Reyna, 1996). Such verbatim mismatches provide reliable, principled bases for correctly rejecting related lures in recognition because they supply a compelling explanation of how events that were not experienced could nevertheless seem familiar. This ability is referred to as *recollection-rejection* (Brainerd et al., 2003).

Fuzzy trace theory's assumptions outlined above make it possible to explore mnemonic discrimination effects that fall out as predictions from the opponent-processes ideas. The most direct prediction is that, as gist traces are entirely composed of shared semantic information, semantic similarity drives mnemonic discrimination errors (Brainerd & Reyna, 2002; Reyna & Lloyd, 1997). These errors are predicted to be observed only where lures are semantically similar to studied items. According to fuzzy trace theory, this relation should hold consistently for semantic overlap but nor for perceptual overlap (Reyna & Lloyd, 1997). Thus, if semantic similarity drives false recognition, it follows that false recognition levels will be higher for more semantically similar items (see Chapter 3 for details).

1.1.2 Activation-monitoring theory

A further theory, in which dissociated memory mechanisms are hypothesized, is the activation monitoring theory. This framework was developed to explain performance on the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). Here, people study lists of associated words (e.g., bed, rest, awake) and then take a free recall or recognition memory test. The typical finding is that people often misrecognise critical lures (e.g., sleep) that were not presented in the study lists. The activation monitoring theory proposes that, at study, non-presented critical lures are often activated by relational processing induced by strong lexical associations between lures and list items (Roediger et al., 2001). The relative degree of item-specific and relational processing at encoding is thought to determine the strength of lure concept activation (Hunt & Einstein, 1981), and the success of monitoring at test. Greater item-specific encoding is thought to reduce later false recognition by increasing the probability that item-specific information is available at test, while greater relational processing contributes to reduced availability of item-specific information (Gallo & Roediger, 2002). If a lure concept was activated at study, false recognition is proposed to result from retrieval monitoring failure, misattributing the activated concept to the studied list.

Fuzzy trace theory makes similar predictions to the activation monitoring theory that increased verbatim (or item-specific) relative to gist (or relational) processing at encoding reduces false recognition in DRM and other recognition paradigms. Thus, the key difference between the theories is the activation monitoring theory's emphasis is on associative rather than semantic processing. Despite these similarities, one difficulty of fuzzy trace theory in trying to explain false recognition is that gist has never been operationally defined (Roediger et al., 2001); that is, given a list of pictures or words, how does one specify the gist representation of the list and quantify how strong it is? According to the activation monitoring theory, one plausible interpretation of gist is backward associative strength, the average tendency for words in the study list to elicit the critical lure on a free association test (Nelson et al., 2004). The evidence suggests that the more the list items are associated to the critical item, the more a gist representation of the critical lure is created (Deese, 1959; Roediger & McDermott, 1995). However, further studies (Brainerd et al., 2008; Cann et al, 2011) showed that the backward associative strength could be further decomposed into multiple semantic relations (see Chapter 2 for details). These findings suggest, counter to the activation monitoring theory, that semantic processing may play a role in accounting for mnemonic discrimination outcomes beyond associative processing. Importantly, as seen above, false recognition can also occur for perceptually similar abstract pictures that have no pre-existing meaning or associations (Koutstaal et al., 1999), as well as in DRM tasks that used phonologically similar, although associatively unrelated, lures (see Chapter 2 for details). These findings do indicate that semantic gist can cause false recognition in the absence of pre-existing associations. However, fuzzy trace theory cannot explain the results that show that perceptual similarity alone can cause false recognition.

1.1.3 Cognitive neuroscience of mnemonic discrimination

A key assumption of fuzzy trace theory is that despite parallel and independent gist and verbatim encoding, one trace can attract greater encoding resources, resulting in a relatively stronger trace (Brainerd & Reyna, 1990). Later recognition is then more likely to rely on the stronger trace, if intact at retrieval (Brainerd & Reyna, 2002). Therefore, a testable prediction is that the neural processes engaged at encoding will differ according to whether later recognition relies on verbatim or gist memory, or both (Kim & Cabeza, 2007) (see Chapter 4 for more detail). Neuroimaging allows to separate investigation of encoding and retrieval phases. In particular, the development of fMRI techniques has facilitated evaluation of such predictions using the subsequent memory paradigm (Otten et al., 2002; Wagner et al., 1998). Although lots is known about encoding mechanisms and neural networks supporting later true recognition (Kim, 2011), brain regions involved in false memory encoding are relatively unexplored (for a review see Dennis et al., 2015).

In recent years, fMRI subsequent memory studies have revealed differences between regions showing encoding-related activity supporting later true recognition, and regions supporting later false recognition of perceptually and semantically similar images (Abe et al., 2013; Garoff et al., 2005) or semantically similar words (Baym & Gonsalves, 2010; Kim & Cabeza, 2007). Such encoding differences are consistent with fuzzy trace theory’s claims that true recognition is more likely to engage verbatim encoding than false recognition (Brainerd & Reyna, 2002). Kim and Cabeza, (2007) also reported overlap in left ventrolateral prefrontal cortex between encoding predicting true and false recognition, supporting assumptions that both memory outcome can rely at least partially on gist encoding. Garoff et al. (2005) found a similar encoding-related activation in the left fusiform gyrus. In the above studies, many of the left-lateralized regions encoding false memories (e.g., left superior frontal gyrus, left middle temporal gyrus, and left ventrolateral prefrontal cortex) have previously been associated with semantic processing (Binder et al., 2009). If the encoding activity in these regions reflects processing of semantic information, findings linking these regions to later false recognition of semantically similar words/images (Baym & Gonsalves, 2010; Garoff et al., 2005; Kim & Cabeza, 2007) would be largely consistent with fuzzy trace theory’s predictions that semantic processing at encoding contributes to later gist-based memory. Such an interpretation was proposed by Kim and Cabeza (2007) and Garoff et al. (2005). Similarly, activation in posterior regions was interpreted in terms of an association between perceptual processing and later accurate recognition, again consistent with suggestions of fuzzy trace theory that verbatim encoding contributes to accurate memory (Brainerd & Reyna, 2002). However, these interpretations rely on reverse inference (Poldrack, 2006), as it cannot be concluded that activity reflects semantic and/or perceptual processing unless multivariate techniques aim to specifically isolate semantic and perceptual processing (see Davis et al., 2020 and Chapter 4) .

Successful mnemonic discrimination depends on semantic and perceptual processing in regions that are known to support these operations, but also within the medial temporal lobe. The hippocampus is often implicated in forming new associative memories, storing memories independently of each other, retrieving memories from partial cues, and flexibly applying stored memories to novel situations. Marr (1971) was the first to suggest a role of the hippocampus in pattern separation and completion. Pattern separation and pattern completion are dissociable neural computations which are thought to be critical to encoding and retrieval of episodic memory and in particular to mnemonic discrimination (Sahay et al., 2011; Wilson et al., 2006; Yassa & Stark, 2011). Pattern separation is defined as the orthogonalisation of overlapping input representations, serving to reduce overlap between incoming information and existing representations in memory. Pattern completion is the process that retrieves a stored memory trace based on a set of sensory cues as input, increasing representational overlap between incoming information and previously stored representations (McNaughton & Morris, 1987; O'Reilly & McClelland, 1994). Without these complementary processes, people would be incapable of storing separate memories due to high levels of similarity among stimuli and we could not learn new memories, experiencing what has been called as catastrophic interference (McClelland et al., 1995; Norman & O'Reilly, 2003). The complementary actions of pattern separation and pattern completion are thought to contribute to mnemonic discrimination of both studied items and similar lures: efficient pattern separation at encoding is thought to better enable successful mnemonic discrimination of studied items from similar lures, while inefficient pattern separation coupled with dominance of pattern completion at encoding can increase the likelihood that similar overlapping lures would trigger false recognition (Guzowski et al., 2004; Sahay et al., 2011; Yassa et al., 2011). Unlike fuzzy trace theory and the activation monitoring theory, the pattern separation account does not make any distinction between the types of similarity likely to influence false as opposed to true recognition. Consequently, mnemonic discrimination errors can be triggered by a failure in discriminating between events that are similar across multiple dimensions, not just semantically (Yassa & Stark, 2011). Thus, this theory can help reconcile previous results that also highlighted a role of perceptual similarity.

1.1.4 Formal models of semantic and perceptual processing

Fuzzy trace theory put emphasis on gist encoding and retrieval as responsible for false recognition. As I mentioned above, gist is assumed to reflect shared semantic information among a set of events. Formal models of semantic memory can allow us to objectively quantify what has been so far only a theoretical construct. One of these models, the *Conceptual Structure Account* (CSA; Tyler & Moss, 2001, see Chapter 2 for details), assume concepts to be organised in a connectionist system based on a number of elementary properties, or features. Thus, the activation of a concept corresponds to processing the concept’s features (McRae et al., 1997; Tyler et al., 2000; Tyler & Moss, 2001; Vigliocco et al., 2004). Within this framework, the statistical characteristics of features have been proposed to determine how the concepts are processed. One of these, feature sharedness, quantifies the fact that some concepts’ features are highly shared (e.g., <has a tail >, < has eyes>) and are found in many other concepts, while others are highly distinctive (e.g., < has an udder >, < has a mane >) and can only be found in a few concepts. Thus, some concepts are more confusable with others in the semantic space (Moss et al., 1997; Moss et al., 1998; Taylor & Tyler, 2012). In Chapters 2, 3, and 4, feature-based models of semantic processing will allow us not only to define semantic similarity as a continuous measure, disentangled from a perceptual measure, but also to use this metric in neuroimaging to overcome the reverse inference problem identified above. Similarly, we can obtain low-level perceptual attributes of each item using computational models, such as the *Hierarchical Model and X* (HMax; Riesenhuber & Poggio, 1999; Serre et al., 2007), which captures properties processed by the ventral visual stream, a hierarchy of brain areas thought to mediate object recognition in cortex. Once we extracts perceptual properties of single objects, we can calculate a metric of perceptual similarity between objects. These models can help us understand the contribution of semantic and perceptual similarity on true and false recognition of common objects.

1.2 Aging and mnemonic discrimination

Age differences in cognitive functions are well-established. Older adults have difficulty with episodic memory that are manifested by a reduced ability to retrieve studied information (Luo & Craik, 2009). Relative to young adults, older adults generally exhibit deficits in recall, and less so in recognition of studied items (Davis et al., 2003; Kvavilashvili et al., 2009).

Importantly, older adults are more likely to falsely recognise highly similar lures than young adults. It is important therefore to understand the mechanisms underlying these age-related memory differences, in the hope that alternative encoding or retrieval strategies, or cognitive interventions can be targeted at reducing its impact. Moreover, age-related differences can provide a cue for a better understanding of the mechanisms involved in mnemonic discrimination in young adults while testing theories of cognitive ageing.

Many studies on aging showed increased levels of false recognition in older adults (Koutstaal & Schacter, 1997; Toner et al., 2009; Yassa et al., 2011). This deficit appears to be more reliably observed than age differences in true recognition of studied items, suggesting that mnemonic discrimination errors of similar lures are a sensitive marker of age-related memory decline (Stark et al., 2013; Toner et al., 2009; Yeung et al., 2013). A review of studies of age-related differences in mnemonic discrimination (Gomes et al., 2014) found that of 40 experiments examining true and false recognition, 83% reported increased false recognition in older adults, while only 45% showed age-related decline in true recognition (but see Fraundorf et al., 2019). Although most of the recognition studies employed associatively-related words in the DRM paradigm, where older adults typically showed heightened false recognition of critical lures (Balota et al., 1999; Lövdén, 2003; Norman & Schacter, 1997; but see Pansuwan et al., 2020 for different results), age-related increases in mnemonic discrimination errors were also reported for images of common objects (Koustaal et al., 1999). These errors were attributed to an increase in gist-based memory (Brainerd & Reyna, 2002; Koutstaal & Schacter, 1997; Tun et al., 1998).

Despite older adults' false recognition is primarily thought to be driven by semantic similarity, other authors that used both verbal and pictorial material have suggested that perceptual similarity may also contribute to false recognition (Koutstaal et al., 1999). Briefly, relative to young adults, older adults' false recognition errors were higher for phonological similar lures (e.g., cork) that were semantically unrelated to studied items (i.e., corn) (Budson et al., 2003; Ly et al., 2013), and for lures related to studied items only by their shared font (Burnside et al., 2017). Age-related increases in false recognition between studied items and lures were also attributed to perceptual similarity objectively measured in terms of luminance and contrasts (Boutet et al., 2019). These results suggest that older adults, like young adults, may be susceptible to perceptual similarity (but see Chapter 4 for more details). Below, I consider how fuzzy trace theory can be applied to aging and how other theories can integrate these results.

1.2.1 Increased reliance on semantic gist

The literature has suggested that aging produces a shift away from reliance on verbatim information toward reliance on memory for meaning content (Brainerd & Reyna, 2005). According to Brainerd and Reyna (2005)'s hypothesis, the *gist-preference conjecture*, older adults place greater emphasis than young adults on semantic gist information during encoding and retrieval, and are less likely to employ stored verbatim traces in support of memory (Brainerd & Reyina, 2005; Koutstaal & Schacter, 1997, Reyna & Lloyd, 1997). This is thought to lead to impaired mnemonic discrimination between items which overlap in gist and therefore to heightened false recognition (Brainerd & Reyna, 2002; Koutstaal & Schacter, 1999). Thus, although originally it was not primarily an aging theory, fuzzy trace theory can be used to explain some of the age-related errors outlined above. Consistent with this, in the categorized picture paradigms older adults manifest increased false recognition relative to young adults (Koustaal et al., 1999; Koutstaal & Schacter, 1997; Koutstaal et al., 2001). Importantly, however, as I explained earlier for young adults, as images were both semantically and perceptually similar to studied items, the possibility that perceptual similarity has a role cannot be ruled out. There have only been few direct investigations of the specific roles that semantic information play in older adults' memory errors (Koutstaal et al., 2003; Pidgeon & Morcom, 2014). These studies found heightened false recognition in older adults for concrete meaningful objects, leading the authors to propose that older people rely more on semantic processing than young adults. Although the concrete objects were not perceptually matched, these results suggests that the ability to exploit pre-existing semantic information is well maintained in the elderly (Matzen & Benjamin, 2013; Nyberg et al., 2003; Umanath & Marsh, 2014).

1.2.2 Monitoring decline in aging

The activation monitoring theory is also a gist-based account and can explain false recognition. However, it suggests that young adults and older adults activate a similar number of critical lures during encoding of DRM study lists, and as such the activation component is not predicted to show significant age-related changes in false recognition (Dehon & Brédart, 2004). Increased false recognition in older adults is instead explained by increased tendency to attribute the source of activated lures to study lists (Dehon & Brédart, 2004). Fuzzy trace theory's recollection-rejection has also in part attributed age-related differences in false recognition to reduced source monitoring ability (Dodson & Schacter, 2002). Older adults would be less reliably able than young adults to benefit from manipulations encouraging the use of verbatim traces to reduce false memory through recollection-rejection (Gallo et al., 2006), interpreted as support for decline in retrieval-based monitoring. While providing insights into the contribution of age-related change in retrieval processes to reduced mnemonic discrimination in older adults, the activation monitoring theory does not provide a complete account of aging and mnemonic discrimination as they do not make specific predictions regarding age-related differences in encoding. Thus, fuzzy trace theory may be better suited than the activation monitoring theory to explaining false recognition of semantically and perceptually similar lures.

1.2.3 Cognitive neuroscience of age-related mnemonic discrimination

Many different theories have been developed in the literature to account for age differences in mnemonic discrimination. Age-related changes in the neural underpinnings of cognitive processes are well documented and largely reflect deterioration from young adulthood (Raz et al., 2005). These changes include volumetric declines of grey matter in prefrontal cortex and hippocampus, two regions associated with memory encoding and retrieval and predictive of mnemonic discrimination of studied items and similar lures. Previous functional neuroimaging studies of encoding and aging have found significant age effects in prefrontal cortex and medial temporal lobe activity. These regions have been associated with different cognitive abilities contributing to mnemonic discrimination. The prefrontal cortex is thought to support executive functions (Miller and Cohen, 2001; Stuss and Alexander, 2000), which refer to cognitive processes that control and regulate other cognitive operations. Executive functions are necessary to control encoding or retrieval operations (Mitchell and Johnson, 2009). In contrast, the medial temporal lobe is assumed to mediate the processes of pattern separation and completion (Yassa & Stark, 2011), and to bind episodes into one coherent representation. Thus, here I focus on a theory emphasizing executive function deficits (linked to prefrontal cortex dysfunction), and a theory underscoring pattern separation deficits (linked to medial temporal lobe dysfunction).

The resource deficit hypothesis (Craik, 1983, 1986; Craik & Byrd, 1982) suggests that aging is associated with a reduction in the amount of attentional resources, which results in deficits in demanding cognitive tasks. A corollary of the resource deficit hypothesis predicts that age-related differences should be smaller when the task provides a supportive environment which reduces attentional demands (e.g., when older adults emphasise semantic information). Successful mnemonic discrimination also depends on evaluation and monitoring processes assumed to depend on prefrontal cortex, and the link between age-related increases in false recognition and prefrontal cortex dysfunction is supported by evidence of increased false memory rates in frontal lobe patients (Curran et al., 1997), and older adults with poorer frontal lobe functioning (Butler, McDaniel, Dornburg, Price, & Roediger, 2004). Fuzzy trace theory and the activation monitoring theory also propose that decline in prefrontal functions contributes to age-related mnemonic discrimination decline due to a monitoring failure at retrieval.

Second, compared to young adults, older adults showed reduced activity in medial temporal lobe regions (Dennis et al., 2006; Gutchess et al., 2005), possibly reflecting a deficit in the formation of new memory traces. This result would be consistent with a role of these regions in pattern separation and pattern completion (Wilson et al., 2006; Yassa et al., 2011; Yassa et al., 2010). Wilson et al.'s (2006) model of neurocognitive ageing predicts that the older hippocampus is less able to efficiently perform pattern separation, and exhibits an increased tendency towards pattern completion. This age-related shift from pattern separation to pattern completion is proposed to lead to reduced ability to mnemonically discriminate between overlapping representations. The model also predicts that older adults require greater reduction in overlap between incoming and existing representations before pattern separation can occur. The resulting reduced ability to uniquely encode novel items via pattern separation, and tendency towards pattern completion in older adults are proposed to lead to greater mnemonic interference from existing memory traces (Wilson et al., 2003; 2004; 2006). Wilson et al. (2006) explicitly assumed that pattern separation contributes to mnemonic discrimination and that retrieval of existing representations via dominance of pattern completion can lead to mnemonic discrimination errors. It is therefore assumed that older adults' tendency towards pattern completion directly contributes to age-related mnemonic discrimination difficulties. Thus, If age-related differences in mnemonic discrimination extends to multiple dimensions of similarity, including perceptual similarity, false recognition could be accounted for by pattern separation accounts proposing generalised mnemonic discrimination decline (Azab et al., 2014; Yassa & Stark, 2011).

1.3 Thesis

The studies comprising this thesis investigated semantic and perceptual contributions to mnemonic discrimination in young adults and older adults. Fuzzy trace theory assumes that the age-related increase in false recognition is driven by semantic similarity between lures and previously encoded items, while other researchers predict that further non-semantic dimensions of similarity can lead to mnemonic discrimination failure and heightened misrecognition. The lack of a way to separate semantic and perceptual similarity has led to difficulty in accurately assessing the role of these dimensions in driving mnemonic discrimination errors. This thesis includes three empirical investigations, detailed in three chapters (Chapters 2, 3, 4), and followed by a general discussion (Chapter 5). Each empirical chapter begins with an overview of the relevant literature, before methods and results are described and findings discussed in relation to the existing literature.

Chapter 2 describes two published behavioural experiments on young adults (Naspi et al., 2020). In these experiments, we employed pictures of common everyday objects and used a recognition memory task in which, after a study phase, we showed previously studied items, related lures (i.e., different exemplars of studied basic-level concepts), and unrelated novel items (i.e., unstudied basic-level concepts of studied categories). Experiment 1, which was exploratory, aimed to understand what dimensions of similarity were associated with the mnemonic discrimination task. Experiment 2, which was pre-registered, aimed to replicate the results obtained in Experiment 1 and to rule out alternative explanations of the observed results. Chapter 3 describes a published fMRI study conducted on young adults at encoding that combined the subsequent memory paradigm with RSA. I investigated whether the strength of perceptual and semantic representations, as defined by feature-based and computational cognitive models, is reflected in pattern of activity predictive of subsequent true and false recognition. This would support the untested assumption that regions engaged in perceptual and semantic processing also contribute to episodic encoding. Chapter 4 describes age-related differences in mnemonic discrimination in the behavioural experiments described in Chapter 2. The experiments investigated fuzzy trace theory’s prediction that semantic similarity is critical to driving the age-related increase in false recognition. Chapter 5 provides an overview of findings in relation to the above aims, and contributions to the literature are discussed. Strengths, limitations and potential avenues for future research are outlined.

Chapter 2: Multiple Dimensions of Semantic and Perceptual Similarity Contribute to Mnemonic Discrimination for Pictures

2.1 Abstract

People often misrecognize objects that are similar to those they have previously encountered. These mnemonic discrimination errors are attributed to shared memory representations (gist) typically characterized in terms of meaning. In two experiments, we investigated multiple semantic and perceptual relations that may contribute: at the concept level, a feature-based measure of concept confusability quantified each concept's tendency to activate other similar concepts via shared features; at the item level, rated item exemplarity indexed the degree to which the specific depicted objects activated their particular concepts. We also measured visual confusability over items using a computational model of vision, and an index of color confusability. Participants studied single (Experiment 1, N = 60) or multiple (Experiment 2, N = 60) objects for each basic-level concept, followed by a recognition memory test including studied items, similar lures, and novel items. People were less likely to recognize studied items with high concept confusability, and less likely to falsely recognize their lures. This points to weaker basic-level semantic gist representations for objects with more confusable concepts because of greater emphasis on coarse processing of shared features relative to fine-grained processing of individual concepts. In contrast, people were more likely to misrecognize lures that were better exemplars of their concept, suggesting that enhanced basic-level semantic gist processing increased errors due to gist across items. False recognition was also more frequent for more visually confusable lures. The results implicate semantic similarity at multiple levels and highlight the importance of perceptual as well as semantic relations.

2.2 Introduction

Memory for unique experiences depends on the ability to discriminate between events that share multiple overlapping features. For example, we may misrecognize an unknown car in a car park as one we have just hired because it is similar in type and color. Memory theory explains these mnemonic discrimination errors in terms of a gist shared between incoming information and previously encoded representations (Reyna & Brainerd, 1995). Gist is assumed to embody essential meaning shared by multiple items, but the representations underpinning gist are poorly understood. Mnemonic discrimination errors are observed for a wide range of materials including pictures, words, and narratives (Brainerd & Reyna, 2002). More direct evidence also suggests that multiple semantic relations may contribute to the tendency to falsely recognize similar items as having been studied (Brainerd et al., 2008; Cann et al., 2011; Coane et al., 2016; Montefinese et al., 2015). Moreover, although gist is typically conceptualized as semantic, shared perceptual information may also be important (Koutstaal, Reddy, Jackson, Prince, Cendan, & Schacter, 2003; Pidgeon & Morcom, 2014). Here, we combined objective measures of semantic and perceptual similarity with a linear mixed modelling approach to tease apart multiple influences on mnemonic discrimination in one exploratory study and one confirmatory, preregistered study.

Studies using categorized pictures suggest that semantic relations at different organizational levels may impact mnemonic discrimination. In the typical task, individuals study pictures of multiple exemplars of basic-level concepts (e.g., several cats). At test, memory probes include studied items (e.g., the same cat) together with lure items belonging to the same studied basic-level concepts (e.g., a different cat), as well as novel items that do not belong to any of the studied concepts (e.g., a snake) (Koustaal et al., 1999; Koutstaal & Schacter, 1997). Participants must decide whether or not they have previously been shown each picture. Koutstaal and Schacter (1997) found that people were about 24% more likely to falsely endorse the related lures as “old” than they were to endorse unrelated novel items, and subsequent studies have found a similar pattern. Intuitively, studied items and lures (e.g., a different cat) tend to be semantically as well as perceptually similar. Koutstaal and Schacter (1997) attributed these memory errors to processing of semantic and/or perceptual gist. Without specifying the relative roles of these kinds of relations, gist has been proposed to reflect properties like similarity that are shared between studied exemplars and lures (e.g., Gutchess & Schacter, 2012; Koutstaal et al., 2003; Pidgeon & Morcom, 2014; Slotnick & Schacter, 2004).

Variants of the categorized picture task have also been used in which studied and lure objects were related at the superordinate-category level (e.g., land animals: a cat, a horse, a cow), rather than at the basic level. In these tasks, lures in the recognition memory test (e.g., a different land animal, such as a lion) are related to studied items by membership of the same superordinate category (Bowman et al., 2019; Seamon et al., 2000). As for the basic-level version of the task, false recognition is more frequent to lures than to unrelated novel items. The findings from both versions of the categorized pictures task suggest that mnemonic discrimination of pictures depends on multiple semantic as well as perceptual relations, although these studies could not distinguish the different influences. Such results are also consistent with studies using verbal materials in which studied items and lures are semantically related either by superordinate category (Brainerd et al., 1995) or at a higher level such as a narrative (Reyna et al., 2016).

According to fuzzy trace theory, lures elicit false recognition errors when a gist memory representation is not opposed by a detailed memory representation of the related studied items (Reyna & Brainerd, 1995). The gist memory is an episodic trace that represents meanings and relations shared by sets of events, but the informational content of these relations is not specified (Brainerd et al., 2008; Roediger et al., 2001). In the Deese-Roediger-McDermott paradigm (DRM) the studied word lists (e.g., bed, rest, awake) are all associated to a non-presented critical lure (e.g., sleep) (Deese, 1959; Roediger & McDermott, 1995). Frequent false recognition of the critical lures is attributed to these backward associations rather than to shared meaning. However, although critical lures are not typically semantically similar to most studied list items (e.g., sleep is not very similar to bed), they do share other semantic relations (e.g., familiarity and meaningfulness; Brainerd et al., 2008). Preliminary data suggest that lure words that are highly similar to studied items or share thematic information may elicit additional errors that cannot be explained by associative strength alone (Cann et al., 2011; Coane et al., 2016; Montefinese et al., 2015). Coane et al. (2016) showed that people were more likely to misrecognize lure words when studied lists shared semantic features and category membership with the lures as well as being associatively related. However, the between-list comparison could not identify specific effects of feature similarity or category membership. Montefinese et al. (2015) more directly investigated the effects of feature similarity using an index of the number of shared semantic features between pairs of concepts derived from norms for production frequency (McRae et al., 2005; Montefinese et al., 2013). After studying sets of categorically related words (e.g., car, bicycle, truck), participants were more likely to falsely endorse as “old” unstudied items that shared more semantic features to their studied items (e.g., bus compared to plane) (see also Montefinese et al., 2018). However, as there was no baseline novel item condition, it is unknown whether the effects of feature similarity on false alarms reflected a real effect on memory or a modulation of response bias.

There is less agreement about the role of perceptual similarity in mnemonic discrimination. While fuzzy trace theory characterizes false recognition of similar lures in terms of semantic gist, other theories explain it in terms of generic similarity and therefore also predict errors due to perceptual relations between studied items and lures. In global matching models, false recognition (like true recognition) reflects feature overlap with stored memory traces, including visual context as well as semantic relations (Arndt & Hirshman, 1998; Arndt, 2010). Likewise, pattern separation/completion accounts describe mnemonic discrimination in terms of complementary computational processes that act to minimize overlap between new and existing memory traces along multiple dimensions of similarity, and to reinstate stored traces in response to partial cues at test (Wilson et al., 2006; Yassa & Stark, 2011). Several studies using verbal material have suggested that perceptually-driven errors can occur when lures rhyme with lists of studied rhyming words (Budson et al., 2003; Watson et al., 2003; Watson et al., 2001). However, in the majority of studies, which have presented words visually (for an exception see Reyna & Kiernan, 1994), these errors cannot be attributed simply to sensory properties of the stimuli but may also reflect similarity of phonological and/or orthographic representations. Further evidence for perceptual influences on mnemonic discrimination comes from studies showing effects of shared perceptual context in the form of distinctive fonts (Arndt & Reder, 2003). Such effects have also been demonstrated when lure words are not also semantically associated with studied items (Burnside et al., 2017).

The effects of semantic and perceptual relations are more difficult to separate for pictorial material, because some semantic features of objects can be directly perceived (e.g., it can be seen from an image whether an object < has legs >, < is red >). To isolate the effects of semantic similarity on mnemonic discrimination of pictures, Koutstaal et al. (2003) compared memory for sets of colored drawings of abstract and concrete objects. While concrete images depicted meaningful objects that were exemplars of the same basic-level concept, as in the standard categorized pictures task, the sets of abstract images were created to be pre-experimentally meaningless but perceptually similar. False recognition was about 10% higher for lures related to studied concrete than abstract categories, supporting a specific role for semantic as opposed to perceptual gist (for a replication, see Pidgeon & Morcom, 2014).

Perceptual effects were suggested by above-zero false recognition of abstract objects, although this conclusion assumed that the perceptual relatedness of the concrete and abstract objects was matched and that the abstract objects were not processed semantically. This second assumption was questioned by subjective reports from Pidgeon and Morcom (2014)'s participants of spontaneous verbal labelling of the abstract images. Therefore, this paradigm could not unambiguously identify distinct semantic and perceptual contributions to mnemonic discrimination errors. Perceptual effects are suggested by findings that people make errors to lures that are rotated photographs of studied stimuli, in old/new (Motley & Kirwan, 2012) as well as forced-choice (Brady et al., 2008) recognition tasks. In such cases the task is to identify the same image, so in that sense the lures are related semantically to the studied object and the two elements difficult to separate. However, increasing the perceptual (rotational) difference does reduce lure errors (Motley & Kirwan, 2012). There is also recent evidence of perceptual-level interference in true recognition in a retrieval-induced forgetting task as a result of similarity of object shape and color (Reppa et al., 2020). In these studies only one dimension is varied at a time, so perceptual and semantic variables have not been shown to influence memory for the same items. Konkle et al. (2010) separated these elements more objectively, with semantic and perceptual distinctiveness ratings in a task using large sets of pictures organized at the basic level. They found that mnemonic discrimination errors in a forced choice recognition task were slightly more frequent for larger sets of objects that had been rated as less semantically distinctive, but perceptual effects were not significant. Their two ratings were relatively uncorrelated, although it cannot be assumed that perceptual properties did not influence semantic ratings or vice versa (see also Pidgeon & Morcom, 2014).

2.2.1 The Current Research

The studies reviewed above have established that semantic relations between studied and unstudied items are a key driver of mnemonic discrimination errors, but it remains unclear what kind of semantic information is critical. Moreover, although gist is usually conceptualized as semantic, perceptual similarity may also be important, particularly for rich pictorial material. The two experiments reported here used objectively quantified dimensions to investigate the effects of semantic and perceptual similarity on mnemonic discrimination in a typical categorized pictures task. In this task, items are pictures of individual objects, so lures are different exemplars of the same basic-level concept. We operationalized semantic similarity at two distinct organizational levels: at the *concept level*, indexing the relations between the basic-level concepts, and at the *item level*, indexing the relations between individual exemplars and their basic-level concepts. Perceptual similarity was also quantified using properties shared between items.

We used feature overlap to measure semantic similarity at the concept level. A large body of experimental evidence supports the view that semantic memory is structured in terms of features. According to distributed feature models, like the *Conceptual Structure Account* (CSA; Tyler & Moss, 2001), concepts are represented in the brain by their features

(e.g., < has legs >, < has eyes >, < has a tail >) in a connectionist system in which the mutual co-activation of the feature nodes determines the semantic processing (McRae et al., 1997; Tyler & Moss, 2001; Vigliocco et al., 2004). The statistical regularities of semantic features, derived from property norms, have proven to be a useful way of characterizing the structure and content of semantic representations (Devereux et al., 2014; Garrard et al., 2001). The most prominent statistical characteristic assumed to structure the semantic space and determine how concepts are processed is concept confusability (Clarke & Tyler, 2014). Concept confusability measures the degree to which a concept's semantic features are shared with other concepts (e.g., many animals < have ears >). Highly shared features (e.g., < has legs >, < has eyes >, < has a tail >) provide coarse information about the superordinate categories to which a concept belongs (e.g., land animals), and support decisions that depend only on this coarse-grained information. In contrast, accessing specific basic-level concepts (e.g., tiger) requires finer-grained semantic processing that includes features more distinctive to the particular concept (e.g. < has stripes >), and support more specific tasks like naming. In the current work, we used concept confusability as an index of information shared across the basic-level concepts.

To complement the examination of concept confusability, we assessed effects of semantic and perceptual relations at the item level, via properties of the individual depicted exemplars. Such properties are likely to be important determinants of people's ability to correctly reject lures which are unstudied exemplars of studied basic-level concepts. Previous studies have shown that basic-level processing is enhanced for objects that are more representative of a stored basic-level conceptual representation. Pictures that are better exemplars are categorized faster at the basic level than pictures that are poorer exemplars (Barry et al., 1997; Snodgrass & Vanderwart, 1980). Therefore, we used ratings of item exemplarity to index how well a picture corresponded to its basic-level representation. We assumed that similarity between a picture and its conceptual representation would facilitate gist processing at the item level, i.e., information shared across exemplars of each concept.

To operationalize perceptual similarity, we drew on an established computational model of perceptual processing. The *Hierarchical Model and X* (HMax; Riesenhuber & Poggio, 1999; Serre et al., 2007) models different hierarchical stages of the ventral processing stream in different layers, progressing from early visual cortex (V1) to posterior inferior temporal cortex (IT). The C1 layers correspond to increasingly position- and scale-invariant early visual cortex (V1/V2) which maintain feature specificity, while C2 layers simulate the extrastriate visual area cells (V4/IT) that integrate visual features from previous layers to represent object shape. Measures based on these two layers have been validated in studies of visual object recognition that have distinguished the time courses and neural correlates of semantic versus visual processing (Clarke & Tyler, 2014; Clarke et al., 2015). Here, we generated measures of the visual confusability of each image with others in the set, defined in terms of the properties indexed by C1 and C2 (e.g., orientation and shape). Lastly, since the HMax model does not represent color, we computed a novel index of color confusability using the CIELab color space, known to be an approximation of human color perception (Rubner et al., 2000).

In a first, exploratory, experiment, we examined mnemonic discrimination in a simple task in which participants studied one exemplar for each basic-level concept and were later tested on a single lure (Bakker et al., 2008; Lacy et al., 2011; Reagh et al., 2016; Stark et al., 2013; Yassa et al., 2011). Then in a second, preregistered, confirmatory experiment, we sought to replicate and explain the findings of the first by increasing the number of studied exemplars of each basic-level concept. Successful mnemonic discrimination requires accurately identifying studied items and rejecting similar lures. Since distinct processes may contribute to memory for these two trial types, performance is typically assessed using separate measures of true and false recognition, and many studies focus mainly on false recognition. To adjust for response criterion, sensitivity measures derived from signal detection theory can be computed from rates of endorsement of studied and lure items as “old”, relative to unrelated unstudied items. Equivalent measures of studied relative to lure item endorsement can also assess overall ability to discriminate studied items from lures in memory. In both experiments, we used generalized linear mixed effects models, which provide parameter estimates to index participants’ memory sensitivity and response bias that are equivalent to those provided by the equal variance signal detection theory framework (DeCarlo, 1998). This allowed us to preserve and take into account variability across items as well as participants, which results in an increased accuracy and generalizability of the parameter estimates (Baayen et al., 2008; Quené & van den Bergh, 2008).

In this task, the studied and lure exemplars of the same basic-level concept can be regarded as sharing a semantic gist and possibly also a perceptual gist across items. We reasoned that concept confusability would impact mnemonic discrimination by weakening the basic-level conceptual representations shared by studied items and their lures. If a concept is highly confusable with other concepts because they share semantic features, the concept’s distinctive, basic-level representation will be less likely to be encoded. Thus, we expected that studied objects whose concepts are more confusable would be less likely to be recognized and their corresponding lures more likely to be successfully rejected. For example, if a cherry is studied – a concept with high confusability due to multiple highly shared features like < does grow on trees >, < is sweet >, < is edible > – a weak representation of “cherry” will be encoded alongside the details of the particular exemplar. In contrast, if a foot is studied – a concept with low confusability due to its few shared features and its distinctive features < has toes >, < is found at the end of a leg > – a strong representation of “foot” will be more likely to be encoded. In a memory test, the same exemplar of “cherry” will thus be less likely to be recognized than the same exemplar of “foot”, since memory for the former is weaker. For the same reason, for lures, a different exemplar of “cherry” will be more likely to be successfully rejected than a different exemplar of “foot”, given weaker encoding of the studied concept. In terms of the item-level metrics, we assumed that objects with higher rated item exemplarity would more strongly engage the basic-level conceptual processing shared by studied items and lures. We therefore predicted that higher exemplarity lures would be more likely to be misrecognized. Lastly, based on similar logic we predicted more frequent false recognition of more visually confusable lures.

2.3 Experiment 1

2.4 Materials and Methods

2.4.1 Participants

The study included sixty participants aged 18-33 years (*M* = 21; *SD* = 2.2; 15 male, 45 female). Eleven further participants were excluded from data analysis: 7 due to errors in stimulus lists or data acquisition issues, 1 due to misunderstanding of instructions, and 3 who did not meet the inclusion criterion for English fluency. The sample size was determined *a priori* using the simR package in R (Version 1.0.4; Green & Macleod, 2016). We powered the study for an interaction of condition (lure vs. new) × concept confusability, based on a pilot study suggesting an effect size equivalent to Cohen's *d* = .17 (*OR* = 1.35). With *N* = 60 we had 87.30 % power to detect such an effect at alpha = .05. Inclusion criteria were fluency in English (spoken since at least the age of 5 years), and normal or corrected-to-normal vision. Data from an additional group of older participants collected at the same time will be included in a separate report. Participants were compensated financially or with course credits. They were contacted by local advertisement and provided informed consent. The study was approved by the University of Edinburgh Psychology Research Ethics Committee (Ref. 278-1617/1).

2.4.2 Stimuli

Stimuli were pictures of objects corresponding to 180 of the 638 basic-level concepts in the Centre for Speech, Language and the Brain property norms (the CSLB norms; Devereux et al., 2014). These 180 basic-level concepts comprised 9 members of each of 20 different superordinate categories (*Appliance*, *Bird*, *Body Part*, *Clothing*, *Container*, *Drink*, *Flower*, *Food*, *Fruit*, *Furniture*, *Invertebrate*, *Kitchenware*, *Land Animal*, *Music*, *Sea Creature*, *Tool*, *Toy*, *Vegetable*, *Vehicle*, *Weapon*) and half were living and half non-living. We sourced two images for each basic-level concept. One set of 180 was a subset of images used by Clarke and Tyler (2014), and the other was compiled from the Bank of Standardized Stimuli (BOSS; Brodeur et al., 2014) and from the Internet. Each study list comprised 120 images of exemplars of different basic-level concepts, selected evenly from the superordinate categories (i.e., 6 different basic-level concepts per superordinate category). Each test list consisted of 180 items: 60 studied images, 60 similar lures (i.e., novel images corresponding to the other 60 studied basic-level concepts), and 60 novel items (i.e., novel images of basic-level concepts that had not been studied). Three filler trials prefaced both study and test phases. We generated 6 different study and test lists which fully counterbalanced the allocation of the basic-level concepts and the two sets of images to conditions (studied, lure, and novel).

2.4.3 Procedure

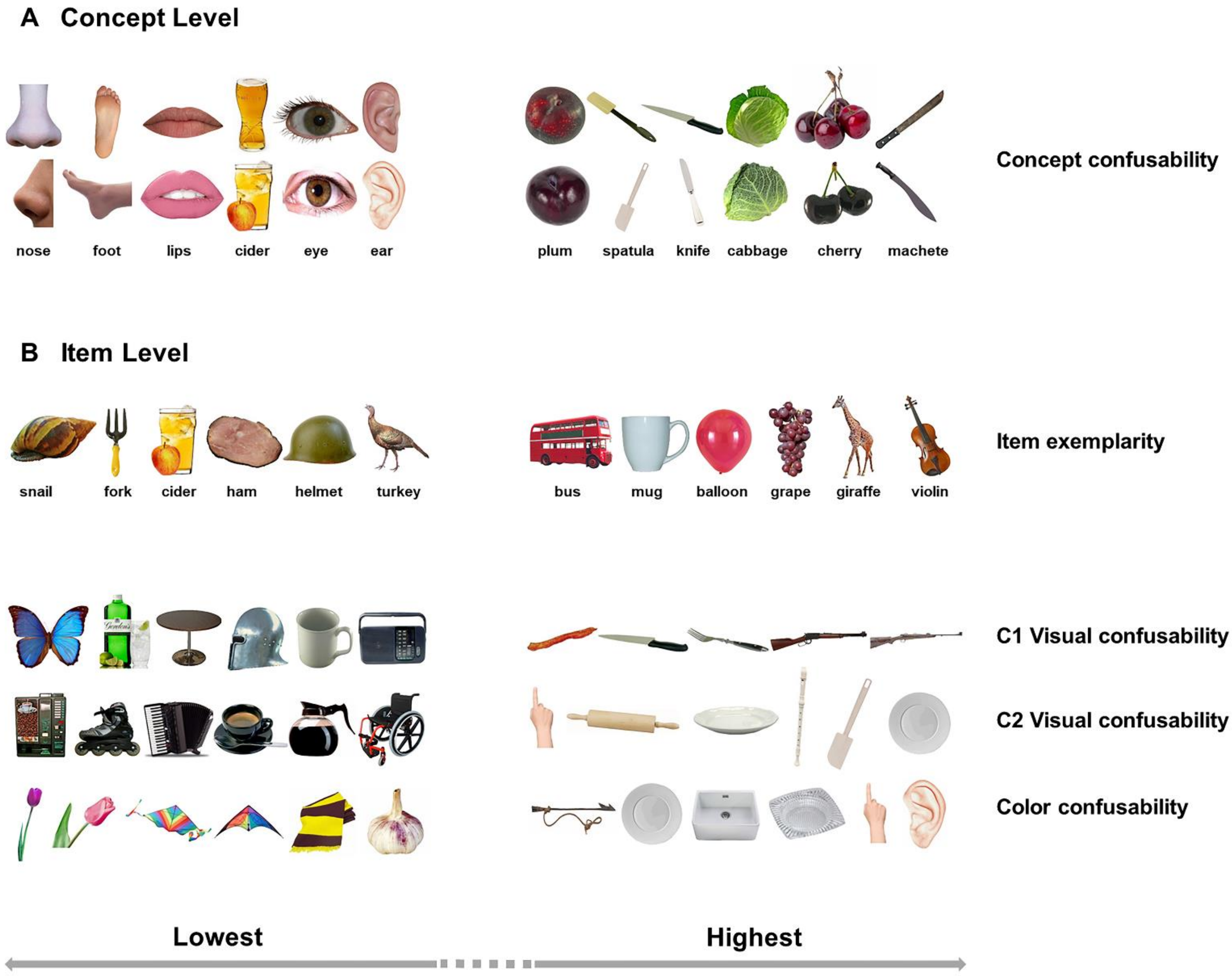
The experiment consisted of a single study phase followed by a recognition test phase. Between study and test phases, participants completed standardized cognitive tests (not reported here) for 15 minutes. Stimuli were presented using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Participants were tested individually in the laboratory. At study, they viewed one image at a time, and rated how pleasant it was from 1 (very unpleasant) to 5 (very pleasant). Participants were not informed of a later memory test. Images were presented against a white background within a 15 × 13 cm area in the center of a computer screen, and viewed at a distance of approximately 50 cm. Trials were self-paced. At test, participants viewed one image at a time every 3 s and judged each as “old” or “new” using the keyboard, indicating at the same time whether they were confident in this judgement. Responses were made using the “Z” and “A” keys when judging an item as “old” with high and low confidence, respectively; the “M” and “K” keys when judging an item as “new” with high and low confidence, respectively, and mappings of responses to hands counterbalanced over conditions. After completing the test phase, each participant also completed the concept familiarity or the item exemplarity ratings. The entire procedure took approximately 50 minutes and participants were debriefed at the end of the experiment.

2.4.4 Variables of Interest

The concept- and-item level variables of interest are illustrated in Figure 1.

Figure 1

Schematic Depiction of Stimuli Used in Both Experiments



Note. Rows show individual exemplars with the highest (right side) and lowest (left side) scores on each concept- and-item level experimental variable. Panel A illustrates concept confusability: the basic-level concept name is given alongside images of both exemplars representing it in Experiment 1. High confusability concepts share more semantic features with other concepts. Panel B illustrates item-level measures for individual images of objects. Item exemplarity is an index of similarity between the depicted exemplar and the concept representation. The perceptual measures define confusability of an item as the similarity with its most similar neighbor in the set. C1 and C2 were obtained from gray-scaled version of the images depicted in Figure 1. For definitions see Variables of Interest section.

2.4.4.1 Concept Level

Concept confusability. Our measure of concept confusability was based on that of Clarke and Tyler (2014), but we used a gentler weighting system and an updated set of property norms (Devereux et al., 2014). The property norms provide a matrix of features associated with 638 objects (e.g., has 4 legs, has stripes and lives in Africa are features of a zebra). These were collected by presenting participants with a written concept name and asking them to produce properties of the concept. Taxonomic features (e.g., < is a bird >) were excluded as they refer to a superordinate category and are not normally regarded as true semantic features (Taylor et al., 2012). Based on the feature norms, each object can be represented by a binary vector. Semantic similarity between concepts was computed as the cosine angle between feature vectors in a semantic feature matrix in which each concept was represented by a binary vector indicating whether each feature was associated with the concept (1) or not (0). Concept confusability with all the other concepts in the set was then calculated by a weighted sum of the similarities in which each weight was the between-concept similarity itself, i.e., the sum of squared similarities (see Figure 1). This measure emphasized feature sharedness of each concept with those concepts with which it shared many semantic features. We also calculated the number of features for each concept from the same norms.

Concept familiarity. Thirty of the 60 participants judged the 360 pictures representing the 180 concepts. Following Snodgrass and Vanderwart (1980), we asked subjects to judge the familiarity of each picture *"according to how usual or unusual the object is in your realm of experience"*. Concept familiarity was defined as *"the degree to which you come in contact with or think about the concept"*. Participants were required to give their response on a 7-point scale (1 = completely unfamiliar, 7 = completely familiar) using the keyboard. For each picture, familiarity scores were averaged over all participants (see Supplemental Materials for its theoretical relevance).

2.4.4.2 Item Level

Item exemplarity. To obtain a measure of item exemplarity we used a rating task in which pictures were presented together with their verbal label. Following Taylor et al. (2012), we asked participants to judge "*how closely each picture resembles your mental image of the object*", giving their response on a 7-point scale (1 = poor picture of concept word, 7 =

excellent picture of concept word). The 30 participants who did not provide familiarity ratings were asked to provide these item exemplarity scores, which were averaged to give a single score per picture (see Figure 1).

Perceptual confusability. Measures capturing the low- and high-level visual attributes of each picture were derived for each of the 360 images (2 per concept) used to represent our basic-level concepts. We calculated two indices for each dimension which embodied two alternative hypotheses about how perceptual similarity between images might affect mnemonic discrimination. Although the second measure was preregistered for Experiment 2, for clarity we only report the results from the first in the main paper, since it gave a coherent picture across the two experiments (see also Discussion). We first extracted HMax estimates of low- and high-level visual object information: a C1 response related to early visual cortex (V1/V2), and a C2 response related to V4/posterior IT. Response vectors from the C1 and C2 layers were computed for grey-scaled versions of each image. Similarity between pictures was then calculated using the Pearson correlation coefficients between vectors. For the main visual confusability measure, for each image we defined confusability as the similarity value with its most similar picture (i.e., the nearest neighbor; see Figure 1). The second metric, graded visual confusability, was analogous to concept confusability, indexing an image's similarity to the full set of images (see Supplemental Figure 4). We calculated a weighted score for each image by summing the squared ranks of Pearson correlations between each image i and all other images j, so pictures with high graded visual confusability scores were those that were similar to many other pictures. To obtain the ranks of all the similarity values, we first extracted the matrix of Pearson correlations between images as a single vector. Then, we computed the corresponding rank for each correlation, and transformed the vector of ranks back into a matrix. This allowed us to deal with negative Pearson correlations, which were assigned the lowest ranks.

