THE EFFECT OF DISTINCTIVENESS ON VISUAL FALSE MEMORIES

by
İREM KÜSMÜŞ

Submitted to the Graduate School of Social Sciences
in partial fulfilment of
the requirements for the degree of Master of Science

Sabancı University
July 2024

THE EFFECT OF DISTINCTIVENESS ON VISUAL FALSE MEMORIES

Approved by:

Asst. Prof. Olesya Blazhenkova . . . .
(Thesis Supervisor)

Prof. Ali İzzet Tekcan . . . .

Asst. Prof. Eren Günseli . . . .

Date of Approval: July 24, 2024

ABSTRACT

THE EFFECT OF DISTINCTIVENESS ON VISUAL FALSE MEMORIES

İREM KÜSMÜŞ

PSYCHOLOGY M.S. THESIS, JULY 2024

Thesis Supervisor: Asst. Prof. OLESYA BLAZHENKOVA

Keywords: visual false memory, recognition memory, distinctiveness, category associates procedure

The current research aimed to investigate false memories by implementing the Category Associates Procedure using a newly created set of pictorial stimuli and to test the effect of distinctiveness on false memories. In Study 1, we examined conceptual distinctiveness by manipulating the backgrounds of studied objects in terms of meaning congruency. In Study 2, we examined perceptual distinctiveness by manipulating the backgrounds of studied objects in terms of color heterogeneity. We hypothesized that the false memory effect would be present, with participants recognizing critical lures more than unrelated items. Furthermore, we hypothesized that distinctiveness would lead to lower rates of false memories. We found a false memory effect for objects (Studies 1 and 2) as well as their backgrounds (Study 1), consistent with previous studies. However, the predicted effect of distinctiveness was not found. In Study 1, we observed no difference between distinctive and non-distinctive conditions in terms of recognition rates of objects. Additionally, we found the opposite of the predicted effect in terms of recognition rates for backgrounds. This effect was further observed when analyzing the subset of data with the most and least congruent lists. For Study 1, we also observed lower memory encoding for the incongruent condition, indicated by signal detection index. In Study 2, we observed no difference between distinctive and non-distinctive conditions neither in recognition rates, RTs and confidence ratings nor in signal detection indices. Our findings suggest that conceptual but not perceptual distinctiveness has a greater effect on memory encoding. However, conceptual incongruity seems to play more disruptive rather than a supporting role for memory. The results are discussed in relation to encoding-based and retrieval-based theories. Additionally, we found some associations between recognition performance and individual differences in imagery measures as well as self-reported strategies.

ÖZET

AYIRT EDİCİLİĞİN GÖRSEL BELLEK YANILGILARI ÜZERİNDEKİ ETKİSİ

İREM KÜSMÜŞ

PSİKOLOJİ YÜKSEK LİSANS TEZİ, TEMMUZ 2024

Tez Danışmanı: Dr. Öğr. Üyesi OLESYA BLAZHENKOVA

Anahtar Kelimeler: görsel bellek yanılgısı, tanıma belleği, ayırt edicilik, kategori içi çağrışımlar prosedürü

Bu çalışmanın amacı, yeni oluşturulan bir görsel uyaran seti ile Kategori İçi Çağrışımlar Prosedürünü uygulayarak bellek yanılgılarını araştırmak ve ayırt ediciliğin bellek yanılgıları üzerindeki etkisini test etmektir. Deney 1'de, çalışılan nesnelerin arka planları anlam uyumu açısından manipüle edilerek kavramsal ayırt edicilik test edildi. Deney 2'de, çalışılan nesnelerin arka planları renk heterojenliği açısından manipüle edilerek algısal ayırt edicilik test edildi. İki deneyde de, bellek yanılgısı etkisinin mevcut olacağı ve ayırt edici koşulların daha düşük bellek yanılgısına yol açacağı öne sürüldü. Önceki çalışmalarla tutarlı olarak, nesnelerin (Deney 1 ve 2) yanı sıra arka planlar (Deney 1) için de bellek yanılgısı etkisi gözlemlendi. Ancak, ayırt ediciliğin öngörülen etkisi gözlemlenmedi. Deney 1'de, nesnelerin tanınma oranları açısından ayırt edici ve ayırt edici olmayan koşullar arasında bir fark gözlemlenmedi. Buna ek olarak, arka planların tanınma oranları açısından öngörülen etkinin tam tersi bulundu. Bu etki, en çok ve en az ayırt ediciliğe sahip veri alt kümesi analiz edildiğinde de gözlemlendi. Ek olarak, Deney 1 için sinyal tespit kuramı ile yapılan analizlerde ayırt edici koşul için daha düşük bellek kodlaması bulundu. Çalışma 2'de, ayırt edici ve ayırt edici olmayan koşullar arasında tanıma belleği oranları, reaksiyon zamanı, eminlik yargıları ve sinyal tespit endeksleri açısından bir fark gözlemlenmedi. Bulgularımız, algısal ayırt ediciliğin değil kavramsal ayırt ediciliğin bellek kodlaması üzerinde daha büyük bir etkiye sahip olduğunu göstermektedir. Bununla birlikte, kavramsal uyumsuzluğun bellek için destekleyici bir rolden ziyade daha yıkıcı bir rol oynadığı öne sürülebilir. Sonuçlar, kodlama temelli ve geri getirme temelli teorilerle ilişkili olarak tartışılmıştır. Ayrıca, tanıma belleği performansı ile imgeleme ölçümlerindeki bireysel farklılıklar ve bildirilen stratejiler arasında bazı ilişkiler bulunmuştur.