For each image, we also generated a nearest neighbor index of color confusability using the color distance package (Weller & Westneat, 2019) in R (version 3.4.3; R Core Team, 2017). This measure represents the degree to which the color of each item resembled that of the most similarly-colored item in set. After converting the RGB channels into CIELab space, we calculated the earth mover's distance between each pair of images (Rubner et al., 2000). We then normalized the distance and transformed the distance matrix in a similarity matrix using the equation $S = 1 - D$ so that similarity values ranged from 0 (lowest similarity) to 1 (highest similarity). Then, for each item, we retained the similarity with its most similar item in the set (see Figure 1). As described above for the HMax measures, we also calculated a measure of graded color confusability. From the similarity matrix, to obtain a single metric for each image, we summed the squared ranks of similarities, so higher values indicate greater color confusability with all the other pictures in the set (see Supplemental Figure 4).

2.4.5 Nuisance Variables

Mnemonic discrimination may be influenced by a range of other visual, phonological, lexical, and semantic factors in addition to the semantic and perceptual confusability measures of interest here. We controlled for the effect of the following nuisance variables, described in more detail in the Supplemental Materials: forward and backward associative strength estimated using a continuous association task (De Deyne & Storms, 2008; Nelson et al., 2004), word frequency (van Heuven et al., 2014), concreteness (Brysbaert et al., 2014), age of acquisition (Brysbaert & Biemiller, 2017), phonological neighborhood density (Baayen et al., 1995), the number of non-white pixels, color entropy (Chouinard & Goodale, 2012), and concept familiarity (derived from our rating task; see also Taylor et al., 2012).

2.4.6 Statistical Analysis

We tested our hypotheses using a series of generalized linear mixed-effect model analyses with the function glmer from the lme4 package in R (version 1.1-17, Bates et al., 2015). The linear mixed-effect model approach affords greater robustness and generalizability of inference compared to the analysis of variance typically used in studies of memory (Baayen et al., 2008). Modelling random effects of items as well as participants is helpful in studies of memory where generalization to other stimulus sets as well as other participant samples is desirable (Clark, 1973). Accuracy was modelled using a multiple linear logistic regression on participants' binary recognition judgments ("new" = 0, "old" = 1) fitted by means of a logit link function. Thus, rather than considering our data in terms of proportions of hits, misses, false alarms, and correct rejections, as in the standard recognition test analysis, we directly predicted behavioral outcomes from item status (i.e., studied, lure, or novel). In this way we estimated how the probability of judging an item as "old" depended on its actual status, and asked which of our variables of interest moderated this effect. The log-odds-ratio coefficients generated by the generalized linear mixed effect logit models are formally proportional to d' in a Gaussian signal detection analysis ($d' \approx .6$ logOR; DeCarlo, 1998; Wright et al., 2009). This relation holds over a wide range of values, so logit and probit models yield equivalent results except at the extremes. In the Results section we therefore report effect sizes in terms of *d'* equivalent for ease of comparison with other studies. We also performed an exploratory linear mixed-effect analysis of response times (RTs) at study to check whether the experimental variables would impact (non-speeded) decisions about item pleasantness. The results did not reveal any significant effects, and are not reported further (but can be found on https://osf.io/ndk83/).

To test specific predictions about memory for studied items and lures, we set the reference level for the condition factor to "novel" so that with simple contrasts we could examine modulations of a) the probability of correctly identifying studied items as "old" relative to novel items (an index of sensitivity for studied items equivalent to *d'*, reflecting true memory), b) the probability of misrecognizing related lures relative to novel items (an index of sensitivity for related lures equivalent to *d'*, reflecting false memory). This also yielded c) the probability of falsely judging novel items as "old" (an index of baseline false alarms, equivalent to the response criterion *c*). So for example, an estimated *d'* = 1.5 for an interaction between a continuous variable and the contrast of lure versus novel items, means that a one SD increase in that continuous variable is associated with a 1.5 x increase in *d'* for lures versus novel items. In a final set of contrasts, we set the reference level for the condition factor to "lure" to assess modulations of d) the probability of endorsing studied items as "old" relative to lures (an index of overall sensitivity equivalent to *d'*). This allowed us to evaluate the effect of the semantic and perceptual variables on overall mnemonic discrimination performance, a net effect of their modulations of true and false memory (Koutstaal & Schacter, 1997; Loiotile & Courtney, 2015).

For each model, we specified *a priori* random intercepts of both participants and concepts (Matuschek et al., 2017). In the fixed part, our variables of interest were condition (studied, lure, novel), two concept-level variables (i.e., concept confusability, number of features), and four item-level variables (i.e., item exemplarity, C1 and C2 visual confusability, color confusability). Within the concept-level and item-level partitions we also included the corresponding interactions with condition. To minimize model complexity, and because the nuisance variables were moderately-to-highly intercorrelated, we performed data reduction of the nuisance variables with principal components analysis (PCA) using the prcomp function in R. PCA with varimax rotation produced a 7 factor solution which accounted for 86.14% of variance (see Supplemental Table 5). We then compared the goodness-of-fit between i) the confounds model with the original nuisance variables and ii) the reduced confounds model with the principal components using the corrected Akaike information criterion (AICc). We also complemented the goodness-of-fit measure provided by AICc with the corresponding Bayesian information criterion (BIC) and the likelihood ratio test (LRT). Model selection revealed that the simpler confounds model with principal components provided a better goodness-of-fit (AIC = 14420.23, BIC = 14492.97) relative to the model with the original nuisance variables (AIC = 14424.65, BIC = 14519.21; for LRT for a difference between models, $\chi 2(3) = 1.59$, $p = 0.662$). Thus, these 7 principal components were included as nuisance variables for our models.

Model selection was carried out to determine the fixed effects structure with the best goodness-of-fit based on sets of theoretically motivated predictors, which included concept-level, item-level, and confound principal component variables. Starting with the most complex model, we used the AICc to compare progressively simpler models. At each step, we verified whether the exclusion of a particular set (in order, concept-level, item-level, and nuisance variables) was justified or not. We also supplemented this measure with the corresponding BIC and the LRT between models. Model comparison was performed using the AICcmodavg package (version 2.3-1) and the anova function in R. *Post-hoc* analyses were conducted, and results interpreted after selecting the best model. All the continuous predictor variables were standardized, and the resulting β coefficients representing log-odd-ratios were used to calculate the corresponding d' coefficients.

2.5 Results

The results of the mixed effects analysis for the winning model are shown in Table 1, and illustrated in Figure 2. Raw recognition responses by item type are reported in Supplemental Figure 5. The initial model comparison suggested that the full model with concept-level, item-level, and confounds principal component variables was the most parsimonious model for the data. This model received the lowest AICc score (AICc = 8821.92), indicating decisive evidence relative to simpler models (see Supplemental Table 6). It received substantial weight (AICc weight = .98) of the total weight of the models considered, with an evidence ratio between the top-ranked model and the second-ranked model of 58.15. That is, the evidence was 58.15 times stronger for the best model. This result was also supported by a significant LRT in favor of the full relative to the model ranked second on the basis of AICc ($\chi 2(7) = 19.59$, $p = .007$). Unlike AICc and LRT, BIC provided evidence in favor of the model with concept-level variables only as the most parsimonious model (BIC = 8991.90). However, when we allowed the free selection of variables rather than the comparison based on our three blocks (i.e., removal of concept-level, item-level, and confounds principal component variables in this order), BIC resulted in a better goodness-of-fit when item exemplarity and C1 visual confusability were included in the model together with the concept-level variables (BIC = 8827.70). Participants' responses were collapsed across confidence judgments as results were qualitatively similar when high confidence mnemonic discrimination was analyzed. Coefficients in Table 1 represent log-odds-ratios with the corresponding *d'* effects. All the *p*-values reported below are FDR-multiple comparison corrected (Benjamini & Hochberg, 1995).

Table 1

Results of Experiment 1 (Novel Items as Baseline)

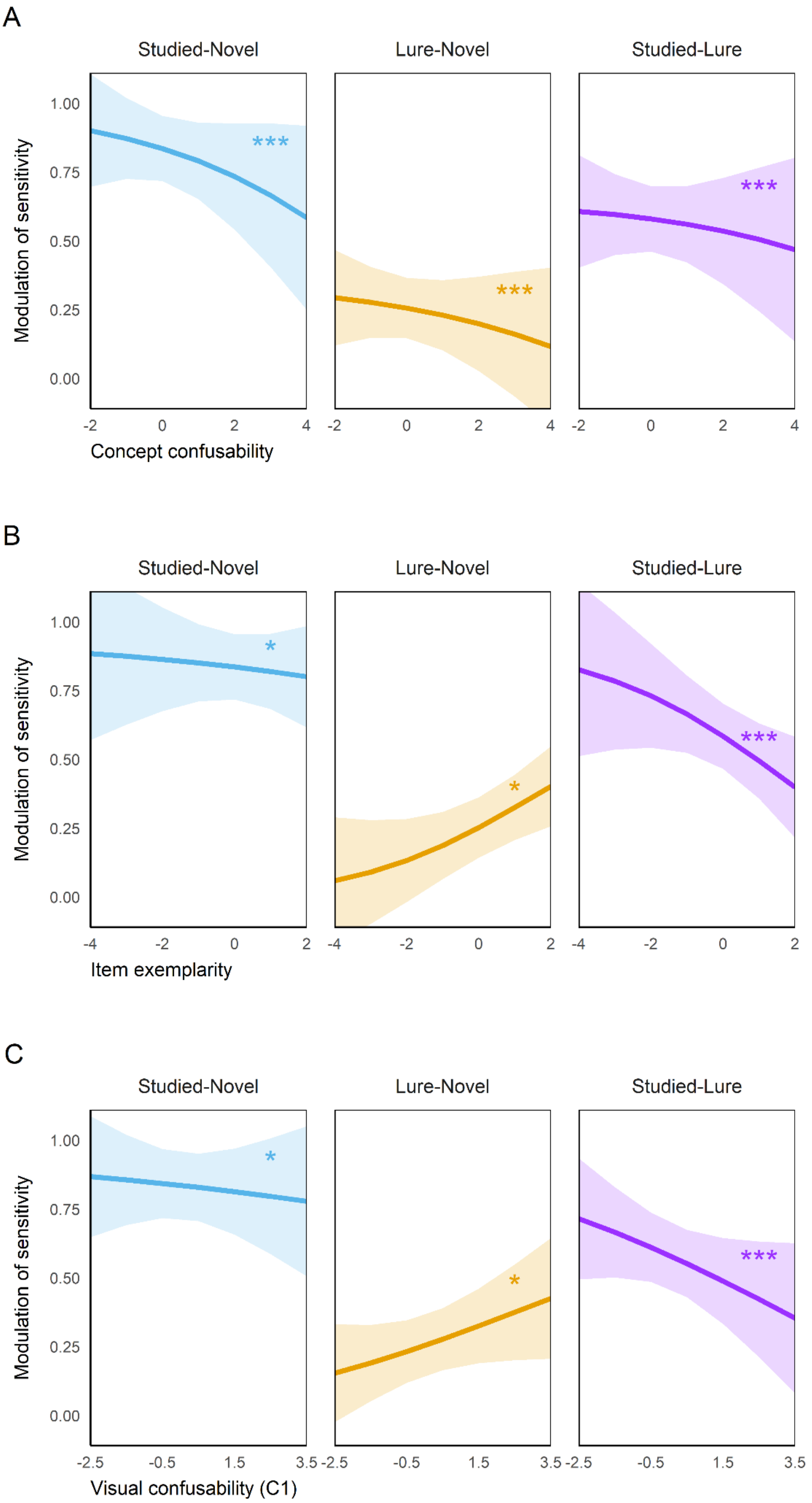
Variable	Estimate	*d'*	*SE*	*z*-value	*p*
(Intercept)	-2.41	-1.37	0.12	-20.02	<.001
Lure	1.71	0.95	0.07	23.09	<.001
Studied	4.85	2.74	0.09	52.00	<.001
Number of Features	0.13	0.07	0.08	1.66	.136
Concept Confusability	0.28	0.15	0.08	3.44	.002
Visual Confusability (C1)	0.06	0.03	0.07	0.85	.438
Visual Confusability (C2)	0.20	0.10	0.09	2.28	.040
Color Confusability	0.04	0.02	0.08	0.47	.635
Item Exemplarity	0.18	0.09	0.07	2.55	.025
Lure × Number of Features	-0.31	-0.17	0.07	-4.25	<.001
Studied × Number of Features	-0.10	-0.05	0.09	-1.11	.332
Lure × Concept Confusability	-0.28	-0.15	0.07	-3.96	<.001

Studied × Concept Confusability	-0.54	-0.29	0.09	-6.30	<.001
Lure × Visual Confusability (C1)	0.16	0.10	0.07	2.28	.040
Studied × Visual Confusability (C1)	-0.19	-0.10	0.08	-2.24	.041
Lure × Visual Confusability (C2)	-0.15	-0.06	0.09	-1.67	.136
Studied × Visual Confusability (C2)	-0.11	-0.05	0.10	-1.07	.332
Lure × Color Confusability	-0.04	-0.02	0.08	-0.48	.635
Studied × Color Confusability	-0.11	-0.06	0.09	-1.17	.317
Lure × Item Exemplarity	0.20	0.13	0.08	2.57	.025
Studied × Item Exemplarity	-0.21	-0.10	0.09	-2.29	.040

Note. The reference level of condition is set to “novel”. Parameter estimates (logOR), *d’* equivalent, standard errors, z-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Nearest neighbor perceptual confusability measures were reported in the model above. See Material and Methods, Variables of Interest, and Results for details. *SE* = Standard Error.

Figure 2

Effects of Semantic and Perceptual Variables on Mnemonic Sensitivity in Experiment 1



Note. Plots show effects of semantic and perceptual variables on modulation of sensitivity. The plot lines represent the effect of the predictor variables on the probabilities of endorsing studied items as “old” relative to novel items (light blue), lures as “old” relative to novel items (orange), and studied items as “old” relative to lures (purple). Panel A, B, and C show the effects of concept confusability, item exemplarity, and C1 visual confusability in Experiment 1. See Material and Methods, Variables of Interest, and Results for details. * $p < .05$; ** $p < .01$; *** $p < .001$ (FDR-corrected).

2.5.1 Sensitivity for Studied Relative to Novel Items

The probability of judging a studied item as “old” was compared to the probability of judging a novel item as “old”. Overall, discrimination of studied from novel items was very good (simple effect on *d’* = 2.74; 95% CI [2.65, 2.83]). The results also showed modulations of concept- and item-level variables on sensitivity for studied items. Images whose concepts were more confusable with other concepts in the set were less likely to be remembered (Figure 2A; interaction of concept confusability with studied items on *d’* = -.29; 95% CI [-.38, -.20]). Pictures judged with high exemplarity were also less likely to be remembered (Figure 2B; interaction of item exemplarity with studied items on *d’* = -.10; 95% CI [-.19, -.01]). Lastly, participants were less likely to recognize items that were visually confusable in terms of their low-level representations (Figure 2C; interaction of C1 with studied items on *d’* = -.10; 95% CI [-.19, -.01]). No other variable significantly modulated sensitivity for studied items.

2.5.2 Sensitivity for Lure Relative to Novel Items

The probability of judging a lure item as “old” was compared to the probability of judging a novel item as “old”. As expected, participants were more likely overall to incorrectly endorse lures than novel items (simple effect on *d’* = .95; 95% CI [.88, 1.03]). The concept-level semantic variables also had substantial effects on lure sensitivity. Fewer errors were observed for lure images whose concepts were more confusable with other concepts in the set (Figure 2A; interaction of concept confusability with lure on *d’* = -.15; 95% CI [-.23, -.07]), as well as for those with a greater number of semantic features (interaction of number of features with lure on *d’* = -.17; 95% CI [-.25, -.10]). At the item-level, false recognition was more likely for lures rated as better exemplars of their concept (Figure 2B; interaction of item exemplarity with lure on *d’* = .13; 95% CI [.05, .21]), as well as for lures whose early visual representations (such as line orientation) were more confusable (Figure 2C; interaction of C1 with lure on *d’* = .10; 95% CI [.03, .18]). No other variable significantly modulated sensitivity for lure items.

2.5.3 Sensitivity for Studied Relative to Lure Items

The results of the analysis that examined the net modulation of participants' ability to discriminate studied items from lures are shown below in Table 2, and illustrated in Figure 2. Overall, performance was fairly good (simple effect on *d'* = 1.79; 95% CI [1.71, 1.87]). Both concept- and-item level variables modulated this effect. Mnemonic discrimination was poorer for items with high concept confusability (Figure 2A; interaction of concept confusability with studied items on *d'* = -.13; 95% CI [-.21, -.06]), but was better for concepts with a larger overall number of semantic features (interaction of number of features with studied items on *d'* = .12; 95% CI [.04, .20]). At the item level, objects judged as having high item exemplarity were also less well discriminated (Figure 2B; interaction of item exemplarity with studied items on *d'* = -.23; 95% CI [-.31, -.15]), as were images with greater low-level (C1) visual confusability (Figure 2C; interaction of C1 with studied items on *d'* = -.20; 95% CI [-.28, -.13]). No other variable significantly modulated sensitivity for studied relative to lure items.

Table 2

Results of Experiment 1 (Lure Items as Baseline)

Variable	Estimate	*d'*	SE	z-value	*p*
(Intercept)	-0.70	-0.42	0.11	-6.39	<.001
Studied	3.14	1.79	0.07	42.02	<.001
Studied × Number of Features	0.21	0.12	0.08	2.80	.005
Studied × Concept Confusability	-0.25	-0.13	0.07	-3.49	<.001
Studied × Visual Confusability (C1)	-0.35	-0.20	0.07	-4.76	<.001
Studied × Visual Confusability (C2)	0.03	0.02	0.08	0.40	.686
Studied × Color Confusability	-0.07	-0.04	0.08	-0.93	.354
Studied × Item Exemplarity	-0.40	-0.23	0.08	-5.07	<.001

Note. The reference level of condition is set to “lure”. Parameter estimates (logOR), *d’* equivalent, standard errors, z-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Nearest neighbor perceptual confusability measures were reported in the model above. See Material and Methods, Variables of Interest, and Results for details. SE = Standard Error.

2.5.4 False Alarms to Novel Items

Overall, participants were good at identifying unstudied novel items as “new” (intercept on *d’* = -1.37; 95% CI [-1.50, -1.25]). This baseline was also modulated by both concept- and item-level semantic variables. False alarms were more frequent for highly confusable novel items (Supplemental Figure 5A; simple effect of concept confusability on *d’* = .15; 95% CI [.07, .23]), and for those with high rated item exemplarity (Supplemental Figure 5B; simple effect of item exemplarity on *d’* = .09; 95% CI [.02, .17]). The baseline probability of false alarms to novel items was also modulated by the C2 index of visual confusability. People were more likely to falsely endorse novel items as “old” if they were highly confusable in terms of late visual representations (such as global shape) (simple effect of C2 on *d’* = .10; 95% CI [.01, .19]). No other variable significantly modulated baseline false recognition of novel items.

2.6 Discussion

In Experiment 1 we investigated the effects of conceptual and perceptual confusability on mnemonic discrimination of objects, using objective measures of similarity within a generalized linear mixed model framework. The findings for both concept-level and item-level variables support the proposal that semantic and perceptual properties shared by studied and lure objects contribute to mnemonic discrimination. At the concept level, studied objects whose concepts were more confusable because they shared features with other concepts were less likely to be remembered, and the corresponding lures were more likely to be correctly rejected. In contrast, at the item level, while studied objects that were better exemplars of their concept were less well recognized, lure objects with this property triggered more frequent errors. Likewise, people also misrecognized more lures that were visually confusable with their nearest neighbor. Overall, these effects resulted in poorer mnemonic discrimination between studied and lure objects with high concept confusability, high item exemplarity, and high visual confusability.

The effects of concept confusability on sensitivity for studied items and lures suggest that when concepts were more confusable with other concepts, memory was weaker for the basic-level information shared by the studied items and lures. Our metric of concept confusability was based on an established model of conceptual structure and existing norms for feature properties of common concepts (Devereux et al., 2014; Tyler & Moss, 2001). This measure reflects conceptual processing of shared information across a set of items. The more confusable concepts shared more features with other concepts, and had fewer distinctive features not shared with other concepts. Our results suggest that if people remember less distinctive information about an object’s concept, they are less likely to remember the object. The same logic applies to lures: lures with more confusable concepts can be more easily rejected as unstudied because memory is weaker for the concept shared with the studied item. This interpretation is supported by Taylor et al.'s (2012) finding that domain-level categorization decisions (living/ non-living judgments) were faster for concepts with more shared features, while basic-level naming was faster for concepts with more distinctive features. Thus, the efficiency of basic-level processing depended on the relative emphasis on coarse, cross-concept processing of shared features relative to fine-grained processing of individual concepts with distinctive features. Here, we did not find any modulation of study phase RT by concept confusability or other experimental variables. This probably reflected the use of self-paced judgments at study and prioritization of accuracy over speed given our primary concern with memory accuracy. From a fuzzy trace theory viewpoint, gist memory contributes positively to both veridical memory and lure false recognition (Brainerd et al., 2008; Brainerd et al., 1995). In this typical categorized pictures task, the gist is at the basic level since studied and lure objects are exemplars of the same basic-level concepts (e.g. a dog was studied, and a different dog appeared as a lure). In Experiment 1, this gist was impoverished for highly confusable concepts, reducing both recognition of the studied items and gist-related errors for the lures. Overall, mnemonic discrimination between these items was poorer.

While people were better able to successfully reject lures with more confusable concepts, false alarms to novel items with more confusable concepts were increased (see Supplemental Figure 5A). These items did not share a basic-level concept with any studied items but did share semantic features with them to varying degrees. This result is similar to that reported by Montefinese et al. (2015) for verbal stimuli. Participants studied sets of categorically related words (e.g., car, truck, scooter for the vehicle category) and were more likely to falsely recognize unstudied words which had more shared semantic features (e.g., tram). In that study, the novel words were related to the studied sets by membership of common superordinate categories, like the novel objects in the current study. Thus, it is possible that this higher-level semantic similarity caused these items to be more difficult to discriminate in memory. However, since both this result and Montefinese et al.'s (2015) finding were modulations of raw false alarms, we cannot rule out the alternative possibility that they reflect an effect of concept confusability on response criterion (Heit et al., 2003; Loiotile & Courtney, 2015).

The results for item exemplarity show that sensitivity for both studied items and lures depended on semantic processing at the item level as well as the concept level. As predicted, higher exemplarity lures were more likely to be falsely identified as studied, consistent with enhanced activation of basic-level representations. This suggests that participants tended to remember having studied the corresponding basic-level concept (e.g., giraffe in Figure 1) even if they did not remember the specific studied exemplar. These data are consistent with Barry et al.'s (1997) finding of facilitated basic-level naming for high exemplarity objects. However, studied items rated as better exemplars of their basic-level concept were more likely to be forgotten. This suggests that better exemplars were more likely to trigger reactivation of basic-level gist information at test, and perhaps less likely to elicit retrieval of specific information. It is not consistent with a simple modulation of basic-level gist at encoding, which should increase true recognition as well as lure errors.

Our initial analysis of the effects of perceptual relations on memory used graded indices of confusability computed across the whole set of stimuli (see Supplemental Figure 4 and Table 7). However, since the graded weighting prioritized similarity over many items, the high confusability items had generic visual properties rather than being similar to any specific items in the set. These observations motivated the analysis using perceptual confusability measures with a stronger weight, restricted to nearest neighbors only. As Figure 1 shows, these metrics had a quite different profile from the across-set measures (compare to Supplemental Figure 4). For C1, more confusable items were those with a distinctive linear orientation, shared with at least one other specific item in the set (the neighbor), although more color-confusable items were relatively uniform as well as bland. The analysis using these measures revealed that lures that shared low-level (C1) visual features with their nearest neighbor were more likely to be misrecognized. However, the corresponding studied items were more likely to be forgotten. We consider these effects of nearest neighbor visual confusability further below, in light of the results of Experiment 2.

The data point to multiple gist-like effects on mnemonic discrimination reflecting both semantic and perceptual dimensions along which studied items and lures were similar to other studied objects. Here, the lures were different exemplars of studied basic-level concepts, so variables increasing emphasis on processing at this level tended to trigger errors, while variables indexing processing shared with other concepts tended to reduce them. We have proposed that these effects reflect strengthening or weakening of basic level conceptual memory representations. However, there is an alternative mechanism by which lures with more confusable concepts could be better discriminated: they might be more easily rejected using the memory editing strategy of recall-to-reject (Gallo, 2004; Brainerd et al., 2003). When a lure triggers recollection of similar studied items, people can avoid gist-related errors by comparing the recollected information with the lure (e.g., they decide that a white dog was not presented, because they remember having studied a black dog). A similar mechanism might also apply to differences in performance for lures that were related to studied items in other ways, at the item level. It was important to establish whether our results would generalize to a situation in which recall-to-reject was prevented.

2.7 Experiment 2

In Experiment 2 we aimed to replicate Experiment 1's novel semantic and perceptual effects on memory, and test whether these modulations of mnemonic discrimination would generalize to a task where the use of a recall-to-reject strategy was prevented. We therefore amended the procedure so that multiple different exemplars of each basic-level concept were studied. Once people have to recall more than 5 different studied items in order to reject a single lure, a recall-to-reject strategy becomes ineffective, particularly if the set size varies within the study list (Gallo, 2004). In Experiment 2, participants studied sets of either two or eight different exemplars of the same basic-level concept. Our preregistered prediction was that if Experiment 1's finding of reduced false recognition of lures with more confusable concepts was due to impoverished representations of these basic-level concepts in memory, Experiment 2 would show a similar effect. With more studied exemplars per concept we might also observe enhanced effects of concept confusability on mnemonic discrimination in Experiment 2 compared to Experiment 1, and for set size 8 compared to set size 2. However, if the recall-to-reject account is correct, concept confusability would either be associated with *increased* lure errors in Experiment 2, or would have no effect. In terms of item-level processing, we expected that effects might be similar, or more pronounced because of the larger studied sets.

2.8 Materials and Methods

The experimental methods were preregistered with the Open Science Framework (https://osf.io/3h7kf).

2.8.1 Participants

The study included sixty adults aged 18-33 years (*M* = 21.2; *SD* = 3.2, 12 male, 48 female). A further ten participants were excluded from data analysis: 9 due to technical issues with recording responses, 1 due to poor performance at test (using the preregistered criterion of *d'* for studied item discrimination of less than 3 *SD* from the mean). The sample size was determined with sensitivity analyses setting alpha to .05. The principal effect of interest was the overall effect of concept confusability on lure false recognition (interaction of concept confusability × lure versus novel baseline; collapsed across set size). Simulations with simR suggested that with *N* = 60 we would have .92 power to detect a small effect (*OR* = .70; equivalent to Cohen's *d* = .20; the effect size in Experiment 1 was equivalent to *d* = .15, although as outlined above, we expected this effect to be larger, if present). For the higher order interaction of concept confusability with study set size condition (lures for set size 8 versus 2), *N* = 60 could detect a medium-sized interaction of concept confusability ×

condition × lure versus novel baseline at .99 power (OR = .58; equivalent to Cohen's *d* = .30). Data from a group of older participants collected at the same time will be included in a separate report. All participants were fluent English speakers (since at least the age of 5), and had normal or corrected to-normal vision. Participants were recruited by local advertisement and provided informed consent. They received either course credit or an honorarium. The study was approved by the University of Edinburgh Psychology Research Ethics Committee (Ref. 278-1617/1).

2.8.2 Stimuli

Except where specified, stimuli were the same as in Experiment 1. For each of the 200 basic-level concepts, 9 different sets of images were obtained for a total of 1800 images. Each study list included 600 items: 480 in the large sized sets (i.e., set size 8) and 120 in the small sized sets (i.e., set size 2). Each test list consisted of 300 items: 120 studied images (60 from set size 8 and 60 from set size 2), 120 similar lures (60 unstudied exemplars of studied basic-level concepts from set size 8 and 60 from set size 2), and 60 novel items whose basic-level concepts had not been studied. Three filler trials prefaced both the study and the test phase. We generated one study and test list for each participant which randomized the allocation of the concepts and their exemplar images to conditions (i.e., studied, lure, and novel) and set size (2 and 8) with the constraint that half the concepts in each condition (item type and set size) were living and half non-living.

2.8.3 Procedure

The experiment consisted of a single study phase followed by a recognition test phase with interspersed standardized cognitive tests (not reported here) for 15 minutes. Stimuli were presented with MATLAB (R2018b, The MathWorks) using PsychToolbox (Kleiner et al., 2007; Version 2.0.14). The procedure was otherwise the same of Experiment 1 except that trials during the study phase were not self-paced, but presented every 3 s.

2.8.4 Statistical Analysis

The variables of interest and nuisance variables (see Supplemental Table 8) were identical to those used in Experiment 1, but the item metrics were recomputed for this larger set of images. Analyses reported here were based on the nearest neighbor metrics of perceptual confusability for reasons noted in Experiment 1, Materials and Methods, Variables of Interest, and the results for graded perceptual confusability metrics are given in the

Supplemental Material (Supplemental Table 11). Item exemplarity ratings were again collected from the participants after the test phase. As the concepts were identical to those used in Experiment 1, we used the same concept familiarity values collected previously. The main preregistered statistical analyses collapsed across the large and small study set sizes were identical to those used in Experiment 1, as were the model selection procedures. We also examined *a priori* the modulatory effects of set size (i.e., set size 8 vs set size 2), but as this variable had no significant effects we focus here on the results collapsed over the two set sizes (see Supplemental Table 10 for the results of the full model). Lastly, we again conducted a linear mixed-effect analysis of study phase RTs (results not reported, see https://osf.io/ndk83/).

2.9 Results

As in Experiment 1, there was decisive evidence in favor of the full model including concept-level, item-level, and confounds principal component variables, relative to simpler models (see Supplemental Table 9). The full model was the most parsimonious, receiving the lowest AICc score (AICc = 17597.96), and substantial weight (AICc weight = .89) relative to the other models, with an evidence ratio between the top-ranked model and the second-ranked model of 7.83. This result was also supported by a significant LRT in favor of the full relative to the model ranked second on the basis of AIC ($\chi2(7) = 18.16$, $p = .011$). Unlike Experiment 1, BIC now provided evidence in favor of the model with both concept- and-item level variables as the most parsimonious model (BIC = 17780.99). The results of the mixed effects analysis of the full model are shown below in Table 3, and illustrated in Figure 3. All the *p*-values are FDR-multiple comparison corrected (Benjamini & Hochberg, 1995). Raw recognition responses by item type are reported in Supplemental Figure 6.

Table 3

Results of Experiment 2 (Novel Items as Baseline)

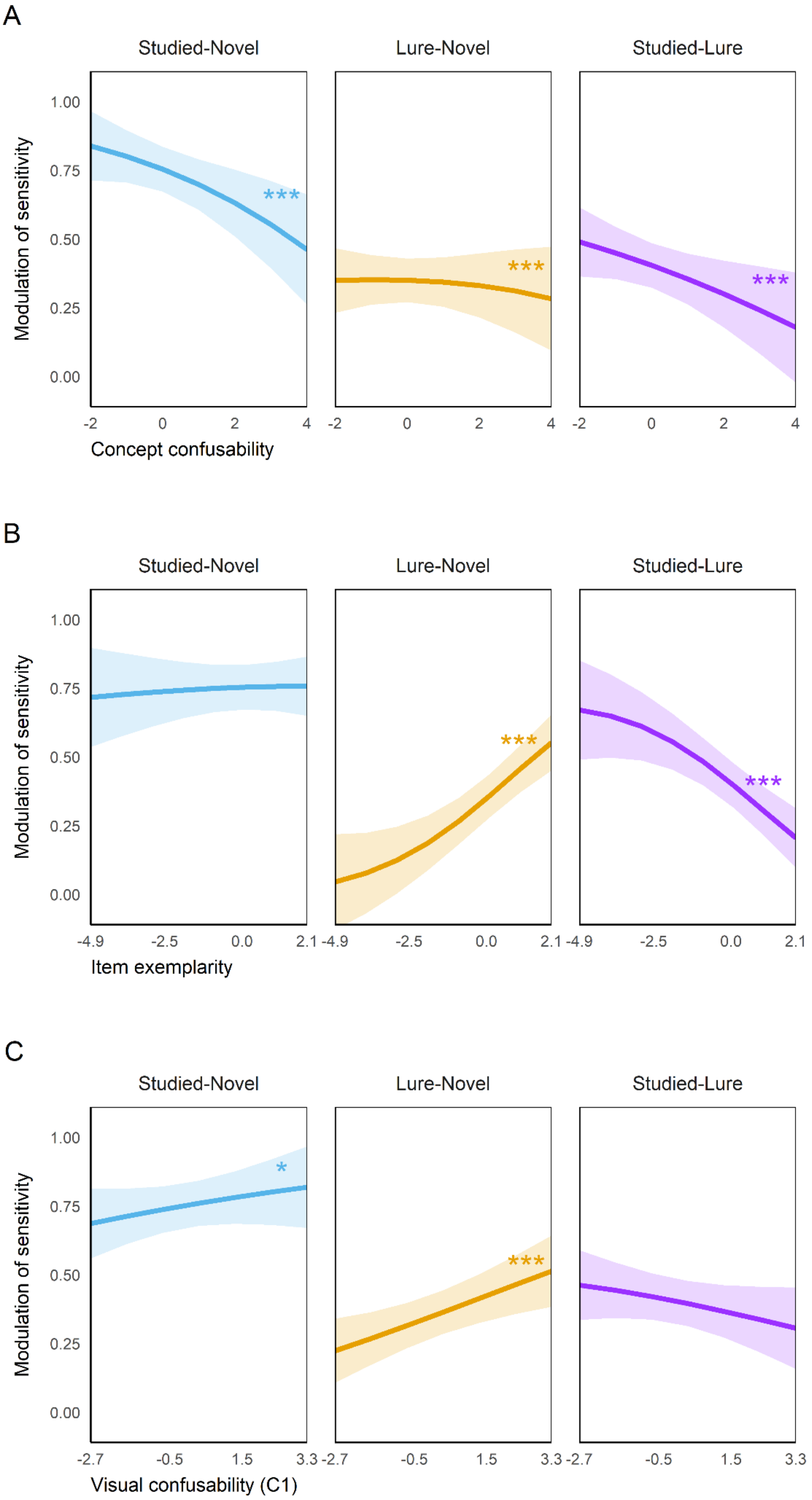
Variable	Estimate	*d’*	*SE*	*z*-value	*p*
(Intercept)	-2.70	-1.55	0.10	-26.65	<.001

Lure	2.34	1.33	0.07	32.00	<.001
Studied	4.20	2.44	0.08	54.42	<.001
Number of Features	0.00	0.00	0.07	-0.03	.978
Concept Confusability	0.31	0.16	0.07	4.30	<.001
Visual Confusability (C1)	-0.08	-0.03	0.07	-1.13	.364
Visual Confusability (C2)	0.22	0.10	0.08	2.83	.011
Color Confusability	0.07	0.03	0.07	1.07	.375
Item Exemplarity	0.14	0.08	0.07	2.11	.067
Lure × Number of Features	-0.01	0.00	0.07	-0.12	.949
Studied × Number of Features	-0.09	-0.05	0.07	-1.28	.301
Lure × Concept Confusability	-0.25	-0.12	0.07	-3.59	<.001
Studied × Concept Confusability	-0.53	-0.29	0.07	-7.32	<.001
Lure × Visual Confusability (C1)	0.26	0.14	0.07	3.66	<.001
Studied × Visual Confusability (C1)	0.20	0.10	0.07	2.66	.016
Lure × Visual Confusability (C2)	-0.16	-0.06	0.08	-1.95	.082
Studied × Visual Confusability (C2)	-0.17	-0.07	0.08	-2.05	.071
Lure × Color Confusability	0.05	0.04	0.07	0.70	.536
Studied × Color Confusability	-0.07	-0.03	0.07	-0.95	.422
Lure × Item Exemplarity	0.29	0.18	0.07	4.03	<.001
Studied × Item Exemplarity	-0.06	-0.03	0.07	-0.82	.483

Note. The reference level of condition is set to “novel”. Parameter estimates (logOR), *d’* equivalent, standard errors, *z*-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Nearest neighbor perceptual confusability measures were included in the model above. See Material and Methods, Variables of Interest, and Results for details. *SE* = Standard Error

Figure 3

Effects of Semantic and Perceptual Variables on Mnemonic Sensitivity in Experiment 2



Note. Plots show effects of semantic and perceptual variables on modulation of sensitivity. The plot lines represent the effect of the predictor variables on the probabilities of endorsing studied items as "old" relative to novel items (light blue), lures as "old" relative to novel items (orange), and studied items as "old" relative to lures (purple). Panel A, B, and C show the effects of concept confusability, item exemplarity, and C1 visual confusability in Experiment 2. See Material and Methods, Variables of Interest, and Results for details. * $p < .05$; ** $p < .01$; *** $p < .001$ (FDR-corrected).

2.9.1 Sensitivity for Studied Relative to Novel Items

As in Experiment 1, discrimination of studied from novel items was very good (simple effect on *d'* = 2.44; 95% CI [2.36, 2.51]). Concept confusability impaired sensitivity for studied items which were more likely to be forgotten (Figure 3A; interaction of concept confusability with studied items on *d'* = -.29; 95% CI [-.36, -.21]). However, the effect of item exemplarity on true recognition was no longer significant (Figure 3B; interaction of item exemplarity with studied items on *d'* = -.03; 95% CI [-.10, .05]). Also, more visually similar studied items were now *more* (rather than less) likely to be correctly recognized (Figure 3C; interaction of C1 with studied items on *d'* = .10; 95% CI [.03, .18]). No other variable significantly modulated sensitivity of studied items.

2.9.2 Sensitivity for Lure Relative to Novel Items

As in Experiment 1, participants were generally more likely to judge lures than novel items as "old" (simple effect on *d'* = 1.33; 95% CI [1.26, 1.41]). The direction of the concept confusability effect was also unchanged: false recognition of lures was again relatively *less* likely for pictures whose concepts shared many semantic features with other concepts (Figure 3A; interaction of concept confusability with lure on *d'* = -.12; 95% CI [-.20, -.05]). At the item level, lure errors were again more frequent for items with higher rated exemplarity (Figure 3B; interaction of item exemplarity with lure on *d'* = .18; 95% CI [.11, .26]), and with higher visual confusability in terms of low-level visual representations (Figure 3C; interaction of C1 with lure on *d'* = .14; 95% CI [.07, .22]). No other variable significantly modulated sensitivity for lure items.

2.9.3 Sensitivity for Studied Relative to Lure Items

The results of the analysis that examined participants' ability to discriminate studied items from lures are shown below in Table 4, and illustrated in Figure 3. Overall, participants were fairly good at discriminating studied items from similar lures (simple effect on *d'* = 1.10; 95% CI [1.06, 1.15]). Both concept- and-item level variables modulated this effect. Similar to Experiment 1, studied items with high concept confusability were less likely to be correctly discriminated from highly confusable lures (Figure 3A; interaction of concept confusability with studied items on *d'* = -.16; 95% CI [-.21, -.12]). However, unlike Experiment 1, more semantic features did not improve discrimination (interaction of number of features with studied items on *d'* = -.05; 95% CI [-.10, .00]). Studied concepts whose pictures were judged to have high exemplarity were again less likely to be discriminated from their lure exemplars (Figure 3B; interaction of item exemplarity with studied items on *d'* = -.21; 95% CI [-.26, -.16]). Lastly, participants were less likely to correctly discriminate studied items with high color confusability (interaction of color confusability with studied items on *d'* = -.07; 95% CI [-.12, -.02]). Unlike Experiment 1, the effect of C1 visual confusability was no longer significant (Figure 3C; interaction of C1 with studied items on *d'* = -.04; 95% CI [-.09, .01]). No other variable significantly modulated sensitivity for studied relative to lure items.

Table 4

Results of Experiment 2 (Lures as Baseline)

Variable	Estimate	*d'*	SE	z-value	*p*
(Intercept)	-0.35	-0.21	0.08	-4.45	<.001
Studied	1.85	1.10	0.04	44.54	<.001
Studied × Number of Features	-0.09	-0.05	0.04	-1.91	.079
Studied × Concept Confusability	-0.28	-0.16	0.04	-6.65	<.001
Studied × Visual Confusability (C1)	-0.06	-0.04	0.04	-1.47	.186
Studied × Visual Confusability (C2)	-0.01	-0.01	0.04	-0.25	.847
Studied × Color Confusability	-0.12	-0.07	0.04	-2.89	.007

Studied × Item Exemplarity	-0.35	-0.21	0.04	-8.32	<.001

Note. The reference level of condition is set to “lure”. Parameter estimates (logOR), *d’* equivalent, standard errors, z-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Nearest neighbor perceptual confusability measures were reported in the model above. See Material and Methods, Variables of Interest, and Results for details. SE = Standard Error.

2.9.4 False Alarms to Novel Items

Participants correctly identified most unstudied novel items as “new” (intercept on *d’* = -1.55; 95% CI [-1.65, -1.44]). As predicted, and as in Experiment 1, items with high concept confusability were more likely to be falsely recognized as “old” (Supplemental Figure 6A; simple effect of concept confusability on *d’* = .16; 95% CI [.09, .24]). At the item level, false alarms to novel items were modulated by C2 visual confusability. People were more likely to misrecognize novel items with more confusable late visual representations (i.e., their overall shape) (simple effect of C2 on *d’* = .10; 95% CI [.02, .17]). No other variable significantly modulated baseline false recognition of novel items.

2.10 Discussion

Experiment 2 closely reproduced Experiment 1’s procedures, except that participants studied multiple exemplar images of each basic-level concept. This allowed us to test whether the variables contributing to mnemonic discrimination were altered when people could not effectively use a recall-to-reject strategy. The effect observed in Experiment 1 for concept confusability was qualitatively the same in Experiment 2: people were again less likely to misrecognize lures with high concept confusability (reductions in *d’* by factors of -.15 and -.12 in Experiments 1 and 2), and more likely to forget more confusable studied items (*d’* reductions of -.29 and -.29 in Experiments 1 and 2). Therefore, the pattern of findings in Experiment 1 cannot be explained by a facilitation of recall-to-reject for lures with more confusable concepts. Instead, the data suggest that processing shared features over concepts weakened the representation of basic-level conceptual information in memory. Thus, concept confusability reduced recognition of both studied and lure exemplars and impeded the ability to discriminate between them in memory.

As in Experiment 1, semantic relations over items also had a marked effect on lure errors, which were again more frequent for lures that were better exemplars of their concept (*d'* increasing by factors of .13 and .18 in Experiments 1 and 2). This suggests that at test, people were more likely to respond “old” to high exemplarity lures because they remembered having studied the corresponding basic-level concepts. However here, unlike in Experiment 1, item exemplarity did not significantly affect true recognition of studied items (*d'* effects of -.10 and -.03 in Experiments 1 and 2). We do not place much emphasis on this null finding, since the between-Experiment interaction between study set size, item exemplarity and old versus novel items was also not significant (see Supplemental Table 12). It may be that there is an effect on true recognition which is too small for us to detect consistently. Alternatively, the effect observed in Experiment 1 may have been weakened by the increase in set size: for example, exemplarity of individual studied items might matter less when multiple exemplars are studied. Despite this, the effect for lures remained robust, as did the effect on overall discrimination between studied items and lures.

For the perceptual item-level variables, as already noted in Experiment 1, the graded measures did not seem to capture our initial intuition that some lures would be highly visually confusable with specific studied items. Therefore, we focused instead on exploratory analyses that better tested our original prediction that such lures would be more frequently misrecognized. The nearest neighbor metrics yielded consistent findings in the two experiments in this regard. For the low-level C1 measure, lures that were more visually confusable with another item were more likely to be misrecognized (*d'* increased by factors of .10 and .14 for Experiments 1 and 2). However, the effects of C1 visual confusability on recognition of studied items differed between the two experiments. While in Experiment 1 more confusable items were less likely to be remembered, in Experiment 2 they were *more* often remembered (*d'* modulations of -.10 and .10). This reversal may reflect a genuine difference due to the increase in study set size in Experiment 2: for example, if more highly similar nearest neighbors were introduced by use of multiple studied exemplars of each basic-level concept. Further data will be required to establish whether this is a robust finding. As a result of this different effect for studied items, visual confusability did not significantly reduce overall mnemonic discrimination of studied items from lures in Experiment 2, unlike for Experiment 1. The current results point to a particular salience of simple visual features like line orientation for false recognition of lures (see Figure 1). We explore the possible reasons for this below.

2.11 General Discussion

In this research, we used objective and model-based measures to show for the first time that multiple semantic and perceptual relations contribute to people's ability to discriminate objects in memory. This approach allowed us to assess simultaneous influences of semantic and perceptual relations between objects while controlling for the potential effects of other variables known to influence mnemonic discrimination. Using generalized linear mixed model analysis we were also able to directly model the effects of predictors on binary memory outcomes, and generalize the results over concepts as well as participants (DeCarlo, 1998; Wright et al., 2009). In both experiments, studied objects that shared semantic features with many other concepts were more difficult to recognize, and lures whose corresponding studied object had many shared features were easier to correctly identify as new. In contrast, misrecognition was more frequent for lure objects that were more representative of their basic-level concept. The findings demonstrate distinct effects of semantic relations among concepts, and semantic relations between concepts and their exemplars. Simple perceptual properties shared with other studied images also contributed to misrecognition of lures, suggesting that image as well as concept properties are important in mnemonic discrimination for pictures.

To assess the effects of semantic similarity on memory we used a concept confusability metric derived from a feature-based model of semantic memory (Devereux et al., 2014; Tyler & Moss, 2001). Previous studies have shown that conceptual structure understood in terms of feature relations between concepts can explain a range of phenomena, such as differences in the processing of living and nonliving concepts in healthy people (McRae et al., 1997; Randall et al., 2004; Taylor et al., 2011; Vigliocco et al., 2004) and specific impairments in neuropsychological patients (Forde & Humphreys, 1999; Humphreys & Forde, 2001; Warrington & Shallice, 1984). For example, Moss et al. (1997) described a post-encephalitic patient who was very poor at differentiating between highly similar objects (e.g., tiger versus panther), but had no difficulty in determining the superordinate category of an object (e.g., land animals; see also Tyler et al., 2004). This dissociation between finer-grained and coarser categorical levels of conceptual processing converges with Taylor et al.'s (2012) finding that, in healthy people, processing highly shared semantic features facilitated domain-level categorization decisions but impeded basic-level naming.

The current study is the first to show that feature-based conceptual structure impacts mnemonic discrimination. The results converge with earlier evidence from Montefinese et al. (2015) that semantic feature similarity increases false alarms to unstudied items (see also Montefinese et al., 2018). Our data confirm that overlapping semantic features impact mnemonic sensitivity indices that adjust false recognition for response criterion, and further show that concept confusability also affects true recognition. The results are also in line with Coane et al. (2016)’s suggestion that specific effects of semantic categorical relations on false recognition reflect feature similarity at least in part (see also Coane et al., 2020). We have interpreted this finding in terms of gist memory traces representing basic-level conceptual information. According to fuzzy trace theory, memory outcomes reflect the relative accessibility of two kinds of memory traces encoded in parallel: verbatim traces containing specific representations of the studied items, and gist traces representing their meaning traces (Brainerd & Reyna, 2002). Memory for studied items is supported by both verbatim and gist traces, and false memory for related lures occurs when gist but not verbatim traces are retrieved. Although the theory does not specify the informational content of gist, the logic of the mnemonic discrimination task suggests that the relevant gist information in any given task is the information that is shared between the studied items and lures. Our results support this assumption, and suggest that different levels of semantic similarity have distinct effects on mnemonic discrimination. In the categorized pictures task used here, studied items and lures shared basic-level concepts. Concept confusability – a similarity metric reflecting shared processing of conceptual features – weakened the representations of individual concepts in memory and therefore reduced gist-like effects on memory for studied items and lures. The net effect on mnemonic discrimination was also negative: for concepts that were more confusable, people were less able to discriminate in memory between studied items and lures.

In contrast, item exemplarity – similarity between an exemplar item and its concept – strengthened activation of the concept and therefore increased gist-like effects, at least for lures. Exemplarity, or image agreement, defines a relation between an image exemplar and its basic-level concept, and is closely related to typicality (Barry et al., 1997; Snodgrass & Vanderwart, 1980). A typical exemplar shares more features with other members of its category (Rosch & Mervis, 1975) and may therefore more strongly elicit a gist memory trace that overlaps with representations of other exemplars. In line with this, we found that people falsely recognized more high exemplarity basic-level lures. Together with the findings for concept confusability, the data suggest that both concept-level and item-level variables impacted mnemonic discrimination via modulations of basic-level memory representations. A few prior studies have examined the specific semantic information that gist is based on. These have used specific stimuli or measures that focused on a single type of relation (Cann et al., 2011; Coane et al., 2016; Montefinese et al., 2015). Our use of multiple measures enabled effects of relations at different levels to be examined at the same time.

As outlined in the Introduction, studies using a range of different task materials also suggest that the content of the memory representations critical for mnemonic discrimination varies. Studies using pictures, words, situational themes, and narratives as to-be-remembered material find that lures that are related to studied items at these different semantic levels are misrecognized as studied, implicating gist representations at different levels. The results for novel items offer converging evidence that the effects of different semantic relations on lure errors may depend on the nature of the lures to be discriminated. In both experiments, while concept confusability improved rejection of lures, it increased false alarms to unrelated novel images. Although neither these objects (e.g., a panther) nor other exemplars of their basic-level concepts had been studied (e.g., a different panther), people had studied other items from the same superordinate categories (e.g., a cat and a dog) as well as further items with which they shared semantic features (e.g. a chair, which < has legs >). Since no basic-level gist memory was ever encoded for these items' concepts, its effects could not be impoverished by emphasis on features shared with other concepts. However, if the coarse semantic activation not only reduced within-concept gist but enhanced across-concept gist, it would lead some novel objects with many shared features to be classified as "old". A similar argument was made by Montefinese et al. (2015) for verbal material. Our finding is also in line with previous studies showing that novel pictures (Bowman, Chamberlain, & Dennis, 2019; Seamon et al., 2000) and words (Brainerd et al., 1995; Coane et al., 2016, 2020) belonging to the same superordinate category as studied items are more likely to be misrecognized than unrelated novel items. The current study and that of Montefinese et al. (2015) are the first to directly demonstrate that shared semantic features can increase false recognition of pictures and words, respectively. However, both these findings could be due to modulations of response criterion rather than effects on memory (see also Montefinese et al., 2018). People’s bias to respond “old” versus “new” can change item-by-item in response to properties of the test probes (Heit et al., 2003; Kent et al., 2018). For example, processing shared semantic features might increase attention to items in the absence of any retrieval of a gist trace. Further studies with an additional unrelated novel item baseline condition will be required to adjudicate between these two possibilities.

We have discussed the findings for both concept confusability and item exemplarity in terms of their effects on concept gist shared by studied items and lures. However, unlike concept confusability, item exemplarity did not affect memory for studied items as would be expected according to fuzzy trace theory if it modulated encoding of gist traces. Recognition of higher exemplarity studied items was numerically *reduced* in both experiments, significantly so in Experiment 1. The effects of exemplarity or typicality on recognition memory are relatively unexplored, but several studies suggest that words that are more typical members of their superordinate category tend to be better recalled, mediated in part by clustering by category (Schmidt, 1996). One possibility is that the restriction of the effect of exemplarity to lure errors derived from the way that exemplarity as an item-level variable assessed gist. Unlike concept confusability, which directly indexed shared features among concepts, item exemplarity indirectly indexed shared features among exemplars via judgments of the strength of the relations between exemplars and their concepts. Prioritization of relations between exemplars and concepts may have increased sensitivity to test phase cueing of memory traces by images eliciting strong basic-level representations. Greater overlap between an exemplar and its basic-level concept may therefore have increased the degree to which a lure image could trigger retrieval of a gist trace, but not the degree to which a studied image could elicit gist encoding at this level. Currently, norms are not available for the semantic features that are shared between individual pictured exemplars of a concept but, based on the assumptions of fuzzy trace theory, we would predict that similarity between studied items and lures on such measures would be associated with increases in recognition of both types of items, and reduced discrimination between the two.

The current data also support the proposal that mnemonic discrimination errors are driven by perceptual as well as semantic similarity (Bowman, Chamberlain, & Dennis, 2019; Brady et al., 2008; Burnside et al., 2017; Koutstaal & Schacter, 1997; Motley & Kirwan, 2012; Seamon et al., 2000). We examined the effects of perceptual relations on memory using item-level measures of visual properties of the images. Lure items with greater low-level (i.e. C1) visual confusability with another image in the set were more likely to be misrecognized as "old". As these nearest neighbor analyses were exploratory, the results need to be treated with some caution, but the metrics yielded consistent results across the two experiments in the direction originally predicted for visual confusability effects. The findings suggest that perceptual effects on false recognition of lures are robust, at least for early visual properties like line orientation. The contrasting lack of consistent effects of perceptual confusability measured with graded metrics points to a dependence of this effect on specific similarity between lures and individual items in the set. We considered whether defining measures of perceptual (and semantic) similarity only between test items and each participant's set of studied items might provide a clearer picture; however, this was not possible due to very high collinearity with the original indices (Pearson's r between .80 and .98 in Experiments 1 and 2).

Our data converge with the handful of previous studies that have directly addressed perceptual effects. It is difficult to draw strong conclusions about color effects since we did not find consistent effects over experiments, and Reppa et al. (2020) (see Introduction) used a very different task and measures from ours. However, the current results are broadly in line with their finding that perceptual similarity (in their case, of shape) can cause interference in memory. Our data also converge with those of Brady et al. (2008) who found that people made errors to lures that were different-view images of studied objects, as well as to different exemplars of studied concepts. They used a forced choice recognition task, in which discrimination performance is generally substantially better than in old/new recognition, and may rely more on familiarity (Migo et al., 2009) and/or processing fluency (Voss et al., 2012). Our findings (and those of Motley and Kirwan, (2012); see Introduction) broadly converge with these earlier data to suggest that perceptual as well as semantic similarity both influence mnemonic discrimination regardless of the recognition task format. Our model-based measures specifically implicate low-level visual similarity in this task, but future studies may show that higher-level attributes like shape and view also contribute independently.

Others have found that perceptually similar but pre-experimentally meaningless lures trigger false recognition (Koutstaal et al., 2003; Pidgeon & Morcom, 2014; Slotnick & Schacter, 2004). For example, Slotnick and Schacter (2004) found that meaningless shapes that were visually related to studied items were more likely to be misrecognized relative to novel unrelated shapes. While visual similarity was not formally measured, lures tended to have similar linear orientation to their studied items. Our data are also in line with fMRI studies that have shown engagement of early visual cortex during false recognition of visually similar picture lures, whether or not these are also semantically related, supporting the interpretation that such activity represents enhanced visual processing during lure misrecognition (Bowman et al., 2019; Garoff-Eaton et al., 2006; Gutchess & Schacter, 2012; Slotnick & Schacter, 2004). Thus, behavioral and neuroimaging results converge on the idea that mnemonic discrimination errors can be elicited by both semantic and perceptual relations. The use of model-based metrics allowed us to go further by indexing perceptual properties directly, and addressing concerns that people may bring conceptual processing to bear even on experimentally unfamiliar stimuli (Pidgeon & Morcom, 2016).

These findings support earlier suggestions that gist-based memory can be perceptual as well as conceptual (Koutstaal & Schacter, 1997). This possibility is consistent with fuzzy trace theory’s opponent processes, if gist is not restricted to information about meaning. A fuller understanding of mnemonic discrimination needs to take into account the nature of the relations and representations involved. Our finding that semantic relations between concepts and items had different effects on mnemonic discrimination suggests that different semantic relations may impact memory in ways that are modulated by task demands. An important question to address in future studies will be whether manipulations previously shown to impact measures of gist reliance (Brainerd & Reyna, 2005) can be shown to have dissociable effects on the contributions of semantic and perceptual similarity to mnemonic discrimination. Our results are also in line with previous data suggesting that semantic influences on memory errors go beyond associative activation (Brainerd et al., 2008; Cann et al., 2011; Coane et al., 2016, 2020) The activation/monitoring theory explains false recognition in terms of associative strength (Roediger & McDermott, 2000; Roediger et al., 2001). On this view, studying a list of words produces an automatic associative activation that spreads through the lexical semantic system, and false recognition occur when lures accrue substantial activation through this process. A strength of the activation/monitoring theory is that the associative relations assumed to determine memory errors are quantifiable (Roediger et al., 2001). However, this account cannot explain the above findings. It is also inconsistent with perceptual effects on lure errors as it specifies that misrecognition stems from activation of lure representations at study. Similar lures which have no pre-existing associations to studied items are unlikely to be spontaneously generated at encoding (Arndt, 2010). Associative activation also cannot explain lure errors reflecting spatial proximity to the location of a studied object (Reagh et al., 2016, 2014).

Global-matching models can explain lure errors that reflect perceptual as well as semantic relations (Arndt & Hirshman, 1998; Arndt, 2010; Hintzman, 1988). According to models like MINERVA2 (Arndt & Hirshman, 1998), there is no gist memory trace, but mnemonic discrimination errors occur to the degree that lures presented at test are globally similar to multiple traces previously stored at encoding, without specification of the nature of the similar features. Such a retrieval-based mechanism may also explain why some variables – here, item exemplarity and visual confusability – impact lure errors but not memory for studied items. In fuzzy-trace theory's conjoint recognition model, lure-specific factors could modulate the probability of similarity responding at test, and therefore lure errors (Brainerd et al., 1999; Brainerd & Wright, 2005), although its processing tree model does not explicitly separate encoding and retrieval processes. An explicit model of encoding and retrieval contributions is offered by the pattern separation and completion processes at the heart of the complementary learning systems model (Marr, 1971; McClelland et al., 1995; McNaughton & Morris, 1987; Norman & O'Reilly, 2003). At study, pattern separation by the hippocampus ensures that specific memory traces are generated. When this separation fails memory traces may become more gist-like, and lures more likely to trigger errors (Wilson et al., 2006; Yassa & Stark, 2011). At test, people may also fail to discriminate lures because the lures trigger pattern completion due to high overlap with studied items (Motley & Kirwan, 2012; Norman

& O’Reilly, 2003; Norman, 2010; Yotsumoto et al., 2007). Hippocampal pattern separation and completion do not specify the types of similarity more likely to influence mnemonic discrimination, and predict influences of perceptual, semantic and contextual properties (Hunsaker & Kesner, 2013; Reagh et al., 2014; Yassa & Stark, 2011). The complementary learning systems model further specifies non-pattern-separated neocortical inputs that may also contribute to semantically-driven mnemonic discrimination errors (Norman & O’Reilly, 2003; Pidgeon & Morcom, 2014; Wilson et al., 2006). These model predictions about specific encoding and retrieval operations are difficult to test with behavioral measures alone, but can be more directly investigated using neuroimaging measures, which enable semantic and perceptual modulations of neural processing during study and test phases to be assessed.

2.12 Conclusions

In this work we used objective measures derived from established models of conceptual structure and low-level vision to show that semantic and perceptual relations can simultaneously contribute to gist-like effects in memory. The results from two experiments implicated relations at multiple representational levels and suggested that similarity between studied and unstudied items does not always impair mnemonic discrimination. The coarse semantic activation elicited by processing shared semantic features across concepts impeded memory for studied objects while reducing false alarms to lure exemplars of the same concepts. In contrast, a strong semantic overlap at the item-level between an object and its basic-level concept was associated with more frequent false recognition of lures, as was strong low-level visual similarity between an object and a studied image. The initial findings were replicated in the second experiment which increased and varied the study set size in order to rule out recall-to-reject as an explanation for the results. Taken together, our findings point to the utility of a more structured and formal approach to understanding the relations underpinning gist-like effects in memory, and highlight the importance of image as well as concept properties in mnemonic discrimination.

Chapter 3: Perceptual and Semantic Representations at Encoding Contribute to True and False Recognition of Objects

When encoding new episodic memories, visual and semantic processing are proposed to make distinct contributions to accurate memory and memory distortions. Here, we used functional magnetic resonance imaging (fMRI) and preregistered representational similarity analysis (RSA) to uncover the representations that predict true and false recognition of unfamiliar objects. Two semantic models captured coarse-grained taxonomic categories and specific object features, respectively, while two perceptual models embodied low-level visual properties. Twenty-eight female and male participants encoded images of objects during fMRI scanning, and later had to discriminate studied objects from similar lures and novel objects in a recognition memory test. Both perceptual and semantic models predicted true memory. When studied objects were later identified correctly, neural patterns corresponded to low-level visual representations of these object images in the early visual cortex, lingual, and fusiform gyri. In a similar fashion, alignment of neural patterns with fine-grained semantic feature representations in the fusiform gyrus also predicted true recognition. However, emphasis on coarser taxonomic representations predicted forgetting more anteriorly in the anterior ventral temporal cortex, left inferior frontal gyrus and, in an exploratory analysis, left perirhinal cortex. In contrast, false recognition of similar lure objects was associated with weaker visual analysis posteriorly in early visual and left occipitotemporal cortex. The results implicate multiple perceptual and semantic representations in successful memory encoding and suggest that fine-grained semantic as well as visual analysis contributes to accurate later recognition, while processing visual image detail is critical for avoiding false recognition errors.

3.1 Significance Statement

People are able to store detailed memories of many similar objects. We offer new insights into the encoding of these specific memories by combining fMRI with explicit models of how image properties and object knowledge are represented in the brain. When people processed fine-grained visual properties in occipital and posterior temporal cortex, they were more likely to recognize the objects later, and less likely to falsely recognize similar objects. In contrast, while object-specific feature representations in fusiform gyrus predicted accurate memory, coarse-grained categorical representations in frontal and temporal regions predicted forgetting. The data provide the first direct tests of theoretical assumptions about encoding true and false memories, suggesting that semantic representations contribute to specific memories as well as errors.

3.2 Introduction

Humans are able to remember objects in great detail and discriminate them in memory from others that are similar in appearance and type (Standing, 1973). To achieve this, highly specific memories must be encoded. Successful object encoding engages diverse cortical regions alongside the hippocampus (Kim, 2011). These areas intersect with networks involved in visual object processing and semantic cognition (Binder et al., 2009; Clarke and Tyler, 2014). However, little is known about the neural operations these regions support during encoding. According to fuzzy-trace theory, the specific memory traces that contribute to true recognition depend on encoding of perceptual features, while semantic gist representations promote both true and false recognition (Brainerd and Reyna, 1990). However, recent data suggest that perceptual relations between studied items and lures can also trigger false recognition (see Chapter 2; Naspi et al., 2020). Here, we used functional magnetic resonance imaging (fMRI) and representational similarity analysis (RSA) to investigate the perceptual and semantic representations engaged that allow people to recognize these same objects later among perceptually and semantically similar lures.

In line with fuzzy-trace theory, a few fMRI studies have shown stronger activation in occipito-temporal regions when people later successfully recognize specific studied objects than when they misrecognize similar lures (Garoff et al., 2005; Gonsalves et al., 2004; Okado and Stark, 2005). However, activation of similar posterior areas has also been associated with later false recognition (Garoff et al., 2005), and activation in left inferior frontal gyrus – a region typically associated with semantic processing – with later true recognition (Pidgeon and Morcom, 2016). Such results appear to challenge any simple mapping between perceptual and semantic processing and true and false recognition (see also Chapter 2; Naspi et al., 2020). However, one cannot infer type of processing based on presence or absence of activation alone. Here, we investigated the underlying processes that give rise to such effects, using RSA to test whether patterns of neural similarity that indicate visual and semantic processing predict subsequent memory performance.

Object recognition involves visual analysis and the computation of meaning, proceeding in an informational gradient along the ventral visual pathway (Clarke and Tyler, 2015). The coarse semantic identity of an object emerges gradually from vision in posterior cortices including lingual, fusiform, parahippocampal, and inferior temporal gyri that integrate semantic features capturing taxonomic relationships (Devereux et al., 2013; Mahon et al., 2009; Tyler et al., 2013). The lingual and fusiform gyri in particular are also engaged when memories of objects are encoded (Kim, 2011). At the apex of the ventral pathway, the perirhinal cortex provides the finer-grained feature integration required to differentiate similar objects (Clarke and Tyler, 2014; Devlin and Price, 2007; Winters and Bussey, 2005), and activation here predicts later memory for specific objects (Chen et al., 2019). Other researchers ascribe this role more broadly to the anterior ventral temporal cortex, considered a semantic hub that integrates modality-specific features into transmodal conceptual representations (Lambon Ralph et al., 2017). Beyond the ventral stream, left inferolateral prefrontal regions supporting controlled, selective semantic processing are also critical for memory encoding (Gabrieli et al., 1998; Kim, 2011).

According to theory, the perceptual and semantic representations encoded in memory traces reflect how items were originally processed (Craik and Lockhart, 1972; Otten and Rugg, 2001). We therefore expected that some of these ventral pathway and inferior frontal representations would be revealed in distinct distributed activity patterns giving rise to later true and false recognition. We quantified perceptual representations in terms of low-level visual attributes of object images, and semantic representations of the objects' concepts in terms of their coarse taxonomic category membership as well as their specific semantic features. We used these models to identify representational similarity patterns between objects at encoding using a novel approach that combined RSA and the subsequent memory paradigm in a single step. This allowed us to test where the strength of perceptual and semantic object representations predicts subsequent accurate memory and false recognition of similar lures.

3.3 Materials and Methods

3.3.1 Participants

Twenty-eight right-handed adults aged 18-35 years underwent fMRI scanning (*M* = 23.07 years, *SD* = 3.54; 18 females, 10 males). Data from a further 4 participants were excluded due to technical failures. All participants also spoke English fluently (i.e., had spoken English since the age of 5 or lived in an English-speaking country for at least 10 years) and had normal or corrected-to-normal vision. Exclusion criteria were a history of a serious systemic psychiatric, medical or a neurological condition, visual issues precluding good visibility of the task in the scanner, and standard MRI exclusion criteria (see https://osf.io/ypmdj for preregistered criteria). Participants were compensated financially. They were contacted by local advertisement and provided informed consent. The study was approved by the University of Edinburgh Psychology Research Ethics Committee (Ref. 116-1819/1). All the following procedures were preregistered unless otherwise specified.

3.3.2 Stimuli

Stimuli were pictures of objects corresponding to 491 of the 638 basic-level concepts in The Centre for Speech, Language and the Brain concept property norms (the CSLB norms; Devereux et al., 2014). These were members of 24 superordinate categories (*Appliance*, *Bird*, *Body Part*, *Clothing*, *Container*, *Drink*, *Fish*, *Flower*, *Food*, *Fruit*, *Furniture*, *Invertebrate*, *Kitchenware*, *Land Animal*, *Miscellaneous*, *Music*, *Sea Creature*, *Tool*, *Toy*, *Tree*, *Vegetable*, *Vehicle*, *Water Vehicle*, *Weapon*), and 238 were living things and 253 non-living things. We sourced two images for each basic-level concept. Of the 982 images, 180 were a subset of the images used by Clarke and Tyler (2014), 180 were compiled from the Bank of Standardized Stimuli (BOSS; Brodeur et al., 2014) and the remaining 622 were taken from the Internet. Each study list included single exemplar images of either 328 or 327 concepts. Of these, half were subsequently tested as old and half were subsequently tested test as lures. Each test list consisted of 491 items: 164 (or 163) studied images, 164 (or 163) similar lures (i.e., images of different exemplars of studied basic-level concepts), and 163 (or 164) novel items (i.e., images of basic-level concepts that had not been studied). Three filler trials prefaced the test phase. For each participant, living and non-living concepts were randomly allocated to the conditions with equal probability, i.e., to be studied/lure or novel items. Each study and test list was presented in a unique random trial order.

3.3.3 Procedure

The experiment comprised a scanned encoding phase followed by a recognition test phase outside the scanner. Stimuli were presented using MATLAB 2019b (The MathWorks Inc., 2019) and PsychToolbox (Version 2.0.14; Kleiner et al., 2007). In the scanner, stimuli were viewed on a back-projection screen via a mirror attached to the head coil. Earplugs were employed to reduce scanner noise, and head motion was minimized using foam pads. During the study phase participants viewed one image at a time, and they were asked to judge whether the name of each object started with a consonant or with a vowel, responding with either index finger via handheld fiber-optic response triggers. By requiring participants to retrieve the object names, we ensured that they processed the stimuli at both visual and semantic levels. Participants were not informed of a later memory test. Images were presented centrally against a white background for 500 ms. This was followed by a black fixation cross with duration sampled from integer values of 2 to 10 s with a flat distribution, and then a red fixation cross of 500 ms prior to the next trial, for a stimulus onset asynchrony (SOA) of 3-11 s ($M = 6$). At test, participants viewed one image at a time for 3 s followed by a black fixation cross for 500 ms, and they judged each picture as “old” or “new” indicating at the same time whether this judgment was accompanied by high or low confidence using one of 4 responses on a computer keyboard. Mappings of responses to hands were counterbalanced at both encoding and retrieval.

3.3.4 fMRI acquisition

Images were acquired with a Siemens Magnetom Skyra 3T scanner at the Queen’s Medical Research Centre (QMRI) at the Royal Infirmary of Edinburgh. T2*-weighted functional images were collected by acquiring multiple echo-time sequences for each echo-planar functional volume (repetition time (TR) = 1700 ms, echo time (TE) = 13 ms (echo-1), 31 (echo-2) ms, and 49 ms (echo-3)). Functional data were collected over 4 scanner runs of 360 volumes, each containing 46 slices (interleaved acquisition; 80 × 80 matrix; 3 mm × 3 mm × 3 mm, flip angle = 73°). Each functional session lasted ~ 10 min. Before functional scanning, high-resolution T1-weighted structural images were collected with TR = 2620 ms, TE = 4.9 ms, a 24-cm field of view (FOV), and a slice thickness of 0.8-mm. Two field map magnitude images (TE = 4.92 ms and 7.38 ms) and a phase difference image were collected after the 2nd functional run. At the end, T2-weighted structural images were also obtained (TR= 6200 ms and TE = 120 ms).

3.3.5 Image preprocessing

Except where stated, image processing followed procedures preregistered at https://osf.io/ypmdj and was conducted in SPM 12 (v7487) in MATLAB 2019b. The raw fMRI time series were first checked to detect artefact volumes that were associated with high motion or were statistical outliers (e.g. due to scanner spikes). We checked head motion per run using an initial realignment step, classifying volumes as artefacts if their absolute motion was > 3 mm or 3 deg, or between-scan relative motion > 2 mm or 2 deg. Outlier scans were then defined as those with normalized mean or standard deviation (of absolute values or differences between scans) > 7 *SD* from the mean for the run. Volumes identified as containing artefacts were replaced with the mean of the neighboring non-outlier volumes, or removed if at the end of a run. If more than half of the scans in a run had artefacts, that run was discarded. Artefacts were also modeled as confound regressors in the first level design matrices. Next, BOLD images acquired at different echo times were realigned and slice time corrected using SPM12 defaults. The resulting images were then resliced to the space of the first volume of the first echo-1 BOLD time series. A brain mask was computed based on preprocessed echo-1 BOLD images using Nilearn 0.5.2 and combined with a grey-and-white matter mask in functional space for better coverage of anterior and ventral temporal lobes (Abraham et al., 2014). The three echo time series were then fed into the Tedana workflow (Kundu et al., 2017), run inside the previously created brain mask. This workflow decomposed the time series into components and classified each component as BOLD signal or noise. The three echo series were optimally combined and noise components discarded from the data. The resulting time series were unwarped to correct for inhomogeneities in the scanner's magnetic field: the voxel displacement map calculated from the field maps was coregistered to the first echo-1 image from the first run, and applied to the combined time series for each run. The preprocessed BOLD time series corresponding to the optimal denoised combination of echoes outputted by the Tedana workflow were then used for RSA analysis, where we used unsmoothed functional images in native space to keep the finer-grained structure of activity. For univariate analysis, the preprocessed BOLD time series were also spatially normalized to MNI space using SPM's non-linear registration tool, DARTEL; spatially normalized images were then smoothed with an 8 mm isotropic full-width half maximum Gaussian kernel.

3.3.6 Experimental design and statistical analysis

3.3.6.1 Sample size

The sample size was determined using effect sizes from two previous studies. Staresina et al. (2012) reported a large encoding-retrieval RSA similarity effect (*d* = 0.87). However, subsequent memory effects are typically more subtle, for example *d* = 0.57 for an activation measure (Morcom et al., 2003). We calculated that, with $N = 28$, we would have .8 power to detect $d = .55$ for a one sample *t*-test at alpha = .05 (G*Power 3.1.9.2).

3.3.6.2 Behavioural analysis

To assess whether differences in task engagement during memory encoding predicted later memory, we modelled the effects of encoding task accuracy (0, 1) on subsequent memory outcomes using two separate generalized linear mixed effect models (GLMM) for studied items tested as old (subsequent hits and misses as predictors), and for studied items tested as lures (subsequent false alarms and correct rejection as predictors). Similarly, to assess any differences in study phase reaction times (RTs) according to subsequent memory status, we used two further linear mixed effect models (LMM). At test, to evaluate the effects on memory of perceptual and semantic similarity between objects, we also applied a generalized linear mixed model following the methods of Chapter 2 (Naspi et al., 2020). This had dependent measures of response at test ("old" or "new") and confusability predictors calculated for each image and concept. C1 visual and color confusability were defined as the similarity value of an image with its most similar picture (i.e., the nearest neighbor) from Pearson correlation and earth's mover distance metrics, respectively. Concept confusability was calculated by a weighted sum of the cosine similarities between objects in which each weight was the between-concept similarity itself, i.e., the sum of squared similarities (see Chapter 2; Naspi et al., 2020). All the analyses described above were carried out data with the lme4 package (Version 1.1-23) in R (Version 4.0.0). Models included random intercepts to account for variation over items and participants.

3.3.6.3 Multivariate fMRI analysis

Overview. The goal of our study was to investigate how perceptual and semantic representations processed at encoding predict successful and unsuccessful mnemonic discrimination. To test this, we used RSA to assess whether the fit of perceptual and semantic representational models to activity patterns at encoding predicted subsequent memory. In two main sets of analyses we examined representations predicting later true recognition of studied items, and representations predicting false recognition of similar lures. We implemented a novel approach that models the interaction of representation similarity with subsequent memory in a single step. Each memory encoding model contrasts the strength of visual and semantic representations of items later remembered versus forgotten (or falsely recognized versus correctly rejected) within the same representational dissimilarity matrix (RDM). In a third set of analyses we also aimed to replicate Clarke and Tyler (2014) key findings regarding perceptual and semantic representations irrespective of memory. All RSA analyses were performed separately for each participant on trial-specific parameter estimates from a general linear model (GLM). We then followed three standard steps: 1) For each theoretical perceptual and semantic model, we created model RDMs embodying the predicted pairwise dissimilarity over items; 2) For each ROI (or searchlight sphere), we created fMRI data RDMs embodying the actual dissimilarity of multivoxel activity patterns over items; 3) We determined the fits between the model RDMs and the fMRI data RDM for each ROI (or searchlight sphere). The implementation of these steps is outlined in the following sections.

RSA first level general linear model. Statistical analysis of fMRI data was performed in SPM12 using the first-level GLM and a Least-Squares-All (LSA) method (Mumford et al., 2012). For each participant, the design matrix included one regressor for each trial of interest, for a total of 327 or 328 regressors (depending on counterbalancing), computed by convolving the 0.5 s duration stimulus function with a canonical hemodynamic response function (HRF). For each run, we also included twelve motion regressors comprising the three translations and three rotations estimated during spatial realignment, and their scan-to-scan differences, as well as individual scan regressors for any excluded scans, and session constants for each of the 4 scanner runs. The model was fit to native space pre-processed functional images using Variational Bayes estimation with an AR(3) autocorrelation model (Penny, Trujillo-Barreto, & Friston, 2005). A high-pass filter with a cutoff of 128 s was applied and data were scaled to a grand mean of 100 across all voxels and scans within sessions. Rather than using the default SPM whole-brain mask (which requires a voxel intensity of 0.8 of the global mean and can lead to exclusion of ventral anterior temporal lobe voxels), we set the implicit mask threshold to 0 and instead included only voxels which had at least a 0.2 probability of being in grey or white matter, as indicated by the tissue segmentation of the participant's T1 scan.

Regions of interest. All regions of interests (ROIs) are shown in Figure 1. We defined six ROIs including areas spanning the ventral visual stream, which have been implicated in visual and semantic feature-based object recognition processes (Clarke and Tyler, 2014; Clarke and Tyler, 2015). We also included the left inferior frontal gyrus, strongly implicated in semantic contributions to episodic encoding (Kim, 2011), and bilateral anterior ventral temporal cortex, which is implicated in semantic representation (Lambon Ralph et al., 2017) and is hypothesized to contribute to false memory encoding, albeit mainly in associative false memory tasks (Chadwick et al., 2016; Zhu et al., 2019). Except where explicitly stated, ROIs were bilateral and defined in MNI space using the Harvard-Oxford structural atlas: 1) the early visual cortex (EVC; BA17/18) ROI was defined using the Julich probabilistic cytoarchitectonic maps (Amunts et al., 2000) from the SPM Anatomy toolbox (Eickhoff et al., 2005); 2) the posterior ventral temporal cortex (pVTC) ROI consisted of the inferior temporal gyrus (occipito-temporal division; ITG), fusiform gyrus (FG), lingual gyrus (LG), and parahippocampal cortex (posterior division; PHC); 3) the perirhinal cortex (PrC) ROI was defined using the probabilistic perirhinal map including voxels with a > 10% probability to be in that region (Devlin and Price, 2007; Holdstock et al., 2009); 4) the anterior ventral temporal cortex (aVTC) ROI included voxels with >30% probability of being in the anterior division of the inferior temporal gyrus and >30% probability of being in the anterior division of the fusiform gyrus; 5) the left inferior frontal gyrus (LIFG; BA44/45) consisted of the pars triangularis and pars opercularis. Lastly, we used univariate analysis as a preregistered method to define additional ROIs for RSA around any regions not already in the analysis that showed significant subsequent memory effects. Based on this analysis, we also included 6) the left inferior temporal gyrus (occipito-temporal division; LITG) (see Results, Univariate fMRI analysis). The LITG has been previously implicated in true and false memory encoding (Dennis et al., 2007; Kim and Cabeza, 2007). The ROIs in Figure 1 are mapped on a pial representation of cortex using the Connectome Workbench (https://www.humanconnectome.org/software/connectome-workbench).

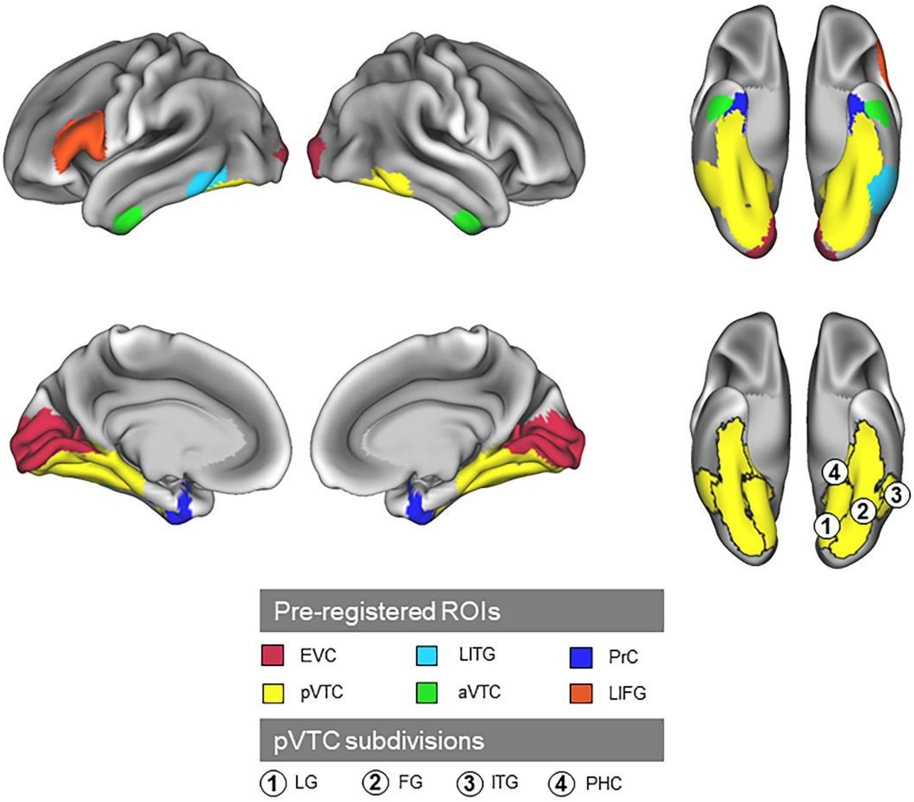


Figure 1. Binary ROIs overlaid on a pial cortical surface based on the normalized structural image averaged over participants. Colored ROIs represent regions known to be important in episodic encoding and in visual or semantic cognition. Circled numbers specify different subregions within pVTC (see Region of Interest for details).

RSA region of interest analysis.

Model RDMs. We created four theoretical RDMs using low-level visual, color, binary-categorical, and specific object semantic feature measures. Figure 2 illustrates the multidimensional scale (MDS) plots for the perceptual and semantic relations expressed by these models, and Figure 3 shows the model RDMs. Memory encoding RDMs are displayed in Figure 3A and 3B, and overall RDMs irrespective of memory in Figure 3C.

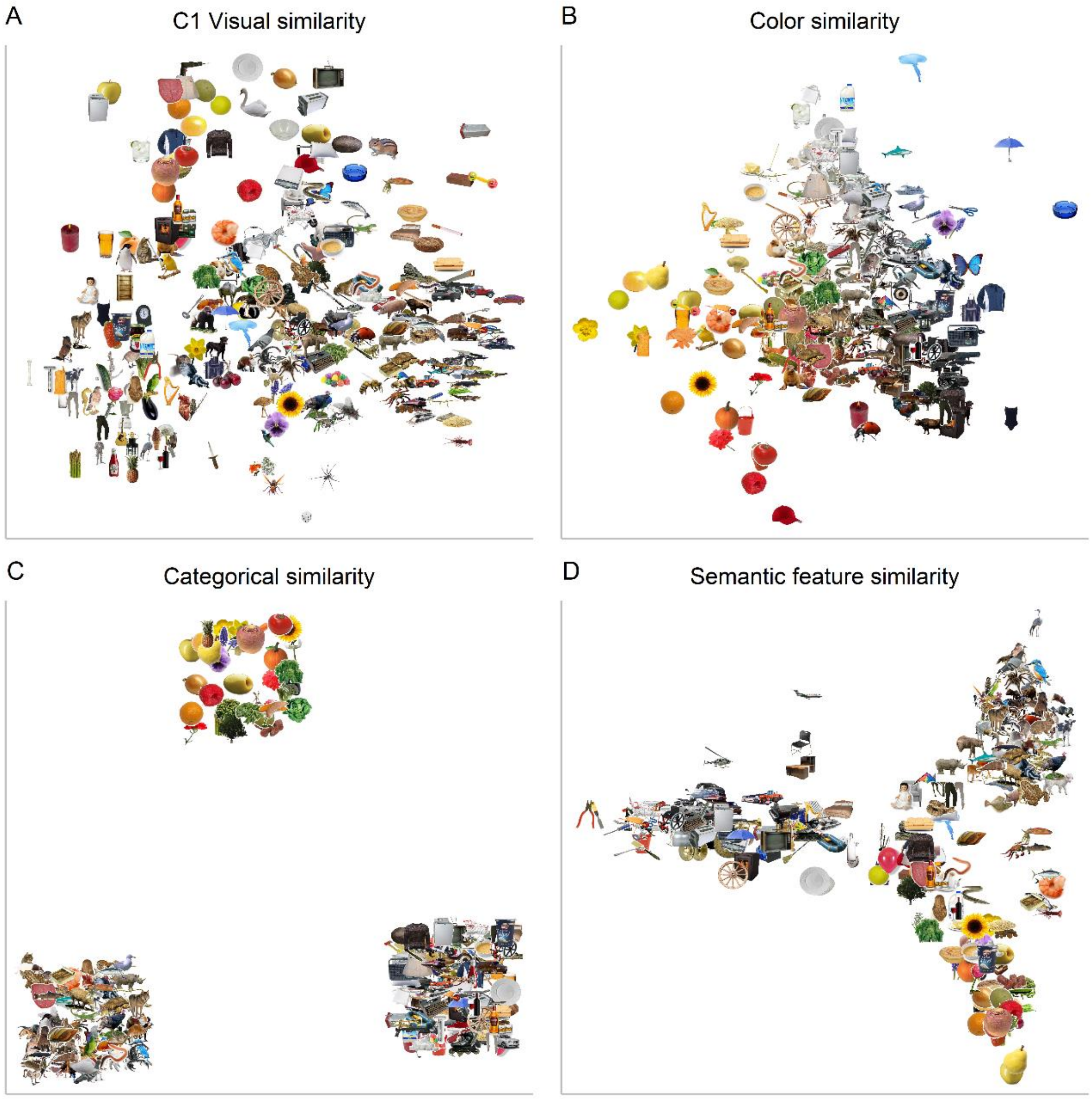


Figure 2. MDS plots for perceptual and semantic similarities for the four models. Pair-wise similarities were calculated to create representational dissimilarity matrices (RDMs). **A**, C1 visual similarity codes for a combination of orientation and shape (e.g., round objects towards the top, horizontal shapes on the right, vertical shapes at the bottom). **B**, Color similarity represents color saturation and size information (i.e., from bright on the left to dark at the bottom, and white towards the top). **C**, Binary categorical semantic similarity codes for domain-level representations distinguishing animals, plants and nonbiological objects (bottom-left, top, bottom-right, respectively). **D**, Semantic feature similarity codes for finer-grained distinctions based on features of each concept (e.g., differences within living things at the bottom, non-living things on the left, and many categories of animal on the top-right). The objects shown are taken from a single subject at encoding.

1) The *early visual RDM* was derived from the HMax computational model of vision (Riesenhuber and Poggio, 1999; Serre et al., 2007) and captured the low-level (V1) visual attributes of each picture in the C1 layer. Pairwise dissimilarity values were computed as 1 - Pearson's correlations between response vectors for gray-scale versions of each image.

2) The *color RDM* was calculated using the color distance package (Version 1.1.0; Weller and Westneat, 2019) in R. After converting the RGB channels into CIELab space we calculated the earth mover's distance between each pair of images (Rubner et al., 2000). We then normalized the distance so that the dissimilarity values ranged from 0 (lowest) to 1 (highest).

3) The *animal-nonbiological-plant RDM* combined the 24 object categories together according to 3 domains: animal, nonbiological, and plants (Clarke and Tyler, 2014). Pairwise dissimilarity values in this RDM were either 0 (same domain) or 1 (different domain).

4) Construction of the *semantic feature RDM* followed Clarke and Tyler (2014), but used updated property norms (Devereux et al., 2014). We first computed pairwise feature similarity between concepts from a semantic feature matrix in which each concept is represented by a binary vector indicating whether a given feature is associated with the concept or not. Pairwise dissimilarity between concepts was computed as 1 – S where S is equal to the cosine angle between feature vectors. This RDM captures both categorical similarity between objects (as objects from similar categories have similar features) and within-category object individuation (as objects are composed of a unique set of features).

For the analyses of memory encoding, model RDMs were split into two, giving one RDM for each subsequent memory analysis. The true subsequent memory RDMs included only items that were subsequently tested as old; these were coded as subsequent hits or subsequent misses (Fig. 3A). The false subsequent memory RDMs included only items that were subsequently tested as lures; these were coded as subsequent false alarms or subsequent correct rejections (Fig. 3B). For true subsequent memory, we computed dissimilarity between all pairs of subsequently remembered items, and all pairs of subsequently forgotten items, omitting pairings of subsequently remembered and subsequently forgotten items. Then, to assess how dissimilarity depended on subsequent memory we weighted the model RDMs so that the sum of the cells corresponding to remembered items equaled 1 and the sum of the cells corresponding to forgotten items equaled -1, so the dissimilarity values for all included trials summed to 0 (i.e., subsequent hits – subsequent misses). Thus, positive correlations of the model RDMs with the fMRI data RDMs indicate that the representations are aligned more strongly with neural patterns for items that are later remembered than forgotten. Conversely, negative correlations indicate greater alignment for items that are later forgotten than remembered items. For false subsequent memory, we followed the same procedure, but subsequent false alarms were substituted for subsequent hits, and subsequent correct rejections for subsequent misses. Although an unequal number of trials can create spurious effects, we have a sufficiently large number of trials for each participant and condition for reliable correlation coefficients. According to one estimate, a correlation needs to have at least 150 observations to be considered stable (Schönbrodt & Perugini, 2013). In our study, only one participant yielded less than 150 similarity values, from a matrix of 17 falsely recognized trials, which gives a vector of 136 unique similarity values. Thus, we can be fairly confident in our results. Analyses were implemented using custom MATLAB 2019b (The MathWorks Inc., 2019) and R (Version 4.0.0; R Core Team, 2017) functions (https://osf.io/ypmdj). For the RSA analyses irrespective of memory, we modeled dissimilarities between all item pairs, treating all trials in the same way (see Fig. 3C).

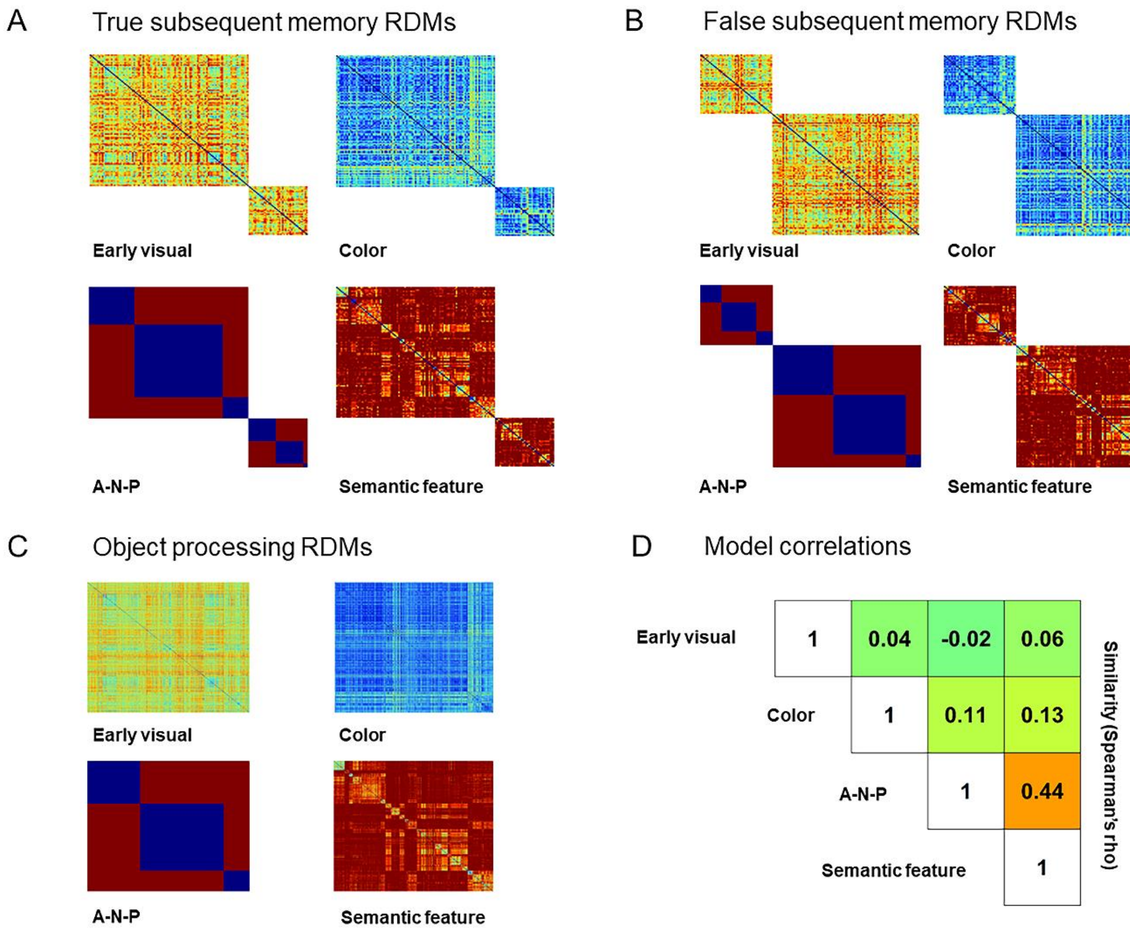


Figure 3. Representational dissimilarity matrices. **A**, Dissimilarity predictions of the four true subsequent memory models which included items that were later tested as old, coding subsequent hits positively (upper-left quadrants) and subsequent misses negatively (bottom-right quadrants). **B**, Dissimilarity predictions of the four false subsequent memory models which included items that were later tested as lures, coding subsequent false alarms positively (upper-left quadrants) and subsequent correct rejections negatively (bottom-right quadrants). **C**, Dissimilarity models of object processing including all the items. **D**, Similarity between theoretical models. The color palettes used for the model correlations in D are the inverse of those used for the model RDMs in A, B, and C. The specific models are unique for each participant. For visualization purposes, similarity values within true and false subsequent memory RDMs have not been scaled. A-N-P = Animal-nonbiological-plant.

fMRI data RDMs. Parameter estimates were extracted from gray matter voxels in each ROI for all trials of interest. For each voxel, these betas were then normalized by dividing them by the standard deviation of its residuals (Walther et al., 2016). As for the model RDMs, we constructed separate fMRI data RDMs for the true and false subsequent memory and overall object processing analyses. For the true subsequent memory analysis, the fMRI data RDM represented activity patterns for concepts subsequently tested as old, and for the false subsequent memory analysis, the fMRI data RDM represented activity patterns for concepts subsequently tested as lures. For the overall analysis, the RDM represented activity patterns for all study trials. For the fMRI data RDMs for the subsequent memory analysis, as for the model RDMs, we computed dissimilarity between all pairings of subsequently remembered (or falsely recognized) items, and between all pairings of subsequently forgotten (or correctly rejected) items, omitting pairings between different trial types. Distance between each item pair was computed as 1 - Pearson's correlation, creating a dissimilarity matrix.

Fitting model to data RDMs. Each fMRI data RDM was compared with each theoretical model RDM using Spearman's rank correlation, and the resulting dissimilarity values were Fisher-transformed. It is important to note that the Spearman's rank correlation is not affected by the weighting procedure for number of trials, since this measure does not depend on the distance between pair of items; thus the order of the ranks is equivalent. For the subsequent memory analysis, we tested for significant positive and negative similarities between model RDM and fMRI data RDMs at the group level using a two-sided Fisher's one-sample randomization (10,000 permutation) test for location with a Bonferroni correction over 6 ROIs. The permutation distribution of the test statistic T enumerates all the possible ways of permuting the correlation signs, positive or negative, of the observed values and computes the resulting sum. Thus, for a two-sided hypothesis, the p-value is computed from the permutation distribution of the absolute value of T, calculating the proportion of values in this permutation distribution that are greater or equal to the observed value of T (Millard and Neerchal, 2001). For the overall analysis we only tested for significant positive similarities between model RDM and fMRI data RDMs (Clarke and Tyler, 2014), using a one-sided test, in which the p-value is evaluated as the proportion of sums in the permutation distribution that are greater than or equal to the observed sum T (Millard and Neerchal, 2001). To find the unique effect of model RDMs, each fMRI data RDM showing a significant effect was also compared with each theoretical model RDM while controlling for effects of all other significant model RDMs (using partial Spearman's rank correlations). While correlations between model RDMs were generally low, the object-specific feature model shared about 19% variance with the category model ($r = 0.44$; Fig 3D), likely reflecting the information about coarse semantic categories as well as individual objects that is carried by feature similarities (Clarke and Tyler, 2014).