ACKNOWLEDGEMENTS

I would like to express my gratitude for my support team in Sabancı University; Büşra, Pelin, Selen, İrem, Duygu and Selma, who all made the campus a better place. Special thanks to Büşra and Pelin for their constant support. I would like to thank Belgin Deryalar, who always motivated me, and showed sincere support and understanding through this process. Also, I want to convey my gratitude to my dearest friend Deniz. Despite living far away from each other now, it still feels like we are roommates, and I know that you will be there when I need you no matter what. Thank you for being such a good listener and supporting me in the last two years. I am lucky enough to have many good friends that I cannot fit all their names here. My friends in Ankara, thank you for all the support, and for all the fun times.

I would like to thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) for funding this research through the 2210/A General Domestic Graduate Scholarship Program. I would like to thank my thesis supervisor Asst. Prof. Olesya Blazhenkova for her guidance and understanding through this process. I learned a lot of new things in the past two years, which I am grateful for. Also, I would like to thank Prof. Ali İ. Tekcan and Asst. Prof. Eren Günseli for accepting to be my thesis jury members.

The biggest thank you goes to my partner Eyüphan Şekeroğlu. I am deeply grateful for everything we share together. Looking back now, I know that this would not be possible without you. Since day one, you have been my biggest source of support, love and happiness. Thank you for believing in me more than I believe in myself and motivating me in days when I feel like giving up. It gives me comfort to know that I'll get to have your support all my life, and I am so excited for our future together.

Lastly, I am deeply grateful for my family members, Hülya Küsmüş, Yusuf Küsmüş and Utku Erdem Küsmüş, and for their love and support. I feel lucky to have such a great family like you. Special thanks to my parents, who always put their children first, and gave importance to our education. This accomplishment would not be possible without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 False Memory

Memory is an essential part of cognition that enables individuals to learn, communicate, and perform daily tasks by retrieving and reorganizing information related to a past experience (Goldstein 2014). We rely on our memory while talking to a friend, grocery shopping or telling a story from our past. Despite holding an indispensable part in our daily lives, our memories are prone to errors. Memory is not a perfect recording of past experiences, rather it is a constructive process which may lead to distortion and error (Schacter 2012). These errors in the process of remembering are often referred to as false memories, which is a concept that has been widely investigated in memory literature. A false memory occurs when an individual claims to recall an event that never happened or recalls it in a distorted way (Roediger and McDermott 1995).

It is essential to study the failures of memory for a better understanding of its structure, since the very nature of memory is also the reason why it sometimes fails (Misirlisoy 2004; Roediger 1996). Therefore, this question remains paramount: Why might false memories occur? In the literature, a vast array of studies has investigated this question, and it was suggested that false memories may occur as a result of factors such as suggestion (Loftus 1997), misleading information or questions (Foster et al. 2012; Roebers and Schneider 2000), and schemas and semantic networks (Gallo 2013; Marsh, Eslick, and Fazio 2008). Additionally, false memories were linked to individual differences in imagery, that is, vivid imagery may lead to a higher rate of false memories (Gonsalves et al. 2004).

1.1.1 Methods Examining False Memories

In false memory research, the mostly used procedures to investigate the formation of false memories are the DRM paradigm (Deese 1959; Roediger and McDermott 1995) and Category Associates Procedure (CAP; Seamon et al. 2000; Verma and Kashyap 2020). In a typical DRM paradigm, there are two phases: the study phase where items are encoded, and the test phase where memory for items is tested. At the study phase, individuals are presented with semantically related words that belong to a particular thematic list. Each thematic list has a "critical lure" item that is not presented during the study phase. This critical lure item is the highest semantic associate of the related thematic list, which is the word that has the most semantic connection to the items of that list (Roediger and McDermott 1995). While creating DRM lists, this semantic relation between the list items and the critical lure item is measured with the Backward Associative Strength (BAS), which is basically the degree to which the list items are semantically associated with the critical lure item (Roediger and Gallo 2017). An example list from the DRM paradigm includes items such as "dream, rest, tired, bed", and the critical lure item of this list is "sleep", which has the highest semantic connectedness to the list items (Roediger and McDermott 1995). After studying all lists and an interval; in the test phase, participants are asked to recognize (respond with "yes" if they think they have seen that item) studied words as well as non-studied "critical lures", and unrelated non-studied items also known as unrelated lures (random objects; Deese 1959; Roediger and McDermott 1995). In the DRM paradigm, the term "false memory" is used for falsely recognizing critical lure items during the test. It is commonly observed that false recognition rates of critical lures may exceed even 50% , usually ranging between 25% and 60%, and in some cases false alarm rates are observed to be at similar rates with recognition of studied items (Huff and Bodner 2013; Lampinen, Neuschatz, and Payne 1999). At the same time, participants recognize critical lure items more often than unrelated lures (Deese 1959; Roediger and McDermott 1995). In the literature, the DRM paradigm was tested in various conditions by changing the encoding task, the retention interval or the presentation duration; as well as across different participants in varying age, cognitive abilities and cultures (Gallo 2010). Overall, studies using the DRM paradigm consistently demonstrated false memory effects (Huff and Bodner 2013; Pardilla-Delgado and Payne 2017; Sugrue and Hayne 2006). However, the presentation modality was found to have an effect on false memory rates (Beauchamp 2002; Smith and Hunt 1998). In their study, Smith and Hunt (1998) found that the false memory rate approaching the hit rate is observed only when the items are presented audibly, but not when they are presented visually. The authors concluded the visual modality provides a basis to discriminate studied items from critical lure items, leading to a larger difference between false alarm rates and hit rates.