Post hoc RSA analyses by memory item type. For regions and models showing significant RSA memory effects, we explored whether representations aligned with each item type were significantly different from zero. To do this, we created four separate model and fMRI data RDMs for items subsequently remembered, forgotten, falsely recognized, and correctly rejected. We then followed the same steps as described for the ROI analysis irrespective of memory, but fit model RDMS to fMRI data RDMs for each trial type separately. Then, we tested for significant positive similarity at the group level using a one-tailed Fisher's one-sample randomization test (10,000 permutations) for location. We applied Bonferroni corrections for the 6 preregistered and the 6 exploratory ROIs.

RSA searchlight analysis. In addition to the targeted ROI analysis, we ran a whole-brain searchlight analysis. This followed the same 3 main steps as the ROI analysis (see RSA region of interest analysis). For each voxel, the fMRI data RDM was computed from parameter estimates for gray matter voxels within a spherical searchlight of radius 7 mm, corresponding to maximum dimensions 5 × 5 × 5 voxels. Dissimilarity was again estimated using 1 - Pearson's correlation. As in the ROI analysis, this fMRI data RDM was compared with the model RDMs, and the resulting dissimilarity values were Fisher transformed and mapped back to the voxel at the center of the searchlight. The similarity map for each model RDM and participant was then normalized to the MNI template space (see Image preprocessing). For each model RDM, the similarity maps were entered into a group-level random-effects analysis and thresholded using permutation-based statistical nonparametric mapping (SnPM; http://www.nisox.org/Software/SnPM13/). This corrected for multiple comparisons across voxels and the number of theoretical model RDMs. As for the ROIs we performed two-tailed tests in the subsequent memory analyses and one-tailed tests for the overall analysis. Variance smoothing of 6 mm FWHM and 10,000 permutations were used in all analyses. We used cluster-level inferences with FWE-correction at $\alpha = .025$ in each direction for the two-tailed tests and $\alpha = .05$ for the one-tailed test, in both cases with a cluster forming threshold of .005 uncorrected. All results are presented on an inflated representation of the cortex using the BrainNet Viewer (Xia et al., 2013, http://www.nitrc.org/projects/bnv/) based on a standard ICBM152 template.

3.3.6.4 Univariate fMRI analysis

In addition to RSA, we used univariate analysis to test whether activation in PrC was related to the conceptual confusability of an object, in a replication of Clarke and Tyler (2014), and whether this activation predicted memory. We also used activations to define additional ROIs (see Regions of interest). The first level GLM for each participant included one regressor of interest for each of the 4 experimental conditions (subsequent hits, misses, false alarms, and correct rejections). For each condition, we also included 4 linear parametric modulator regressors representing concept confusability values for each concept with other concepts in the CSLB property norms (Devereux et al., 2014). We first computed a semantic similarity score between each pair of concepts (see RSA region of interest analysis, Model RDMs). The concept confusability score of each concept was then equal to the sum of squared similarities between it and the other concepts in the set. This was equivalent to a weighted sum of pair-wise similarities in which each weight was the between-concept similarity itself, a measure used in our recent behavioral study in Chapter 2 (Naspi et al., 2020). As also specified in the preregistration, since the results of the concept confusability analysis diverged from those of Clarke and Tyler (2014), we ran an additional analysis using a measure of concept confusability with a stronger weighting scheme equivalent to theirs. They defined concept confusability as the exponential of the ranked similarities of all the paired concepts, which is very close to a nearest neighbor scheme in which each concept's similarity is equal to its similarity to the most similar concept in the set. Due to our larger number of items the exponential weighting produced extremely large weights, so we substituted the simpler nearest neighbor scheme (the two measures were correlated at $r = .98$). We used an explicit mask including only voxels which had at least a .2 probability of being in grey matter as defined using the MNI template. To permit inferences about encoding condition effects across participants, contrast images were submitted to a second-level group analysis (one sample *t*-test) to obtain *t*-statistic maps. The maps were thresholded at $p < .05$, FWE-corrected for multiple comparisons at the voxel level using SPM (the preregistration specified 3dClustSim in AFNI, but this function had since been updated (Cox et al., 2017) so for simplicity we used the SPM default). Only regions whose activations involved contiguous clusters of at least 5 voxels were retained as ROIs for subsequent RSA analysis.

3.3.7 Code accessibility

All analyses were performed using custom code and implemented either in MATLAB or R. All code and the data for the behavioral and the fMRI analyses are available through https://osf.io/z4c62/.

3.5 Results

3.5.1 Memory task performance

In the study phase, participants correctly identified most of the time whether concepts began with a consonant or vowel on the incidental encoding task (M proportion = 0.78). Analysis on task engagement (see Materials and Methods, Behavioral data) using a GLMM showed that accuracy at encoding did not differ according to whether items that were tested as studied were later remembered relative to forgotten (β = .110, SEM = .242, z = .456, p = .649), or whether items that were tested as lures were later falsely recognized relative to correctly rejected (β = .051, SEM = .202, z = .251, p = .802). Similarly, a linear mixed model did not reveal any difference in RTs related to subsequent old items that were later remembered relative to forgotten (β = .002, SEM = .017, t = .123, p = .902), or subsequent lures that were later falsely recognized relative to correctly rejected (β = -.013, SEM = .015, t = -.873, p = .383). Thus, the fMRI subsequent memory effects are not attributable to differences in accuracy or time on task at encoding.

At test, as a simple check on the overall level of performance we used the discrimination index *Pr*, i.e., the difference between the probability of a hit to studied items and the probability of a false alarm to novel items. All participants passed the preregistered inclusion criterion of $Pr > 0.1$. Overall, discrimination collapsed across confidence was very good (M = .649, SD = .131, $t_{(27)}$ = 26.259, $p < .001$). Discrimination was also above chance for high confidence (M = .771, SD = .152, $t_{(27)}$ = 26.868, $p < .001$) and low confidence judgments (M = .330, SD = .145, $t_{(27)}$ = 12.014, $p < .001$). This suggests that low confidence responses at test carried veridical memory, so we followed our preregistered plan to include trials attracting both high and low confidence responses in the subsequent memory analysis. Following an analogous procedure for false recognition of similar lures corrected by subtracting the proportion of false alarms to novel items, we also found that this was significantly above chance for judgments collapsed across confidence (M = .271, SD = .090, $t_{(27)}$ = 15.996, $p < .001$), and for both high confidence (M = .293, SD = .133, $t_{(27)}$ = 11.618, p

< .001) and low confidence (*M* = .157, *SD* = .160, $t_{(27)}$ = 5.187, *p* < .001) considered separately.

We then used a GLMM to quantify the influence of perceptual and semantic variables on memory performance according to item status. Our variables of interest were condition (studied, lure, or novel), concept confusability, C1 visual confusability, and color confusability (see Behavioral data for details). Results revealed modulations of memory by perceptual and semantic variables in line with our recent behavioral study in Chapter 2 (Naspi et al., 2020). People were less likely to recognize studied items for which the low-level visual representations (C1) were more similar to those of their nearest neighbor (β = -.166, SEM = .064, *z* = -2.584, *p* = .015), and also less likely to recognize studied items with high concept confusability relative to novel items (β = -.533, SEM = .067, *z* = -7.963, *p* < .001). As expected, concept confusability also had a substantial effect on false recognition of similar lures relative to novel items, whereby images whose concepts were more confusable with other concepts in the set were less likely to be falsely recognized (β = -.273, SEM = .064, *z* = -4.292, *p* < .001).

3.5.2 Preregistered RSA analysis in regions of interest

3.5.2.1 Perceptual and semantic representations predict true recognition

To examine representations engaged during successful encoding we compared the fit of early visual, color, animal-nonbiological-plant, and semantic feature models for studied items tested as old that were subsequently remembered (number of trials, *M* = 61.41; range = 60-146) versus forgotten (number of trials, *M* = 19.93; range = 17-104) (Fig. 4A). These comparisons were bidirectional, since engagement of perceptual and/or semantic processing in a region might either support or be detrimental to later memory. Thus, we used a two-sided Fisher's randomization test *T*. In posterior ROIs, engagement of both perceptual and finer-grained semantic representations tended to predict successful later recognition. In EVC, the early visual model strongly predicted later true recognition of studied items (*M* = .07, 95% CI [.05, .09], *T* = 1.86, *p* < .001). Thus, when the neural patterns at study were representing visual information, items were more likely to be correctly recognized. Both the early visual and semantic feature models also predicted true recognition in pVTC (*M* = .03, 95% CI [.02, .04], *T* = .82, *p* < .001, and *M* = .02, 95% CI [.01, .04], *T* = .67, *p* = .007, respectively). In contrast, taxonomic semantic representations coded more anteriorly were associated with later forgetting. In aVTC and in the LIFG, model fit for categorical semantic information represented by the animal-nonbiological-plant domain was less for remembered than forgotten studied items (*M* = -.01, 95% CI [-.02, -.01], *T* = .35, *p* = .001, and *M* = -.02, 95% CI [-.04. -.01], *T* = .65, *p* = .004, respectively). Thus, when neural patterns in these regions were aligned with items' taxonomic categories, participants were less likely to successfully recognize them. No other results were significant.

We also checked which representations showed unique effects that predicted memory after controlling for effects of other significant models using partial correlation. In pVTC, only the early visual model uniquely predicted successful recognition memory for studied items (*M* = .02, 95% CI [.01, .03], *T* = .64, *p* = .004) (but see Exploratory ROI analysis).

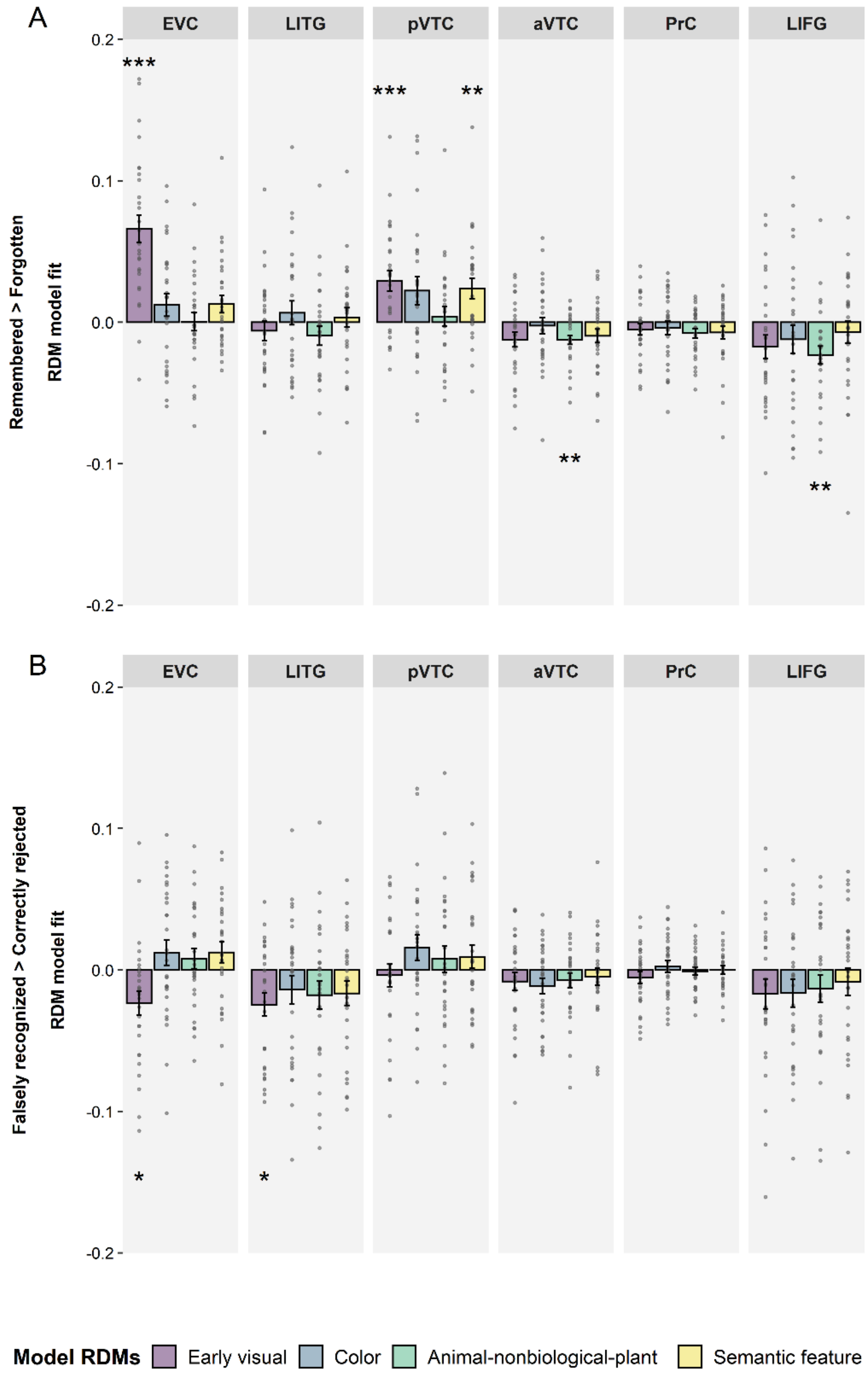


Figure 4. Perceptual and semantic representations predicting subsequent memory for *a priori* ROIs and models. Plots show the relative difference in the strength of perceptual and semantic representations at the group level associated with: **A**, true subsequent memory (greater representational similarity for remembered than forgotten items, positive bars); **B**, false subsequent memory (greater representational similarity for falsely recognized than correctly rejected items, positive bars). Error bars represent the standard error of the mean

(SEM) across participants. Asterisks indicate models for which Spearman's rho differed significantly from zero at the group level (two-sided Fisher's randomization test for location; Bonferroni correction calculated by multiplying the uncorrected *p*-value by the number of preregistered ROIs, i.e., 6). * $p < .05$, ** $p < .01$, *** $p < .001$

3.5.2.2 Weak perceptual representations predict false recognition

To examine how the perceptual and semantic representations embodied in our theoretical models contributed to subsequent memory for lures, we compared RSA model fit for items that were later falsely recognized (number of trials, $M = 30.71$; range = 26-107) versus correctly rejected (number of trials, $M = 50.61$; range = 54-131) (Figure 4B). In posterior regions, weaker low-level visual representations of pictures predicted subsequent false recognition of lures. We observed this pattern in both the EVC and the LITG ($M = -.02$, 95% CI [-.04, -.01], $T = 0.66$, $p = 0.047$, and $M = -.02$, 95% CI [-.04, -.01], $T = .69$, $p = .026$, respectively). Thus, when neural patterns in these regions were not aligned with the early visual model, items were more likely to be falsely recognized. No other results were significant.

3.5.2.3 Perceptual and semantic object processing irrespective of memory

Replicating Clarke and Tyler (2014), we also examined the perceptual and semantic representations of objects that were reflected in fMRI activity patterns regardless of memory encoding. The results (Fig. 5) showed that while visual information is broadly represented posteriorly, activity patterns in the aVTC, PrC, and LIFG reflect finer-grained semantic information. Posteriorly, EVC showed a strong relationship with the low-level visual model ($M = .08$, 95% CI [.06, .10], $T = 2.21$, $p < .001$), and a weaker but significant relation with the semantic feature model ($M = .01$, 95% CI [.00, .01], $T = .20$, $p = .032$). More anteriorly, the low-level visual and semantic feature models were both significantly related to activity patterns in pVTC ($M = .04$, 95% CI [.03, .04], $T = 1.00$, $p < .001$, and $M = .02$, 95% CI [.02, .03], $T = .60$, $p < .001$, respectively) and in LITG ($M = .01$, 95% CI [.00, .02], $T = .26$, $p < .038$, and $M = .02$, 95% CI [.01, .02], $T = .45$, $p < .001$, respectively). At the apex of the ventral visual pathway, semantic feature information was coded in both the bilateral aVTC ($M = .01$, 95% CI [.00, .01], $T = .17$, $p = .006$) and in bilateral PrC ($M = .01$, 95% CI [.00,

.01], *T* = .19, *p* < .001). These findings replicated those of Clarke and Tyler (2014). The specific semantic properties of objects were also represented in the LIFG (*M* = .01, 95% CI [.01, .02], *T* = .30, *p* = .001).

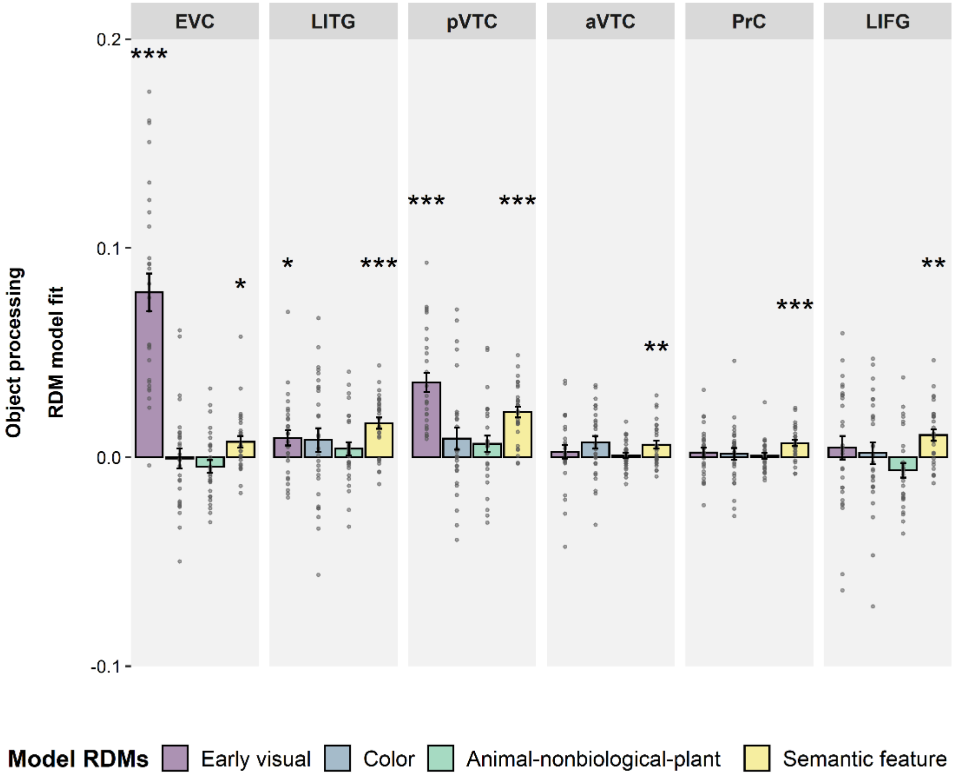


Figure 5. Semantic and perceptual representations represented in ROIs regardless of memory encoding. Plots show the strength of perceptual and semantic representations at the group-level within patterns of activity along the ventral stream and frontal regions. Error-bars are standard error of the mean (SEM) across subjects. Asterisks above and below the bars depict *p*-values for tests of whether each individual Spearman's correlation is greater than zero (one-sided Fisher's randomization test for location; Bonferroni correction calculated by multiplying the uncorrected *p*-value by the number of preregistered ROIs, i.e., 6). * *p* < .05, ** *p* < .01, *** *p* < .001

We then ran a partial correlation on those ROIs showing significant effects for multiple models. As expected, patterns of activity in the EVC were uniquely related to the early visual model (*M* = .08, 95% CI [.06, .10], *T* = 2.20, *p* < .001), replicating Clarke and Tyler's (2014) results. Thus, the semantic feature model was no longer significant when the early visual model was controlled for. More anteriorly, the pattern of activity in the pVTC had unique relations to both low-level visual and semantic feature information (*M* = .03, 95% CI [.03, .04], *T* = .96, *p* < .001, and *M* = .02, 95% CI [.01, .02], *T* = .54, *p* < .001, respectively). However, after controlling for the low-level visual model, activity patterns in the LITG were only uniquely associated with semantic feature representations (*M* = .02, 95% CI [.01, .02], *T* = .44, *p* < .001). Thus, like Clarke and Tyler (2014), we found that visual information is represented in early visual regions. We also replicated their finding that semantic feature similarity information was coded more anteriorly in the PrC, and found further, also anterior, regions that showed a similar pattern, in the aVTC and the LIFG (see also RSA searchlight fMRI analysis).

3.5.3 Exploratory RSA analysis in regions of interest

3.5.3.1 Perceptual and semantic representations in pVTC subdivisions predict true recognition

In the preregistered analyses reported above, our large pVTC ROI showed evidence of both visual and semantic feature representations predicting memory success. We therefore explored whether four subdivisions of this large bilateral region showed distinct effects: the LG, ITG, FG, and PHC (see Regions of interest). Moreover, given our strong *a priori* prediction of involvement of PrC in subsequent memory, we ran exploratory analyses in left perirhinal cortex (LPrC) and right perirhinal cortex (RPrC), separately. The results are shown below in Figure 6. Posteriorly, in bilateral LG, perceptual information related to the early visual model predicted later recognition of studied items (*M* = .03, 95% CI [.01, .04], *T* = .74, *p* = .002), as it did in the EVC ROI. In contrast, more anteriorly, activity patterns in the FG related to both the low-level visual and semantic feature models predicted subsequent true recognition (*M* = .03, 95% CI [.02, .05], *T* = .87, *p* = .002, and *M* = .04, 95% CI [.02, .05], *T* = 1.01, *p* < .001, respectively), as did categorical semantic information represented by the animal-nonbiological-plants model in the PHC (*M* = .02, 95% CI [.01, .03], *T* = .55, *p* = .019). Lastly, activity related to the categorical semantic model in the LPrC predicted subsequent forgetting (*M* = -.01, 95% CI [-.02, .00], *T* = .28, *p* = .023).

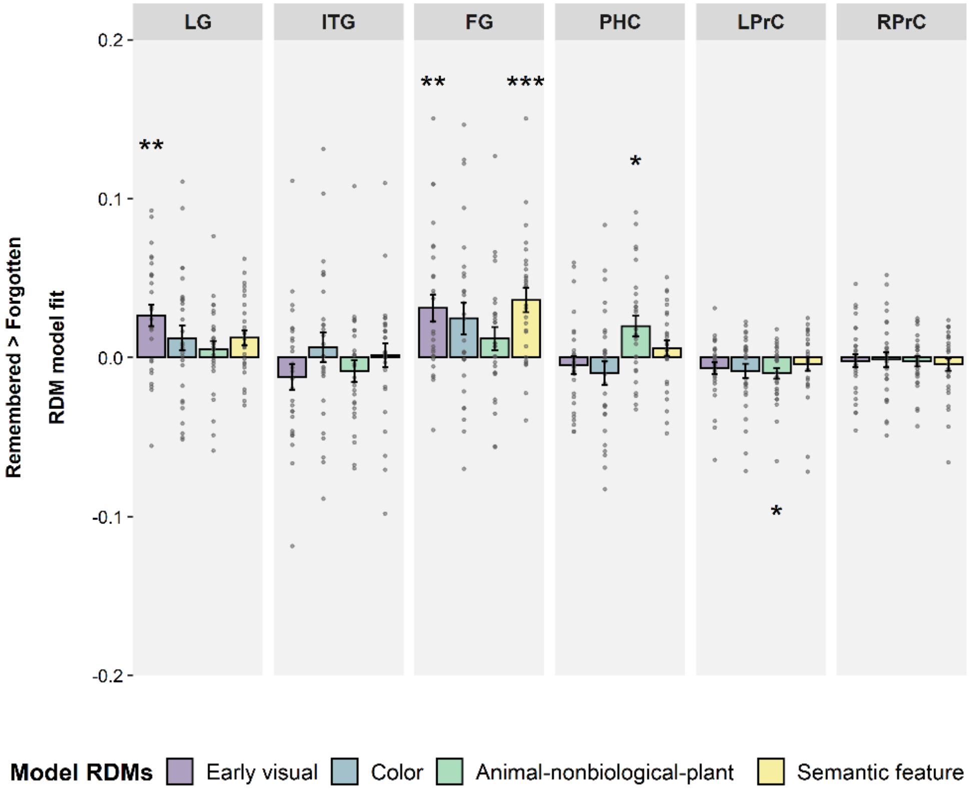


Figure 6. Perceptual and semantic representations predicting true subsequent memory in exploratory ROIs. Plots show the relative difference in the strength of perceptual and semantic representations at the group-level associated with true subsequent memory (greater representational similarity for remembered than forgotten items, positive bars). Error bars represent the standard error of the mean (SEM) across participants. Asterisks indicate significance of tests of group level differences of Spearman's rho from zero (two-sided Fisher's randomization test for location; Bonferroni correction calculated by multiplying the uncorrected *p*-value by the number of exploratory ROIs, i.e., 6). * $p < .05$, ** $p < .01$, *** $p < .001$

A partial correlation analysis for the FG (which showed effects of multiple models) confirmed that both the early visual and semantic feature models were uniquely associated with later true recognition ($M = .02$, 95% CI [.01, .03], $T = .58$, $p = .034$, and $M = .03$, 95% CI [.01, .04], $T = .71$, $p = .002$, respectively). Thus, both simple visual and object-specific semantic information contributed to memory after controlling for each other.

3.5.3.2 Perceptual representations in the early visual cortex predict true recognition

Lastly, following our main analyses of true and false memory encoding, we wanted to check for evidence that the key representations predicting later memory differed according to the type of memory (true or false). Thus, we compared the fit of our theoretical models for studied items tested as old and subsequently remembered versus those tested as lures and subsequently falsely recognized. Results showed that low-level visual information mapped in EVC was stronger for true than false recognition ($M = .04$, 95% CI [.02, .06], $T = 1.16$, $p < .001$). No other results were significant at the Bonferroni-corrected threshold of 6 ROIs, but without a correction the theoretically important object-specific semantic representations in FG were also stronger for true than false recognition ($M = .02$, 95% CI [.00, .04], $T = .61$, $p = .030$).

3.5.3.3 Post hoc RSA analyses by memory item type

Where the RSA analyses showed that representational similarity differed significantly according to subsequent memory, we explored which trial types – hits or misses, and falsely recognized or correctly rejected – carried representations of the relevant information. To do this, we asked whether representational similarity was significantly different from zero for each trial type separately (Table 1). For all models and ROIs where the alignment of neural patterns with a perceptual or semantic model positively predicted true memory, significant representational similarity was present only for subsequently remembered items. Examples were low-level visual representations in EVC and pVTC, and fine-grained semantic representations in pVTC. In contrast, for almost all models and ROIs in which the alignment of neural patterns with the model predicted forgetting, significant representational similarity was present only for forgotten items. This pattern was found in aVTC and LIFG, and in the exploratory analysis, in LPrC. Lastly, in the visual regions where the alignment of neural patterns with low-level visual representations predicted correct rejection of lures in early and late visual regions, there was significant representational similarity only on correct rejection trials.

Table 1. Post hoc analysis in regions associated with true and false subsequent memory in the preregistered and exploratory analysis.

	True subsequent memory (Sub hits > Sub misses)							
Preregistered ROIs	Early visual model		Color model		A-N-P model		Semantic feature model	
	M	*CI*	*M*	*CI*	*M*	*CI*	*M*	*CI*
EVC								
Sub hits	0.07***	0.05-0.09						
Sub misses	0.02	0.01-0.03						
pVTC								
Sub hits	0.05***	0.04-0.06					0.04***	0.04-0.05
Sub misses	0.01	-0.01-0.02					0.01	-0.01-0.03
aVTC								
Sub hits					0.00	0.00-0.01		
Sub misses					0.02**	0.01-0.03		
LIFG								
Sub hits					-0.01	-0.01-0.00		
Sub misses					0.02*	0.01-0.03		
Exploratory ROIs								
LG								
Sub hits	0.04***	0.03-0.05						
Sub misses	0.00	-0.02-0.02						
FG								
Sub hits	0.04***	0.03-0.05					0.04***	0.03-0.05
Sub misses	0.00	-0.02-0.02					0.00	-0.02-0.01
PHC								
Sub hits					0.04***	0.03-0.05		
Sub misses					0.01*	0.00-0.03		
LPrC								
Sub hits					0.00	0.00-0.01		
Sub misses					0.03***	0.01-0.04		
	False subsequent memory (Sub FAs > Sub CRs)							
Preregistered ROIs	Early visual		Color		A-N-P		Semantic feature	
	M	CI	M	CI	M	CI	M	CI
EVC								
Sub FAs	0.01	-0.01-0.03						
Sub CRs	0.04***	0.03-0.06						
LITG								
Sub FAs	0.00	-0.01-0.01						
Sub CRs	0.03***	0.02-0.05						

Mean estimate (M) and confidence intervals (CI) are reported in the table for each trial type. Asterisks indicate models for which Spearman's rho differed significantly from zero at the group level (one-sided Fisher's randomization test with Bonferroni corrections for preregistered and exploratory ROIs, i.e., 6 each). A-N-P = animal-nonbiological-plant; FAs = false alarms; CRs = correct rejections. * $p < .05$; ** $p < .01$; *** $p < .001$

3.5.4 Preregistered RSA searchlight analysis

3.5.4.1 Perceptual and semantic representations associated with memory encoding

The RSA searchlight analysis tested for any further brain regions coding for perceptual and semantic information associated with memory encoding (Fig 7 and Table 2). The true subsequent memory models showed significant fit to activity patterns in several areas beyond the *a priori* ROIs. The color similarity model was related to patterns in the right parietal opercular cortex, superior frontal gyrus, and precentral gyrus, and this representation at encoding predicted later successful recognition of studied items. Fine-grained semantic features represented in the right lateral occipital cortex (LOC) also predicted true recognition. Coarse categorical semantic representations in right inferior frontal gyrus (RIFG; BA44/45/47) and frontal pole (FP) were associated with later forgetting, paralleling the findings for the *a priori* ROI in LIFG (BA44/45).

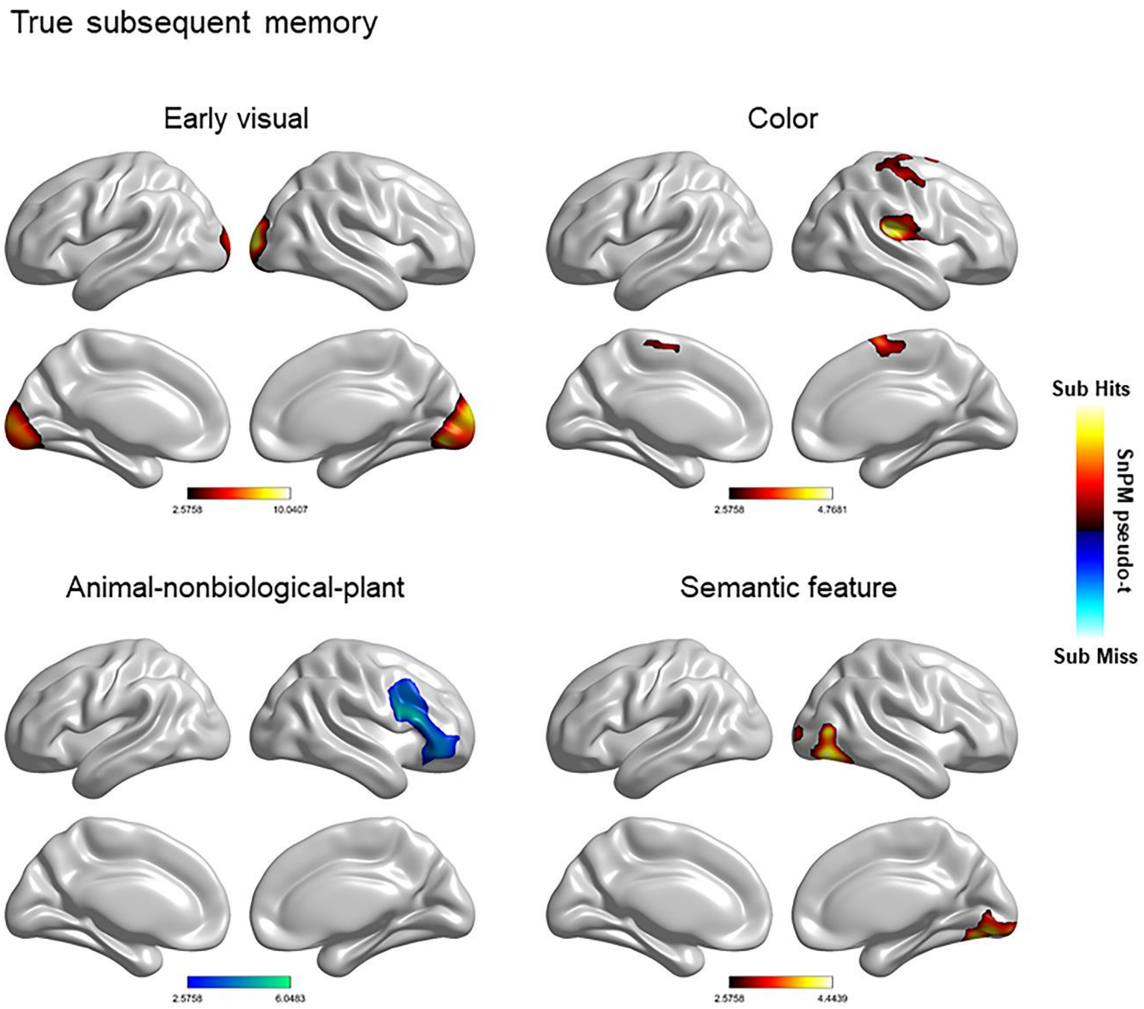


Figure 7. RSA searchlight results for perceptual and semantic models. The figure shows regions in which multivoxel activity pattern predicted successful subsequent true recognition (hot map) and unsuccessful true recognition (i.e., subsequent forgetting, cool map). All significant clusters are shown at the FWE-corrected threshold used for analysis (see Materials and Methods: RSA searchlight analysis). No suprathreshold voxels survived for the subsequent false recognition models. Similarity maps are presented on an inflated representation of the cortex based on the normalized structural image averaged over participants.

Table 2. RSA searchlight results showing perceptual and semantic effects on successful true memory encoding

Regions	Cluster extent	Cluster-level *p*(FWE)	Pseudo-*t*	x	y	z
Early visual						
R occipital pole	2493	.005	10.04	18	-93	9
R lingual gyrus			8.91	15	-78	-6
L occipital pole			7.20	-12	-96	6
Color						
R parietal operculum cortex	1756	.010	4.77	48	-21	24
R superior frontal gyrus			3.91	9	3	66
R precentral gyrus			3.58	18	-18	69
Animal-nonbiological-plant						
R inferior frontal gyrus (BA44)	1405	.012	6.05	54	15	27
R inferior frontal gyrus (BA45)			5.27	52	24	18
R frontal pole			4.35	51	39	3
R inferior frontal gyrus (BA47)			3.34	33	30	-18
Semantic feature						
R lingual gyrus	1230	.018	4.44	12	78	-12
R lateral occipital cortex			4.32	42	-75	-12
R occipital fusiform gyrus			4.29	39	-72	-12
R inferior temporal gyrus (OT)			3.43	45	-60	-15

MNI coordinates and significance levels are shown for the peak voxel in each cluster. Anatomical labels are provided for peak locations in each cluster; Effects in clusters smaller than 20 voxels not shown; OT = Occipito-temporal division.

3.5.4.2 Perceptual and semantic object processing irrespective of memory

Searchlight analysis was also conducted for the perceptual and semantic model RDMs across all trials regardless of memory encoding (Fig. 5 and Table 3). The models showed significant fit to multivoxel activity patterns in several areas beyond the *a priori* ROIs. In particular, the effects for the color model were largely restricted to the right lateral occipital cortex, right middle temporal gyrus, and intracalcarine cortex, but also extended into the left lateral occipital cortex and supramarginal gyrus. Categorical semantic representations represented by the animal-nonbiological-plant domain were largely restricted to posterior parts of the ventral stream, highlighting the coarse nature of object information represented in the posterior ventral temporal cortex. This included the right temporal fusiform cortex, the right lingual gyrus, and the posterior division of parahippocampal cortex, but also extended into the middle temporal lobe. In contrast, representation of finer-grained semantic properties of objects extended more anteriorly in the ventral pathway beyond the preregistered ROIs, into bilateral hippocampus, temporal pole and ventromedial frontal regions.

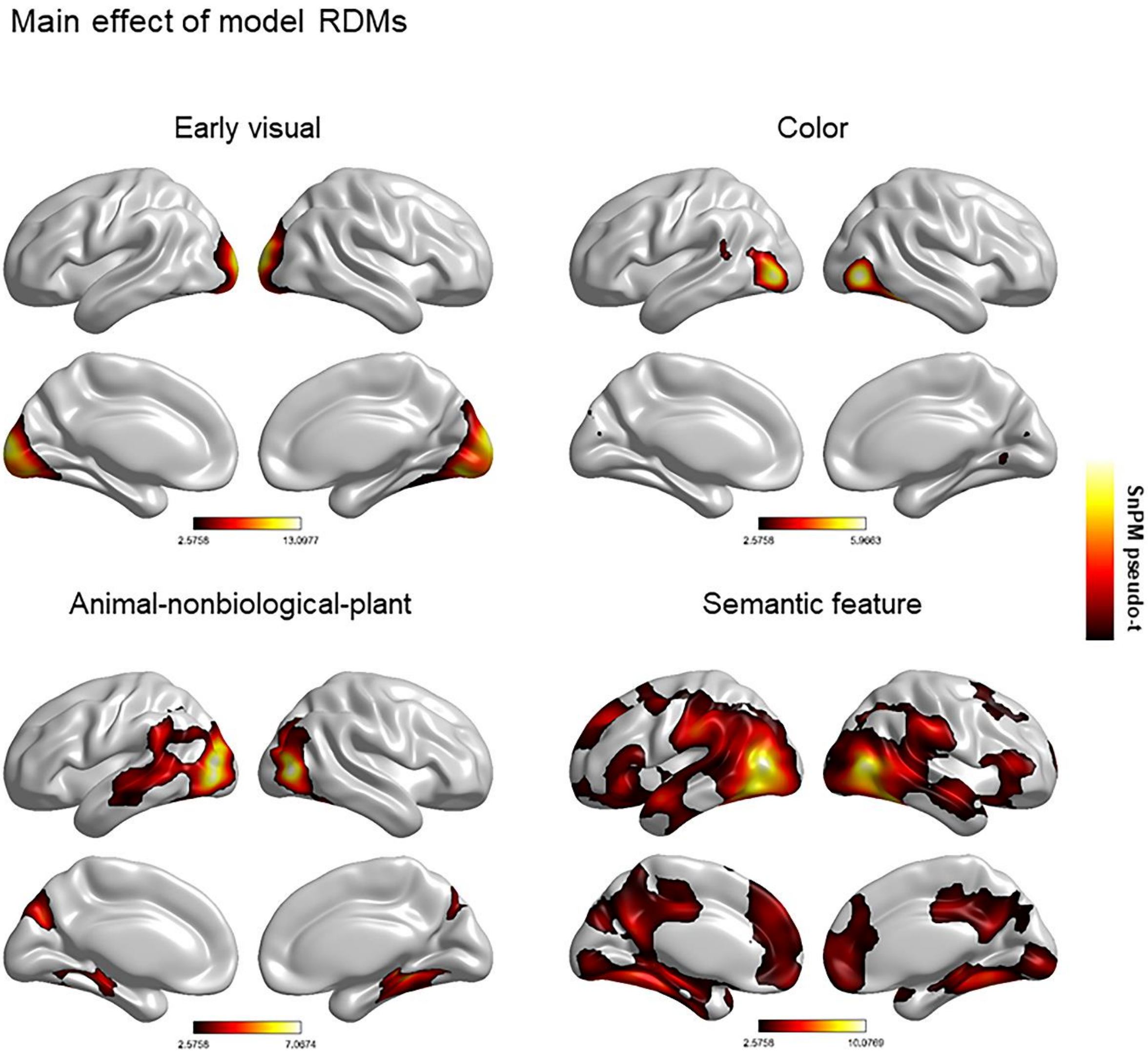


Figure 8. RSA searchlight results for perceptual and semantic models. The figure shows regions in

which multivoxel activity pattern was associated with object processing (i.e., irrespective of memory encoding). All significant clusters are shown at the FWE-corrected threshold used for analysis (see Materials and Methods: RSA searchlight analysis). Similarity maps are presented on an inflated representation of the cortex based on the normalized structural image averaged over participants.

Table 3. RSA results showing perceptual and semantic effect of object processing

Regions	Cluster extent	Cluster-level *p*(FWE)	Pseudo-*t*	x	y	z
Early visual						
R occipital pole	4844	.002	13.10	18	-96	12
L occipital pole			13.02	-15	-99	6
R occipital fusiform gyrus			11.53	18	-78	-12
Color						
R lateral occipital cortex	1121	.019	5.97	45	-75	-3
R middle temporal gyrus			3.45	36	-57	15
R intracalcarine cortex			3.45	21	-72	3
L lateral occipital cortex	714	.044	5.67	-42	-81	-3
L supramarginal gyrus			3.66	-60	-48	15
Animal-nonbiological-plant						
R lateral occipital cortex	2110	.005	6.05	45	-78	6
R lingual gyrus			5.98	30	-39	-6
R temporal fusiform cortex			4.18	39	-54	-18
L parahippocampal cortex	3865	.002	5.14	-18	-39	-21
L middle temporal gyrus			4.58	-63	-42	0
L supramarginal gyrus			4.43	-60	-42	30
Semantic feature						
L lateral occipital cortex	28111	.000	10.08	-48	-75	9
R lateral occipital cortex			9.69	51	-72	6
R temporal fusiform cortex			8.12	42	-51	-15
L temporal fusiform cortex			7.10	-45	-60	-15
L middle temporal gyrus			6.46	-60	0	-18
L hippocampus			5.50	-33	-27	-12
L perirhinal cortex			4.58	-27	-12	-36
R inferior frontal gyrus (BA45)			4.20	51	27	0

R inferior frontal gyrus (BA44)	4.10	51	18	9
R ventromedial prefrontal cortex	4.08	9	51	-12
L ventromedial prefrontal cortex	4.03	-6	51	-12
L inferior frontal gyrus (BA44)	4.02	-51	18	12
L ventral anterior temporal lobe	3.84	-45	-9	-39
L inferior frontal gyrus (BA45)	3.62	-51	27	0
L temporal pole	3.60	-36	3	-36
R hippocampus	3.34	33	-12	-18

MNI coordinates and significance levels shown for the peak voxel in each cluster. Anatomical labels are provided for locations in each cluster. Effects in clusters smaller than 20 voxels not shown.

3.5.5 Preregistered univariate fMRI analysis

3.5.5.1 Encoding activity predicting true and false recognition

Univariate analysis was run to derive ROIs for RSA based on subsequent memory effects in regions where prior literature is suggestive, but not clear, regarding their involvement. This showed significant activation for subsequently remembered > subsequently forgotten items in the LITG (cluster size: $k = 13$, $p < 0.05$ FWE). No significant activation was revealed for subsequently falsely recognized > subsequently correctly rejected items after FWE correction.

3.5.5.2 Parametric effect of concept confusability

Finally, we were interested in the specific role of the PrC, and possibly aVTC, in processing conceptually confusable objects. These regions were not related to parametric changes in concept confusability regardless of memory encoding. Therefore, we did not replicate Clarke and Tyler (2014)'s finding of increased activation for more conceptually confusable objects (uncorrected $p = .139$ and $p = .05$ for PrC and aVTC, respectively). Subsequent memory effects were also not significant at the preregistered FWE-corrected threshold. However, at an uncorrected threshold, activity associated with concept confusability was greater for subsequently forgotten than remembered items in right PrC (cluster size: $k = 12$, $p < .005$) and bilateral aVTC (right cluster size: $k = 19$, $p < .001$; left cluster size: $k = 6$, $p < .001$).

Activity associated with concept confusability was also greater for subsequently falsely recognized than correctly rejected items in bilateral PrC (right cluster size: $k = 35$, $p < .005$; left cluster size: $k = 11$, $p < .005$), and right aVTC (cluster size: $k = 22$, $p < .005$), and for subsequently falsely recognized than remembered items in bilateral PrC (right cluster size: $k = 25$, $p < .005$; left cluster size: $k = 12$, $p < .005$), and right aVTC (cluster size: $k = 16$, $p < .005$).

3.6 Discussion

Our results show that semantic and perceptual representations play distinct roles in true and false memory encoding. By combining explicit models of prior conceptual knowledge and image properties with a subsequent memory paradigm, we probed their separate contributions to encoding of objects. Fine-grained perceptual and semantic processing in the ventral visual pathway both predicted later recognition of studied objects, while coarser-grained categorical semantic information processed more anteriorly predicted forgetting. In contrast, only weak low-level visual representations in posterior regions predicted false recognition of similar objects. The data provide the first direct tests of fuzzy-trace theory's assumptions about how memories are encoded, and suggest that semantic representations may contribute to specific as well as gist memory phenomena (Brainerd and Reyna, 2002).

Our results for the early visual model in the ROI and searchlight analyses converge with studies showing univariate subsequent memory effects in the same regions (Kim and Cabeza, 2007; Kirchhoff et al., 2000; Pidgeon and Morcom, 2016; Wagner et al., 1998). Distributed low-level visual representations in EVC predicted successful later recognition of specific studied objects. The C1 HMax representations embody known properties of primary visual cortex relating to local edge-orientations in images (Kamitani and Tong, 2005), and this model clustered our object images by overall shape and orientation (Fig. 2). These results converge with Davis et al. (2020)'s recent finding that RSA model fit for an early layer of a deep convolutional neural network (DNN) in early visual cortex predicted later memory for pictures. Our data point to specific lower-level properties available in the presented images that contribute to memory. The searchlight analysis showed that these properties also include color (Fig 7). The roles of the regions with significant memory effects are not clear, but overall, color information was represented in LOC as expected.

In late visual regions, such as LG and FG, activity patterns fitting the early visual model also predicted true recognition (Fig. 5 and 7), as hypothesized based on activation studies (Garoff et al., 2005; Kim, 2011; Kirchhoff et al., 2000; Stern et al., 1996; Vaidya et al., 2002). We also found that object semantic features coded in FG predicted true recognition. These pVTC regions receive low-level properties as input to compute complex shape information (Kanwisher, 2001). Emerging data suggest that the FG processes visible (but also verbalizable) semantic features, supporting extraction of meaning from vision. Devereux et al. (2018) combined deep visual and semantic attractor networks to model the transformation of vision to semantics, revealing a confluence of late visual representations and early semantic feature representations in FG (see also Tyler et al., 2013). This converges with Martin et al.'s (2018) finding that FG patterns aligned with rated visual object features. Davis et al. (2020) reported that in FG the mid-layer of a visual DNN predicted memory for object names when the objects were forgotten, while semantic features of the object images predicted memory for the images when the names were forgotten. Our findings clarify that both image-based visual codes and non-image-based semantic feature codes are represented here during successful encoding. Together, the data further suggest that this initial extraction of visual semantic features is important for the effective encoding of memories of specific objects, but not false recognition of similar objects.