Another method to investigate false memories, highly similar to the DRM paradigm, is called Category Associates Procedure (Seamon et al. 2000; Verma and Kashyap 2020). Despite similarities, while in the DRM paradigm thematic lists that consist of semantically related items from different themes are used, and they may be represented by different parts of speech; in the CAP categorical lists are used, and they are typically represented by nouns. A categorical list includes items that are exemplars of an upper category title. For example; two example items from a list can be banana and strawberry, where the category they belong to is "fruits". In this case, the item that first comes to mind given the category title (the most representative exemplar of each category) will be the "critical lure" of its list. For instance, the critical lure item of the fruits category is apple, while the critical lure item of the clothes category is shirt (Misirlisoy 2004; Van Overschelde, Rawson, and Dunlosky 2004). As with the DRM paradigm, it is possible to observe a false memory effect using CAP. First, it was demonstrated by Seamon et al. (2000), and then it was replicated using different manipulations such as presenting items in different modalities (Beauchamp 2002) or testing different age groups (Intons-Peterson et al. 1999).

When comparing CAP and DRM paradigms, the former seems more suitable for studying visual false memory since most categorical lists include items that are easier to objectively present as pictures, while thematic lists from the DRM paradigm may include abstract items, adjectives and verbs. Therefore, CAP seems the most appropriate method to investigate false memories by implementing a visual version of the procedure. In the literature, a category associates procedure was sometimes implemented to study false memory using visual stimuli and yield comparable effects to visual DRM (Seamon et al. 2000; Verma and Kashyap 2020; Wang et al. 2018). In these studies, the main approach was to use the black and white line drawings that were created by Battig and Montague (1969). However, it has been stated that black and white line drawings may not suffice to represent the items since the match between the drawing and item might be low (Smith and Hunt 2020). Moreover, often the verbal labels are also presented together with the drawings or read out loud (Gallo et al. 2007; Howe 2008; Israel and Schacter 1997; Smith and Hunt 2020), which may affect the false memory rates (Smith and Hunt 1998).

1.2 Theories Explaining the Formation of False Memories

1.2.1 Activation Monitoring Theory

One of the theories that brings an explanation to the occurrence of false memories is the Activation Monitoring Theory (Roediger et al. 2001). This theory suggests that when encoded items are from the same thematic or categorical list, a semantic network is activated in memory, and this leads to the activation of non-presented but semantically/thematically related words. This spread of activation during the encoding then leads to the confusion that the unpresented critical words were presented and interferes with the monitoring process during the retrieval, where individuals decide whether they have seen an item before or not. In the case of CAP, for example, while the participant perceives different types of flowers (either presented as picture, as word, or audibly), a semantic network of flowers is activated in memory, and this network reaches to the most representative item of that category as well. In the case of the flower category, this item is a rose (Misirlisoy 2004). Therefore, in the test phase, when participants are asked whether they have seen the object rose in the study phase, they rely on the information of the activation of rose during encoding and this information is mistaken for actually seeing a rose. According to this theory, the associative strength between the list items and the critical lure affects false memory rates (Coane, Huff, and Hutchison 2016). In short, Activation Monitoring Theory necessitates the activation of the critical lure item during encoding, which will then be erroneously interpreted during the monitoring process when retrieving (Gallo and Roediger 2002; Misirlisoy 2004).

1.2.2 Fuzzy Trace Theory

Fuzzy Trace Theory (Reyna and Brainerd 1995) is another popular theory in the explanation of false memories, according to which it is not necessary to have the activation of the critical lure item during the encoding. This theory argues that during the encoding, verbatim traces and gist traces are captured. Verbatim traces are related to the features and sensory details of the studied items, while gist traces are related to the overall meaning associated with the studied items. Reyna and Brainerd (1995) suggested that, when studying items from a related list, a strong gist memory which is related to the overall theme/category is created, more so than verbatim memory. Moreover, verbatim traces tend to become inaccessible faster, while gist traces are more persistent (Reyna and Brainerd 1995). As a result, because the gist of the presented items are captured during presentation, individuals rely on these gists during retrieval while rejecting or accepting the items. For example, when a participant is presented with daisy, clove and violet, the gist of these presentations is “flower”. Therefore, later during the test phase when rose is presented, because it matches the flower category, it is not rejected even though it was not presented in the study phase. This theory emphasizes that individuals do not explicitly recall the activation of rose, rather they rely on their gist - that they have seen flowers.

1.3 Distinctiveness and False Memories

False memories are frequently observed in a variety of settings, some of which may have serious consequences, such as during eyewitness testimonies (Kaplan et al. 2016; Loftus and Pickrell 1995; Roebers and Schneider 2000) or in therapeutic settings (Loftus 1997; Otgaar et al. 2022). Thus, as well as conducting research on the occurrence of false memories, it is essential to investigate how to reduce the formation of them. Research suggested that one of the ways to reduce false memories is to increase the distinctiveness (see Hunt and Worthen 2006, for a review). In the majority of the literature, an intuitive definition of the term distinctiveness is embraced; distinctiveness implies that stimuli that are salient, unusual or surprising captivates more attention, and this higher allocation of attention leads to enhanced memory performance (Hunt and Worthen 2006). However, the definition of distinctiveness has been inconsistent in the literature. Although the use of a more intuitive definition is common as mentioned above, a more outlined definition of distinctiveness was suggested by Schmidt (1991), by pointing out two types of distinctiveness: primary and secondary. Primary distinctiveness refers to the situations where the distinct item is standing out by being in contrast with the environment, therefore it can be also referred to as perceptual distinctiveness. For example, among white balls, a red ball will be distinct due to primary distinctiveness. Secondary distinctiveness occurs when one's expectations about real life become violated, by perceiving something that does not align with one's previous expectations, which can be considered as conceptual distinctiveness. For example, seeing a person behaving in an unusual way in the supermarket could be an example of secondary distinctiveness. In this example, that man's behavior is likely to be remembered more saliently than any other customer at the supermarket that day (Michelon and Snyder 2006).

In previous studies, various manipulations were used to tap distinctiveness. In their pioneering study, Israel and Schacter (1997) found that making the study lists more distinct by presenting words together with pictures results in reduction of false memories compared to a word-only condition. Subsequently, a range of studies have implemented various manipulations of distinctiveness such as presenting words in unique fonts (Arndt and Reder 2003), instructing the use of mental imagery during encoding of word lists (Oliver, Bays, and Zabrucky 2016; Robin 2010), or generating words from audio anagrams (McCabe and Smith 2006), and generally observed a decrease in false memory rates for more distinct stimuli. For example, Huff and Aschenbrenner (2018) led participants to engage in distinctive encoding by instructing them to think of a unique feature of each studied item from a categorized list which separates that item from others in that list (item-specific group); and this condition was compared with the group instructed to focus on the shared characteristics of items within a list (relational-encoding group), and a read-only group. The diminishing effect of distinctiveness on false memories was also replicated in this experiment, with the item-specific group showing lower false memory.

It should also be noted that there are inconsistencies in the literature about the effect of distinctiveness. As mentioned above, it is a widely accepted finding that pictorial representation (which is assumed to be more distinct than verbal) reduces false recognition when compared to verbal or auditory representation (Israel and Schacter 1997; Smith and Hunt 2020). However, there are also other findings where a decrease in false recognition was not observed when using pictorial stimuli in DRM paradigm (Howe 2008) and CAP (Israel and Schacter 1997; Koutstaal, Schacter, and Brenner 2001), especially when real life pictures of list items were used rather than line drawings (Wang et al. 2018). Moreover, it can be observed in the literature that in DRM paradigm experiments where the presentation modality was pictures, the researchers used the black and white line drawings generated by Israel and Schacter (1997) which also include the word labels underneath the drawings. Therefore, a further test of the use of real-life pictures without any labels seems necessary.

Considering these reviewed studies, it can be seen that primary distinctiveness manipulation – making the studied items more standing out during the encoding – was the most popular while less studies focused on secondary distinctiveness.

In addition to the operational definition of distinctiveness, the mechanism behind its effect on false memories has been an ongoing debate in the literature, with two main theories of focus: encoding-based and retrieval-based theories.

1.3.1 Encoding-Based Account Explaining Distinctiveness Effect on False Memory: The Impoverished Relational Encoding

On one side of the distinctiveness debate, the encoding-based account, argues that distinctiveness reduces false memories due to the processes that occur during encoding. This account can be understood in relation to both Activation Monitoring Theory (Roediger et al. 2001) and Fuzzy Trace Theory (Reyna and Brainerd 1995). As mentioned earlier, Activation Monitoring Theory argues that false memories occur due to the activation of the critical lure and the semantic meaning of the studied list during encoding. Later, this activation is interpreted as having seen the critical lure item. It is suggested that distinctiveness promotes item-specific processing at the expense of relational encoding (Arndt and Reder 2003; Einstein and Hunt 1980); that is, the spreading activation of the critical lure is prevented, which in turn reduces false memories. Similarly, in terms of Fuzzy Trace Theory, it was suggested that the gist representations encoded during the learning process lead to false memories. Distinctiveness is suggested to interfere with the relational encoding of the lists, and therefore the occurrence of gist representations. False memories are reduced since the participants cannot rely on strong gist representations. Based on the explanations of aforementioned theoretical frameworks, Hege and Dodson (2004) coined the term "impoverished relational encoding" (that refers to the disruption of relational encoding due to distinctiveness) to explain the mechanisms that occur during false memory formation, supporting the encoding-based side of the distinctiveness debate.

1.3.2 Retrieval-Based Account Explaining Distinctiveness Effect on False Memory: The Distinctiveness Heuristic

According to the other side of the debate – the retrieval-based account – it is suggested that the effect of distinctiveness occurs due to a strategy implemented during retrieval, which can be referred to as "distinctiveness heuristic" (Schacter and Wiseman 2006; Schacter, Israel, and Racine 1999). According to this memory monitoring strategy, after encoding the items in the distinctive condition, participants expect to have a more vivid and confident recollection of those items during retrieval. As a result, they embark on a more stringent criterion while deciding to recognize or reject an item. If an item does not satisfy their expectations, it is rejected. One important claim of the distinctiveness heuristics theory is that when a within-subject design is used; that is, participants study some items in the distinctive condition and some in the non-distinctive condition, this criterion is applied to all items regardless of condition (Schacter and Wiseman 2006). Since this criterion is implemented by rejecting all items that do not come with distinguishing information (such as visual details), the items that lack this information (all unpresented lure), will be rejected (Schacter and Wiseman 2006). In their study, Arndt and Reder (2003) supported this argument by testing the same encoding manipulation in between and within subjects. In the between subject design, when half of the participants read the word lists and the other half said the list items out loud, there was a decrease in false recognition for the second group, as saying words out loud during study brings distinctiveness. However, the authors were not able to observe a decrease in false recognition when a within-subject design was used, i.e., when participants read the word lists both silently and out loud. Again, it was shown that the participants relied on a distinctiveness heuristic since they showed overall decreased false recognition when compared to the non-distinctive condition from the between subject design study. The authors concluded that when items are encoded both distinctively (read out loud) and non-distinctively (read silently), distinctiveness heuristics is not used to distinguish said items from only read items and applied to items from both conditions, leading to no difference in false recognition rates for the distinctive and non-distinctive conditions.