More anteriorly, taxonomic categorical representations in aVTC and LIFG, as well as (in the searchlight analysis) RIFG, predicted forgetting of studied items. In an exploratory result, LPrC showed a similar pattern. These findings support the idea that coarse-grained domain-level semantic processing is detrimental to memory for specific objects. Bilateral IFG typically shows strong univariate subsequent memory effects for nameable object stimuli (Kim, 2011). It is thought to support selection and control processes involved in elaborative semantic encoding (Jackson et al., 2015; Prince et al., 2007). Object-specific semantic information was also represented in this region, but did not predict recognition. In contrast, taxonomic semantic information was not represented on average across trials, but was present only for forgotten items, suggesting that processing this information at encoding was detrimental to memory. One possibility is that domain-level taxonomic processing impeded selection of specific semantic information. Another possibility, in line with the levels of processing principle, is that the object naming encoding task did not strongly engage semantic control operations that promote subsequent memory (Craik and Lockhart, 1972; Otten and Rugg, 2001). Object naming depends on basic-level object-specific processing in the FG, consistent with the current findings (Taylor et al., 2012). Future studies can test this by manipulating cognitive operations at encoding to determine whether the representations promoting later memory are also task-dependent.

The absence of any association between object-specific representations in PrC and encoding was unexpected, although we replicated Clarke and Tyler (2014)’s central finding that PrC represents object-specific semantic features. The PrC encodes complex conjunctions of visual (Barense et al., 2012; Bussey et al., 2002) and semantic features (Bruffaerts et al., 2013; Clarke and Tyler, 2014) that enable fine-grained object discrimination and may contribute to later item memory (Brown and Aggleton, 2001; Yonelinas et al., 2005). As the object-specific semantic model fit embodied both shared and distinctive feature information, we ran a further, univariate analysis to examine the directional effect of shared features (concept confusability). We did not replicate Clarke & Tyler's (2014) finding that PrC activation was higher overall for more confusable objects, interpreted in terms of feature disambiguation. However, we found preliminary evidence that in both PrC and aVTC, activity correlating with concept confusability predicted forgetting of studied objects. This result is consistent with our finding that concept confusability strongly impairs true recognition, as well as discrimination between studied objects and lures (see also Chapter 2; Naspi et al., 2020), results replicated here. These data also suggest an interpretation of Davis et al.'s (2020) report that semantic feature model fit in PrC predicted later true recognition of object concepts when their pictures were forgotten, which may correspond to nonspecific encoding.

An important and novel feature of our study is the investigation of the representational content associated with encoding of false memories. Our results revealed that weak visual representations coded in EVC and extending to LITG predicted later false recognition (Fig. 5), and model fit differed significantly from true recognition. This supports fuzzy-trace theory’s proposal that visual detail is encoded in specific memory traces that confer robustness to later true recognition (Brainerd and Reyna, 2002). Several univariate fMRI studies of memory retrieval have shown greater early and late visual cortex activation for true than false memories of objects (Dennis et al., 2012; Karanian and Slotnick, 2017, 2018; Schacter and Slotnick, 2004). Of the few encoding studies, two have found occipital activation predicting true but not false recognition (Dennis et al., 2008; Kirchhoff et al., 2000; Pidgeon and Morcom, 2016; but see Garoff et al., 2005). Here, we not only show that visually specialized regions are engaged more when encoding true than false memories, but also characterize the visual features involved. Thus, insufficient early visual analysis at encoding leads to poor mnemonic discrimination of similar lures. This may prevent later recollection of details of the studied item that would allow people to reject the similar lures (recollection rejection; Brainerd et al., 2003). The RSA result is also consistent with the behavioral increase in false recognition for more visually confusable objects (see also Chapter 2; Naspi et al., 2020).

We did not find any evidence here that semantic processing contributes to false memory encoding, and in FG, semantic feature representations impacted true memory encoding more strongly. Clearly, we cannot place weight on the null result, and our models did not comprehensively address all potential semantic processes but focused on concept-level processes we have shown to contribute behaviorally in this task (see Chapter 2; Naspi et al., 2020). Lateral and ventral temporal regions previously implicated in false memory encoding in verbal tasks did not show significant effects here (Chadwick et al., 2016; Dennis et al., 2007). These areas may support higher-level verbal semantics linking studied items to lures. Nonetheless, both in the current task and following deep semantic judgments at encoding (see Chapter 2; Naspi et al., 2020), concept confusability reduced lure false recognition relative to novel objects as well as true recognition. An intriguing possibility is that the semantic processes reducing lure false recognition operates at retrieval rather than at encoding. This hypothesis will be tested using RSA analysis of retrieval phase brain activity in this task.

In conclusion, we have revealed some of the visual and semantic representations that allow people to form memories of specific objects and later reject similar novel objects. This is the first – to our knowledge – preregistered study of neural representations in memory encoding, and the first probe of representations predicting false recognition. Using previously validated representational models, we were able to disentangle low-level image properties from semantic feature processing. The data provide novel support for theoretical assumptions implicating visual detail in specific memory encoding, but suggest that semantic information may contribute to specific as well as gist memory. Our approach offers a path by which future studies can evaluate the respective roles of encoding and retrieval representations in true and false memory.

Chapter 4: Reducing Emphasis on Shared Semantic Information Affects Object Mnemonic Discrimination in Older Adults

4.1 Abstract

Older adults are prone to misrecognize objects that are semantically and perceptually similar to those they have previously encountered. These errors are typically attributed to shared semantic gist, but may also reflect inefficient encoding of perceptual features. In two experiments, we investigated age-related differences in the effects of multiple objective semantic and perceptual measures that have been found to be important determinants of mnemonic discrimination in young adults. We quantified semantic similarity at two different levels – concept confusability, indexing each concept's tendency to activate other similar concepts via shared features, and item exemplarity, indexing the strength with which depicted objects activated their concepts – as well as low-level visual confusability. In Experiment 1 (*N*= 60, 60), participants studied single object images for each basic level concept, followed by a recognition memory test including studied items, similar lures, and novel items. As in young adults, true recognition was lower for more confusable concepts, while images with higher rated exemplarity were more likely to be falsely recognized. The two groups did not differ significantly in true or false recognition. However, in Experiment 2 (*N* = 60, 60) when multiple exemplar objects were studied for each concept, true and false recognition were both greater in the older than the young group. Critically, the negative effects of concept confusability on both true and false recognition were attenuated in the older relative to the young group, although the effects of item exemplarity and visual confusability of individual items did not differ according to age. The results point to reduced spontaneous processing of object-specific conceptual features in older people, which would be consistent with a reduction in general cognitive resources, but inconsistent with the assumption that gist, or relational processing, is always increased as a result.

4.2 Introduction

Difficulty with memory for details is one of the major complaints of older adults (Hertzog, 2002; Lineweaver & Hertzog, 1998). In the laboratory, this difficulty is manifested particularly by a reduction in mnemonic discrimination between previously studied items and similar lures (Fraundorf et al., 2019; Koutstaal & Schacter, 1997), as well as by impaired recollection of details and contextual information (Spencer & Raz, 1995). Substantial age-related differences are also seen in tasks involving working memory, attention, inhibition, and processing speed, all of which can be considered as types of executive function (Grady, 2012). Nevertheless, some aspects of cognition are maintained with age, notably semantic memory (Umanath & Marsh, 2014). It has therefore been proposed that older adults rely more than young adults on semantic memory to support declining episodic memory, and are more likely to falsely recognize semantically similar lures in mnemonic discrimination tasks (Koutstaal et al., 2003; Reder et al., 2007; Umanath & Marsh, 2014). Without operationalizing semantic processing, it is difficult to draw clear conclusions on its contribution to support of memory since perceptual attributes may also play a role. The current study investigated the degree to which age-related differences in mnemonic discrimination for objects are attributable to differences in their processing of semantic, or perceptual information, using objective measures of these dimensions that we recently showed to be important determinants of young adults' memory performance (see Chapter 2; Naspi et al., 2020).

Older adults are more prone to false recognition than young adults when they are asked to remember specific visually-presented objects that are semantically and perceptually similar to the lures encountered at test (Koustaal et al., 1999; Koutstaal & Schacter, 1997; Pidgeon & Morcom, 2014; Yassa & Stark, 2011). In the typical categorized pictures task, in which lures represent different exemplars of studied basic level concepts (e.g., a different cat), age-related differences are prominent, with older adults false recognizing up to .60–.70 of lures compared with .25–.30 in younger adults (Koustaal et al., 1999; Koutstaal & Schacter, 1997; Koutstaal et al., 2001). Similar effects are often observed for verbal stimuli (Balota et al., 1999; LaVoie & Faulkner, 2000; Norman & Schacter, 1997; but see Burnside et al., 2017; Gallo 2006; Pansuwan et al., 2020). This effect of age has been attributed to an increased reliance on semantic gist, i.e., encoded semantic properties of studied items that are shared with semantically related lures (Brainerd & Reyna, 2002; Koutstaal et al., 2003; Tun et al., 1998) However, perceptual properties may also contribute to age-related differences in mnemonic discrimination (Koutstaal & Schacter, 1997; Pidgeon & Morcom, 2014).

As mentioned earlier, one broad proposal is that older adults' memory depends more on semantic knowledge than younger adults' memory. This may reflect not only preservation of knowledge in aging, but the acquisition of more knowledge over the lifespan (Reder et al., 2007; Umanath & Marsh, 2014; Zacks & Hasher, 2006). This knowledge can support older people's memory (Castel, 2005; Castel et al., 2013; Matzen & Benjamin, 2013; McGillivray & Castel, 2017), but can also lead to confusions among items that are semantically similar (Koutstaal et al., 2003). According to this view, older adults' performance should be driven more by semantic similarity between studied items and lures. Koutstaal et al. (2003)'s second experiment supported this idea since older adults were more likely than young adults to falsely recognize meaningful lures that were semantically and perceptually similar to studied familiar objects than abstract lures that were perceptually similar to studied unfamiliar abstract shapes (for a replication, see Pidgeon & Morcom, 2014). Although this theory can account well for some memory errors made by older adults, there are conflicting data. One challenge is that prior knowledge does not always benefits older people's memory more (Badham et al., 2012; Badham et al., 2016). For example, in a series of experiments that elicited emphasis on pre-existing semantic knowledge (i.e., encoding of familiar versus novel proverbs, or meaningful versus meaningless scenes), Badham et al. (2016) found that young and older adults can utilise prior knowledge to the same extent, and that the performance in an old-new recognition memory test benefited both age groups similarly, at least for true recognition.

A more specific proposal to explain age-related errors in mnemonic discrimination is that older adults particularly emphasize basic-level semantic information when processing pictures, by focusing on the concept name (Koutstaal et al., 2003). This semantic categorization account suggests that what is critical is meaning overlap between studied and lure exemplar items that share the same basic-level concept. It is assumed that semantic categorization at this level detracts from the encoding of further, perceptual features. This proposal has not been systematically tested: initial evidence from Koutstaal et al.'s (2003) first experiment suggested that older adults were more likely to false alarm to lure pictures of ambiguous objects when these pictures were accompanied by disambiguating verbal labels at study and at test, but not when the pictures were presented alone. However, it was not established that this effect on lure recognition survived a correction for group differences in response bias. One way to test this account further is to examine the effects of age on influences of semantic relations at multiple different levels in the same task, as we outline below.

Any semantically-based account of age effects on mnemonic discrimination is further challenged by a few studies that have shown increased false recognition in older people for perceptually similar but semantically unrelated lures (Burnside et al., 2017; Pidgeon & Morcom, 2014; but see Pansuwan et al., 2020). In one study, Boutet et al. (2019) found increased false alarms in older adults for perceptually similar lures representing different exemplars of studied chairs or inverted faces. A strength of this study was that perceptual similarity was objectively measured in terms of luminance and contrasts, and semantic similarity assumed to be matched. These findings, as well as the age effects for semantically similar lures, can also be explained by the alternative view that older adults are more sensitive to generic similarity. Wilson et al. (2006) proposed that aging leads to inefficient pattern separation by the hippocampus. Pattern separation is a process whereby stimuli produce distinct neuronal representations of similar input at encoding to support later mnemonic discrimination between studied and novel items (Stark et al., 2013; Wilson et al., 2006). In aging, overlapping neuronal memory representations are thought to be due to multiple dimensions of similarity, without a special focus on semantic similarity (Wilson et al., 2006; Yassa et al., 2011; Yassa et al., 2011). Greater overlap of representations may account for increased false recognition in older adults (Toner et al., 2009). Neuroimaging studies in humans, as well as studies with animals, support this proposal (Reagh et al., 2018; Yassa et al., 2011; Yassa et al., 2011). At the behavioural level, a critical test of this account is whether the effects of perceptual as well as semantic similarity on mnemonic discrimination increase in aging.

4.2.1 The Current Research

The studies reviewed above suggested that semantic and/or perceptual relations between studied items and lures are a key driver of age-related increases in false recognition and reductions in mnemonic discrimination. The two experiments reported here used objectively quantified multiple dimensions of semantic and perceptual similarity to investigate their impact on age-related differences (see Chapter 2; Naspi et al., 2020). We included variables that we previously showed to impact memory in young adults, examining the effects of semantic similarity on memory at two distinct levels, as well as low-level visual confusability. The first, concept confusability, indexed semantic similarity at the *concept level* in terms of the overlap of a concept's semantic features with the features of other concepts. These shared semantic features are informative about domain-level and superordinate category membership, but not so helpful for identifying individual concepts. For example, processing a lion's properties < has legs >, < has eyes >, and < has a tail > would emphasize its land animal category as these features are shared with other animals like tiger and dog. This measure derives from the Conceptual Structure Account's distributed feature-based model of semantic processing (Tyler & Moss, 2001). In contrast, item exemplarity is a measure of semantic overlap at the *item level,* in this case between individual exemplars and their basic level concepts. This measure was derived from participant ratings of how good an exemplar was of its concept. Lastly, visual confusability was measured using an index from the HMax computational model of vision that captures low-level visual properties of objects (Riesenhuber & Poggio, 1999; Serre et al., 2007). In this typical mnemonic discrimination task, studied items and lures are different exemplar images of the same basic-level concept and share semantic and perceptual similarity at the item level.

Our previous analysis in young adults (see Chapter 2; Naspi et al., 2020) showed that concept confusability, which captured generic processing across concepts, reduced both recognition of studied items and false recognition of similar lures. In contrast, both item exemplarity and low-level visual confusability captured shared characteristics of studied and lure items, and both increased false recognition. We had three alternative hypotheses for the age effects. If semantic processing of objects is emphasized in general in older compared to young adults, we predicted older adults would show enhanced impact of both concept confusability and item exemplarity on mnemonic discrimination. Alternatively, based on the semantic categorization account (Koutstaal, et al., 2003), we reasoned that age effects at concept- and-item levels would differ because the memoranda are nameable objects. If older adults focus more on processing individual basic-level concepts, they should show an increase in false recognition relative to the young for higher exemplarity lures. In contrast, the greater reliance on item-level processing may *reduce* processing of shared semantic relations across concepts, so the older group should show reduced effects of concept confusability. Lastly, if age effects on mnemonic discrimination reflect impaired hippocampal pattern separation, we predicted that older adults would be more prone to false recognition of lures that were more similar to their studied items at the item level, whether the similarity was semantic or perceptual. Therefore, they should show greater effects of both item exemplarity and visual confusability than the young, suggesting a generalized impairment in the specificity of memory representations for individual items. In a first, exploratory, experiment we presented a single exemplar of each basic level concept at study, and participants were later tested on a single lure at test. This did not reveal any age-related differences, so in a second experiment we increased the number of studied exemplars of each basic level concept.

4.3 Experiment 1

4.4 Materials and Methods

4.4.1 Participants

The study included one hundred and twenty participants. 60 young adults aged 18-33 years and 60 older adults aged 65-80 years. Demographic characteristics are summarized along with baseline cognitive test results in Table 1. The young adult data is the same as in Chapter 2 (Naspi et al., 2020). We excluded eleven further young adults from data analysis: 7 due to errors in stimulus lists or data acquisition issues, 1 due to misunderstanding of instructions, and 3 who did not meet the inclusion criterion for English fluency. Three further older adults were also excluded: 1 due to data acquisition issues, and 2 due to poor memory performance. The sample size was determined *a priori* using the simR package in R (Version 1.0.4; Green & Macleod, 2016). We powered the study for an interaction of aging (older vs. young) × condition (lure vs. new) × concept confusability, based on results on young adults suggesting an effect size equivalent to Cohen’s $d = 0.17$ ($OR = 1.35$). With $N = 120$ (60 young adults, 60 older adults) we had 87.50 % power to detect such an effect at alpha = .05. Inclusion criteria were fluency in English (spoken since at least the age of 5 years), normal or corrected-to-normal vision, and a good self-reported health. Participants were compensated financially or with course credits. They were contacted by local advertisement and provided informed consent. The study was approved by the University of Edinburgh Psychology Research Ethics Committee (Ref. 278-1617/1).

Table 1

Demographic and Cognitive Test Data for Experiment 1 and Experiments 2

	Experiment 1		Experiment 2	
	Young Adults (*N* = 60)	Older Adults (*N* = 60)	Young Adults (*N* = 60)	Older Adults (*N* = 60)
Age	21 (2.2)*	71 (3.6)*	21.2 (3.2)*	71.2 (3.6)*
Sex	45 F; 15 M*	34 F; 26 M*	48 F; 12 M	43 F; 17 M
Years of Education	15.63 (1.89)	16.27 (2.99)	15.97 (2.41)	15.75 (3.65)
Trail Making Test	22.46 (14.56)	27.65 (18.17)	25.59 (15.11)	29.43 (19.33)
Digit Span Forward	6.48 (1.17)*	7.18 (1.41)*	6.97 (1.44)	7.23 (1.50)
Digit Symbol	87.42 (12.42)*	70.52 (15.37)*	89.55 (14.09)*	68.15 (13.77)*
Standardized WTAR	116.80 (9.73)*	120.88 (6.21)*	115.50 (9.23)*	121.00 (5.22)*
Semantic Fluency	27.42 (5.64)	26.75 (7.03)	23.80 (6.16)	24.98 (6.03)
FAS Fluency	40.97 (11.46)*	56.15 (15.20)*	42.17 (10.29)*	51.70 (14.69)*

Note. Mean (standard deviation). * denotes significant within-experiment age difference ($p < .05$). The Trial Making Test score reflects the average difference in seconds between part B and A; the Digit Span Forward score represents the number of correct digits produced forwards; the Digit Symbol score consists of the number of items completed in 120 s in a substitution task; the Wechsler Test of Adult Reading (WTAR) yields a total number of irregular words correctly pronounced. This raw score is transformed to an age-adjusted standard score, which predicts IQ ($M = 100$; $SD = 15$); the Semantic and FAS Fluency tests were scored summing the total words produced for the animal category and over three letters, respectively; F = Female; M = Male; See Cognitive Tests section for details of statistical analyses.

4.4.2 Stimuli

The stimuli were identical to those reported in Chapter 2 (Naspi et al., 2020). Briefly, each study list comprised 120 images of exemplars of different basic level concepts. Each test list consisted of 180 items: 60 studied images, 60 similar lures (i.e., different exemplars of previously studied basic level concepts), and 60 novel items (i.e., novel basic level concepts not previously studied). Three filler trials prefaced both study and test phases. We generated 6 different study and test lists which fully counterbalanced the allocation of the basic level concepts and the two sets of images to conditions (studied, lure, and novel).

4.4.3 Procedure

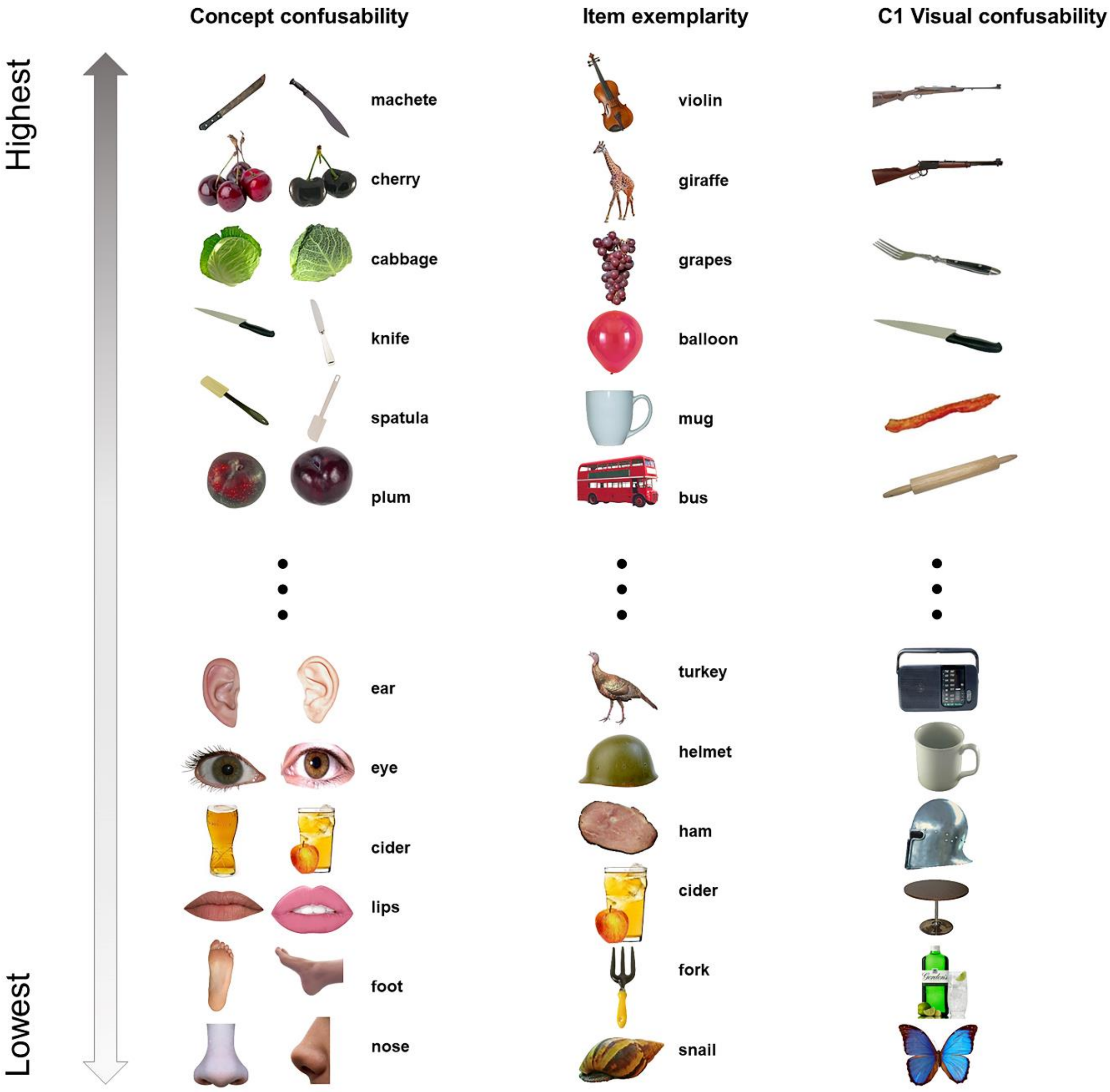
The experiment consisted of a single study phase followed by a recognition test phase. Between study and test phases, all participants completed the following standardized cognitive tests (see Table 1) for 15 minutes: the Trail Making Tests (Tombaugh, 2004), the Digit Span Forward and Digit Symbol Coding of the Wechsler Adult Intelligence Scale IV (WAIS-IV^{UK}; Wechsler, 2008), and the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001). Raw WTAR scores were converted to standard scores based on the UK standardization sample. Lastly, we administered the phonemic (FAS) and semantic (animal naming) fluency tests (Tombaugh et al., 1999). Participants viewed one image at a time during the study phase and rated their pleasantness from 1 (very unpleasant) to 5 (very pleasant) using a button press response. Trials were self-paced. At test, we presented one image at a time every 3 s and participants judged each as “old” or “new” indicating the confidence of their judgment. For full details, see Chapter 2 (Naspi et al., 2020).

4.4.4 Variables of Interest

The concept- and-item level variables of interest are illustrated in Figure 1 and described below. For further background on these measures, see Chapter 2 (Naspi et al., 2020).

Figure 1

Schematic Depiction of Stimuli Used in Both Experiments



Note. Columns show individual exemplars with the highest (top) and lowest (bottom) similarity scores on each concept- and-item level experimental variable. For concept confusability, the basic level concept name is given alongside images of both exemplars representing it in Experiment 1. High confusability concepts share more semantic features with other concepts. Item exemplarity is an index of similarity between the depicted exemplar and the concept representation. C1 visual confusability defines confusability of an item as the similarity with its most similar neighbor in the set and was obtained from gray-scaled version of the images depicted in Figure 1. For definitions see Variables of Interest section.

4.3.4.1 Concept Level

Concept confusability. This measure emphasizes feature sharedness between concepts that share many semantic features (Clarke & Tyler, 2014; Devereux et al., 2014; Naspi et al., 2020). Semantic similarity between concepts was computed as the cosine angle between feature vectors in a semantic feature matrix in which each concept was represented by a binary vector indicating whether each feature was associated with the concept (1) or not (0). Concept confusability with all the other concepts in the set was then calculated by a weighted sum of the similarities in which each weight was the between-concept similarity itself, i.e., the sum of squared similarities (see Figure 1). 30 participants also collected concept familiarity (not reported here).

4.3.4.2 Item Level

Item exemplarity. We obtained a measure of item exemplarity using a rating task in which pictures were presented together with their verbal label. Following Taylor et al. (2012), we asked participants to judge "*how closely each picture resembles your mental image of the object*", giving their response on a 7-point scale (1 = poor picture of concept word, 7 = excellent picture of concept word). 30 participants provided the item exemplarity scores, which were averaged to give a single score per picture (see Figure 1 of Chapter 2; Naspi et al., 2020).

Visual confusability. The low-level visual attributes of each picture, for a total of 360 images (2 per concept) were used to represent our basic level concepts. We first extracted HMax estimates of low-level visual object information: a C1 response related to early visual cortex (V1/V2). The response vector from the C1 layer was computed for grey-scaled versions of each image. Similarity between pictures was then calculated using the Pearson correlation coefficients between vectors. For each image we defined confusability as the similarity value with its most similar picture (i.e., the nearest neighbor; see Figure 1).

4.4.5 Nuisance Variables

Mnemonic discrimination is influenced by a range of other visual, phonological, lexical, and semantic factors in addition to the semantic and perceptual confusability measures of interest here. We controlled for the effect of the following nuisance variables, described in more detail in the Supplemental Material of Chapter 2: forward and backward associative strength estimated using a continuous association task (De Deyne & Storms, 2008; Nelson et al., 2004), word frequency (van Heuven et al., 2014), concreteness (Brysbaert et al., 2014), age of acquisition (Brysbaert & Biemiller, 2017), phonological neighborhood density (Baayen et al., 1995), the number of non-white pixels, color entropy (Chouinard & Goodale, 2012), and concept familiarity (derived from our rating task; see also Taylor et al., 2012).

4.4.6 Statistical Analysis

We tested our hypotheses using a generalized linear mixed-effect model analyses with the function glmer from the lme4 package in R (version 1.1-17, Bates et al., 2015). This analysis, that we used in Chapter 2 (Naspi et al., 2020) on the same data of young adults, generates log-odds-ratio coefficients that are formally proportional to *d'* in a Gaussian signal detection analysis ($d' \approx .6$ logOR; DeCarlo, 1998; Wright et al., 2009). In the Results section we therefore report effect sizes in terms of *d'* equivalent for ease of comparison with other studies. In the analyses, we tested for age-related differences in mnemonic discrimination driven by concept-and item level variables that predicted memory performance in the initial analysis in the same data of young adults (see Chapter 2; Naspi et al., 2020). These predictors of interest were concept confusability, item exemplarity, and C1 visual confusability. To test specific predictions about age-related differences for studied items and lures, we set the reference level for the condition factor to "novel" and for the age group to "young". Thus, we assessed age-related differences in memory sensitivity (*d'* equivalent) using simple contrasts that yielded a) modulations of older adults' sensitivity for studied items relative to the young, reflecting true memory), and b) modulations of the older adults' sensitivity for related lures relative to the young, reflecting false memory). The novel items provided c) an index of modulations of baseline false alarms for the older relative to the young group, equivalent to differences in response criterion (DeCarlo, 1998). In a final set of contrasts, we set the reference level for the condition factor to "lure", keeping the reference level for age group as "young", to assess modulations of the d) the overall older adult's sensitivity relative to the young equivalent to *d'*). This allowed us to evaluate the effect of the semantic and perceptual variables on overall mnemonic discrimination performance in older relative to young adults, to determine the net effect of their modulations of true and false memory (Koutstaal & Schacter, 1997; Loiotile & Courtney, 2015; Naspi et al., 2020).

Consistent with our previous analysis on the same data of young adults in Chapter 2 (Naspi et al., 2020), we specified *a priori* random intercepts of both participants and concepts

(Matuschek et al., 2017). In the fixed part, our variables of interest were age group (young, older), condition (studied, lure, novel), one concept-level variable (i.e., concept confusability), and two item-level variables (i.e., item exemplarity and C1 visual confusability). Within the concept-level and item-level partitions we also included the corresponding interactions with age and condition. To minimize model complexity, and because the nuisance variables were moderately-to-highly intercorrelated, we used the 7 principal components (see Supplemental Material of Chapter 2, Table 5) that we obtained in the previous analysis on the same group of young adults with principal components analysis (PCA) using the prcomp function in R. PCA with varimax rotation produced a 7 factor solution which accounted for 86.14% of variance (this approach was the same as used in Chapter 2; Naspi et al., 2020). All the continuous predictor variables were standardized, and the resulting β coefficients representing log-odd-ratios were used to calculate the corresponding *d’* coefficients.

4.5 Results

4.5.1 Cognitive Tests

Cognitive test results for participants in Experiment 1 are summarized in Table 1. Young and older adults did not differ in years of education ($t(118) = 1.39$, $p = .168$), Trail Making Test ($t(118) = 1.72$, $p = .087$), and Semantic Fluency ($t(118) = .57$, $p = .568$. As expected, young adults outperformed older adults on the Digit Symbol Coding task ($t(118) = 6.62$, $p < .001$), while older adults showed a better performance on the Digit Span Forward ($t(118) = 2.96$, $p = .004$), WTAR ($t(118) = 2.74$, $p = .007$), and the FAS Fluency ($t(118) = 6.18$, $p < .001$). Chi-squared test of independence revealed that the sex distribution differed between age groups ($\chi2(1) = 4.48$, $p = .034$), i.e., in the young group there were more female participants than in the older group.

4.5.2 Memory Performance

The results of the mixed effects analysis are shown below in Table 2, and illustrated in Figure 2. The results for effects including age-related differences are listed in Table 2, and analysis outcomes for young and older groups separately are reported within the text. Age-invariant effects are given in Supplemental Material of Chapter 4, Table 13. The pattern of findings was qualitatively similar when only high confidence mnemonic discrimination was analyzed, so these results are not reported (https://osf.io/ndk83/). Participants' responses were collapsed across confidence judgments as results. Coefficients in Table 2 represent log-odds-ratios with the corresponding *d'* effects. All the *p*-values reported below are FDR-multiple comparison corrected (Benjamini & Hochberg, 1995).

Table 2

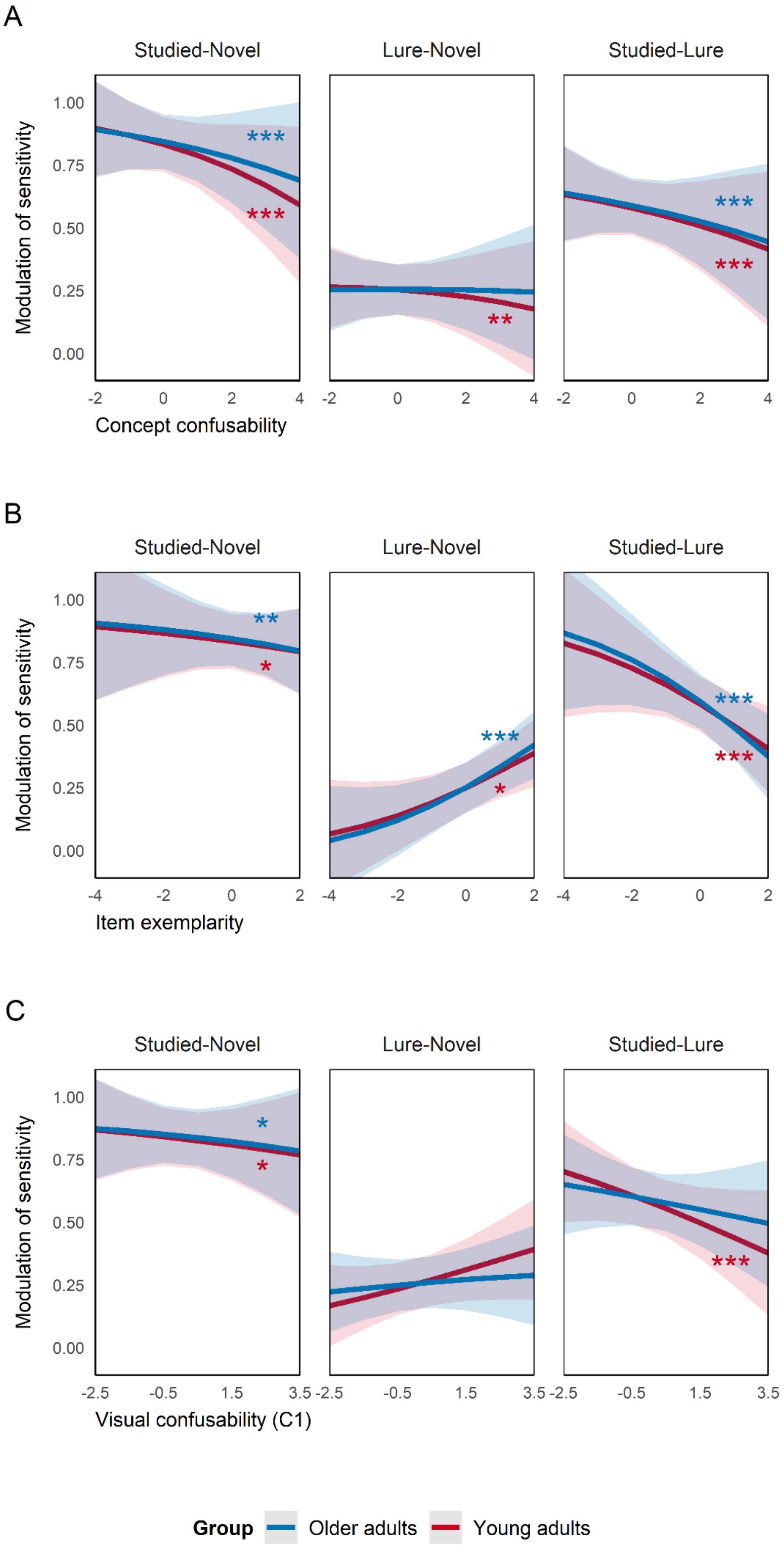
Age-Related Differences in Experiment 1

Variable	Estimate	*d'*	*SE*	z-value	*p*
• Young Adults/Novel Items as Baseline					
(Intercept)	-2.38	-1.35	0.11	-21.65	<.001
Older	-0.20	-0.11	0.14	-1.44	.249
Older × Lure	0.15	0.08	0.10	1.42	.249
Older × Studied	0.16	0.09	0.12	1.36	.261
Older × Concept Confusability	-0.10	-0.06	0.08	-1.30	.263
Older × Visual Confusability	0.09	0.04	0.08	1.12	.303
Older × Item Exemplarity	-0.11	-0.06	0.09	-1.24	.263
Older × Lure × Concept Confusability	0.10	0.06	0.10	1.05	.320
Older × Studied Concept Confusability	0.14	0.09	0.11	1.23	.263
Older × Lure × Visual Confusability	-0.18	-0.09	0.10	-1.88	.132
Older × Studied × Visual Confusability	0.00	0.01	0.11	0.03	.974
Older × Lure × Item Exemplarity	0.13	0.07	0.11	1.25	.263
Older × Studied × Item Exemplarity	-0.04	-0.02	0.12	-0.33	.770
• Young Adults/Lure Items as Baseline					
(Intercept)	-0.69	-0.42	0.10	-6.96	<.001
Older × Studied	0.02	0.01	0.10	0.19	0.890
Older × Studied × Concept Confusability	0.04	0.02	0.09	0.37	0.773
Older × Studied × Visual Confusability	0.18	0.10	0.10	1.89	0.121
Older × Studied × Item Exemplarity	-0.18	-0.10	0.11	-1.65	0.168

Note. In the upper part of the table, the reference level of age group is set to "young adults", and condition is set to "novel items". In the lower part of the table, the reference level of age group is set to "young adults", and condition is set to "lure items". Parameter estimates (logOR), *d'* equivalent, standard errors (SE), z-values, and false discovery rate (FDR) corrected *p*-values are listed for condition, concept-level, and item-level variables in the linear mixed model.

Figure 2

Effects of Semantic and Perceptual Variables on Mnemonic Sensitivity in Experiment 1


A
Studied-Novel
Lure-Novel
Studied-Lure
Modulation of sensitivity
1.00
0.75
0.50
0.25
0.00
***
***
**
***
***
-2
0
2
4
Concept confusability
B
Studied-Novel
Lure-Novel
Studied-Lure
Modulation of sensitivity
1.00
0.75
0.50
0.25
0.00
**
*
***
*
***
***
-4
-2
0
2
Item exemplarity
C
Studied-Novel
Lure-Novel
Studied-Lure
Modulation of sensitivity
1.00
0.75
0.50
0.25
0.00
*
*
***
-2.5
-0.5
1.5
3.5
Visual confusability (C1)
Group
Older adults
Young adults

Note. The plot lines represent the effect of the predictor variables on the probabilities of endorsing studied items as “old” relative to novel items, lures as “old” relative to novel items, and studied items as “old” relative to lures in young (red lines) and older adults (blue lines). Panel A, B, and C show the effects of concept confusability, item exemplarity, and C1 visual confusability in Experiment 1 in young and older adults. Coloured asterisks indicate effects that were significant within each age group separately. Black asterisks at the top within the panels indicate significant age-related differences. See Material and Methods, Variables of Interest, and Results for details. * $p < .05$; ** $p < .01$; *** $p < .001$ (false discovery rate-corrected).

4.5.2.1 Sensitivity for Studied Relative to Novel Items

We did not find any significant age-related effects on sensitivity for studied items in Experiment 1 (Table 2, Figure 2). There was no difference in overall sensitivity (for interaction of older group with studied items on $d' = .09$; 95% CI [-.03, .21]). None of the concept-and item level experimental variables modulated the age-related differences in sensitivity for studied items, and separate comparisons for each group (Figure 2) showed similar modulations of concept- and item-level variables on sensitivity for studied items. Images whose concepts were more confusable with other concepts in the set were less likely to be remembered in both groups (Figure 2A; interaction of concept confusability with studied items on $d' = -.18$; 95% CI [-.27, -.10] for older adults; $d' = -.28$; 95% CI [-.36, -.20] for young adults). Pictures judged with high exemplarity were also less likely to be remembered in both older (Figure 2B; interaction of item exemplarity on $d' = -.13$; 95% CI [-.22, -.04]) and young adults (Figure 2B; interaction of item exemplarity with studied items on $d' = -.10$; 95% CI [-.19, -.01]). Lastly, both groups were less likely to recognize more visually confusable items (Figure 2C; interaction of C1 with studied items on $d' = -.10$; 95% CI [-.18, -.01] for older adults; $d' = -.12$; 95% CI [-.20, -.04] for young adults).

4.5.2.2 Sensitivity for Lure Relative to Novel Items

In line with the findings for true recognition, there were no significant age-related effects on sensitivity for lures in Experiment 1 (Table 2, Figure 2). We found no group difference in overall sensitivity (interaction of older group with lure items on $d' = .08$; 95% CI [-.03, .19]). Again, none of the concept-and item level experimental variables modulated the age-related differences in sensitivity for lure items (Table 2; Figure 2 for differences in slopes). In older adults, concept confusability did not significantly modulate sensitivity for lure items, unlike in the young (Figure 2A; interaction of concept confusability with lure on $d' = -.04$; 95% CI

[-.12, .04]; in young group, modulation of *d'* = -.11; 95% CI [-.18, -.03]). At the item-level, false recognition was more likely for lures rated as better exemplars of their concept in both older and young adults (Figure 2B; interaction of item exemplarity with lure on *d'* = .19; 95% CI [.11, .28] for older adults and *d'* = .13; 95% CI [.05, .22] for young adults). Unexpectedly, C1 visual confusability did not modulate sensitivity for lure items in either group (Figure 2C).

4.5.2.3 Sensitivity for Studied Relative to Lure Items

We did not observe age-related differences in sensitivity for studied items relative to lures (Table 2; interaction of older adults with studied items on *d'* = -.01; 95% CI [-.09, .12]). Moreover, none of the concept-and item level experimental variables modulated the age-related overall differences in sensitivity for studied relative to lure items (Table 2; Figure 2). Both concept- and-item level variables modulated this effect in each group (Figure 2). Mnemonic discrimination was poorer for items with high concept confusability in both groups (Figure 2A; interaction of concept confusability with studied items on *d'* = -.14; 95% CI [-.22, -.07] for older adults; *d'* = -.17; 95% CI [-.24, -.10] for young adults). At the item level, objects judged as having high item exemplarity were also less well discriminated in both groups (Figure 2B; interaction of item exemplarity with studied items on *d'* = -.32; 95% CI [-.40, -.24] for older adults; *d'* = -.23; 95% CI [-.31, -.15 for young adults). However, C1 visual confusability did not modulate older adults' discrimination of studied items from similar lures (Figure 2C; interaction of C1 with studied items on *d'* = -.08; 95% CI [-.16, -.01]), while it did so in the young adults (Figure 2C; interaction of C1 with studied items on *d'* = -.20; 95% CI [-.27, -.12]).

4.6 Discussion

In Experiment 1, we investigated the effects of semantic and perceptual similarity on age-related differences in mnemonic discrimination, using objective measures that in Chapter 2 we have found to be important predictors in the same data of young adults (Naspi et al., 2020). Unexpectedly, we did not observe significant age-related differences in memory, or differences in the effects of the semantic and perceptual similarity. This result contrasts with previous studies that used a single studied exemplar, and reported age-related increases in false recognition (Koutstaal et al., 2003; Pidgeon & Morcom, 2014; Reagh et al., 2016; Yassa et al., 2011), and in some cases also decreases in true recognition (Koutstaal & Schacter, 1997).

One factor that is known to account for some of the variance in cognitive ability among older adults is education. More education has been associated with a slower longitudinal age-related decline in memory (Colsher & Wallace, 1991; Evans et al., 1993; Rönnlund et al., 2005). Although the groups in Experiment 1 did not differ significantly in education, the older group had numerically slightly more years of education than the young group. A higher baseline ability in the older group may be more clearly indicated by the older adults' significantly higher forward digit span scores. This test is not thought to index executive aspects of working memory and is typically not sensitive to advancing age (Babcock & Salthouse, 1990; Grégoire & Van der Linden, 1997; but see Hester et al., 2004). The other age-related differences in standardized cognitive test performance were as expected: the older participants showed evidence of cognitive slowing on the Digit Symbol Coding test (Salthouse, 1992) despite their (also typically) higher crystallized ability as indicated by the WTAR verbal IQ estimate and the verbal fluency (FAS) tests (Tomer & Levin, 1993; Yuspeh et al., 1998; but see Bryan & Luszcz, 2000). In light of the null finding for memory differences and the suggestion that the groups might not be perfectly matched, we ran a check analysis comparing a subgroup of 48 young and 48 older participants who were perfectly matched on education, and did not show a significant difference in forward digit span. Even when education and forward digit span were completely matched, we could not find any age-related difference in mnemonic discrimination.

Despite not finding age-related differences, the analysis on the two groups separately did not show identical results. Unlike the young adults, processing shared semantic information across concepts did not modulate older adults' probability to correctly reject similar lures. Although the group difference was not significant, this hints at a possibility that older adults may engage *less* in cross-concept processing of pre-existing knowledge and emphasize more basic level information. However, like the young adults, the older group showed a significant (though numerically slightly smaller) reduction in recognition of studied items when concepts were more confusable with other concepts (Naspi et al., 2020). The older adults' sensitivity for lures also depended on semantic processing at the item level. Similar to young adults, lures judged as more similar to the participants' internal representations were more likely to be misrecognized in older adults, consistent with enhanced activation of item-level semantic representations in both groups. Lastly, although more visually confusable studied items were more likely to be forgotten in both groups, visual confusability did not modulate either young or older adults' ability to discriminate similar lures. This result in the young is inconsistent with the results of the original model, in which this effect was significant (see Chapter 2; Naspi et al., 2020). We consider this finding further in the General Discussion, in light of the results of Experiment 2. In the second experiment we aimed to enhance any underlying age-related differences by examining the effects of our semantic and perceptual variables using a manipulation under which older adults should show a clear overall decrement in mnemonic discrimination.

4.7 Experiment 2

To provide a stronger test of age-related effects of semantic and perceptual confusability on memory, we increased the studied category set size in Experiment 2. Previous studies have shown that increasing the number of studied exemplars of each basic level concept often exacerbates the age differences in false recognition of categorized picture lures (Koutstaal et al., 2003; Koutstaal & Schacter, 1997). Increasing set size is assumed to enhance gist-based memory by emphasizing processing of the studied concepts. However, Pidgeon and Morcom (2014) also observed this effect when increasing the number of abstract pictures without pre-existing semantic information, suggesting that increasing set size might also enhance shared perceptual processing. A second motivation for this manipulation was to clarify the possible role of a recall-to-reject strategy in the negative or null concept confusability effects on memory (see also Chapter 2; Naspi et al., 2020). When only one exemplar is studied per concept, people can avoid gist-related errors when lures trigger recollection of similar studied items, by comparing the recollected information with the lure (e.g., they decide that a white dog was not presented, because they remember having studied a black dog). It was important to establish whether our results would generalize to a situation in which recall-to-reject was prevented. In Experiment 2, participants therefore studied sets of either two or eight different exemplars of the same basic level concept. Our predictions for age-related effects were the same as in Experiment 1.

4.8 Materials and Methods

4.8.1 Participants

The study included one hundred and twenty participants: 60 young adults aged 18-33 years (*M* = 21.2; *SD* = 3.2, 12 male, 48 female) that consisted of the same data included in Chapter 2 (Naspi et al., 2020), and 60 older adults aged 65-80 (*M* = 71.2; *SD* = 3.6, 17 male, 43 female). A further ten young adults were excluded from data analysis: 9 due to technical issues with recording responses, 1 due to poor performance at test (using the preregistered criterion of *d'* for studied item discrimination of less than 3 SD from the mean). Five further older adults were also excluded: 2 due to data acquisition issues, 2 due to poor memory performance, and 1 due to misunderstanding of instructions. Similarly to Experiment 1, we powered the study for an interaction of aging (older vs. young) × condition (lure vs. new) × concept confusability, based on results on young adults suggesting an effect size equivalent to Cohen's $d = .17$ ($OR = 1.35$). With $N = 120$ (60 young adults, 60 older adults) we had now 99.80 % power to detect such an effect at alpha = .05. All participants were fluent English speakers (since at least the age of 5), had normal or corrected to-normal vision, and a good self-reported health. Participants were recruited by local advertisement and provided informed consent. They received either course credit or an honorarium. The study was approved by the University of Edinburgh Psychology Research Ethics Committee (Ref. 278-1617/1).

4.8.2 Stimuli

Except where specified, stimuli were the same as in Experiment 1 (see Chapter 2 for details). Briefly, each study list included 600 items: 480 in the large sized sets (i.e., set size 8) and 120 in the small sized sets (i.e., set size 2). Each test list consisted of 300 items: 120 studied images (60 from set size 8 and 60 from set size 2), 120 similar lures (60 different exemplars of previously studied basic level concepts from set size 8 and 60 from set size 2, and 60 novel basic level concepts not previously studied). Three filler trials prefaced both the study and the test phase. We generated one study and test list for each participant which randomized the allocation of the concepts and their exemplar images to conditions (i.e., studied, lure, and novel) and set size (2 and 8) with the constraint that half the concepts in each condition (item type and set size) were living and half non-living.