1.3.3 Comparing Encoding-Based and Retrieval-Based Accounts

To sum up, the main difference between the two sides of the Encoding-Based and Retrieval-Based debate is that the impoverished relational encoding theory argues that in the distinctive encoding condition, less memorized information related to the critical lure is encoded, which in turn leads to a decrease in false memory. On the other hand, the distinctiveness heuristic account suggests that there is no difference in terms of the encoded memorial information about the critical lure between the distinctive and non-distinctive conditions. The decrease in false memories is observed because a retrieval strategy is applied during the test, where all items without the anticipated distinctive information are rejected deliberately (Hege and Dodson

2004).

There is still no consensus as to whether encoding-based or retrieval-based processes overpower one another or operate conjointly supporting the effect of distinctiveness on false memory (Smith and Hunt 2020). There is also not enough evidence whether the effect could depend on ways of operationalizing distinctiveness and modality of stimuli (visual vs. verbal). In literature, it is possible to observe contradictory and mixed findings favoring both theories; and some of them will be discussed below.

Schacter et al. (2001), using the DRM paradigm, tested whether the use of a retrieval strategy due to distinctiveness can be turned off during the test. Pictorial vs. verbal encoding was compared by including an additional condition of instruction type. In one condition, participants were given "inclusion instructions" directing them to classify any item that is similar to the studied items and relevant to the studied themes as "old". The other group received no specific instructions. It was observed that the reduction in false recognition rates due to pictorial encoding was still present in no instruction group while it was not present in the inclusion instruction group, leading the authors to conclude that the effect of distinctiveness emerges because participants embark a retrieval strategy, which can be turned off when the retrieval instructions change. This finding is in favor of a distinctiveness heuristic, because if the effect of distinctiveness was due to the automatic encoding related processes as suggested by impoverished relational encoding, it would not be possible to observe different results due to a change in retrieval strategy.

On the other hand, a finding that is in favor of the impoverished relational encoding account came from the study of Arndt and Reder (2003), also the DRM paradigm, where participants studied some lists in the distinctive condition, where each item of a list was presented with a unique font. The other lists were presented in the non-distinctive condition, with each item of a list presented with the same writing font. It was expected that studying each item of a list with a unique font would promote item-specific processing, which would then decrease false recognition. The expectation of the authors was confirmed with the distinctive DRM lists leading to fewer false alarms, which is a finding that would not have been observed if distinctiveness heuristics was in operation, since it includes a retrieval strategy that is applied to all items regardless of the condition. Therefore, the authors concluded that in their case, the impoverished relational encoding is the factor that reduces false memories.

In the studies mentioned above, it can be seen that distinctiveness was investigated by focusing on the studied items themselves or manipulating the encoding instructions. The effect of distinctiveness can also be observed by manipulating the semantic context in which the items are encoded (Alakbarova, Hicks, and Ball 2021). If we refer to the definitions of primary and secondary distinctiveness, semantic context manipulation falls under the definition of secondary distinctiveness.

However, little research has focused on this aspect of distinctiveness so far. One example comes from the study of Alakbarova, Hicks, and Ball (2021) where they embedded DRM list items into sentences that are converging on the theme of the list vs. diverging from the theme of the list. Lower false memory rates were observed in the divergent condition, leading the authors to conclude that the semantic context had an effect on false memories, and a possibility is that the sentences in the divergent condition were more distinctive (Alakbarova, Hicks, and Ball 2021). This finding was a replication of studies that used the same sentence manipulation before, using only children or comparing adults and children (Dewhurst, Pursglove, and Lewis 2007; Howe and Wilkinson 2011; Thomas and Sommers 2005). It can be seen that the manipulation of the semantic context was done verbally in the prior studies, using sentences or stories. To our knowledge, only one study used realistic pictures of objects as list items and natural scene backgrounds to manipulate semantic context and to investigate the effect of distinctiveness. In particular, Howe (2008) manipulated semantic context to make some items more distinctive, by placing them on a semantically mismatching background (such as a fruit in the context of bathroom) where the items and backgrounds were both in the pictorial modality. It was expected that the incongruent background condition would yield to lower false memory rates; however, it was not the case. The decrease in false recognition rates were only observed when backgrounds were distinct perceptually but not conceptually. That is, when the author investigated distinctiveness by making the backgrounds homogeneous or heterogeneous within the list in terms of color (i.e. presenting all list items in one-color background or presenting each list item on an individual-color background; Howe 2008), they were able to observe a decrease in false memories for the heterogenous background condition. In this study, Howe (2008) concluded that children benefit from distinctiveness when perceptual but not conceptual features of studied items are manipulated. It remains unclear whether this finding would apply to adults or young adults, since it is known that children show less reliance on semantic connections and may prefer to use perceptual information instead of conceptual (Bjorklund 2005; Howe 2008). It should also be noted that in this study, the item names were read out loud by the experimenter during the encoding process, which makes the experiment not entirely in the visual modality.

1.4 Our Research

Considering the above concerns about the lack of studies investigating conceptual distinctiveness and the effect of semantic context, we aimed to fill in this gap in with our research. Also, it is essential to better understand the effects of conceptual distinctiveness and to compare it with perceptual distinctiveness, and to shed light on its mechanisms through encoding-based and retrieval-based theories. Considering that most studies use verbal material while investigating distinctiveness, and there is a lack of studies that used visual CAP, we aimed to test the distinctiveness effect using the visual CAP, which also enabled us to present objects and backgrounds in the same modality.

In two studies, we aimed to study the effect of distinctiveness on visual false memories using our newly created stimuli set for the pictorial version of the Category Associates Procedure (CAP). In the literature, when pictures are used while presenting DRM lists or categorical lists, in most of the studies the general approach is to also include the words on the screen or present the words auditorily in addition to the pictures (Gallo, Weiss, and Schacter 2004; Gallo et al. 2007; Ghetti, Qin, and Goodman 2002; Israel and Schacter 1997; Schacter, Israel, and Racine 1999; Smith and Hunt 2020). However, as stated by Smith and Hunt (1998), additional auditory presentation has an influence on the visual formats of presentation, which may influence the false memory rates. Also, studies testing false memories with pictorial stimuli generally used black and white line drawings of DRM and CAP word lists (Israel and Schacter 1997; Smith and Hunt 2020). It has been suggested that black and white line drawings may not completely depict the item that it is supposed to represent (Smith and Hunt 2020), and there were inconsistencies between the results of experiments that used line drawings and real-life pictures (Israel and Schacter 1997; Koutstaal, Schacter, and Brenner 2001; Smith and Hunt 2020; Smith, Hunt, and Dunlap 2015). In studies that used real life pictures, most of the time authors created their stimuli set by picking items from DRM lists that can be presented pictorially (Shimane and Itoh 2022). However, for the use of real-life pictures, using category lists within the CAP seems a more reliable solution since DRM lists may include abstract items or items that cannot be objectively presented as real-life pictures (Deese 1959; Roediger and McDermott 1995). Using this stimuli set, we aimed to investigate the effect of distinctiveness on visual false memories. To promote distinctiveness, the context (visual background) in which the studied items (visual objects) are encoded was manipulated. Conceptual and perceptual distinctiveness was compared in two studies by changing the encoding context (i.e., objects' background) in terms of meaning congruency between the object and background scene where it is placed (Study 1; conceptual distinctiveness) and in terms of background color heterogeneity (Study 2; perceptual distinctiveness). We also aimed to shed light on encoding-based and retrieval-based theories and how they might explain distinctiveness effects on false memory. We tested these hypotheses:

H1. False Memory Effect

H1.1 False recognition of objects (tested in Studies 1 and 2)

Individuals are more likely to recognize critical lure objects than other non-presented (unrelated lure) objects. If expected results are found, the false memory effect commonly obtained with DRM and CAP lists will be replicated using the visual CAP with our newly created stimuli set.

H1.2 False recognition of backgrounds (tested in Study 1)

Individuals are more likely to recognize critical lure background items than other non-presented (unrelated lure) background items. If expected results are found, the common false memory effect will be observed using background pictures from different categories.

H2. The Effect of Conceptual Distinctiveness

H2.1 Distinctiveness effect for objects

Individuals have lower false memory rates for incongruent lists (i.e., more distinct) than for congruent lists. That is, the proportion of yes responses given to critical lures will be lower for incongruent lists than congruent lists.

H2.2 Distinctiveness effect for backgrounds

Individuals have lower false memory rates for incongruent backgrounds (i.e., more distinct) than for congruent backgrounds. That is, the proportion of yes responses given to critical lure backgrounds will be lower for incongruent lists than congruent lists.

H3. The Effect of Perceptual Distinctiveness

H3.1 Homogeneous versus Heterogeneous backgrounds (Distinctiveness as unexpected context)

The heterogeneous condition will lead to lower false memories than the homogeneous condition since in the heterogeneous condition item-specific processing will be in operation, instead of relational processing (Howe 2008). Specifically, we expected heterogeneity to be more surprising, thus leading to higher distinctiveness. If this hypothesis is confirmed, the finding of Howe (2008) will be extended, pointing to similar processes for children and adults in terms of the effect of perceptual distinctiveness on false memories.

H3.2 Neutral (Gray) versus Color (Distinctiveness as visual saliency)

Individuals have lower false memory rates for lists from the colored background condition than for lists from the neutral background condition. Specifically, we expected a colored background item to perceptually stand out more, thus providing higher distinctiveness.

2. STUDY 1

Study 1 aimed to investigate the effect of *conceptual distinctiveness* on false memories using a pictorial Category Associates Procedure. The list items were presented on congruent versus incongruent background pictures to manipulate conceptual distinctiveness. Distinctiveness was manipulated by presenting items in either congruent or incongruent backgrounds. The congruent condition involves natural and expected pairings of items with their backgrounds, while the incongruent condition involves unnatural and unexpected pairings. We also tested the false memory effect using our newly created stimuli set (shared at OSF: https://osf.io/nbfru/).

2.1 Method

2.1.1 Participants

Fifty-two participants joined the experiment. The participants were Sabancı University students (40 females, 12 males) recruited from Sabancı University SONA System (Systems n.d.), and they received course credit for participating in the study. The mean age of the participants was $M = 21.1$ ($SD = 1.62$). The ethical approval was received from the Sabancı University Research Ethics Council (SUREC) prior to the study. All the participants received informed consent.