4.8.3 Procedure

The experiment consisted of a single study phase followed by a recognition test phase with interspersed standardized cognitive tests for 15 minutes (see Table 3). Stimuli were presented with MATLAB (R2018b, The MathWorks) using PsychToolbox (Kleiner et al., 2007; Version 2.0.14). The procedure was otherwise the same of Experiment 1 except that trials during the study phase were not self-paced, but presented every 3 s.

4.8.4 Statistical Analysis

The variables of interest and nuisance variables were identical to those used in Experiment 1, but the item metrics were recomputed for this larger set of images. Item exemplarity ratings were again collected from the young adults after the test phase. As the concepts were identical to those used in Experiment 1, we used the concept familiarity values collected previously. We also examined *a priori* the modulatory effects of set size (i.e., set size 8 vs set size 2), but as this variable had no significant effects, we focus here on the results collapsed over the two set sizes.

4.9 Results

4.9.1 Cognitive Tests

Cognitive test results for participants in Experiment 2 are summarized in Table 1. Young and older adults did not differ in years of education ($t(118) = .38$, $p = .702$), and unlike in Experiment 1, the numerical values were closely similar. The groups also did not differ in performance on the Trail Making Test ($t(118) = 1.21$, $p = .228$), Digit Span Forward ($t(118) = .90$, $p = .322$), and Semantic Fluency ($t(118) = 1.06$, $p = .290$. As expected, and as in Experiment 1, young adults outperformed older adults on the Digit Symbol Task ($t(118) = 8.41$, $p < .001$), while older adults showed better performance on the WTAR ($t(118) = 4.02$, $p < .001$), and the FAS Fluency ($t(118) = 4.12$, $p < .001$). A chi-squared test of independence revealed that the sex distribution did not differ between age groups ($\chi2(1) = 1.14$, $p = .286$).

4.9.2 Memory Performance

The results of the mixed effects analysis are shown below in Table 3, and illustrated in Figure 3. Table 3 only reports age-related differences. Separate young and older adults' effects are compared within text. Age-invariant effects are reported in Supplemental Material of Chapter 4, Table 14. Coefficients in Table 1 represent log-odds-ratios with the corresponding *d'* effects. All the *p*-values reported below are FDR-multiple comparison corrected (Benjamini & Hochberg, 1995).

Table 3

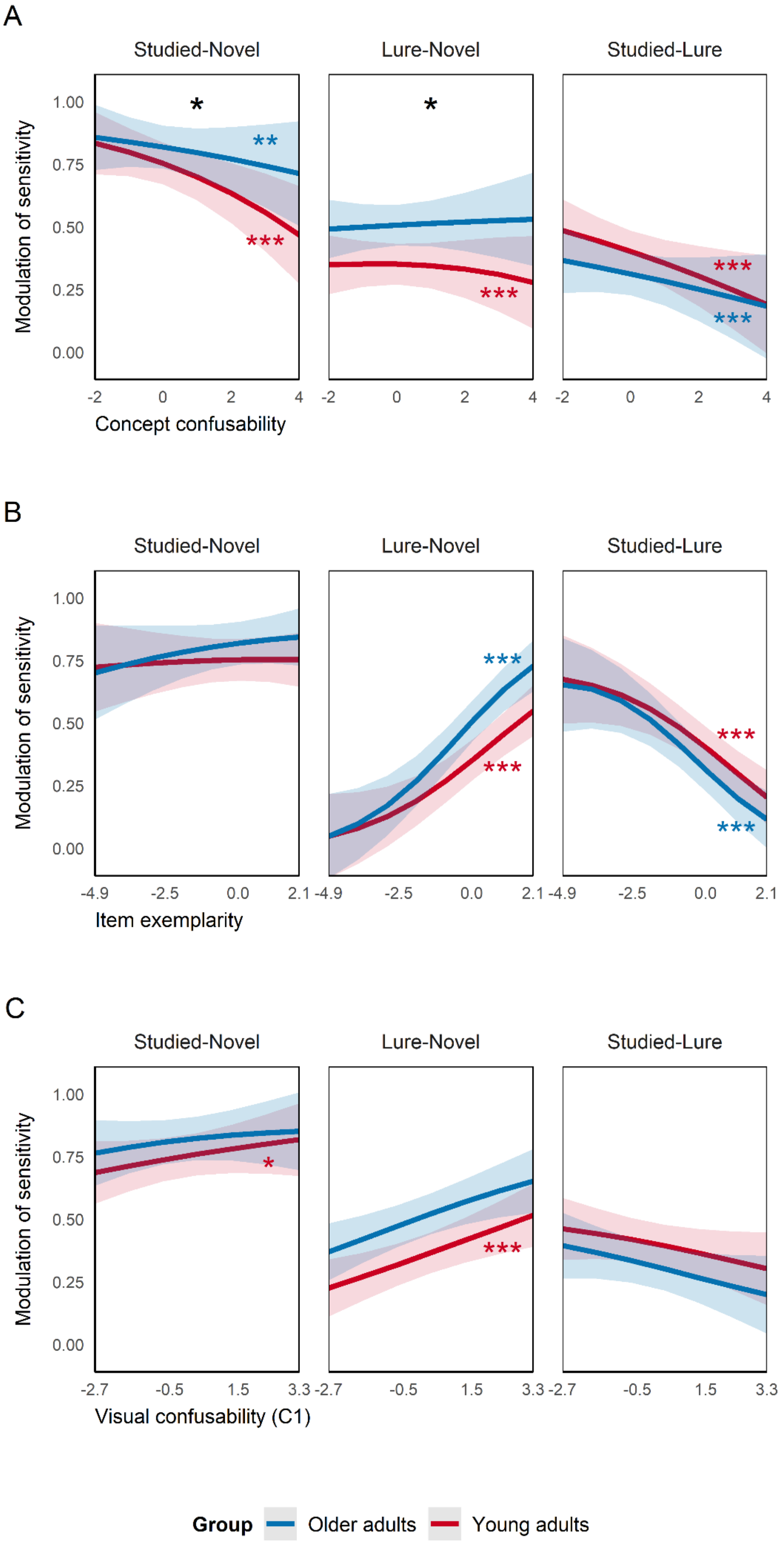
Age-Related Differences in Experiment 2

Variable	Estimate	*d'*	*SE*	z-value	*p*
• Young Adults/Novel Items as Baseline					
(Intercept)	-2.69	-1.55	0.10	-26.43	<.001
Older	-0.09	-0.04	0.14	-0.63	.530
Older × Lure	0.69	0.41	0.10	6.79	<.001
Older × Studied	0.56	0.30	0.11	5.20	<.001
Older × Concept Confusability	-0.26	-0.12	0.09	-2.63	.017
Older × Visual Confusability (C1)	0.16	0.08	0.09	1.74	.124
Older × Item Exemplarity	-0.06	-0.03	0.10	-0.68	.520
Older × Lure × Concept Confusability	0.22	0.11	0.10	2.25	.039
Older × Studied × Concept Confusability	0.28	0.15	0.10	2.78	.012
Older × Lure × Visual Confusability (C1)	-0.13	-0.06	0.10	-1.23	.274
Older × Studied × Visual Confusability (C1)	-0.10	-0.04	0.11	-0.89	.405
Older × Lure × Item Exemplarity	0.17	0.10	0.10	1.62	.140
Older × Studied × Item Exemplarity	0.18	0.09	0.11	1.67	.134
• Young Adults/Lure Items as Baseline					
(Intercept)	-0.35	0.21	0.08	-4.29	<.001
Older × Studied	-0.13	-0.10	0.06	-2.23	.042
Older × Studied × Concept Confusability	0.06	0.04	0.06	1.06	.346
Older × Studied × Visual Confusability (C1)	0.03	0.01	0.06	0.50	.675
Older × Studied × Item Exemplarity	0.01	0.00	0.06	0.15	.883

Note. In the upper part of the table, the reference level of age group is set to "young adults", and condition is set to "novel items". In the lower part of the table, the reference level of age group is set to "young adults", and condition is set to "lure items". Parameter estimates (logOR), *d'* equivalent, standard errors (*SE*), *z*-values, and false discovery rate (FDR) corrected *p*-values are listed for condition, concept-level, and item-level variables in the linear mixed model.

Figure 3

Effects of Semantic and Perceptual Variables on Mnemonic Sensitivity in Experiment 2


A
Studied-Novel
Lure-Novel
Studied-Lure
Modulation of sensitivity
1.00
0.75
0.50
0.25
0.00
*
**
***
*
***
***
***
-2
0
2
4
Concept confusability
B
Studied-Novel
Lure-Novel
Studied-Lure
Modulation of sensitivity
1.00
0.75
0.50
0.25
0.00
***
***
***
***
-4.9
-2.5
0.0
2.1
Item exemplarity
C
Studied-Novel
Lure-Novel
Studied-Lure
Modulation of sensitivity
1.00
0.75
0.50
0.25
0.00
*
***
-2.7
-0.5
1.5
3.3
Visual confusability (C1)
Group
Older adults
Young adults

Note. The plot lines represent the effects of the predictor variables on the probabilities of endorsing studied items as “old” relative to novel items, lures as “old” relative to novel items, and studied items as “old” relative to lures in young (red lines) and older adults (blue lines). Panels A, B, and C show the effects of concept confusability, item exemplarity, and C1 visual confusability in Experiment 1 in young and older adults. Black asterisks at the top within the panels indicate significant age-related differences. See Material and Methods, Variables of Interest, and Results for details. * $p < .05$; ** $p < .01$; *** $p < .001$ (false discovery rate-corrected).

4.9.2.1 Sensitivity for Studied Relative to Novel Items

Unlike in Experiment 1, we found greater mnemonic sensitivity for studied items in the older relative to young adults (interaction of older adults with studied items on *d’* = .30; 95% CI [.20, .41]). Moreover, concept confusability modulated this overall age-related increase in sensitivity so that relative to the young, older adults were significantly more likely to recognize highly confusable concepts (Table 3; Figure 3A; interaction of age and concept confusability with studied items on d’ = .15; 95% CI [.05, .25]). This represents an attenuation of the negative effect of concept confusability on true recognition in the young group. As in Experiment 1, concept confusability reduced sensitivity for studied items in both groups (Figure 3A; interaction of concept confusability with studied items on *d’* in older = -.12; 95% CI [-.20, -.05]; in young, modulation of *d’* = -.28; 95% CI [-.35, -.20]). The effect of item exemplarity on true recognition was no longer significant in either group (Figure 3B; interaction of item exemplarity with studied items on *d’* = .05; 95% CI [-.03, .13] for older adults; *d’* = -.04; 95% CI [-.11, .04] for young adults). Also, while older adults’ sensitivity for studied items was not significantly influenced by C1 visual confusability (Figure 3C; interaction of C1 with studied items on *d’* = .06; 95% CI [-.02, .14]), young adults were now *more* (rather than less) likely to correctly recognize studied items that were visually confusable with their lures (Figure 3C; interaction of C1 with studied items on *d’* = .09; 95% CI [.02, .17]).

4.9.2.2 Sensitivity for Lure Relative to Novel Items

Unlike in Experiment 1, sensitivity for lure items was increased in the older relative to the young group, reflecting an overall age-related increase in lure false recognition (interaction of group with lure items on *d'* = .41; 95% CI [.31, .51]). Moreover, concept confusability modulated the age-related increase in sensitivity for lure items in older relative to young people. Thus, in parallel with the results for true recognition, older adults were more likely than young adults to falsely recognize more confusable concepts (Table 3; Figure 3A; interaction of age and concept confusability with lure items on *d'* = .11; 95% CI [.02, .21]). This represented an attenuation of the negative effect of concept confusability on sensitivity for lures in the young (interaction of concept confusability with lure on *d'* = -0.12; 95% CI [-.19, -.05]; *d'* = .00; 95% CI [-.07, .07] for older group). At the item level, lure errors were again more frequent for items with higher rated exemplarity in both groups, with no significant difference (Table 3; Figure 3B; interaction of item exemplarity with lure on *d'* = .18; 95% CI [.10, .25] for young, *d'* = .27; 95% CI [.20, .35] for older adults). Higher visual confusability only significantly increased false recognition in the young adults, not in the older group taken separately, but the groups did not differ significantly (Table 3; Figure 3C; interaction of C1 with lure on *d'* = .14; 95% CI [.06, .21] in young, *d'* = .09; 95% CI [.01, .16] in older group).

4.9.2.3 Sensitivity for Studied Relative to Lure Items

Unlike Experiment 1, results showed an overall age-related reduction in sensitivity for studied items relative to lures, reflecting reduced overall mnemonic discrimination (Table 3; interaction of older adults with studied items on *d'* = -.10; 95% CI [-.17, -.04]). However, none of the concept-and item level experimental variables significantly modulated this overall age-related difference (Table 3; Figure 3). In the older group the semantic variables had similar effects to those reported in Chapter 2 (Naspi et al., 2020) for the young, and to those in Experiment 1. Studied items and lures with high concept confusability were less likely to be correctly discriminated in both groups, and – unlike in the contrasts with novel items as baseline – this effect was numerically slightly larger in the older group so there was no hint of an age-related reduction (Figure 3A; interaction of concept confusability with studied items on *d'* = -.12; 95% CI [-.17, -.07] for older adults; *d'* = -.15; 95% CI [-.20, -.11] for young adults). In both groups, lure objects with high rated exemplarity were again less likely to be discriminated from their corresponding studied exemplars (Figure 3B; interaction of item exemplarity with studied items on *d'* = -.22; 95% CI [-.27, -.17] for older adults; *d'* = -.21; 95% CI [-.26, -.17] for young adults). The effect of C1 visual confusability was not significant in either group (Figure 3C; interaction of C1 with studied items on *d'* = .01; 95% CI [-.06, .08] for older adults; *d'* = -.04; 95% CI [-.09, .01] for young adults).

4.10 Discussion

In Experiment 2, when participants studied multiple exemplars of each basic-level concept, we found the expected age-related increase in both true and false recognition. At the same time, there was an attenuated effect of concept confusability in older relative to young adults, again impacting both true and false recognition (Figure 3A). These results suggest that older adults focused more on processing basic-level conceptual information than features shared over concepts. As a consequence, shared semantic features across concepts had less of a negative effect than in their younger counterparts. This result is consistent with the predictions of the semantic categorization account, and its implications will be considered further in the General Discussion.

All three alternative hypotheses also predicted that the older group would show larger effects of item exemplarity on false recognition. We assumed that exemplars that are more similarly related to people's internal basic-level representations would more strongly reactivate previously studied concepts at test (see also Chapter 2; Naspi et al., 2020). Although the item exemplarity effect on lure false recognition was numerically larger in the older group than in the young, we did not find any significant age-related difference. The lack of an age-related difference may derive from the way that exemplarity as an item-level variable assessed semantic similarity.

Similarly, although visual confusability did not influence older adults' levels of false recognition, and did significantly increase the levels of false recognition in the young, we did not find any age-related difference. According to the pattern separation account (Wilson et al., 2006), which predicts an impairment of mnemonic discrimination that generalizes across multiple dimensions of similarity, we should observe an increased in false recognition in older people that extends beyond the semantic dimension. However, visual confusability did not provide evidence for the theory. We explore the possible reasons of the observed results below.

4.11 General Discussion

In this research, we investigated the contribution of semantic and perceptual similarity to age-related differences in true and false recognition. We compared performance of young and older adults in two different experiments in which we varied the number of exemplars of the same basic level concepts at study. According to the view that aging is associated with a greater reliance on processing any kind of semantic information (Umanath & Marsh, 2014), we predicted that both concept- and item-level semantic variables would impact on mnemonic discrimination more in older than young people. Alternatively, based on the semantic categorization account by Koutstaal et al. (2003), we hypothesized that older adults would emphasize information at the basic level and underscore the similarity between different exemplars of the same concept more than young adults. Lastly, if aging is associated with a general impairment in processing different dimensions of similarity (Wilson et al., 2006), then we would observe a generalized increased impact of perceptual and semantic similarity in aging. In both experiments, more conceptually confusable lures were more likely to be classified as new for the young adults, but not for the older group. In Experiment 2, this age-related difference was associated with a significant increase in false recognition for more confusable concepts in the older group. Similarly, in Experiment 2, more confusable studied objects were more likely to be successfully recognized in older relative to young adults, suggesting a reduced tendency to process relations among concepts that were detrimental to memory for items in this task. In contrast, irrespective of age-group, false recognition was more frequent for lure objects that were more similar to their basic-level concepts, a result that does not provide full support to the semantic categorization account (Koutstaal et al., 2003). Lastly, simple visual features shared with other studied images increased young adults', but not older adults, levels of false recognition, suggesting that aging is not associated with a general impairment across multiple dimensions of similarity.

Experiment 1 did not show significant age-related differences in mnemonic discrimination despite previous research has consistently shown increased false recognition in older adults when a single basic level concept was studied and tested with a similar exemplar, at least for concrete object stimuli (Koutstaal & Schacter, 1997; Pidgeon & Morcom, 2014; Reagh et al., 2016). The literature also suggests that although true recognition is often reduced in older relative to young adults (Koustaal et al., 1999; Koutstaal & Schacter, 1997) it is also frequently age-invariant (e.g., Craik & McDowd, 1987; Danckert & Craik, 2013; Pidgeon & Morcom, 2014; Schonfield & Robertson, 1966). One problem with presenting a single studied item and lure at test is that older people, as well as the young adults, can employ strategies in order to recognize previously studied items, but also to reject similar lures. Although young adults are sometimes able to use these strategies more efficiently, older adults are equally able to adopt monitoring strategies with rich pictorial material (Gallo et al., 2007). Such monitoring strategies could reduce the influence of semantic and perceptual similarity on mnemonic discrimination. In Experiment 2, where recall-to-reject strategies were made ineffective by the increased number of studied items per concept, we found evidence of increased false as well as true recognition in older relative to young adults for those concepts that were more confusable with all the other concepts in the set. In this regard these findings are convergent with Pidgeon and Morcom (2014), but differ from those of Koutstaal et al. (2003), who reported non-significant age effects on true recognition, with a non-significant reduction in true recognition in older adults. Here, we showed that these age-related differences in true as well as false recognition may in part be driven by an attenuated effect of concept confusability in older adults. Concept confusability is an index of semantic similarity across concepts reflecting the activation of coarse semantic representations via processing of shared features. Our results suggest that older adults tend to emphasize the processing of individual basic level concepts over this shared information in this task (Koutstaal et al., 2003). This attenuation of the concept confusability effect in older people was also associated with a more pronounced gist-like modulation at the basic level compared to Experiment 1, with increases in both true and false recognition relative to the young.

Parallel effects of semantic processing on true and false recognition are predicted by fuzzy trace theory’s gist-preference conjecture (Brainerd & Reyna, 2005), which assumes that gist traces engender both true and false memory, but that verbatim traces support accurate recognition of studied items and rejection of lures (Koutstaal & Schacter, 1997; Verfaellie et al., 2002). According to gist accounts, including the semantic categorization account (Koutstaal et al., 2003), encoding of multiple meaningful exemplars leads to stronger semantic gist representations (Koustaal et al., 1999). If older adults rely to a greater extent on basic-level processing, and to a lesser extent on semantic information shared across concepts, this should result in an overall tendency to endorse more studied items as well as semantically similar lures as “old” on the basis of semantic gist shared across exemplars. While lending support to the semantic categorization account, our results provide evidence against the broad claim that older people always rely more than young people on prior semantic relations in order to support declining episodic memory (Reder et al., 2007; Umanath & Marsh, 2014). In previous studies, there was no way to index processing of concept-level relations, thus both the general semantic theories and the semantic categorization account could explain the increase in basic-level false recognition in this task. In this study, we provided the first direct evidence based on objective measures that older adults’ processing of cross-concept relations is attenuated relative to young adults in this categorized pictures task.

Previous research has suggested that false memory susceptibility may in part be a result of overreliance on relational processing at encoding (Thomas & McDaniel, 2013; Thomas & Sommers, 2005), defined as processing involving the relations between studied items (Hunt & Einstein, 1981). On the other hand, item-specific processing is defined by the encoding of individuating, non-overlapping information specific to each item and thus improves memory by making item memories highly distinctive. At face value this seems to conflict with our results, which link false and true recognition to reduced (rather than enhanced) processing of shared semantic relations. However, these results can be reconciled if we consider that in this typical categorized pictures task – especially when multiple exemplars of each concept are studied – the task-relevant gist is at the basic level, because the studied items and their lures share the same concept (e.g. several cats were studied, and the lure is a different cat). However, in other versions of the task (Bowman et al., 2019; Seamon et al., 2000), studied items and lures represent different basic level concepts related by membership of the same superordinate category (e.g., a cow, dog, and deer were studied, and a lion appeared as a lure). In the latter version of the task, processing shared semantic features across concepts may make lure errors more likely when the lures are related to studied items in this way (i.e., at the superordinate level). Thus, whether older adults engage in more or less cross-concept processing than younger adults may depend on the nature of the representations elicited by the current task. Thus, the selective disengagement from cross-concept processing in the current task is consistent with the idea that older people need to select that amount of information to be processed due to limited resources, and to favour basic-level processing.

According to the resource deficit hypothesis (Craik, 1986; Luo & Craik, 2008), older adults show a reduction in attentional control resources relative to young adults. Craik and Broadbent (1983) hypothesized that these age-related deficits reflect impaired prefrontal cortex function (see also West, 1992). Researchers have suggested that relational and item-specific processing may compete for finite resources (e.g., DeLosh & McDaniel, 1996; Serra & Nairne, 1993). Thus, even in young people, paradigms that elicit the activation of shared semantic representations should diminish the encoding of basic-level information relative to conditions that do not promote the cross-concept processing. If older adults operate with reduced resources, then they may over-rely on the directly task-relevant process to resolve the task. For example, our pleasantness decision task together with multiple presentations of different exemplars in Experiment 2 are likely to enhance demands on concept-specific rather than shared semantic information compared to a task in which processing is unconstrained by instructions (Smith & Hunt, 1998). Thus, the generalization of semantic information across exemplars arises because of a resource deficit, which is more likely to occur when the orienting task focus is on basic-level processing, and perhaps also when there are multiple exemplars at study (Koutstaal et al., 2003).

The reduction of the effect of concept confusability in older adults is therefore consistent with the semantic categorization account (Koutstaal et al., 2003). In contrast, we did not find support for the broad claim that older adults' memory is generally more reliant upon semantic knowledge than younger adults' memory (Reder et al., 2007; Umanath & Marsh, 2014). Given that concept confusability supports the semantic categorization account, we were also expecting that older adults would process information more at the item level. However, item exemplarity did not affect older adults' memory for lures as would be expected according to either of the semantic gist-related theories. One possibility is that we failed to detect the effect in older adults due to the way that exemplarity as an item-level variable assessed emphasis on semantic information at the item level. Unlike concept confusability, which directly indexed shared features among concepts, item exemplarity indirectly indexed shared features among exemplars via judgments of the strength of the relations between exemplars and their concepts. Currently, norms for the semantic features that are shared between individual exemplars of a basic level concept are not available but, based on the assumptions of fuzzy trace theory and the semantic categorization account, we would predict that similarity between studied items and lures on such measures would be associated with increases in older adults' levels of false recognition, and reduced discrimination between the two (see also Chapter 2; Naspi et al., 2020)

In the introduction, a third alternative hypothesis we considered was that older adults are more susceptible than young adults to false recognition of items whose similarity to studied items extends beyond the semantic domain, i.e., perceptually similar lures. In young adults, the increase in errors for visually confusable lures was consistent in both experiments, but was particularly strong in Experiment 2 after presenting multiple exemplars at study, presumably a consequence of the strengthened perceptual processing (see also Chapter 2; Naspi et al., 2020). However, here the older adults' false recognition was not modulated by increasing levels of visual confusability, and in Experiment 2 this effect was numerically reduced relative to young adults. Thus, there was little support for theories suggesting age-related elevation in false recognition for more visually confusable lures. According to the pattern separation account, older adults less efficiently encode and later use the rich and detailed perceptual information provided by visual stimuli (Wilson et al., 2006). Our results converge with previous studies that have directly addressed perceptual effects on young and older adults using novel abstract stimuli that did not belong to any pre-existing semantic or conceptual categories (Koutstaal et al., 1999; Schacter et al., 1997). These experiments found that the two age groups showed only modest, and non-significant, differences in overall false recognition for these stimuli, providing support for the semantic categorization account. Overall these studies suggest that the effects of perceptual similarity on false recognition, measured in terms of shared low-level visual attributes, are equivalent in young and older adults.

The picture is not clear-cut, however. Although our measure of visual confusability derived from HMax (HMax; Riesenhuber & Poggio, 1999; Serre et al., 2007) did not detect age-related differences attributable to perceptual processing, others have found that perceptually similar but pre-experimentally meaningless lures can trigger false recognition. For instance, using Koutstaal et al.'s (2003) abstract stimuli, Pidgeon and Morcom (2014) showed that older adults were more likely to confidently falsely recognize abstract lures for which they had studied multiple visually similar exemplars. However, as the authors pointed out, it is possible that these meaningless abstract images attracted some semantic processing, due to resemblance to real-world objects. This was supported by participants' subjective reports of spontaneous verbal labelling of the abstract images. Boutet et al. (2019) more directly investigated the contribution of perceptual factors to age-related differences in false recognition by assessing recognition of different exemplars of four different categories of basic level concepts: upright faces, inverted faces, chairs, and houses. This study had the advantage that perceptual similarity was operationally defined as the similarity in the low-level physical properties of the stimuli (contrast, luminance) using an objective measure that parallels the response profile of neurons in the visual cortex (Zeki, 1978). Older adults showed elevated false alarm levels for inverted faces and chairs relative to young adults.

Support for increased perceptually-driven false recognition in aging also comes from several studies using verbal stimuli. In some, multiple studied words are related to phonologically similar lures (Budson et al., 2003; Rankin & Kausler, 1979; Sommers & Huff, 2003; Watson et al, 2001). Recently, Burnside et al. (2017) extended these findings by demonstrating greater false recognition by older adults of visually similar lures that were related to studied items only by a shared font. These results support the notion that older adults are more sensitive to recognition based on perceptual similarity between sets of related but, unlike in the case of the phonologically similar lures, pre-experimentally unfamiliar features. However, the results of a recent study by Pansuwan et al. (2020) diverged from those of Burnside et al. (2017) since young and older people were equally likely to falsely recognise the visually similar lures presented with the same font. It is therefore not clear why we did not detect such an effect, and whether an underlying effect exists. It is possible that other aspects of visual processing can contribute to older adult's false recognition, for example even lower-level visual characteristics than the simple visual features examined here (e.g., luminance and contrast; Boutet et al., 2019). For example, Gellersen et al., (2021) found reduced perceptual representational qualities in healthy aging, providing evidence that older adults do not only find difficulties with controlled retrieval, but they also struggle with processing on less distinctive stimulus representations. Further studies are needed to clarify what, if any, perceptual attributes of images are important determinants in age-related differences in mnemonic discrimination.

4.12 Conclusions

In this work we used objective measures derived from established models of conceptual structure and low-level vision to investigate semantic and perceptual contributions to mnemonic discrimination in young and older adults. The results from two experiments suggest that older adults emphasize the processing of shared semantic information less than young adults, and show a greater relative focus on basic-level semantic information. In Experiment 2, we showed an increase in older adults' true and false recognition relative to the young for more confusable concepts, but no age-related increase in false recognition attributable to item exemplarity, although this remained an important determinant of false recognition in both age groups. We also found no evidence that low-level visual attributes of objects contribute to false recognition more in older than young adults. This result support the semantic categorization account, but not the generalized semantic gist or the pattern separation account.

Chapter 5: General Discussion

This thesis examined semantic and perceptual contributions to mnemonic discrimination in healthy young and older adults. Mnemonic discrimination is the ability to differentiate similar lures from previously studied items to avoid false recognition. In young adults, behavioural (Chapter 2) and fMRI (Chapter 3) studies employing the categorized picture paradigm supported the prediction that both perceptual and semantic processing impacted mnemonic discrimination. In older adults (Chapter 4), I showed that the disengagement from processing shared semantic information increased their levels of true and false recognition for items with high concept confusability. Below I summarise these results briefly and discuss their broader implications.

In Chapter 2, I reported two different experiments (Experiment 1 and 2) that showed that semantic relations can impact memory at the concept and-item level. Similarity at the concept level, concept confusability, indexed shared semantic information across concepts, and impacted both true and false recognition. Specifically, more confusable studied concepts were more likely to be forgotten, while the lures more likely to be correctly rejected. Similarity at the item level, item exemplarity, indexed the ability to activate the corresponding individual basic-level concept by individual exemplars. In both experiments, high exemplarity lures were more likely to be falsely recognized.

In Chapter 3, I tested the prediction that the perceptual and semantic representational content of objects at encoding differentially contributed to true and false recognition. In a novel approach, I combined RSA with the subsequent memory paradigm in a single step. This allowed me to contrast the relative strength of perceptual and semantic representations for items later remembered versus those later forgotten (and for items later falsely recognized versus those correctly rejected). While object-specific visual and semantic representations in posterior regions were associated with successful encoding of studied items, coarse taxonomic semantic representations more anteriorly in the ventral visual pathway and frontal regions hindered it. I also showed that inefficient visual processing of low-level visual attributes in early and late visual cortices led to false recognition. These results, together with the findings discussed in Chapter 2, support my proposal that multiple kinds of perceptual and semantic representations influence mnemonic discrimination.

In Chapter 4, I investigated whether the same variables that are important determinants for mnemonic discrimination in young adults (Chapter 2) would show a differential impact in the older adults. Experiment 1, in which I presented a single exemplar for each basic-level concept, did not reveal any age-related differences in memory, or differences in the effects of the semantic and perceptual similarity. However, predictions that older adults show impairments in false recognition were supported in Experiment 2, when I presented multiple exemplars of the same basic-level concepts at study. Unexpectedly, these findings also revealed heightened true recognition in older relative to young adults. Age-related increases in both true and false recognition were in part modulated by a reduced effect of concept confusability in the older group. I therefore concluded that, unlike young adults, older adults in this task may emphasize more the basic-level processing of individual concepts and less their cross-concept relations.

5.1 Implications for theories of mnemonic discrimination

One of the predictions of fuzzy trace theory introduced in Chapter 1 was that semantic overlap between study and test items would contribute positively to false as well as true recognition (Brainerd & Reyna, 2002). The behavioural experiments in Chapter 2 partially supported these predictions (Naspi et al., 2020). Our results in young adults suggest that different levels of semantic similarity have distinct effects on mnemonic discrimination that reflect the fact that in the categorized pictures task used here, studied items and lures shared basic-level concepts. Thus, semantic relations that emphasize shared features across the entire set of concepts – indexed by concept confusability – appeared to weaken the representations of individual concepts in memory and therefore reduced gist-like effects for studied items and lures. In contrast, semantic relations that emphasize the representations of individual concepts – indexed by item exemplarity – appeared to strengthen the activation of the basic-level concept and therefore increased gist-like effects, triggering false recognition. This is the suggested mechanism underlying both the reduction of true and false recognition for more confusable concepts, and the increase of false recognition for those lures that were more similar to people's internal representations. However, item exemplarity did not affect memory for studied items as I was expecting if it modulated encoding of a gist trace at this level. Thus, although the data fit well with fuzzy trace theory's prediction that semantic overlap between studied items and lures leads to false recognition, the data do not fully support the theory when it comes to true recognition.

The fMRI study in Chapter 3 also provides partial support for fuzzy trace theory's prediction that semantic overlap is critical for both false and true recognition (see Chapter 3; Naspi et al., 2021). However, these results are only in part convergent with the behavioural findings of Chapter 2. For example, as noted above, I interpreted the behavioural findings with regard to concept confusability in Chapter 2 in terms of reduced emphasis on basic-level concepts. However, at the neural level, object-specific semantic information processed at encoding was not associated either positively or negatively with later false recognition. Here, the idea was that fine-grained semantic processing was likely to facilitate the activation of the basic-level concept at encoding through modulation of gist, promoting false recognition. But this is not what I found. This null result is inconsistent with fuzzy trace theory's prediction that semantic representations contribute to false recognition, even though I did find that the same object-specific semantic representations contributed to true recognition as predicted. Beyond the possibility of a false negative (see Chapter 3), it is unclear why the fMRI study did not reveal any semantic representation at encoding associated with false recognition. Later I will provide a possible explanation when I discuss the relative roles of encoding and retrieval processes in mnemonic discrimination (Section 5.2).

Although fuzzy trace theory is not a theory of aging per se, it has been used to explain age-related differences in memory. My data on older adults does not provide direct support to fuzzy trace theory's prediction that greater reliance on semantic gist increases both false and true recognition (Brainerd & Reyna, 2005), at least in the theory's original specification. In Experiment 2 of Chapter 4, the older people showed increased false and true recognition specifically for those items that shared more semantic features with all the other items in the set, driven in part by attenuated effects of concept confusability. This suggested that the older group might be less prone than the young adults to process semantic relations across concepts and focus relatively more on basic-level information, enhancing a gist-like effect at this level. However, the lack of age-related differences in true and false recognition driven by item exemplarity prevents me from drawing clear-cut conclusions that would be compatible with fuzzy trace theory's gist-preference conjecture in this way. One possible interpretation of the age-related increase in false recognition, comes from the semantic categorization account (Koutstaal et al., 2003). According to this theory, older adults may primarily focus on specific aspects of the semantic information (e.g., the names of objects). This might arise because this type of information is more readily or automatically accessed (Brainerd & Reyna, 1998; Koutstaal & Schacter, 1997; Schacter et al., 1998). Although this theory does not explicitly consider true recognition, an emphasis on the concept word could support false and true recognition by underscoring the meaning overlap between concept names. Thus, these age-related effects are consistent with fuzzy trace theory's gist-preference conjecture introduced in Chapter 1 provided that the concept of gist is not purely restricted to shared semantic information across items, but extends to include the semantic of the word (Brainerd & Reyna, 2005).

Although a seemingly self-evident claim about false memory responses is that they will increase as their similarity to experienced material increases, according to fuzzy trace theory, this relation should hold consistently only for semantically, but not perceptually, similar material (Brainerd & Reyna, 2002). Thus, as outlined in Chapter 1, fuzzy trace theory also predicts that because verbatim traces are representations of surface information including perceptual aspects of the studied items, increasing perceptual similarity between studied and lure items should make the latter better retrieval cues for verbatim traces of the former, thereby decreasing false recognition (also see Reyna & Lloyd, 1997). However, retrieval of verbatim traces should also promote true recognition of studied items. The findings of Chapter 2 are less clear in this regard. In Chapter 2 perceptual similarity increased lure false recognition relative to novel items. This effect was weak in Experiment 1, but stronger in Experiment 2 when multiple perceptually similar exemplars were presented at study. Similarly, the effect of perceptual similarity on true recognition was incosistent across experiments. Therefore these findings neither provide a clear-cut picture of the importance of perceptual similarity in mnemonic discrimination nor are wholly consistent with fuzzy trace theory, which cannot explain lure false recognition due to perceptually similar items.

The fMRI study of Chapter 3 provides a clearer result regarding the role of perceptual representations in young adults that are not necessarily incompatible with the behavioural findings of Chapter 2 or with fuzzy trace theory (Brainerd & Reyna, 2002). At encoding, insufficient visual processing in the early and late visual cortex predicted false recognition. Therefore, it is likely that in the behavioural studies of Chapter 2, young adults failed to efficiently encode low-level perceptual details of studied items, allowing perceptually similar lures at test to elicit false recognition. At the neural level, RSA indexes the strength with which the representational content of items is encoded and can reveal increases or decreases in this strength. The idea that false recognition due to perceptual similarity in Chapter 2 is due to insufficient encoding of visual attributes is also supported in the fMRI study by the fact that successful visual processing at encoding was associated with later true recognition, and also the rejection of similar lures. Thus, the fact that perceptual similarity provided inconsistent results for true recognition in Chapter 2 does not necessarily mean that verbatim encoding and retrieval do not support veridical memory, as predicted by fuzzy trace theory. Inconsistency of results across the experiments may simply reflect unsuccessful encoding of perceptual information in one experiment, but not in the other. In the fMRI study, the lack of any positive perceptual representation associated with later false recognition is compatible with fuzzy trace theory that assumes that false recognition is exclusively driven by semantic information (although not found in the current study). What fuzzy trace theory does not explain is the increase of false recognition for perceptually similar items in Chapter 2.

Although behavioural and neuroimaging results in young adults suggested that perceptual information per se plays a role in false recognition, I did not find evidence of age-related differences in the influence of perceptual similarity on mnemonic discrimination. Fuzzy trace theory’s gist-preference conjecture suggests that the shift from verbatim processing toward gist processing is more in the nature of an encoding and retrieval preference than a deficit in verbatim memory (Brainerd & Reyna, 2005). That is, for some reason, as we age, we come to base our memory more on the meaning content of experience and less on its surface form, despite the fact that perceptual information is attended to and stored. Instead, the semantic categorization account suggest that processing pre-existing semantic information like the concept name truncates, precludes, or preempts further verbatim processing (Koutstaal, Reddy, Jackson, Prince, Cendan, Schacter, et al., 2003). Thus, while fuzzy trace theory suggests that perceptual details are indeed encoded by the older adults but are less effectively used in the recognition context requiring their deliberate, controlled use, the semantic categorization account does not. In this case my data supports fuzzy trace theory. If older adults were not able to encode perceptual information because the semantic meaning impaired it, we would also observe age-related differences in perceptual false recognition due to their inability to recall verbatim information at test to reject perceptually similar lures, at least in Experiment 1 where a single exemplar was studied. But this is not what I observed. However, this conclusion cannot be given for granted since the previous literature also showed that older people's levels of false recognition are sensitive to perceptual similarity (Boutet et al., 2019; Pidgeon & Morcom, 2014; Stark et al., 2013). Thus, a more sensitive measure that better capture perceptual processing in aging or a better model may be able to detect age-related differences in false recognition (see Section 5.5).

Together, the results challenge fuzzy trace theory's assumption that only semantic similarity is implicated in false recognition. As discussed in Chapter 2, fuzzy trace theory cannot readily account for false recognition due to perceptual similarity, unless the notion of gist is extended to incorporate visual information. Several authors borrowing heavily from fuzzy trace theory have not discounted the notion that other kinds of similarity can contribute to recognition outcomes, particularly in relation to aging (Koutstaal & Schacter, 1997; Koutstaal et al., 1999; Pidgeon & Morcom, 2014). Research in vision suggests that observers can process a perceptual gist representation corresponding to the average of two scenes' global perceptual features (Oliva, 2005; Oliva & Torralba, 2006). These data therefore partially support fuzzy trace theory if gist and verbatim memory are not distinguished by a semantic versus perceptual basis. However, to really investigate fuzzy trace theory's assumptions and test the impact of these dimensions on separate verbatim and gist memory components, these components should be formally measured using Brainerd et al's. (1999) conjoint recognition paradigm and associated multinomial model. In the conjoint recognition paradigm, the test list (as here) contains three types of items: studied items, related lures that share a semantic gist with the studied items, and unrelated novel items. The memory test is administered to three groups of participants with different instructions: Under the T instruction, participants are asked to accept as old only studied items; under the R instruction, participants are to accept as old only related lures; and, finally, under the T + R instruction, participants are to accept both studied items and related lures. From the proportions of studied items, related lures, and unrelated novel items obtained in the three different groups, the parameters of a model are estimated that provide measures of verbatim and gist memory. The model's parameters can be interpreted as the probability of the cognitive processes that they represent. Stahl and Klauer (2008) also proposed a simplified conjoint recognition paradigm in which all participants are presented with the same test list, and a single instruction is presented asking the same multiple choice question for all items. Briefly, people are instructed to respond “target” if they believe that the item has been presented in the study phase, “related” if they believe that the item is a related lure, and “new” if they consider it to be an unrelated novel item. Thus, in the simplified conjoint recognition paradigm, valid measures of gist and verbatim memory can be obtained from a single group of participants. Future studies that combine the conjoint recognition paradigm together with predictors representing semantic and perceptual similarity can reveal more about the contributions of these variables to gist and verbatim memory.

Although the data of the current thesis fits relatively well with fuzzy trace theory, other theories introduced in Chapter 1 may also explain the same results, and perhaps elucidate whether different accounts are mutually exclusive or not. As outlined in Chapter 1, the pattern separation account predicts decline in discrimination ability across multiple dimensions of similarity. Although the current thesis did not assess neural pattern separation per se, the data in Chapter 2 and 3 is consistent with a failure in discriminating between events that are similar across multiple dimensions, not just semantically (Motley & Kirwan, 2012; Reagh et al., 2014; Yassa et al., 2011). Thus, fuzzy trace theory and the pattern separation account are not necessarily mutually exclusive, at least in young adults. However, the neurocognitive model of aging proposed by Wilson et al. (2006) does predict age-related differences in false recognition due to increasing levels of perceptual as well as semantic similarity, which I failed to find evidence for here. From a pattern separation perspective, presence of multiple overlapping representations in memory (as when multiple exemplars have been encoded in Experiment 2), should also result in increased likelihood of erroneous pattern completion and therefore mnemonic discrimination errors, particularly in older adults (Wilson et al., 2006). Therefore, according to this theory, I should have observed increased false recognition in older adults in Experiment 2 than Experiment 1 (Chapter 4). The data partially support this view because, although older adults’ false recognition emerged in Experiment 2, there was no between-experiment difference.

Similar to fuzzy trace theory, the activation monitoring theory hypothesises multiple memory mechanisms to explain false memory (Roediger et al., 2001). This theory relies on the assumption that associative activation at study, followed by retrieval-based monitoring failure, jointly lead to false recognition. However, Cann et al. (2011) demonstrated that thematic semantic relations between list items and the critical lures may be a better determinant of false recognition than backward associative strength (see also Brainerd et al. (2008) and Introduction of Chapter 2). The assumed association between associative activation at encoding and false recognition in young adults also means the activation monitoring theory cannot explain Chapter 2 's findings that false recognition could be driven by perceptual similarity. The activation monitoring theory and source monitoring account – like fuzzy trace theory – also specify a role of retrieval-based monitoring, thought to be largely mediated by prefrontal cortex functions, in determining the encoding source of recognised information (Johnson et al., 1993; Roediger et al., 2001). Source monitoring failure is thought to be associated with misattribution of the source of information, and is proposed to lead to false recognition of lures (Johnson, 1997). It has also been suggested that age-related increases in false recognition are explained by monitoring decline (Henkel et al., 1998; West, 1996). However, the source monitoring alone does not predict the effects of the semantic and perceptual relations at encoding and thus it is not sufficient to explain the results of the current thesis. Instead, fuzzy trace theory assumes a prominent role at encoding in processing semantic gist. However, with the behavioural measures (which I also used in the test phase of the fMRI study) I was not directly able to discern whether semantic and perceptual contribution to mnemonic discrimination were due to encoding or retrieval processes. Therefore, in the next section, I will discuss the encoding contribution to mnemonic discrimination and a possible role of retrieval processing.

5.2 Encoding and retrieval processes in mnemonic discrimination

One of the major advantages of the fMRI study is that it could better separate encoding and retrieval and directly test the neural operations involved in mnemonic discrimination. As noted above, fuzzy trace theory assumes that the brain stores two independent memory traces: a verbatim and a gist trace. If verbatim traces are episodically instantiated representations of the surface forms of experienced items, including contextual and perceptual details, to some degree the perceptual representations are equivalent to verbatim traces. Similarly, if gist traces are representations in space organised in terms of object-specific semantic features, to some degree the semantic representations are equivalent to gist traces. It is useful to remember from the Introduction in Chapter 1 that verbatim and gist retrieval are assumed to be opponent processes in false recognition, with gist retrieval supporting lure mnemonic discrimination errors and verbatim retrieval suppressing them. However, verbatim and gist retrieval are convergent processes in true recognition because they both support true recognition of studied items.

The fMRI data partially supports the first assumption. Indeed, while strong perceptual representations at encoding were later associated with correct rejection of similar lures, it is not clear why no semantic representations at encoding predicted false recognition. One possibility is that pictures attract surface-processing attention at encoding, so corresponding gist traces were weak. Alternatively, showing a single exemplar during the encoding phase may be not enough to create a strong gist representation detectable with RSA. The opponent-processes idea predicts that levels of false recognition will increase if studied items that share the same semantic gist are repeated (Brainerd & Reyna, 2005). This prediction follows because the memorial basis for false recognition (gist traces) is being strengthened. However, these lines of reasoning do not explain why object-specific semantic representations at encoding are successfully used to support true recognition. Overall, across the results of the behavioural and fMRI studies is clear that semantic operations that enhance basic-level processing support false recognition responses because high concept confusability tends to reduce them. Less clear is whether object-specific semantic processing at encoding directly contributes to false recognition. It may simply be that the fewer number of falsely recognized items overall decreased the RSA sensitivity in detecting pattern of activity reflecting semantic information specific to false recognition, i.e., a lack of power. Thus, it is important to note that this null finding does not necessarily mean that the effect of semantic representations on memory does not exist. Future studies that employ words, known to tap more strongly on gist processing than pictures, and thus increasing the number falsely recognised trials, can better test whether semantic representations at encoding contribute to later false recognition. Although the first prediction of fuzzy trace theory was only partially supported, the fMRI data are consistent with fuzzy trace theory’s principle that verbatim and gist traces are convergent processes when it comes to true recognition.

As briefly discussed earlier, the fMRI study highlighted the fact that inefficient visual processing at encoding could trigger false recognition. Although this is not incompatible with fuzzy trace theory, this theory has little to say about the encoding mechanisms that subserve false recognition. The importance of encoding operations in processing verbatim information is explained by Molitor et al. (2014) who used a continuous recognition memory task with eye-tracking measures to investigate the cognitive processes underlying false recognition. The authors assumed that mnemonic discrimination errors for similar lures result from the insufficient encoding of study items and predicted that fewer fixations would be made to falsely recognized lures compared to hits and correctly rejected lures during the first presentations of items. The eye-tracking data supported this prediction during first presentations when encoding presumably occurred. Based on previous research that has shown a relationship between the number of fixations during encoding and subsequent memory (Kafkas & Montaldi, 2011; Loftus, 1972), this finding implies that the original studied items were insufficiently encoded. A similar logic can be applied to the finding in Chapter 3 that reduced early visual processing predicted false recognition. This results is compatible with fuzzy trace theory if we assume that insufficient visual processing at encoding may reinforce the representation of a raw perceptual gist which then could seem familiar in a novel exemplar that shares those same generic features, as suggested to explain my findings in Chapter 2 (Naspi et al., 2020). This result is also consistent with the idea that false recognition would exemplify a failure of pattern separation during the first presentation of items (Schacter et al., 1998), rather than an isolated pattern completion process during the presentation of the lures.

This thesis therefore provides some support for the idea that encoding processes can contribute to mnemonic discrimination and that some perceptual and semantic operations at encoding can be responsible for true and false recognition. However, it is of course both encoding and retrieval processing which ultimately determine recognition outcomes, and a comprehensive account of mnemonic discrimination must also clarify the retrieval processes involved and their interactions with encoding (Otten, 2007). This is particularly important given that gist retrieval may also play a substantial role in mnemonic discrimination. Retrieval processes also form a major component of some of the mnemonic discrimination accounts discussed in this thesis. According to fuzzy trace theory, while the relative strength of verbatim and gist representations formed at encoding influences the availability and strength of these representations at test, and thus influences recognition outcomes, retrieval-based factors such as retrieval cues play a major role (Brainerd & Reyna, 2002). Use of studied items as retrieval cues is thought to lead to a tendency to retrieve verbatim traces more often than gist, while presentation at test of lure items sharing gist with studied items leads often to retrieval of gist memory. Thus, it is likely that stronger reactivations of semantic representations for falsely recognised items are elicited at retrieval than encoding. Behaviorally, this effect was suggested for both young and older adults in Chapters 2 and 4 by item exemplarity. Although I found that items more representative of their basic-level concepts were more likely to be falsely recognized, studied items with high item exemplarity were not more likely to be remembered as I expected if item exemplarity modulated semantic gist at encoding. Thus, I argued that this modulation of false recognition was probably due to a retrieval effect. If so, fMRI at retrieval would reveal a better fit of the object-specific semantic model for falsely recognized lures.