2.1.2 Procedure

Participants completed the experiment using the lab computer in a standard lab cubicle. Prior to the beginning of the experiment, they signed consent forms and were given instructions about the experiment. All phases within the experiment were explained to the participants before the experiment, and then the experimenter left the room and the participants completed the experiment on their own. The experiment was carried out using PsychoPy software (Peirce et al. 2019). The duration of the study was 40-50 min.

2.1.3 Materials

2.1.3.1 False visual memory task

We implemented category associates procedure (Brainerd, Reyna, and Kneer 1995; Hintzman 1988) to investigate the effect of distinctiveness on false memory formation, by manipulating the congruence of backgrounds where the category items are presented.

2.1.4 Stimuli

Encoding phase

We selected 14 categories of objects, each including 12 category items (i.e., objects that belong to the category). The congruent backgrounds were determined by the researchers for each category, according to how these category items tend to appear in the natural settings. The incongruent backgrounds were selected by the researchers according to how unrelated or unnatural the background and the category items are when presented together. For example, for the furniture category, in the congruent condition, the furniture items were presented with a living room as the background picture. For the incongruent condition of the same category, the items were presented with a cave above the sea as the background picture. Using this method, congruent and incongruent background pictures were determined for each item of each category (see Figure 2.1).

Figure 2.1 Congruent vs. incongruent conditions

Congruent condition   Incongruent condition

Both category items and backgrounds were presented as colored pictures. The categories and category items were obtained from two sources: Misirlisoy (2004) and Van Overschelde, Rawson, and Dunlosky (2004). In both articles, the category items were presented as words. The categories whose items are eligible to be presented as pictures were selected for our study. Namely; the items for the categories of flowers, four-legged animals, fruits, vehicles, instruments, fish, vegetables, birds and watercraft were obtained from Misirlisoy (2004). The other five category items; carpenter's tools, an article of furniture, an article of clothing, non-alcoholic beverages, and toys were retrieved from Van Overschelde, Rawson, and Dunlosky (2004). For each category, 12 items were presented and for categories that did not have enough visually presentable items in one source, additional items were obtained from the other source. For example, for the carpenter's tools category whose items were mostly retrieved from Van Overschelde, Rawson, and Dunlosky (2004), some other items were used from the carpenter's tools category list from Misirlisoy (2004) to be able to present a total of 12 items. In some cases, some items from the original lists were outdated and it was difficult to find good quality pictures, such as trolleybus from the vehicles category. These kinds of items were again replaced with items from the secondary source; for the trolleybus case, it was replaced with a bicycle from the category list "a transportation vehicle" of Van Overschelde, Rawson, and Dunlosky (2004). The stimuli were presented in 960x540 dimensions, with objects placed in the middle of the backgrounds.

Recognition Phase

The recognition of both, the encoded objects and backgrounds, was tested. The stimuli for objects' recognition included three item types: 1) studied items, 2) critical lure items, 3) unrelated filler items. From each category list, a portion of the items presented during encoding (1th, 4th, 7th and 10th items of each list) was selected as studied items used in the recognition phase, making a total of 56 items. The critical lure object from each category was included in the recognition list, making a total of 14 critical lure items. In addition to those, forty-two unrelated unstudied objects were included in the recognition list as filler items. In total, a recognition test with 112 items was created. Each object picture was standardized using a custom code so that all of them are presented in dimensions of 600x600.

The stimuli for the recognition of backgrounds were also created in the same manner; the test list included a portion of studied backgrounds (4 background items from each category, making a total of 56 studied backgrounds), one lure item for each category of backgrounds (14 lure backgrounds), and 42 unrelated unstudied background items. The only difference from the object recognition test was that for the background recognition test, the lure items were backgrounds that are similar (i.e., belong to the same category) to the studied backgrounds. The background pictures were presented in 960x540 dimensions.

All objects and backgrounds were individually obtained from stock picture websites iStock and Adobe Stock (Stock n.d.; *Stock images, Royalty-Free images, illustrations, vectors and videos* N.d.). The created stimuli set is shared at Open Science Framework (https://osf.io/nbfru/).

2.1.5 Task Design

The task consisted of an encoding phase, a distraction phase and a recognition phase (see Figure 2.2). Afterwards, participants answered strategy questions.