Further retrieval processes also contribute to recognition. According to the source monitoring theory, lure mnemonic discrimination involves evaluation of the match between current and previous representations, with false recognition occurring when overlap is attributed to the lure having been studied (Johnson, 1997). This is similar to suggestions that recall-to-reject mediates comparison of the match between previous and existing representations (Brainerd et al., 2003). However, this explanation cannot explain the results obtain in Experiment 2 of Chapter 2, in which recall-to-reject was prevented. According to the pattern separation account, pattern separation at encoding supports later pattern separation and mnemonic discrimination at test, while dominance of pattern completion at encoding increases the likelihood of dominance of pattern completion at test and thus false recognition. While encoding processes influence test outcomes, dominance of pattern separation or completion at retrieval is critical. At processing of lures, pattern separation enables unique encoding of item-specific information despite overlap with previous representations, enabling avoidance of false recognition (Sahay et al., 2011; Yassa & Stark, 2011). It has been suggested that this dominance of pattern separation, if accompanied by retrieval of overlapping representations via pattern completion, enables comparison of the current and existing representations and therefore identification of lures as novel (Kirwan & Stark, 2007). When lures elicit pattern completion instead of pattern separation, this is assumed to lead to retrieval of overlapping previous episodes, and thus lure false recognition (Wilson et al., 2006). This explanation is not mutually exclusive with encoding mechanisms, and it is likely that the combine interaction of encoding and retrieval determines the final recognition outcome.

5.3 Thesis strengths

A major strength of the thesis is that it provides the first objective evidence that mnemonic discrimination is influenced by multiple semantic and perceptual dimensions in young and older adults. The current thesis explored different measures of semantic and perceptual relations that prior literature only assumed to be relevant. Previous behavioural studies that used the categorized picture task assumed that false recognition was exclusively driven by semantic and/or perceptual similarity between different exemplars of the same concept on the basis of their shared gist (Koutstaal & Schacter, 1997). Moreover, previous research did not take into consideration the potential contribution of semantic and perceptual relations between different items within the same study and test list (e.g., between “cat”, “chair”, and “ladybug” that shared the feature “has legs”). Multiple associative effects have previously been investigated in the context of the DRM paradigm, in which increasing the strength of connections from list items to critical items and increasing the strength of interconnections among the list items produced an increase in false memories in recognition (McEvoy et al., 1999). However, this is the first study that simultaneously accounts for multiple measures of semantic and perceptual processing in a categorised picture paradigm. The current thesis also made explicit an important fact about pre-existing semantic knowledge: not all kinds of semantic processing contribute to true and false recognition. It all depends on whether the currently relevant gist information between study and test items is strengthened or not.

A second strength of the current thesis is that all the studies discussed were pre-registered in Open Science Framework (https://osf.io/dashboard). This practice is particularly important in neuroimaging and, to my knowledge, mine was the first published preregistered study of episodic memory encoding. The reproducibility problem in fMRI has been recently discussed in a paper that showed that about 10% of the published results may be false positive (Eklund et al., 2016; Mumford, 2012). This problem can arise due to the use of different fMRI software (e.g., AFNI, FSL, SPM) that has different pipelines, but sometimes is due to practices that dramatically increases the number of false positive findings that encourage the retrospective formulation of hypotheses as the basis of scientific knowledge and theory (Yamada, 2018). Thus, the submission of research papers for which experimental and analytic methods have been designed and described completely prior to collection of actual data can alleviate this problem. This include the software used, a detailed pipeline of each step taken, but also the author's motivation and hypothesis. For example, although not currently used in my fMRI study, the used of a standardised pipeline for pre-processsing could reduce variability and increase transparency that, consequently, would promote reproducibility of results (Esteban et al., 2019).

As I explained just now, one of the advantages of using objective measures is that I was able to show that the relevant perceptual information that can elicit false recognition is not exclusively restricted to low-level attributes shared between studied items and their specific lures. Unlike concept confusability, the pre-registered measure of visual confusability calculated across the entire set of images did not influence performance on the mnemonic discrimination task. Thus, using an exploratory measure, I hypothesised that the effect of perceptual similarity on false recognition was simply driven by the most similar exemplar, or the nearest neighbour. This is what I found. Although this finding has to be taken with caution, this measure being exploratory, it is a potentially powerful result. At least anecdotally, it shows that visual confusability of a similar lure is not necessarily related to the corresponding studied exemplar that depicts the same basic-level concept (see Chapter 2). For example, one subject in my sample falsely recognized a picture of a "rifle". Its most similar neighbour in terms of low-level visual representation (i.e., in this case their elongated orientation) is the image of an "arrow" that I presented during the study phase. This is a novel aspect of our experiments since previous studies that emphasised the importance of perceptual similarity in eliciting false recognition only assumed this dimension to be at play between studied items and similar lures (Koutstaal et al., 2003; Pidgeon & Morcom, 2014). More generally, the use of objective measures allowed me to begin to disentangle effects of semantic and perceptual similarity on mnemonic discrimination.

Lastly, previous studies that investigated the influence of semantic and perceptual similarity simply operationalized these dimensions as category or thematic membership (Cann et al., 2011), presuming that different exemplars or concepts share semantic or perceptual features, but without actually measuring the degree of semantic or perceptual similarity (but see Boutet et al., 2019; Montefinese et al., 2015). The used of objective measures of similarity allowed me to overcome this important limitation. Similarly, feature-based and computational models in the fMRI study allowed me to directly overcome the reverse inference problem of previous mass univariate analysis studies that interpreted activation in regions associated with semantic or perceptual processing necessarily as a reflection of semantic or perceptual operations (e.g., Kim & Cabeza, 2007)

5.4 Limitations

One of the most important limitations of the current thesis is the fact that the models used to capture perceptual and semantic processing are incomplete and imperfect. This is true for all the measures used across the experiments. Especially in the field of the semantic cognition, for example, it is difficult to capture all the semantic aspects of an object and to define all the existing relations with all the other concepts. The semantic memory model used in the current thesis (CSA; Tyler & Moss, 2001) is optimal for concrete concepts that can be defined in terms of features, but it is incapable of describing how the brain represents in space all the other abstract concepts for which the semantic meaning is less describable in terms of features. Although the current thesis only used concrete concepts, other problems still emerged. One major issue, for example, is the lack of norms that quantify semantic similarity between different exemplars of the same basic-level concept. In the behavioural experiments I was expecting that people could show elevated levels of false recognition for those items that were rated as more similar to the participant's internal representations of those concepts. In both young and older adults, I found that these items were more likely to trigger mnemonic discrimination errors when used as lures, but we did not observe any age-related difference. Moreover, this variable did not impact on true recognition in the hypothesised direction. While exemplarity (as I argued above) may have had its effects at retrieval, it is also possible that this measure was insensitive to potential basic-level gist effects on encoding (see Chapter 2). An improved measure could be derived if feature norms were available for the different depicted exemplars of the same basic-level concepts. This limitation also extends to the fMRI study which by the same argument also lacked a representational model that captured semantic processing at this level.

A second limitation is that although I controlled for associative effects of backward and forward associative strength on memory, they were not included as main predictors in the behavioural models. This is important in light on a recent review that show that semantic memory includes both “taxonomic” (dog-bear), and “thematic” (dog-leash), relations (Mirman et al., 2017). These semantic systems are dissociable and therefore may independently contribute to mnemonic discrimination. In Experiments 1 and 2 of both Chapter 2 and Chapter 4, backward and forward associative strength together accounted for a single principal component. An exploratory examination of statistical results for these components (not reported) suggested a strong simple effect on overall “old” responses to novel items only in the case of the multiple studied exemplars (Experiment 2 of each Chapter). However, as models (to limit the number of predictors) did not include interactions with memory, this result does now allow any inference about the relative contributions of the associative measures to mnemonic discrimination. However, this was not the main purpose of the current thesis, which focused on the modulation by perceptual and semantic variables, understood here in terms of features, on mnemonic discrimination. The results showed that the effects of these variables can go beyond the controlled effect of backward and forward associative strength. I therefore do not argue that associative strength is unimportant in false memory because clearly in the real world some concepts are more likely to co-occur with other concepts in certain situations (e.g., doctor and nurse). Co-occurrence probabilities influence processing at many levels, and, as such, they influence, for example, the probability of remembering an event. My point is that associative strength is not necessary nor by itself sufficient to explain false memory effects. Rather, what is more important are certain types of semantic relations that are processed and extracted from experiences with the world and then later used in support to memory. From my perspective, associative relations embed many semantic relations (Brainerd et al., 2008; Cann al., 2011) and thus the debate of whether semantic relations versus associative strength is more important in false memory may be after all not the best way to make progress.

As far as neuroimaging is concerned, a major limitation is the absence of a study of neural correlates at retrieval. As noted above, a comprehensive account of mnemonic discrimination requires specification of both encoding and retrieval processes. Collecting the data for both phases would allow the researcher to also perform further analysis, e.g., encoding-retrieval similarity between these activity patterns that is not possible with the actual data. Although encoding-retrieval similarity is not an optimal technique when it comes to revealing the kinds of information (perceptual or semantic) reflected in specific brain regions, it can be used to clarify whether false memories are due to a reinstatement of the studied information. A fully reinstated pattern of activity associated with high confidence judgments could help disambiguate whether these representations are associated with a vivid sense of recollection instead of familiarity. For example, fuzzy-trace theory suggests people can experience a phenomenon known as phantom recollection (Brainerd et al., 2003), where people mistake the strong familiarity that results from a test item strongly matching a gist memory trace for the experience of recollecting. This gives rise to the subjective experience of recollecting that it is based upon retrieval of a gist trace.

Lastly, two more general methodological limitations apply to the aging studies. One problem often encountered is that of selection bias in older samples (Hultsch et al., 2002), whereby the sample under study is not representative of the population. In ageing studies this often manifests as older adults being of greater cognitive ability than the population average at the same age (Bootsma-van der Wiel et al., 2002). This is most problematic where volunteers are sought via direct advertisements rather than randomly selected from populations, as typically older adults of greater cognitive ability are more likely to volunteer for research studies (Ganguli et al., 1998). This especially applies to Experiments 1 and 2 of Chapter 4, where older adults were recruited via email invitations to members of a research volunteer panel. Although in Experiment 1 the levels of educational attainment were equivalent between the two groups, older people showed on average slightly higher numerical years of education. This can in turn influence executive abilities, and the finding in Experiment 1 that older adults showed greater Forward Digit Span than young adults is consistent with a high-performing sample. The absence of increased false recognition and decreased true recognition in Experiment 1 can thus in some circumstances reflect these differences. However, after matching the years of education, the observed results still persisted. This suggests that young and older groups were largely comparable, despite older adults' greater Forward Digit Span in Experiment 1. Effects of selection bias also apply often to young samples due to the tendency to recruit young adults from student populations, as here. A second problem is that ageing studies using the cross-sectional design employed in Experiments 1 and 2 of Chapter 4 are, unlike longitudinal studies, susceptible to cohort effects, and provide less clear support for inferences about changes as a result of age rather than differences in environmental, social or other factors between young and older adults (Lindenberger, 2014; Rönnlund et al., 2005).

5.5 Future research

A major way in which this work could be developed in future is by improving these initial models of semantic and perceptual processing to obtain more comprehensive measures of the similarity relations that affect mnemonic discrimination. For example, to address the fundamental question of how the visual properties of an object elicit semantic meaning, researchers have started to combine computationally explicit model of vision with computationally explicit model of semantics. As a model of visual processing one step could be to shift from HMax to the deep convolutional neural network (DNN), which some view as the state-of-the-art in machine vision research, and has remarkable power to capture visual information in images and to label objects accurately (Devereux et al., 2018). A DNN consists of a series of hierarchical layers, where the nodes of each layer correspond to filters that are sensitive to particular patterns in the preceding layer. Nodes in the earliest layers are sensitive to relatively local patterns of low-level properties of the stimulus (e.g. pixel values or line orientations) whilst nodes at later layers are sensitive to higher-level complex visual features which exhibit invariance with respect to lower-level detail, such as position. Although these models have been developed with engineering goals rather than neurocognitive plausibility in mind, recent neuroimaging studies have shown a remarkable correspondence between the layers of DNNs and activation patterns in the visual system predicting memory success (Davis et al., 2020; Deng et al., 2021). However, to my knowledge no work has yet been done on false memories with these recent newly established models of vision. If the model of perception can benefit from advances computational tools, the same applies to the semantic model, particularly in the lack already identified of finer-grained measures that explore within-concept feature similarity. It is difficult to capture for example semantic differences between two similar exemplars that belong to the same concept. Indeed, some of these features that make two objects semantically similar are sometimes only verbalised through the observation of a picture, which necessarily ends up overlapping with the perception itself. Future studies could address this concern by asking participants to rate a pair of stimuli based on their kind (e.g., semantic similarity) similar to Konkle et al. (2010). Moreover, these novel measures should be applied to the aging studies at encoding and retrieval to better test whether they can predict true and false recognition, as explained above.

Further studies on the neural correlates at encoding and retrieval related to the older population are essential to draw clear conclusions of age-related differences in the informational content that drives true and false recognition in this group. For example, there has been accumulating evidence that activation patterns elicited by different types of visual stimuli are less distinct in older than in young adults (Abdulrahman et al., 2017; Koen et al., 2019; St-Laurent et al., 2014). This phenomenon, known as age-related neural dedifferentiation, suggests that aging impairs the quality of representations during visual perception. If inefficiencies in perceptual processing determines poor mnemonic representations (Davidson et al., 2019; Park et al., 2004), we would be able to test whether this phenomenon is associated with the age-related increases in the levels of false recognition. In contrast to perceptual aspect, semantic knowledge is largely preserved in older adults. According to general semantic views, because semantic processing is generally spared in older adults, they can use this knowledge to compensate for impaired cognitive deficits (Umanath & Marsh, 2014). Thus, we could test whether semantic model in older people provides a better fit than in young adults within areas known to be responsible for semantic processing.

Ultimately, progress in neuroscience requires better understanding of the relations between the processes subserved by brain regions and their representations. Most prior fMRI studies investigating the neural bases of these operations have focused primarily on processes (operations performed on information) and only rarely examined age effects on representations (the nature of the information processed) (Cowell et al., 2019). Now, we are observing the opposite trend. Would it be possible to unify these line of research? For instance, although univariate activation in the left inferior frontal gyrus associated with semantic retrieval and elaboration usually predicts later true recognition, my results revealed that coarse semantic information in this area hindered it. How can these kinds of results be integrated together? Answering these questions will be of primary importance in order to disclose the function of human memory.

5.6 Conclusion

This thesis examined the contribution of semantic and perceptual similarity to later mnemonic discrimination in young and older adults. Moreover, it investigated the encoding contribution to true and false recognition in young people. More broadly the results of the current thesis support three general conclusions. First, fuzzy trace theory's idea that semantic information is crucial for both true and false recognition, provided that the relevant gist information is shared between studied items and lures. The results are consistent with an updated view of fuzzy trace theory that also allows for shared perceptual information to contribute to gist. My results, although incompatible with the activation monitoring theory, are also broadly compatible with the pattern separation account which does not put emphasis on any kind of similarity likely to trigger lure mnemonic discrimination errors. It has been noted before that fuzzy trace theory and the pattern separation account are not mutually exclusive. Second, the data partially supports the semantic categorization account that emphasise a primary role of pre-existing semantic knowledge in eliciting false, but also true recognition, in the elderly through processing of the basic-level concept. If gist is not restricted to shared semantic information, but also embeds the meaning of the words common to study and test items, then the results would also be compatible with fuzzy trace theory's gist-preference conjecture. Lastly, I provided evidence that at least in part these effects are due to perceptual and semantic operations that occur at encoding. In conclusion, the used of objective measures allowed me to disentangle the relative role of multiple semantic and perceptual dimensions and to test their contribution to mnemonic discrimination. It is hoped that the findings of this thesis will contribute to future research into encoding strategies or interventions to avoid mnemonic discrimination errors in the young and elderly. Future investigations using fMRI and RSA are also needed to clarify the role of perceptual and semantic processes at retrieval, to determine whether semantic representations may influence true and false recognition during this phase. These lines of enquiry, together with neuroimaging studies in aging that will investigate how differences in the perceptual and semantic representational content at encoding and retrieval contribute to mnemonic discrimination in older people, are likely to improve our understanding of memory.

References

Abdulrahman, H., Fletcher, P. C., Bullmore, E., & Morcom, A. M. (2017). Dopamine and memory dedifferentiation in aging. *NeuroImage, 153*, 211–220. https://doi.org/10.1016/j.neuroimage.2015.03.031

Abe, N., Fujii, T., Suzuki, M., Ueno, A., Shigemune, Y., Mugikura, S., … Mori, E. (2013). Encoding- and retrieval-related brain activity underlying false recognition. *Neuroscience Research*, *76*(4), 240–250. https://doi.org/10.1016/j.neures.2013.05.006

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., … Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics, 8*, 1–10. https://doi.org/10.3389/fninf.2014.00014

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., & Zilles, K. (2000). Brodmann's areas 17 and 18 brought into stereotaxic space - Where and how variable? *NeuroImage, 11*(1), 66–84. https://doi.org/10.1006/nimg.1999.0516

Arndt, J, & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language*, *39*, 371–391.

Arndt, Jason. (2010). The role of memory activation in creating false memories of encoding context. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(1), 66–79. https://doi.org/10.1037/a0017394

Arndt, Jason, & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, *48*(1), 1–15. https://doi.org/10.1016/S0749-596X(02)00518-1

Azab, M., Stark, S. M., & Stark, C. E. L. (2014). Contributions of human hippocampal subfields to spatial and temporal pattern separation. *Hippocampus, 24*(3), 293–302. https://doi.org/10.1002/hipo.22223

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The celex Lexical database (Version 2). Linguistic Data Consortium, University of Pennsylvania.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. *Psychology and Aging*, *5*(3), 421–428. https://doi.org/10.1037/0882-7974.5.3.421

Badham, S. P., Estes, Z., & Maylor, E. A. (2012). Integrative and semantic relations equally alleviate age-related associative memory deficits. *Psychology and Aging*, *27*(1), 141–152. https://doi.org/10.1037/a0023924

Badham, S. P., Hay, M., Foxon, N., Kaur, K., & Maylor, E. A. (2016). When does prior knowledge disproportionately benefit older adults memory? *Aging, Neuropsychology, and Cognition*, *23*(3), 338–365. https://doi.org/10.1080/13825585.2015.1099607

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640–1642. https://doi.org/10.1126/science.1152882

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer’s type. *Cognitive Neuropsychology*, *16*(3–5), 361–384. https://doi.org/10.1080/026432999380834

Barense, M. D., Groen, I. I. A., Lee, A. C. H., Yeung, L. K., Brady, S. M., Gregori, M., … Henson, R. N. A. (2012). Intact memory for irrelevant information impairs perception in amnesia. *Neuron*, *75*(1), 157–167. https://doi.org/10.1016/j.neuron.2012.05.014

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*(3), 560–585. https://doi.org/10.1080/783663595

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baym, C. L., & Gonsalves, B. D. (2010). Comparison of neural activity that leads to true memories, False memories, and forgetting: An fMRI study of the misinformation effect. *Cognitive, Affective and Behavioral Neuroscience*, *10*(3), 339–348. https://doi.org/10.3758/CABN.10.3.339

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate : A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bootsma-van der Wiel, A., van Exel, E., de Craen, A. J. ., Gussekloo, J., Lagaay, A. ., Knook, D. ., & Westendorp, R. G. . (2002). A high response is not essential to prevent selection bias. *Journal of Clinical Epidemiology, 55*(11), 1119–1125. https://doi.org/10.1016/s0895-4356(02)00505-x

Boutet, I., Dawod, K., Chiasson, F., Brown, O., & Collin, C. (2019). Perceptual similarity can drive age-related elevation of false recognition. *Frontiers in Psychology, 10*, 1–12. https://doi.org/10.3389/fpsyg.2019.00743

Bowman, C. R., Chamberlain, J. D., & Dennis, N. A. (2019). Sensory representations supporting memory specificity: Age effects on behavioral and neural discriminability. *Journal of Neuroscience, 39*(12), 2265–2275. https://doi.org/10.1523/JNEUROSCI.2022-18.2019

Bowman, C. R., Chamberlain, J. D., Dennis, N. A., Bowman, X. C. R., Chamberlain, X. J. D., & Dennis, X. N. A. (2019). Sensory representations supporting memory specificity: Age effects on behavioral and neural discriminability. *Journal of Neuroscience, 39*(12), 2265–2275. https://doi.org/10.1523/JNEUROSCI.2022-18.2019

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

Brainerd, C. J., Payne, D. G., Wright, R., & Reyna, V. F. (2003). Phantom recall. *Journal of Memory and Language, 48*(3), 445–467. https://doi.org/10.1016/S0749-596X(02)00501-6

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10*(1), 3–47. https://doi.org/10.1016/0273-2297(90)90003-M

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to "remember" than things that were. *Psychological Science, 9*(6), 484–489. https://doi.org/10.1111/1467-9280.00089

Brainerd, C. J., & Reyna, V. F. (2005). The science of false memory. Oxford University Press.

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*(3), 343–382. https://doi.org/10.1037/0033-2909.134.3.343

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, Vol. 106, pp. 160–179. https://doi.org/10.1037//0033-295x.106.1.160

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection : False-memory editing in children and adults. *Psychological Review*, *110*(4), 762–784. https://doi.org/10.1037/0033-295X.110.4.762

Brainerd, C. J., & Wright, R. (2005). Forward association, backward association, and the false-memory Illusion. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(3), 554–567. https://doi.org/10.1037/0278-7393.31.3.554

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008a). Semantic processing in "associative" false memory. *Psychonomic Bulletin and Review*, *15*(6), 1035–1053. https://doi.org/10.3758/PBR.15.6.1035

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008b). Semantic processing in "associative" false memory. *Psychonomic Bulletin & Review*, *15*(6), 1035–1053. https://doi.org/10.3758/PBR.15.6.1035

Brainerd, C.J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brainerd, Charles J., Reyna, V. F., & Kneer, R. (1995). False recognition reversal: When similarity is distinctive. *Journal of Memory and Language*, *34*, 157–185.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase ii: 930 new normative photos. *PLoS ONE*, *9*(9).

https://doi.org/10.1371/journal.pone.0106953

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*(1), 51–61. https://doi.org/10.1038/35049064

Bruffaerts, R., Dupont, P., Peeters, R., De Deyne, S., Storms, G., & Vandenberghe, R. (2013). Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. *Journal of Neuroscience, 33*(47), 18597–18607. https://doi.org/10.1523/JNEUROSCI.1548-13.2013

Bryan, J., & Luszcz, M. A. (2000). Measurement of executive function: Considerations for detecting adult age differences. *Journal of Clinical and Experimental Neuropsychology, 22*(1), 40–55. https://doi.org/10.1076/1380-3395(200002)22:1;1-8;FT040

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, *49*(4), 1520–1523. https://doi.org/10.3758/s13428-016-0811-4

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Buckner, R. L., Kelley, W. M., & Petersen, S. E. (1999). Frontal cortex contributes to human memory formation. *Nature Neuroscience*, *2*(4), 311–314. https://doi.org/10.1038/7221

Budson, A. E., Daffner, K. R., & Schacter, D. L. (2001). Perceptual false recognition in Alzheimer’s disease. *Neuropsychology*, *15*(2), 230–243. https://doi.org/10.1037//0894-4105.15.2.230

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer’s disease. *Brain and Cognition, 51*(3), 251–261. https://doi.org/10.1016/S0278-2626(03)00030-7

Burnside, K., Hope, C., Gill, E., & Morcom, A. M. (2017). Effects of perceptual similarity but not semantic association on false recognition in aging. *PeerJ*, *5*, e4184. https://doi.org/10.7717/peerj.4184

Bussey, T. J., Saksida, L. M., & Murray, E. A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *European Journal of Neuroscience*, *15*(2), 365–374. https://doi.org/10.1046/j.0953-816x.2001.01851.x

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin and Review*, *11*(5), 921–925. https://doi.org/10.3758/BF03196722

Cann, D. R., Mcrae, K., & Katz, A. N. (2011). False recall in the Deese-Roediger-McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology*, *64*(8), 1515–1542. https://doi.org/10.1080/17470218.2011.560272

Cann, D. R., McRae, K., & Katz, A. N. (2011). False recall in the Deese–Roediger–McDermott paradigm: The roles of gist and associative strength. *The Quarterly Journal of Experimental Psychology*, *64*(8), 1515–1542. https://doi.org/10.1080/17470218.2011.560272

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, *20*(4), 718–721. https://doi.org/10.1037/0882-7974.20.4.718

Castel, A. D., McGillivray, S., & Worden, K. M. (2013). Back to the future: Past and future era-based schematic support and associative memory for prices in younger and older adults. *Psychology and Aging*, *28*(4), 996–1003. https://doi.org/10.1037/a0034160

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences*, *113*(36), 10180–10185. https://doi.org/10.1073/pnas.1610686113

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. https://doi.org/10.1038/13217

Chen, H., Zhou, W., & Yang, J. (2019). Dissociation of the perirhinal cortex and hippocampus during discriminative learning of similar objects. *The Journal of Neuroscience*, *39*(31), 6190–6201. https://doi.org/10.1523/JNEUROSCI.3181-18.2019

Chouinard, P. A., & Goodale, M. A. (2012). FMRI-adaptation to highly-rendered color photographs of animals and manipulable artifacts during a classification task. *NeuroImage*, *59*(3), 2941–2951. https://doi.org/10.1016/j.neuroimage.2011.09.073

Clark, H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359. Retrieved from file:///Users/baumann/Documents/Mendeley Desktop/Clark_1973_LanguageAsAFixedEffectFallacy.pdf

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *Journal of Neuroscience*, *34*(14), 4766–4775. https://doi.org/10.1523/JNEUROSCI.2828-13.2014

Clarke, Alex, Devereux, B. J., Randall, B., & Tyler, L. K. (2014). Predicting the time course of individual objects with MEG. *Cerebral Cortex*, *25*(10), 1–11. https://doi.org/10.1093/cercor/bhu203

Clarke, Alex, & Tyler, L. K. (2015). Understanding what we see: How we derive meaning from vision. *Trends in Cognitive Sciences*, *19*(11), 677–687. https://doi.org/10.1016/j.tics.2015.08.008

Coane, J. H., McBride, D. M., Termonen, M. L., & Cutting, J. C. (2016). Categorical and associative relations increase false memory relative to purely associative relations. *Memory and Cognition*, *44*(1), 37–49. https://doi.org/10.3758/s13421-015-0543-1

Coane, J. H., McBride, D. M., & Xu, S. (2020). The feature boost in false memory: The roles of monitoring and critical item identifiability. *Memory*, *28*(4), 481–493. https://doi.org/10.1080/09658211.2020.1735445

Colsher, P. L., & Wallace, R. B. (1991). Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology*, *1*(3), 215–230. https://doi.org/10.1016/1047-2797(91)90001-S

Connolly, A. C., Swaroop Guntupalli, J., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., … Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Cowell, R. A., Barense, M. D., & Sadil, P. S. (2019). A roadmap for understanding memory: Decomposing cognitive processes into operations and representations. *ENeuro*, *6*(4), 1–19. https://doi.org/10.1523/ENEURO.0122-19.2019

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). FMRI clustering in AFNI: False-positive rates redux. *Brain Connectivity*, *7*(3), 152–171. https://doi.org/10.1089/brain.2016.0475

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human Memory and Cognitive Capabilities* (pp. 409–422). https://doi.org/10.4324/9781315440446

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), The handbook of aging and cognition (pp. 51–110). Lawrence Erlbaum Associates, Inc.

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

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 474–479. https://doi.org/10.1037/0278-7393.13.3.474

Curran, T., Schacter, D. L., Norman, K. A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia*, *35*(7), 1035–1049. https://doi.org/10.1016/S0028-3932(97)00029-8

Dale, E., & O'Rourke, J. (1981). The living word vocabulary. World Book Childcraft International.

Danckert, S. L., & Craik, F. I. M. (2013). Does aging affect recall more than recognition memory? *Psychology and Aging*, *28*(4), 902–909. https://doi.org/10.1037/a0033263

Davidson, P. S. R., Vidjen, P., Trincao-Batra, S., Collin, C. A., & Gutchess, A. (2019). Older adults' lure discrimination difficulties on the mnemonic similarity task are significantly correlated with their visual perception. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *74*(8), 1298–1307. https://doi.org/10.1093/geronb/gby130

Davis, H. P., Small, S. A., Stern, Y., Mayeux, R., Feldstein, S. N., & Keller, F. R. (2003). Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex*, *39*(4–5), 1063–1091. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14584567

Davis, S. W., Geib, B. R., Wing, E. A., Wang, W.-C., Hovhannisyan, M., Monge, Z. A., & Cabeza, R. (2020). Visual and semantic representations predict subsequent memory in perceptual and conceptual memory tests. *Cerebral Cortex*, *00*(00), 1–19. https://doi.org/10.1093/cercor/bhaa269

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*(1), 198–205. https://doi.org/10.3758/BRM.40.1.198

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychol. Methods*, *3*(2), 186–205. https://doi.org/10.1037/1082-989X.3.2.186

Deese, J. (1959a). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *35*(1), 2309. https://doi.org/10.13646/j.cnki.42-1395/u.2014.07.005

Deese, J. (1959b). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dehon, H., & Brédart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology and Aging*, *19*(1), 191–197. https://doi.org/10.1037/0882-7974.19.1.191

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(5), 1136–1146. https://doi.org/10.1037/0278-7393.22.5.1136

Deng, L., Davis, S. W., Monge, Z. A., Wing, E. A., Geib, B. R., Raghunandan, A., & Cabeza, R. (2021). Age-related dedifferentiation and hyperdifferentiation of perceptual and mnemonic representations. *Neurobiology of Aging*, *106*, 55–67.

https://doi.org/10.1016/j.neurobiolaging.2021.05.021

Dennis, N. A., Bowman, C. R., & Turney, I. C. (2015). Functional neuroimaging of false memories. In *The Wiley Handbook on the Cognitive Neuroscience of Memory* (pp. 150–171). https://doi.org/10.1002/9781118332634.ch8

Dennis, N. A., Bowman, C. R., & Vandekar, S. N. (2012). True and phantom recollection: An fMRI investigation of similar and distinct neural correlates and connectivity. *NeuroImage*, *59*(3), 2982–2993. https://doi.org/10.1016/j.neuroimage.2011.09.079

Dennis, N. A., Hayes, S. M., Prince, S. E., Madden, D. J., Huettel, S. A., & Cabeza, R. (2008). Effects of aging on the neural correlates of successful item and source memory encoding. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(4), 791–808. https://doi.org/10.1037/0278-7393.34.4.791

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia*, *45*(14), 3157–3166. https://doi.org/10.1016/j.neuropsychologia.2007.07.003

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, *33*(48), 18906–18916. https://doi.org/10.1523/JNEUROSCI.3809-13.2013

Devereux, B. J., Clarke, A., & Tyler, L. K. (2018). Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-28865-1

Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The Centre for Speech, Language and the Brain (CSLB) concept property norms. *Behavior Research Methods, 46*(4), 1119–1127. https://doi.org/10.3758/s13428-013-0420-4

Devlin, J. T., & Price, C. J. (2007). Perirhinal contributions to human visual perception. *Current Biology*, *17*(17), 1484–1488. https://doi.org/10.1016/j.cub.2007.07.066

Dewhurst, S. A., Hitch, G., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recogntion and recall. *Journal of Experimental Psychology*, *24*(2), 284–

298.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *PNAS*, *113*(33), 7900–7905. https://doi.org/10.1073/pnas.1612033113

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Evans, D. A., Beckett, L. A., Albert, M. S., Hebert, L. E., Scherr, P. A., Funkenstein, H. H., & Taylor, J. O. (1993). Level of education and change in cognitive function in a community population of older persons. *Annals of Epidemiology*, *3*(1), 71–77. https://doi.org/10.1016/1047-2797(93)90012-S

Forde, E. M. E., & Humphreys, G. W. (1999). Category-specific recognition impairments: A review of important case studies and influential theories. *Aphasiology*, *13*(3), 169–193. https://doi.org/10.1080/026870399402172

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: A meta-analysis. *Psychological Bulletin*, *145*(4), 339–371. https://doi.org/10.1037/bul0000185

Gabrieli, J. D. E., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 906–913. https://doi.org/10.1073/pnas.95.3.906

Gallo, D. A. (2004). Using recall to reduce false recognition: diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(1), 120–128. https://doi.org/10.1037/0278-7393.30.1.120

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*(7), 833–848. https://doi.org/10.3758/MC.38.7.833

Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology and Aging*, *22*(1), 209–213. https://doi.org/10.1037/0882-7974.22.1.209

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, *47*(3), 469–497. https://doi.org/10.1016/S0749-596X(02)00013-X

Gallo, D., Bell, D., Beier, J., & Schacter, D. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, *14*(6), 730–741. https://doi.org/10.1080/09658210600648506

Ganguli, M., Lytle, M. E., Reynolds, M. D., & Dodge, H. H. (1998). Random versus volunteer selection for a community-based study. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *53*(1), 39–46. https://doi.org/10.1093/gerona/53A.1.M39

Garoff-Eaton, R. J., Slotnick, S. D., & Schacter, D. L. (2006). Not all false memories are created equal: The neural basis of false recognition. *Cerebral Cortex*, *16*(11), 1645–1652. https://doi.org/10.1093/cercor/bhj101

Garoff, R. J., Slotnick, S. D., & Schacter, D. L. (2005). The neural origins of specific and general memory: The role of the fusiform cortex. *Neuropsychologia*, *43*(6), 847–859. https://doi.org/10.1016/j.neuropsychologia.2004.09.014

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality , distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174. https://doi.org/10.1080/02643290125857

Gellersen, H. M., Trelle, A. N., Henson, R. N., & Simons, J. S. (2021). Executive function and high ambiguity perceptual discrimination contribute to individual differences in mnemonic discrimination in older adults. *Cognition*, *209*(December 2020), 104556. https://doi.org/10.1016/j.cognition.2020.104556

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20. https://doi.org/10.3758/BF03198438

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science, 15*(10), 655–660. https://doi.org/10.1111/j.0956-7976.2004.00736.x

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*(1), 23–29. https://doi.org/10.1037/h0040561

Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience, 13*(7), 491–505. https://doi.org/10.1038/nrn3256

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Grégoire, J., & Van der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition*, *4*(2), 140–149. https://doi.org/10.1080/13825589708256642

Gutchess, A. H., & Schacter, D. L. (2012). The neural correlates of gist-based true and false recognition. *NeuroImage*, *59*(4), 3418–3426. https://doi.org/10.1016/j.neuroimage.2011.11.078

Gutchess, A. H., Welsh, R. C., Hedden, T., Bangert, A., Minear, M., Liu, L. L., & Park, D. C. (2005). Aging and the neural correlates of successful picture encoding: Frontal activations compensate for decreased medial-temporal activity. *Journal of Cognitive Neuroscience*, *17*(1), 84–96. https://doi.org/10.1162/0898929052880048

Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron, 44*(4), 581–584. https://doi.org/10.1016/j.neuron.2004.11.003

Heit, E., Brockdorff, N., & Lamberts, K. (2003). Adaptive changes of response criterion in recognition memory. *Psychonomic Bulletin and Review*, *10*(3), 718–723. https://doi.org/10.3758/BF03196537

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring:

Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127*(3), 251–268. https://doi.org/10.1037/0096-3445.127.3.251

Hertzog, C. (2002). Metacognition in older adults: implications for application. In *Applied Metacognition* (pp. 169–196). https://doi.org/10.1017/cbo9780511489976.009

Hess, T. M. (1990). Aging and schematic influences on memory. In T. M. Hess (Ed.), Aging and cognition: Knowledge organization and utilization (pp. 93–160). North-Holland. https://doi.org/10.1016/S0166-4115(08)60157-0

Hester, R. L., Kinsella, G. J., & Ong, B. (2004). Effect of age on forward and backward span tasks. *Journal of the International Neuropsychological Society*, *10*(4), 475–481. https://doi.org/10.1017/S1355617704104037

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Hoffman, P., Binney, R. J., & Lambon Ralph, M. A. (2015). Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex*, *63*, 250–266. https://doi.org/10.1016/j.cortex.2014.09.001

Holden, H. M., Hoebel, C., Loftis, K., & Gilbert, P. E. (2012). Spatial pattern separation in cognitively normal young and older adults. *Hippocampus*, *22*(9), 1826–1832. https://doi.org/10.1002/hipo.22017

Holdstock, J. S., Hocking, J., Notley, P., Devlin, J. T., & Price, C. J. (2009). Integrating visual and tactile information in the perirhinal cortex. *Cerebral Cortex, 19*(12), 2993–3000. https://doi.org/10.1093/cercor/bhp073

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Maitland, S. B., & Dixon, R. A. (2002). Sampling and generalisability in developmental research: Comparison of random and convenience samples of older adults. *International Journal of Behavioral Development*, *26*(4), 345–359. https://doi.org/10.1080/01650250143000247

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, *5*(1), 67–103. https://doi.org/10.1080/02643298808252927

Humphreys, G W, & Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object recognition: “category-specific” neuropsychological deficits. *The Behavioral and Brain Sciences*, *24*(3), 453–476; discussion 476-509. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11682799

Humphreys, Glyn W, Price, J., & Riddoch, M. J. (1999). From objects to names : A cognitive neuroscience approach. *Psychological Research*, *62*, 118–130.

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*(1), 36–58. https://doi.org/10.1016/j.neubiorev.2012.09.014

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497–514. https://doi.org/10.1016/S0022-5371(81)90138-9

Jackson, R. L., Hoffman, P., Pobric, G., & Ralph, M. A. L. (2015). The nature and neural correlates of semantic association versus conceptual similarity. *Cerebral Cortex*, *25*(11), 4319–4333. https://doi.org/10.1093/cercor/bhv003

Johnson, M. K. (1997). Source monitoring and memory distortion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1362), 1733–1745.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, Vol. 114, pp. 3–28. https://doi.org/10.1037//0033-2909.114.1.3

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989. https://doi.org/10.1080/17470218.2011.588335

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kanwisher N, Downing P, Epstein R, Kourtzi Z (2001) Functional neuroimaging of

visual recognition. In: Handbook of Functional Neuroimaging of Cognition (Cabeza R, Kingstone A, ed), pp. 109–151. The MIT Press, Cambridge, Massachusetts.

Karanian, J. M., & Slotnick, S. D. (2017). False memories for shape activate the lateral occipital complex. *Learning and Memory*, *24*(10), 552–556. https://doi.org/10.1101/lm.045765.117

Karanian, J. M., & Slotnick, S. D. (2018). Confident false memories for spatial location are mediated by V1. *Cognitive Neuroscience*, *9*(3–4), 139–150. https://doi.org/10.1080/17588928.2018.1488244

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. https://doi.org/10.1038/nature06713

Kent, C., Lamberts, K., & Patton, R. (2018). Cue quality and criterion setting in recognition memory. *Memory and Cognition*, *46*(5), 757–769. https://doi.org/10.3758/s13421-018-0796-6

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage*, *54*(3), 2446–2461. https://doi.org/10.1016/j.neuroimage.2010.09.045

Kim, H., & Cabeza, R. (2007). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex*, *17*(9), 2143–2150. https://doi.org/10.1093/cercor/bhl122

Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *Journal of Neuroscience*, *20*(16), 6173–6180. https://doi.org/10.1523/jneurosci.20-16-06173.2000

Kirwan, C. B., & Stark, C. E. L. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, *14*(9), 625–633. https://doi.org/10.1101/lm.663507

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007). What's new in psychtoolbox-3. Perception 36:1–16.

Klippenstein, J. L., Stark, S. M., Stark, C. E. L., & Bennett, I. J. (2020). Neural substrates of mnemonic discrimination: A whole-brain fMRI investigation. *Brain and Behavior*, *10*(3), 1–11. https://doi.org/10.1002/brb3.1560

Koen, J. D., Hauck, N., & Rugg, M. D. (2019). The relationship between age, neural differentiation, and memory performance. *Journal of Neuroscience, 39*(1), 149–162. https://doi.org/10.1523/JNEUROSCI.1498-18.2018

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165

Koustaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging*, *14*(2), 220–237. https://doi.org/10.1037/0882-7974.14.2.220

Koutstaal, W. (2003). Older adults encode - but do not always use - perceptual details: Intentional versus unintentional effects of detail on memory judgments. *Psychological Science*, *14*(2), 189–193. https://doi.org/10.1111/1467-9280.01441

Koutstaal, W. (2006). Flexible remembering. *Psychonomic Bulletin & Review*, *13*(1), 84–91. https://doi.org/10.3758/BF03193817

Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False Recognition of Abstract Versus Common Objects in Older and Younger Adults: Testing the Semantic Categorization Account. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(4), 499–510. https://doi.org/10.1037/0278-7393.29.4.499

Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., Schacter, D. L., … Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: Testing the semantic categorization account. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(4), 499–510. https://doi.org/10.1037/0278-7393.29.4.499

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *583*(37), 555–583. Retrieved from http://www.sciencedirect.com/science/article/pii/S0749596X97925298

Koutstaal, W., Schacter, D. L., & Brenner, C. (2001). Dual task demands and gist-based false recognition of pictures in younger and older adults. *Journal of Memory and Language*, *44*(3), 399–426. https://doi.org/10.1006/jmla.2000.2734

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C., & Jackson, E. M. (1999). Perceptually based false recognition of novel objects in amnesia: Effects of category size and similarity to category prototypes. *Cognitive Neuropsychology*, *16*(3–5), 317–341. https://doi.org/10.1080/026432999380816

Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., & Bandettini, P. A. (2017). Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *NeuroImage*, *154*(March), 59–80. https://doi.org/10.1016/j.neuroimage.2017.03.033

Kvavilashvili, L., Kornbrot, D. E., Mash, V., Cockburn, J., & Milne, A. (2009). Differential effects of age on prospective and retrospective memory tasks in young, young-old, and old-old adults. *Memory*, *17*(2), 180–196. https://doi.org/10.1080/09658210802194366

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. L. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using highresolution fMRI and variable mnemonic similarity. *Learning and Memory*, *18*(1), 15–18. https://doi.org/10.1101/lm.1971111

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

LaVoie, D. J., & Faulkner, K. (2000). Age differences in false recognition using a forced choice paradigm. *Experimental Aging Research*, *26*(4), 367–381. https://doi.org/10.1080/036107300750015750

Leal, S. L., & Yassa, M. A. (2014). Effects of aging on mnemonic discrimination of emotional information. *Behavioral Neuroscience*, *128*(5), 539–547. https://doi.org/10.1037/bne0000011

Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science*, *346*(6209), 572–578. Retrieved from http://science.sciencemag.org/%0Ahttp://science.sciencemag.org/%0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=ovftp&AN=00007529-201410310-00032

Lineweaver, T. T., & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, and Cognition*, *5*(4), 264–296. https://doi.org/10.1076/anec.5.4.264.771

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology, 3*(4), 525–551. https://doi.org/10.1016/0010-0285(72)90021-7

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

Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language*, *49*(2), 268–283. https://doi.org/10.1016/S0749-596X(03)00069-X

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hearing*, *19*(1), 1–36.

Luo, L., & Craik, F. I. F. (2008). Aging and memory: A cognitive approach. *Canadian Journal of Psychiatry-Revue Canadienne De Psychiatrie*, *53*(6), 346–353. https://doi.org/10.1177/070674370805300603

Luo, L., & Craik, F. I. M. (2009). Age differences in recollection: Specificity effects at retrieval. *Journal of Memory and Language*, *60*(4), 421–436. https://doi.org/10.1016/j.jml.2009.01.005

Ly, M., Murray, E., & Yassa, M. A. (2013). Perceptual versus conceptual interference and pattern separation of verbal stimuli in young and older adults. *Hippocampus*, *23*(6), 425–430. https://doi.org/10.1002/hipo.22110

MATLAB (2019b). version 9.7.0.1216025 (R2019b). Natick, Massachusetts: The MathWorks Inc.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron, 63*(3), 397–405. https://doi.org/10.1016/j.neuron.2009.07.012

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *262*(841), 23–81.

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

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94,* 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Matzen, L. E., & Benjamin, A. S. (2013). Older and wiser: Older adults' episodic word memory benefits from sentence study contexts. *Psychology and Aging, 28*(3), 754–767. https://doi.org/10.1037/a0032945

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

McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(5), 1177–1194. https://doi.org/10.1037/0278-7393.25.5.1177

McGillivray, S., & Castel, A. D. (2017). Older and younger adults' strategic control of metacognitive monitoring: the role of consequences, task experience, and prior knowledge. *Experimental Aging Research, 43*(3), 233–256. https://doi.org/10.1080/0361073X.2017.1298956

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage. *Trends in Neural Sciences*, *10*(10), 408–415.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

McRae, K., De Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99–130. https://doi.org/10.1037/0096-3445.126.2.99

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*(6), 1198–1215. https://doi.org/10.1080/17470210802391599

Millard, S. P., & Neerchal, N. K. (2001). Environmental Statistics with S-PLUS. In *CRC Press, Boca Raton, FL*.

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499–520. https://doi.org/10.1037/bul0000092

Molitor, R. J., Ko, P. C., Hussey, E. P., & Ally, B. A. (2014). Memory-related eye movements challenge behavioral measures of pattern completion and pattern separation. *Hippocampus*, *24*(6), 666–672. https://doi.org/10.1002/hipo.22256

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). Semantic memory : A feature-based analysis and new norms for Italian. *Behavior Research Methods*, *45*(2), 440–461. https://doi.org/10.3758/s13428-012-0263-4

Montefinese, M., Vinson, D., & Ambrosini, E. (2018). Recognition memory and featural similarity between concepts: The pupil's point of view. *Biological Psychology*, *135*, 159–169. https://doi.org/10.1016/j.biopsycho.2018.04.004

Montefinese, M., Zannino, G. D., & Ambrosini, E. (2015). Semantic similarity between old and new items produces false alarms in recognition memory. *Psychological Research*, *79*(5), 785–794. https://doi.org/10.1007/s00426-014-0615-z

Morcom, A. M., Good, C. D., Frackowiak, R. S. J., & Rugg, M. D. (2003). Age effects on the neural correlates of successful memory encoding. *Brain*, *126*(1), 213–229. https://doi.org/10.1093/brain/awg020

Moss, H. E., Tyler, L. K., & Jennings, F. (1997). When leopards lose their spots: Knowledge of visual properties in category-specific deficits for living things. *Cognitive Neuropsychology*, *14*(6), 901–950. https://doi.org/10.1080/026432997381394

Moss, Helen E., Tyler, L. K., Durrant-Peatfield, M., & Bunn, E. M. (1998). "Two eyes of a see-through": Impaired and intact semantic knowledge in a case of selective deficit for living things. *Neurocase*, *4*(4–5), 291–310. https://doi.org/10.1093/neucas/4.4.291-a

Motley, S. E., & Kirwan, C. B. (2012). A parametric investigation of pattern separation processes in the medial temporal lobe. *Journal of Neuroscience*, *32*(38), 13076–13084. https://doi.org/10.1523/JNEUROSCI.5920-11.2012

Mumford, J. A. (2012). A power calculation guide for FMRI studies. *Social Cognitive and Affective Neuroscience*, *7*(6), 738–742. https://doi.org/10.1093/scan/nss059

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classi fi cation analyses. *NeuroImage*, *59*(3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076

Naspi, L., Hoffman, P., Devereux, B., & Morcom, A. (2021). Perceptual and semantic representations at encoding contribute to true and false recognition of objects. https://osf.io/2hgbw/

Naspi L, Hoffman P, Devereux B, Thejll-Madsen T, Doumas L, Morcom AM (2020) Multiple dimensions of semantic and perceptual similarity contribute to mnemonic discrimination for pictures. https://psyarxiv.com/qt5wc/. 10.31234/osf.io/qt5wc.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Nelson, D. L., McEvoy, C L.., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407. Retrieved from http://www.usf.edu/FreeAssociation/

Norman, K A, & O’Reilly, R. C. (2003). Modeling hippocampal and noecortical contributions to recognition memory: A complementary-learning systems approach. *Psychological Review*, *110*(4), 611–646.

Norman, Kenneth A. (2010). How hippocampus and cortex contribute to recognition memory: Revisiting the complementary learning systems model. *Hippocampus, 20*(11), 1217–1227. https://doi.org/10.1002/hipo.20855

Norman, Kenneth A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory and Cognition,* *25*(6), 838–848. https://doi.org/10.3758/BF03211328

Nyberg, L., Rönnlund, M., Dixon, R. A., Maitland, S. B., Bäckman, L., Wahlin, Å., & Nilsson, L. G. (2003). Selective adult age differences in an age-invariant multifactor model of declarative memory. *Psychology and Aging*, *18*(1), 149–160. https://doi.org/10.1037/0882-7974.18.1.149

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus, 4*(6), 661–682. https://doi.org/10.1002/hipo.450040605

Odegard, T. N., & Lampinen, J. M. (2005). Recollection rejection: Gist cuing of verbatim memory. *Memory and Cognition*, *33*(8), 1422–1430. https://doi.org/10.3758/BF03193375

Okado, Y., & Stark, C. E. L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning and Memory, 12*(1), 3–11. https://doi.org/10.1101/lm.87605

Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature*, *202*, 1031–1032. Retrieved from https://www.nature.com/articles/201464a0.pdf

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. In *Progress in Brain Research* (Vol. 155, pp. 23–36). https://doi.org/10.1016/S0079-6123(06)55002-2

Otten, L. J. (2007). Fragments of a larger whole: Retrieval cues constrain observed neural correlates of memory encoding. *Cerebral Cortex*, *17*(9), 2030–2038. https://doi.org/10.1093/cercor/bhl111

Otten, L. J., Henson, R. N. A., & Rugg, M. D. (2002). State-related and item-related neural correlates of successful memory encoding. *Nature Neuroscience, 5*(12), 1339–1344. https://doi.org/10.1038/nn967

Otten, L. J., & Rugg, M. D. (2001). Task-dependency of the neural correlates of episodic encoding as measured by fMRI. *Cerebral Cortex, 11*(12), 1150–1160. https://doi.org/10.1093/cercor/11.12.1150

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of kousta et al. (2011). *Journal of Experimental Psychology: General*, *142*(1), 282–287. https://doi.org/10.1037/a0027004

Paivio, A., & Begg, I. (1971). Imagery and comprehension latencies as a function of sentence concreteness and structure. *Perception & Psychophysics*, *10*(6), 408–412. https://doi.org/10.3758/BF03210323

Pansuwan, T., Breuer, F., Gazder, T., Lau, Z., Cueva, S., Swanson, L., … Morcom, A. M. (2020). Evidence for adult age-invariance in associative false recognition. *Memory*, *28*(2), 172–186. https://doi.org/10.1080/09658211.2019.1705351

Park, D. C., Polk, T. A., Park, R., Minear, M., Savage, A., & Smith, M. R. (2004). Aging reduces neural specialization in ventral visual cortex. *PNAS*, *101*(35), 13091–13095. https://doi.org/10.1073/pnas.0405148101

Penny, W. D., Trujillo-Barreto, N. J., & Friston, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, *24*(2), 350–362. https://doi.org/10.1016/j.neuroimage.2004.08.034

Pidgeon, L. M., & Morcom, A. M. (2014). Age-related increases in false recognition: The role of perceptual and conceptual similarity. *Frontiers in Aging Neuroscience*, *6*, 1–17. https://doi.org/10.3389/fnagi.2014.00283

Pidgeon, L. M., & Morcom, A. M. (2016). Cortical pattern separation and item-specific memory encoding. *Neuropsychologia*, *85*, 256–271. https://doi.org/10.1016/j.neuropsychologia.2016.03.026

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Prince, S. E., Tsukiura, T., & Cabeza, R. (2007). Distinguishing the neural correlates of episodic memory encoding and semantic memory retrieval. *Psychological Science*, *18*(2), 144–151. https://doi.org/10.1111/j.1467-9280.2007.01864.x

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. https://doi.org/10.1016/j.jml.2008.02.002

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: Behavioral and computational studies. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(2), 393–406. https://doi.org/10.1037/0278-7393.30.2.393

Rankin, J. L., & Kausler, D. H. (1979). Adult age differences in false recognitions. *Journal of Gerontology, 34*(1), 58–65.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., … Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex, 15*(11), 1676–1689. https://doi.org/10.1093/cercor/bhi044

Reagh, Z. M., Ho, H. D., Leal, S. L., Noche, J. A., Chun, A., Murray, E. A., & Yassa, M. A. (2016). Greater loss of object than spatial mnemonic discrimination in aged adults. *Hippocampus, 26,* 417–422. https://doi.org/10.1002/hipo.22562

Reagh, Z. M., Noche, J. A., Tustison, N. J., Delisle, D., Murray, E. A., Yassa, M. A., … Murray, E. A. (2018). Functional imbalance of anterolateral entorhinal cortex and hippocampal dentate/CA3 underlies age-related object pattern separation deficits. *Neuron*, *97*(5), 1187–1198. https://doi.org/10.1016/j.neuron.2018.01.039

Reagh, Z. M., Roberts, J. M., Ly, M., Diprospero, N., Murray, E., & Yassa, M. A. (2014). Spatial discrimination deficits as a function of mnemonic interference in aged adults with and without memory impairment. *Hippocampus, 24*(3), 303–314. https://doi.org/10.1002/hipo.22224

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: A computational model of the encoding/retrieval trade-off with familiarity. *Psychology of Learning and Motivation - Advances in Research and Theory, 48*(07), 271–312. https://doi.org/10.1016/S0079-7421(07)48007-0

Reppa, I., Williams, K. E., Greville, W. J., & Saunders, J. (2020). The relative contribution of shape and colour to object memory. *Memory and Cognition, 48*(8), 1504–1521. https://doi.org/10.3758/s13421-020-01058-w

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

Reyna, Valerie F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*(1), 1–9. https://doi.org/10.1016/j.jarmac.2015.12.003

Reyna, Valerie F., Holliday, R., & Marche, T. (2002). Explaining the development of false memories. *Developmental Review*, *22*(3), 436–489. https://doi.org/10.1016/S0273-2297(02)00003-5

Reyna, Valerie F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*(2), 178–191.

Reyna, Valerie F., & Kiernan, B. (1995). Children's memory and metaphorical interpretation. *Metaphor and Symbol*, *10*(4), 309–331. https://doi.org/10.1207/s15327868ms1004

Reyna, Valerie F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*(2), 95–123. https://doi.org/10.1016/S1041-6080(97)90002-9

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.686.709&rep=rep1&type=pdf

Roediger, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, *9*(4), 123–127. https://doi.org/10.1111/1467-8721.00075

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: a multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, *20*(1), 3–18. https://doi.org/10.1037/0882-7974.20.1.3

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth Mover ' s Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, *40*(2), 99–121.

Sahay, A., Wilson, D. A., & Hen, R. (2011). Pattern separation: A common function for new neurons in hippocampus and olfactory bulb. *Neuron*, *70*, 582–288. https://doi.org/10.1016/j.neuron.2011.05.012

Salthouse, T. A. (1992). What do adult age differences in the Digit Symbol Substitution Test reflect? *Journals of Gerontology*, *47*(3), 121–128. https://doi.org/10.1093/geronj/47.3.P121

Schacter, D. L., Koustaal, W., & Norman, K. A. (1997). False memories aging. *Trends in Cognitive Sciences*, *1*(6), 229–236.

Schacter, D. L., Norman, K. A., & Koustaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*, 289–318. https://doi.org/10.1007/s11229-015-0783-4

Schacter, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron*, *44*(1), 149–160. https://doi.org/10.1016/j.neuron.2004.08.017

Schmidt, S. R. (1996). Category typicality effects in episodic memory: Testing models of distinctiveness. *Memory and Cognition*, *24*(5), 595–607. https://doi.org/10.3758/BF03201086

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Seamon, J. G., Luo, C. R., Schlegel, S. E., Greene, S. E., & Goldenberg, A. B. (2000). False memory for categorized pictures and words: the category associates procedure for studying memory errors in children and adults. *Journal of Memory and Language, 42*, 120–146.

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition, 21*(1), 34–40. https://doi.org/10.3758/BF03211162

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research, 165*(06), 33–56. https://doi.org/10.1016/S0079-6123(06)65004-8

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(3), 411–426. https://doi.org/10.1109/TPAMI.2007.56

Slotnick, S. D., & Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience, 7*(6), 664–672. https://doi.org/10.1038/nn1252

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin and Review, 5*(4), 710–715. https://doi.org/10.3758/BF03208850

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Sommers, M. S., & Huff, L. M. (2003). The effects of age and dementia of the Alzheimer's type on phonological false memories. *Psychology and Aging, 18*(4), 791–806. https://doi.org/10.1037/0882-7974.18.4.791

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10*(4), 527–539. https://doi.org/10.1037/0882-7974.10.4.527

St-Laurent, M., Abdi, H., Bondad, A., & Buchsbaum, B. R. (2014). Memory reactivation in healthy aging: Evidence of stimulus-specific dedifferentiation. *Journal of Neuroscience, 34*(12), 4175–4186. https://doi.org/10.1523/JNEUROSCI.3054-13.2014

Stahl, C., Henze, L., & Aust, F. (2016). *False memory for perceptually similar but* *conceptually distinct line drawings*. 1–28.

Stahl, Christoph, & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(3), 570–586. https://doi.org/10.1037/0278-7393.34.3.570

Standing, L. (1973). Learning 10,000 pictures. *Quart.J.Exp.Psychol.*, *25*(2), 207–222. https://doi.org/10.1080/14640747308400340

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic reinstatement in the medial temporal lobe. *Journal of Neuroscience*, *32*(50), 18150–18156. https://doi.org/10.1523/JNEUROSCI.4156-12.2012

Stark, S. M., Yassa, M. A., & Stark, C. E. L. (2010). Individual differences in spatial pattern separation performance associated with healthy aging in humans. *Learning & Memory*, *17*(6), 284–288. https://doi.org/10.1101/lm.1768110

Stark, Shauna M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Stern, C. E., Corkin, S., González, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., … Rosen, B. R. (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(16), 8660–8665. https://doi.org/10.1073/pnas.93.16.8660

Taylor, K. I., & Tyler, L. K. (2012). *Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual object. Supplementary material*.

Taylor, K. I., Devereux, B. J., Acres, K., Randall, B., & Tyler, L. K. (2012). Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. *Cognition*, *122*(3), 363–374. https://doi.org/10.1016/j.cognition.2011.11.001

Taylor, K. I., Devereux, B. J., & Tyler, L. K. (2011). Conceptual structure: Towards an integrated neurocognitive account. *Language and Cognitive Processes*, *26*(9), 1368–1401. https://doi.org/10.1080/01690965.2011.568227

Thomas, A. K., & McDaniel, M. A. (2013). The interaction between frontal functioning and encoding processes in reducing false memories. *Aging, Neuropsychology, and Cognition*, *20*(4), 443–470. https://doi.org/10.1080/13825585.2012.736468

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language*, *52*(1), 71–86. https://doi.org/10.1016/j.jml.2004.08.001

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*(2), 203–214. https://doi.org/10.1016/S0887-6177(03)00039-8

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167–177. https://doi.org/10.1016/S0887-6177(97)00095-4

Tomer, R., & Levin, B. E. (1993). Differential effects of aging on two verbal fluency tasks. *Perceptual and Motor Skills*, *76*(2), 465–466. https://doi.org/10.2466/pms.1993.76.2.465

Toner, C. K., Pirogovsky, E., Kirwan, C. B., & Gilbert, P. E. (2009). Visual object pattern separation deficits vary in cognitively normal older adults. *Society for Neuroscience Abstract Viewer and Itinerary Planner*, *16*, 338–342. https://doi.org/10.1101/lm.1315109.subsequent

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging*, *13*(2), 230–241. https://doi.org/10.1037/0882-7974.13.2.230

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language*, *75*(2), 195–231. https://doi.org/10.1006/brln.2000.2353

Tyler, Lorraine K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*(6), 244–252. https://doi.org/10.1016/S1364-

6613(00)01651-X

Tyler, Lorraine K, Chiu, S., Zhuang, J., Randall, B., Devereux, B. J., Wright, P., … Taylor, K. I. (2013). Objects and categories : Feature statistics and object processing in the ventral stream. *Journal of Cognitive Neuroscience*, *25*(10), 1723–1735. https://doi.org/10.1162/jocn

Tyler, Lorraine K, Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. E. (2004). Processing objects at different levels of specificity. *Journal of Cognitive Neuroscience*, *16*(3), 351–362. https://doi.org/10.1162/089892904322926692

Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*, *9*(4), 408–426. https://doi.org/10.1177/1745691614535933

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. E. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia*, *40*(12), 2136–2143. https://doi.org/10.1016/S0028-3932(02)00053-2

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Verfaellie, M., Schacter, D. L., & Cook, S. P. (2002). The effect of retrieval instructions on false recognition: Exploring the nature of the gist memory impairment in amnesia. *Neuropsychologia*, *40*(13), 2360–2368. https://doi.org/10.1016/S0028-3932(02)00074-X

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*(4), 422–488. https://doi.org/10.1016/j.cogpsych.2003.09.001

Voss, J. L., Lucas, H. D., & Paller, K. A. (2012). More than a feeling: pervasive influences of memory without awareness of retrieval. *Cognitive Neuroscience*, *3*(3–4), 193–207. https://doi.org/10.1080/17588928.2012.674935

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., … Buckner,

R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281*(5380), 1188–1191. https://doi.org/10.1126/science.281.5380.1188

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

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*(3), 829–853. https://doi.org/10.1093/brain/107.3.829

Watson, J. M., Balota, D. A., & Roediger, H. L. (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language, 49*(1), 95–118. https://doi.org/10.1016/S0749-596X(03)00019-6

Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, *15*(2), 254–267. https://doi.org/10.1037/0894-4105.15.2.254

Weller, H. I., & Westneat, M. W. (2019). Quantitative color profiling of digital images with earth mover’s distance using the R package colordistance. *PeerJ, 7*, 1–31. https://doi.org/10.7717/peerj.6398

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272–292. https://doi.org/10.1037/0033-2909.120.2.272

Wilson, I. A., Gallagher, M., Eichenbaum, H., & Tanila, H. (2006). Neurocognitive aging: prior memories hinder new hippocampal encoding. *Trends in Neurosciences*, *29*(12), 662–670. https://doi.org/10.1016/j.tins.2006.10.002

Wingfield, A., & Stine, E. A. L. (1991). Expert systems in nature: Spoken language processing and adult aging. In J. D. Sinnott & J. C. Cavanaugh (Eds.), Bridging paradigms: Positive development in adulthood and cognitive aging (pp. 237–258). Praeger Publishers.

Winters, B. D., & Bussey, T. J. (2005). Transient inactivation of perirhinal cortex disrupts encoding, retrieval, and consolidation of object recognition memory. *Journal of Neuroscience*, *25*(1), 52–61. https://doi.org/10.1523/JNEUROSCI.3827-04.2005

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods*, *41*(2), 257–267. https://doi.org/10.3758/BRM.41.2.257

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer : A network visualization tool for human brain connectomics. *PLoS ONE*, *8*(7), 1–15. https://doi.org/10.1371/journal.pone.0068910

Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in Psychology*, *9*(SEP), 1–3. https://doi.org/10.3389/fpsyg.2018.01831

Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. L. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences*, *108*(21), 8873–8878. https://doi.org/10.1073/pnas.1101567108

Yassa, Michael A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, *21*(9), 968–979. https://doi.org/10.1002/hipo.20808

Yassa, Michael A., Muftuler, L. T., & Stark, C. E. L. (2010). Ultrahigh-resolution microstructural diffusion tensor imaging reveals perforant path degradation in aged humans in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(28), 12687–12691. https://doi.org/10.1073/pnas.1002113107

Yassa, Michael A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yeung, L.-K., Ryan, J. D., Cowell, R. A., & Barense, M. D. (2013). Recognition memory impairments caused by false recognition of novel objects. *Journal of Experimental Psychology: General*, *142*(4), 1384–1397. https://doi.org/10.1037/a0034021

Yonelinas, A. P., Otten, L. J., Shaw, R. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of*

Neuroscience, *25*(11), 3002–3008. https://doi.org/10.1523/JNEUROSCI.5295-04.2005

Yotsumoto, Y., Kahana, M. J., Wilson, H. R., & Sekuler, R. (2007). Recognition memory for realistic synthetic faces. *Memory and Cognition*, *35*(6), 1233–1244. https://doi.org/10.3758/BF03193597

Yuspeh, R. L., Vanderploeg, R. D., & Kershaw, D. A. J. (1998). Normative data on a measure of estimated premorbid abilities as part of a dementia evaluation. *Applied Neuropsychology*, *5*(3), 149–153. https://doi.org/10.1207/s15324826an0503_6

Zacks, R. T., & Hasher, L. (2006). Aging and long-term memory: Deficits are not inevitable. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan Cognition: Mechanisms of Change* (pp. 162–177). https://doi.org/10.1093/acprof:oso/9780195169539.003.0011

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J. Physiol.*, *277*, 273–290.

Zhu, B., Chen, C., Shao, X., Liu, W., Ye, Z., Zhuang, L., & Zheng, L. (2019). Multiple interactive memory representations underlie the induction of false memory. *PNAS*, *116*(9), 3466–3475. https://doi.org/10.1073/pnas.1817925116

Appendices

Supplemental material of Chapter 2

Nuisance Variables

Word Associations

The strength of associations between words is an established modifier of mnemonic discrimination (Deese, 1959; Roediger et al., 2001). Both forward associative strength (FAS, from an unstudied lure concept to a studied concept) and backward associative strength (BAS, from a studied concept to an unstudied lure concept) may contribute to misrecognition and false recall of lures (C. J. Brainerd & Wright, 2005; Roediger et al., 2001). However, the current English language norms (Nelson et al., 1998) only provide data for 156 of the 200 concepts with semantic feature norms used in this study (Devereux et al., 2014). The Nelson et al. (1998) norms were also gathered decades ago (from 1973) in the US. To ensure that association scores were relevant for our participants, we gathered our own association data using a method similar to De Deyne and Storms (2008).

Participants

Two hundred and six participants contributed (Age: M = 24.7, SD = 8.7, 147 female, 59 male). Participants were recruited using social media and Mechanical Turk. They were required to live in the UK and be aged 18-50 years. Non-native English speakers self-evaluated their English ability using a scale from 0 (none) to 5 (fluent) on four dimensions: expression, comprehension, reading, and writing. A score of at least 16 out of 20 was required. Forty further participants were removed who did not meet these criteria. Each concept was seen by an average of 32.4 participants (SD = 2.5).

Stimuli and Procedure

Participants each rated 30 randomly chosen concepts from the 200 used in the study. For each concept, participants were asked to list the first three words that came to mind when reading the name of the concept, ranked such that the most prevalent (or first thought of) was listed first. Participants could also mark a word as unknown or provide less than three answers. There was no time limit to complete the task.

Calculation of Mean BAS and Mean FAS

Associations for a concept were manually sorted and counted to take account of different spellings, spelling mistakes, and capitalizations. Following Nelson et al. (2004), conjugated words and plural/singular words were transformed into the most frequent instance. Following De Deyne and Storms (2008), idiosyncratic words (associations only produced once) were removed in the calculation of frequencies. Then, the associations were tallied for each concept and divided by the number of associations produced for that concept to create relative frequencies of each association.

A useful metric is how much on average a specific concept is associated with the other concepts used in the study (De Deyne & Storms, 2008). This is referred to as mean backward associative strength (MBAS) and mean forward associative strength (MFAS). To calculate these, we followed procedures used by De Deyne and Storms (2008), and Montefinese, Zannino, and Ambrosini (2015). Association data were entered in a 200 x 200 matrix in which rows corresponded to concepts and columns to their generated associations with other concepts. We entered each association frequency in the corresponding cell. The MFAS was then calculated by averaging over frequencies in a given concept’s row, and MBAS by averaging over a given concept’s column.

Word frequency

Frequently encountered words are processed faster than words that are rarely encountered (Oldfield & Wingfield, 1964). Low-frequency words are correctly recognised and correctly rejected more often than high-frequency words (Glanzer & Adams, 1985). We used log-transformed word frequencies from SUBTTLEX-UK, based on British English television subtitles (van Heuven et al., 2014).

Concreteness

Concreteness refers to the extent to which the concept denoted by a word refers to a perceptible entity (Paivio & Begg, 1971). More concrete words are easier to remember than more abstract words (Gorman, 1961; Paivio, 2013). We used norms from Brysbaert et al.

(2014). Participants rated how concrete the meaning of each word was by using a 5-point scale.

Age of Acquisition

Age of acquisition (AoA) is known to influence memory: words learned early in life are less well recognized (Dewhurst et al., 1998). We used Brysbaert & Biemille's (2017) update of Dale and O'Rourke's (1981) norms.

Phonological Neighborhood Density

The phonological neighbourhood density (PND) of a word is the number of words in the language which differ only by the addition, subtraction or substitution of a single phoneme from the target word (Luce & Pisoni, 1998). Items with high PND tend to be processed less quickly and/or accurately. We used norms from the CELEX database (Baayen et al., 1995).

Visual Complexity

The visual complexity measures reflect superficial visual characteristics of images. More complex stimuli may be more easily recognized (Snodgrass & Vanderwart, 1980). We included two different measures of image complexity: the number of non-white pixels in the image, and color entropy, a measure of the color variability of an image. Images with a large proportion of pixels sharing the same color should be less visually complex (Chouinard & Goodale, 2012). Color entropy was computed by finding the relative frequency of all colors that occur in the non-white pixels in the image and calculating the entropy of this probability distribution.

Concept Familiarity

We included concept familiarity as an attribute of images. In recognition memory, more unfamiliar pictures are better recognised than familiar pictures (Snodgrass & Vanderwart, 1980).

Each covariate we considered in the study was standardized and the statistics that follow were calculated over the new set of z-score values.

Experiment 1

Figure 4

Schematic Depiction of Additional Graded Perceptual Confusability Measures

Note. Rows show individual exemplars with the highest (right side) and lowest (left side) scores on each metric of perceptual confusability. The graded perceptual confusability measures define confusability indexing an image's overall similarity to the full set of images. C1 and C2 were obtained from gray-scaled version of the images depicted in Figure 4. For definitions see Variables of Interest section.

Table 5

Principal Component Analysis of Nuisance Variables in Experiment 1

	Components						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Nb of non-white pixel					0.997		
Color entropy	0.154		0.701			-0.272	0.264
Mean BAS	-0.536						-0.334
Mean FAS	-0.811						0.272
PND							-0.822
Concept familiarity	-0.156		0.710			0.258	-0.243
Word co-occurrence				0.992			
Concreteness		-0.700					

Age of acquisition						0.921	
Word Frequency		-0.708					

Note. The top 7 principal components from the PCA with varimax rotation of nuisance variables. Cumulative variance explained is 86.14 %.

Table 6

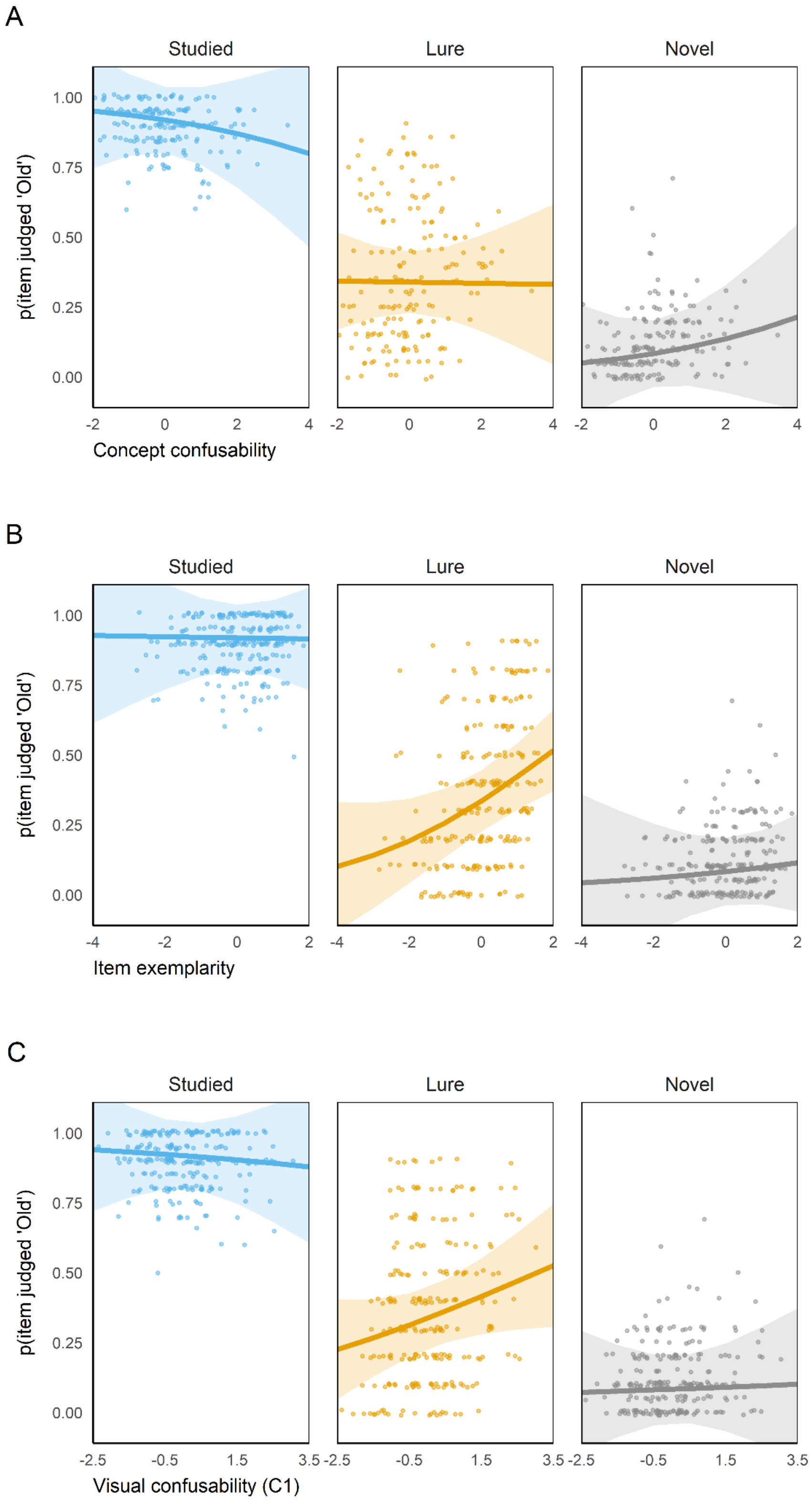
Model Selection for Experiment 1

Model	AICc	ΔAICc	AICcWt	Cum.Wt	LL
Concept + Item + PCs	8821.92	0.00	0.98	0.98	-4380.87
Concept + Item	8830.05	8.13	0.02	1.00	-4391.97
Item + PCs	8876.02	54.10	0.00	1.00	-4413.95
Item	8886.94	65.02	0.00	1.00	-4426.44
Concept + PCs	8899.91	77.98	0.00	1.00	-4431.92
Concept	8911.89	89.96	0.00	1.00	-4444.93
PCs	14420.23	5598.30	0.00	1.00	-7200.10
Null	14430.75	5608.83	0.00	1.00	-7212.38

Note. Summary of AICc results for models including concept-level, item-level, and confounds principal component variables; PCs = Principal components.

Figure 5

Effects of Semantic and Perceptual Variables on Raw Recognition Measures in Experiment 1



Note. Plots show effects of semantic and perceptual variables on raw recognition responses by item type. The plot lines represent the raw probabilities of endorsing studied items as “old” (light blue), lures as “old” (orange), and novel items as “old” (grey). Panel A, B, and C show the effects of concept confusability, item exemplarity, and C1 visual confusability in Experiment 1. Note that for concept-level variables (concept confusability) there are data points for each concept, and for item-level variables (item exemplarity and C1 visual confusability) there are data points for each exemplar image. The clustering around discrete values of p("old") reflects the small numbers of observations for individual exemplars (see Experiment 1, Materials and Methods, Stimuli).

Table 7

Results of Experiment 1 Using Graded Perceptual Confusability Metrics

Variable	Estimate	*d'*	SE	z-value	*p*
(Intercept)	-2.42	-1.38	0.12	-20.09	<.001
Lure	1.73	0.96	0.07	23.37	<.001
Studied	4.86	2.75	0.09	52.24	<.001
Number of Features	0.15	0.08	0.08	1.94	.084
Concept Confusability	0.29	0.16	0.08	3.61	<.001
Visual Confusability (C1)	-0.14	-0.08	0.07	-1.89	.084
Visual Confusability (C2)	0.14	0.06	0.08	1.67	.117
Color Confusability	0.13	0.07	0.08	1.69	.117
Item Exemplarity	0.18	0.10	0.07	2.61	.022
Lure × Number of Features	-0.32	-0.18	0.07	-4.43	<.001
Studied × Number of Features	-0.12	-0.06	0.09	-1.40	.190
Lure × Concept Confusability	-0.27	-0.14	0.07	-3.74	<.001
Studied × Concept Confusability	-0.57	-0.30	0.08	-6.73	<.001
Lure × Visual Confusability (C1)	0.14	0.08	0.08	1.88	.084
Studied × Visual Confusability (C1)	0.20	0.11	0.09	2.27	.041
Lure × Visual Confusability (C2)	-0.08	-0.03	0.08	-0.92	.355
Studied × Visual Confusability (C2)	-0.11	-0.05	0.10	-1.11	.297
Lure × Color Confusability	-0.08	-0.04	0.07	-1.05	.307
Studied × Color Confusability	-0.26	-0.14	0.09	-3.00	.007
Lure × Item Exemplarity	0.19	0.12	0.08	2.44	.028
Studied × Item Exemplarity	-0.23	-0.11	0.09	-2.52	.025

Note. The reference level of condition is set to "novel". Parameter estimates (logOR), *d'* equivalent, standard errors, z-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Graded perceptual confusability measures were reported in the model above. See Material and Methods, and Variables of Interest for details. SE = Standard Error.

Experiment 2

Table 8

Principal Component Analysis of Nuisance Variables in Experiment 2

	Components						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Nb of non-white pixel	-0.203		-0.675	-0.116		-0.229	0.112
Color entropy					-0.989		
Mean BAS	0.602						-0.238
Mean FAS	0.744						0.257
PND							-0.924
Familiarity	0.198		-0.733	0.108		-0.207	
Word co-occurrence				0.992			
Concreteness		0.706					
Age of acquisition						-0.947	
Word frequency		0.707					

Note. The top 7 principal components from the PCA with varimax rotation of nuisance variables. Cumulative variance explained is 86.54 %.

Table 9

Model Selection for Experiment 2

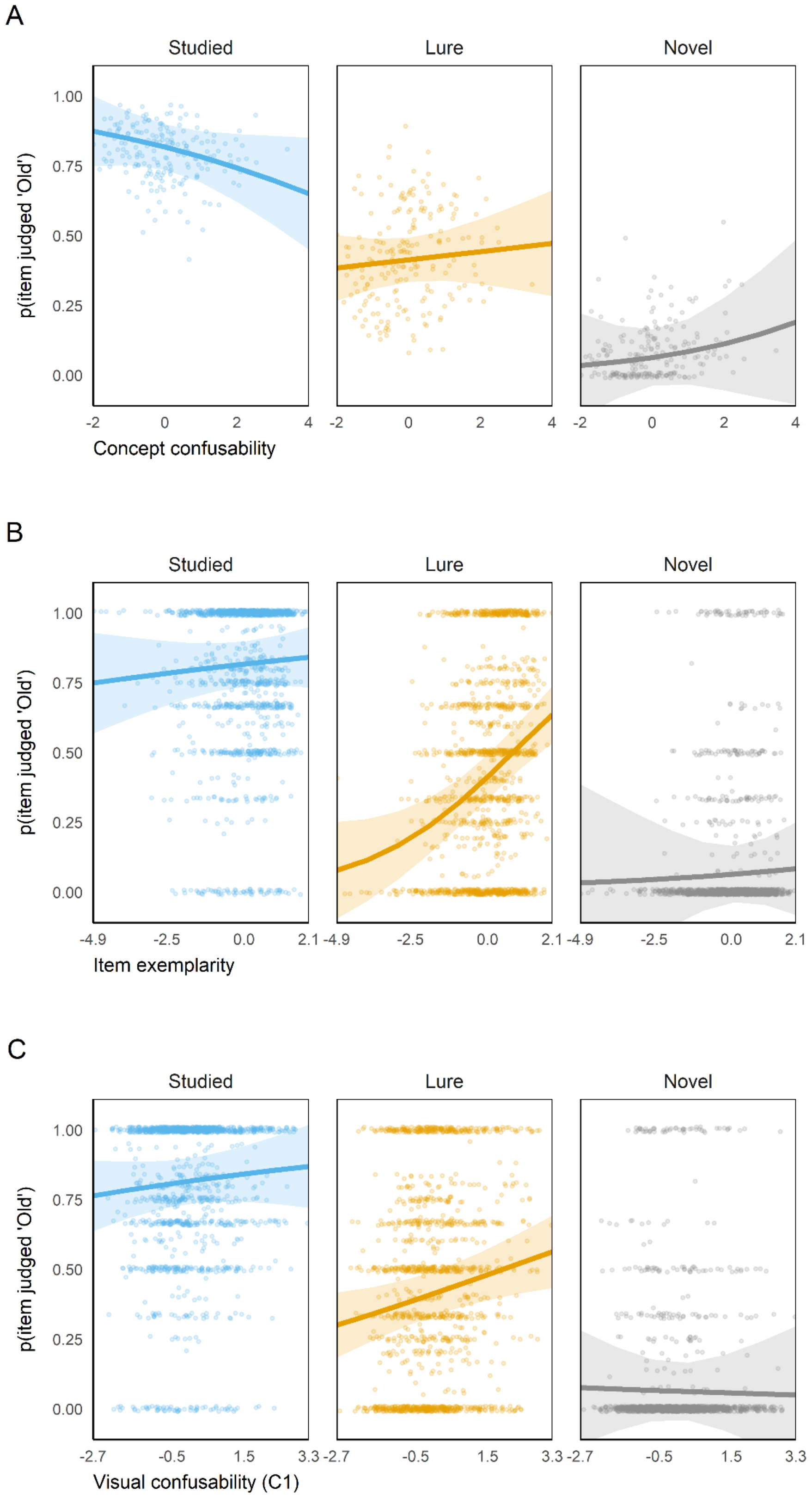
Model	AICc	ΔAICc	AICcWt	Cum.Wt	LL
Concept + Item + PCs	17597.96	0.00	0.89	0.89	-8768.93
Concept + Item	17602.07	4.12	0.11	1.00	-8778.00
Item + PCs	17660.69	62.73	0.00	1.00	-8806.31
Item	17665.99	68.03	0.00	1.00	-8815.98
Concept + PCs	17854.98	257.02	0.00	1.00	-8909.47
Concept	17862.95	264.99	0.00	1.00	-8920.47
PCs	23836.19	6238.23	0.00	1.00	-11908.09

Null	23856.48	6258.53	0.00	1.00	-11925.24

Note. Summary of AICc results for models including concept-level, item-level, and confounds principal component variables; PCs = Principal components.

Figure 6

Effects of Semantic and Perceptual Variables on Raw Recognition Measures in Experiment 2



Note. Plots show effects of semantic and perceptual variables on raw recognition responses by item type. The plot lines represent the raw probabilities of endorsing studied items as “old” (light blue), lures as “old” (orange), and novel items as “old” (grey). Panel A, B, and C show the effects of concept confusability, item exemplarity, and C1 visual confusability in Experiment 2. Note that for concept-level variables (concept confusability) there are data points for each concept, and for item-level variables (item exemplarity and C1 visual confusability) there are data points for each exemplar image. The clustering around discrete values of p(“old”) reflects the small numbers of observations for individual exemplars, as 300 of 1800 images were randomly allocated to the 3 test conditions for each participant (see Experiment 2, Materials and Methods, Stimuli).

Table 10

Results Including Study Set Size Effects for Experiment 2

Variable	Estimate	*d’*	SE	z-value	*p*
(Intercept)	-0.86	-0.51	0.09	-9.94	<.001
Novel	-1.85	-1.03	0.08	-23.48	<.001
Lure-Set8	0.98	0.59	0.05	17.90	<.001
Studied-Set2	2.18	1.30	0.06	37.26	<.001
Studied-Set8	2.57	1.53	0.06	40.54	<.001
Number of Features	-0.02	-0.01	0.05	-0.30	.890
Concept Confusability	0.09	0.06	0.05	1.71	.162
Visual Confusability (C1)	0.13	0.07	0.04	2.94	.010
Visual Confusability (C2)	0.06	0.03	0.05	1.31	.304
Color Confusability	0.08	0.05	0.04	1.81	.135
Item Exemplarity	0.43	0.25	0.05	9.52	<.001
Novel × Number of Features	0.01	0.00	0.08	0.11	.941
Lure-Set8 × Number of Features	-0.01	0.00	0.06	-0.11	.941
Studied-Set2 × Number of Features	-0.03	-0.02	0.06	-0.45	.813
Studied-Set8 × Number of Features	-0.16	-0.09	0.07	-2.28	.050
Novel × Concept Confusability	0.21	0.10	0.08	2.81	.013
Lure-Set8 × Concept Confusability	-0.08	-0.05	0.06	-1.37	.286
Studied-Set2 × Concept Confusability	-0.31	-0.19	0.06	-5.28	<.001
Studied-Set8 × Concept Confusability	-0.33	-0.19	0.06	-5.12	<.001
Novel × Visual Confusability (C1)	-0.20	-0.10	0.08	-2.54	.026
Lure-Set8 × Visual Confusability (C1)	0.14	0.09	0.06	2.53	.026
Studied-Set2 × Visual Confusability (C1)	0.02	0.01	0.06	0.27	.890

Studied-Set8 × Visual Confusability (C1)	-0.03	-0.01	0.06	-0.42	.813
Novel × Visual Confusability (C2)	0.16	0.07	0.09	1.88	.123
Lure-Set8 × Visual Confusability (C2)	0.01	0.01	0.06	0.16	.941
Studied-Set2 × Visual Confusability (C2)	-0.05	-0.03	0.06	-0.92	.503
Studied-Set8 × Visual Confusability (C2)	0.05	0.03	0.06	0.84	.539
Novel × Color Confusability	0.00	-0.01	0.08	-0.01	.988
Lure-Set8 × Color Confusability	0.06	0.03	0.06	1.03	.440
Studied-Set2 × Color Confusability	-0.06	-0.04	0.06	-1.09	.255
Studied-Set8 × Color Confusability	-0.09	-0.05	0.06	-1.46	.417
Novel × Item Exemplarity	-0.28	-0.17	0.08	-3.54	.001
Lure-Set8 × Item Exemplarity	0.04	0.03	0.06	0.63	.689
Studied-Set2 × Item Exemplarity	-0.38	-0.22	0.06	-6.22	<.001
Studied-Set8 × Item Exemplarity	-0.30	-0.18	0.06	-4.83	<.001

Note. The reference level of condition is set to “Lure-Set2”. Parameter estimates (logOR), *d’* equivalent, standard errors, z-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Nearest neighbour perceptual confusability measures were reported in the model above. See Material and Methods, and Variables of Interest and for details. SE = Standard Error.

Table 11

Results of Experiment 2 Using Graded Perceptual Confusability Metrics

Variable	Estimate	*d’*	SE	z-value	*p*
(Intercept)	-2.71	-1.55	0.10	-26.66	<.001
Lure	2.36	1.34	0.07	31.98	<.001
Studied	4.20	2.44	0.08	54.11	<.001
Number of Features	-0.01	-0.01	0.07	-0.16	.917
Concept Confusability	0.29	0.15	0.07	4.05	<.001
Visual Confusability (C1)	0.06	0.03	0.07	0.80	.589
Visual Confusability (C2)	0.30	0.14	0.08	3.96	<.001
Color Confusability	0.02	0.00	0.07	0.30	.896
Item Exemplarity	0.17	0.09	0.07	2.41	.030
Lure × Number of Features	0.00	0.00	0.07	-0.04	.969
Studied × Number of Features	-0.08	-0.05	0.07	-1.10	.405
Lure × Concept Confusability	-0.22	-0.10	0.07	-3.09	.005
Studied × Concept Confusability	-0.51	-0.27	0.07	-6.97	<.001

Lure × Visual Confusability (C1)	-0.05	-0.03	0.07	-0.75	.597
Studied × Visual Confusability (C1)	-0.02	0.00	0.08	-0.23	.908
Lure × Visual Confusability (C2)	-0.25	-0.11	0.08	-3.12	.005
Studied × Visual Confusability (C2)	-0.23	-0.10	0.08	-2.84	.009
Lure × Color Confusability	-0.04	-0.01	0.07	-0.57	.699
Studied × Color Confusability	-0.11	-0.05	0.07	-1.54	.218
Lure × Item Exemplarity	0.27	0.18	0.07	3.68	<.001
Studied × Item Exemplarity	-0.09	-0.04	0.08	-1.17	.390

Note. The reference level of condition is set to “novel”. Parameter estimates (logOR), *d’* equivalent, standard errors, z-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Graded perceptual confusability measures were reported in the model above. See Material and Methods, and Variables of Interest for details. SE = Standard Error.

Experiment 1 and Experiment 2

Table 12

Results for Between-Experiment comparison

Variable	Estimate	*d’*	SE	z-value	*p*
(Intercept)	-2.31	-1.33	0.10	-22.25	<.001
Experiment 2	-0.39	-0.21	0.14	-2.71	.023
Experiment 2 × Lure	0.71	0.40	0.10	6.93	<.001
Experiment 2 × Studied	-0.45	-0.24	0.11	-3.96	<.001
Experiment 2 × Number of Features	-0.12	-0.07	0.09	-1.35	.322
Experiment 2 × Concept Confusability	0.05	0.02	0.09	0.59	.725
Experiment 2 × Visual Confusability (C1)	-0.15	-0.07	0.09	-1.74	.174
Experiment 2 × Visual Confusability (C2)	0.08	0.03	0.11	0.79	.604
Experiment 2 × Color Confusability	0.05	0.01	0.10	0.50	.766
Experiment 2 × Item Exemplarity	-0.06	-0.02	0.09	-0.65	.700
Experiment 2 × Lure × Number of Features	0.29	0.17	0.10	2.97	.011
Experiment 2 × Studied × Number of Features	0.02	0.00	0.11	0.14	.907
Experiment 2 × Lure × Concept Confusability	0.02	0.02	0.10	0.25	.872
Experiment 2 × Studied × Concept Confusability	-0.02	-0.01	0.11	-0.21	.872
Experiment 2 × Lure × Visual Confusability (C1)	0.09	0.04	0.10	0.91	.522
Experiment 2 × Studied × Visual Confusability (C1)	0.37	0.20	0.11	3.42	.003
Experiment 2 × Lure × Visual Confusability (C2)	-0.03	0.00	0.12	-0.28	.872

Experiment 2 × Studied × Visual Confusability (C2)	-0.06	-0.01	0.13	-0.47	.766
Experiment 2 × Lure × Color Confusability	0.05	0.04	0.11	0.49	.766
Experiment 2 × Studied × Color Confusability	0.03	0.03	0.12	0.24	.872
Experiment 2 × Lure × Item Exemplarity	0.11	0.06	0.10	1.05	.444
Experiment 2 × Studied × Item Exemplarity	0.13	0.07	0.11	1.19	.408

Note. The reference levels of condition and experiment are set to “novel” and “Experiment 1”. Parameter estimates (logOR), *d’* equivalent, standard errors, z-values, and FDR-corrected *p*-values are listed for condition, concept-level, and item-level variables in the winning (full) linear mixed model selected with AIC. Nearest neighbor perceptual confusability measures were used in the model above. See Material and Methods, and Variables of Interest for details. SE = Standard Error.

Supplemental material of Chapter 4

Experiment 1

Table 13

Age-Invariant Effects in Experiment 1

Variable	Estimate	*d’*	SE	z-value	*p*
- Novel Items as Baseline					
(Intercept)	-2.48	-1.41	0.09	-28.82	<.001
Lure	1.76	0.98	0.05	33.65	<.001
Studied	4.88	2.69	0.06	75.11	<.001
Concept Confusability	0.21	0.11	0.06	3.28	.001
Visual Confusability	0.13	0.07	0.05	2.55	.013
Item Exemplarity	0.11	0.05	0.05	2.05	.044
Lure × Concept Confusability	-0.16	-0.08	0.05	-3.39	.001
Studied × Concept Confusability	-0.45	-0.23	0.06	-7.99	<.001
Lure × Visual Confusability	0.02	0.03	0.05	0.41	.685
Studied × Visual Confusability	-0.20	-0.10	0.06	-3.49	<.001
Lure × Item Exemplarity	0.25	0.16	0.06	4.53	<.001
Studied × Item Exemplarity	-0.26	-0.12	0.07	-3.98	<.001

• Lure Items as Baseline					
(Intercept)	-0.72	-0.43	0.08	-9.21	<.001
Studied	3.12	1.78	0.05	60.14	<.001
Studied × Concept Confusability	-0.29	-0.15	0.05	-6.06	<.001
Studied × Visual Confusability	-0.22	0.14	0.05	-4.43	<.001
Studied × Item Exemplarity	-0.51	-0.28	0.06	-9.00	<.001

Note. In the upper part of the table, the reference level of condition is set to “novel items”. In the lower part of the table, the reference level of condition is set to “lure items”. Parameter estimates (logOR), *d’* equivalent, standard errors (SE), z-values, and false discovery rate (FDR) corrected *p*-values are listed for condition, concept-level, and item-level variables in the linear mixed model.

Experiment 2

Table 14

Age-Invariant Effects in Experiment 2

Variable	Estimate	*d’*	SE	z-value	*p*
• Novel Items as Baseline					
(Intercept)	-2.75	-1.57	0.08	-35.00	<.001
Lure	2.71	1.55	0.05	52.57	<.001
Studied	4.48	2.59	0.05	81.57	<.001
Concept Confusability	0.22	0.11	0.06	3.93	<.001
Visual Confusability	0.02	0.01	0.05	0.35	.796
Item Exemplarity	0.13	0.07	0.05	2.65	.011
Lure × Concept Confusability	-0.16	-0.07	0.05	-3.30	.001
Studied × Concept Confusability	-0.40	-0.21	0.05	-7.81	<.001
Lure × Visual Confusability	0.19	0.11	0.05	3.64	<.001
Studied × Visual Confusability	0.14	0.08	0.05	2.56	.013
Lure × Item Exemplarity	0.35	0.22	0.05	6.81	<.001
Studied × Item Exemplarity	0.00	0.00	0.05	-0.04	<.969
• Lure Items as Baseline					
(Intercept)	-0.05	-0.03	0.06	-0.72	.469
Studied	1.78	1.05	0.03	58.56	<.001
Studied × Concept Confusability	-0.24	-0.14	0.03	-8.02	<.001

Studied × Visual Confusability	-0.05	-0.03	0.03	-1.61	.128
Studied × Item Exemplarity	-0.36	-0.22	0.03	-11.80	<.001

Note. In the upper part of the table, the reference level of condition is set to “novel items”. In the lower part of the table, the reference level of condition is set to “lure items”. Parameter estimates (logOR), *d’* equivalent, standard errors (SE), z-values, and false discovery rate (FDR) corrected *p*-values are listed for condition, concept-level, and item-level variables in the linear mixed model.