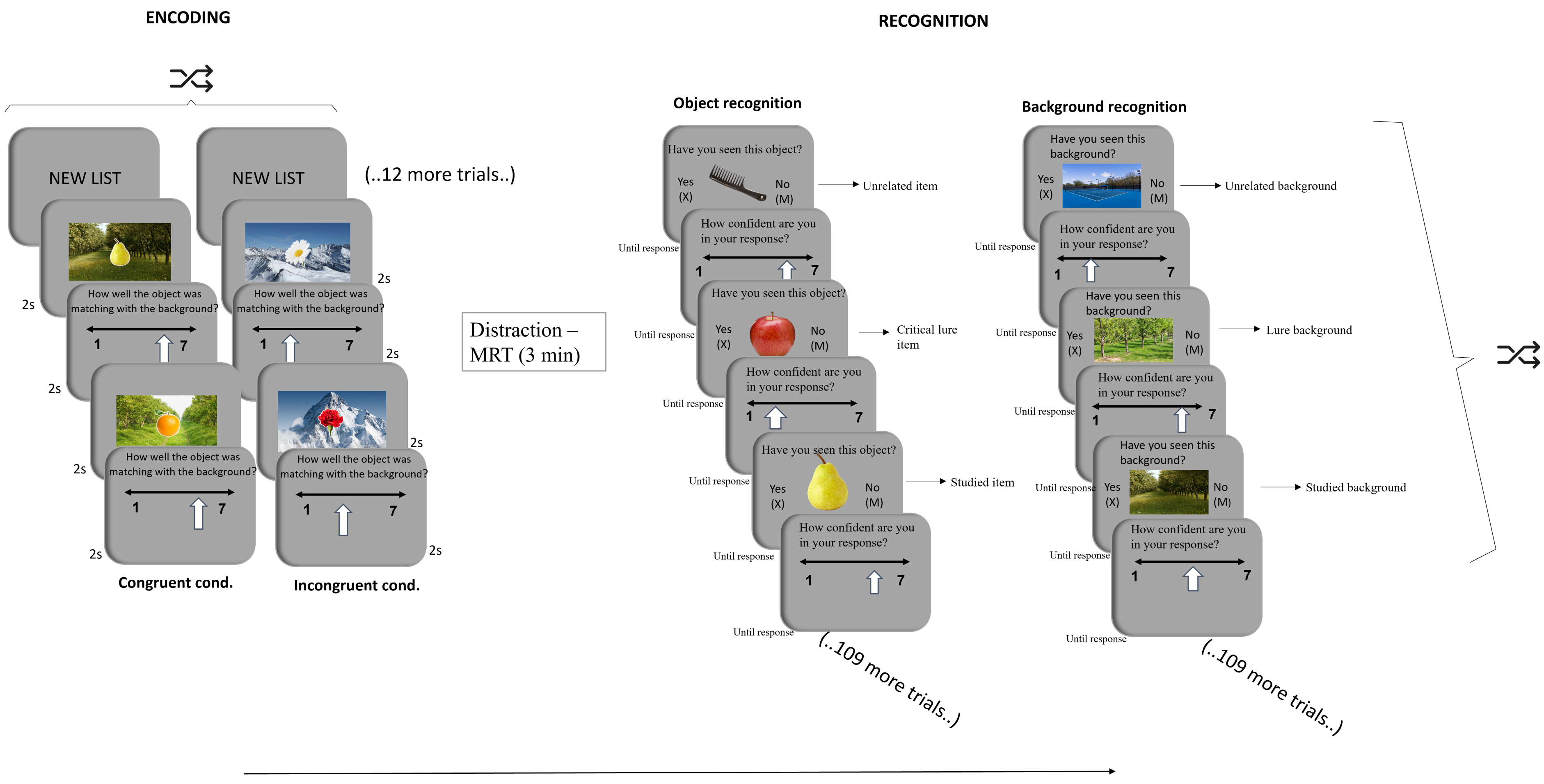
During the encoding phase, fourteen category lists were randomly divided into two; and half of the participants saw the first 7 categories in congruent condition while the other half of the participants saw the last 7 categories in congruent condition, for counterbalancing purposes. Items of each category were presented in a blocked manner, and background congruence was also manipulated in a blocked fashion. That is, items of a category list were either presented on a congruent or incongruent background for one participant.

To make sure that participants were paying attention to the objects as well as the backgrounds and as a measure for the congruence of backgrounds; after each picture, participants rated how matching was the object with the background, on a Likert scale from 1 to 7 (1 = “Not matching at all”, 7 = “Perfectly matching”). The title “New List” was presented between each category list.

During the distraction phase, the unrelated visual-spatial task was performed. Mental Rotation Task (MRT; Peters et al. 1995; redrawn by Vandenberg and Kuse 1978) was used to prevent participants from rehearsing the studied items as well as an individual difference measure. MRT is used to assess the ability to perform spatial visualizations. The test uses 12 drawings of 3D cube constructions. Each question includes one standard cube construction item and four options that include differently rotated cube construction items. The task is to find the two rotated items that are identical with the standard item. For each item, a participant gets a point for correctly identifying the two true options, and a zero point if only one or zero options are correct. The MRT score of a participant is the sum of all points. In the experiment, participants first read the instructions of the MRT, then completed 4 exercise questions, and then moved on to the main task where they were presented with test questions, each on individual pages. The duration of the task was 3 minutes, and then the task automatically ended and moved on to the next phase.

At the recognition phase participants performed a recognition test for both objects and backgrounds. Participants first completed the object recognition test (See Figure 2.2). Each item appeared on the screen until a response was made, and with each item a question also appeared on screen: “Have you seen this object before?” and participants responded using the keyboard, X indicating “Yes” and M indicating “No”. After they made their decision, they rated how confident they were about their response, on a scale between 1 to 7 (1= “Not confident at all”, 7= “Strongly confident”). The same procedure applied for the background recognition test (See Figure 2.2).

Figure 2.2 Flow of Experiment 1



Afterwards, participants were asked 7 questions related to their general strategies while performing the experiment. The responses were obtained on a 7-point Likert scale, 1 indicating "Not true at all", 7 indicating "Certainly true". The questions were as follows:

1) While memorizing the objects, I paid attention to its visual details (color, shape, etc.).

2) While memorizing the objects, I named the objects to myself.

3) I paid more attention to the object than the background.

4) It was more difficult to memorize the object when the background was mismatching.

5) I indicated having seen an object (said yes) because I actually remembered exactly seeing it.

6) I indicated having seen an object (said yes) because it felt familiar.

7) I indicated having seen an object (said yes) because I remembered its category, but not the exact appearance.

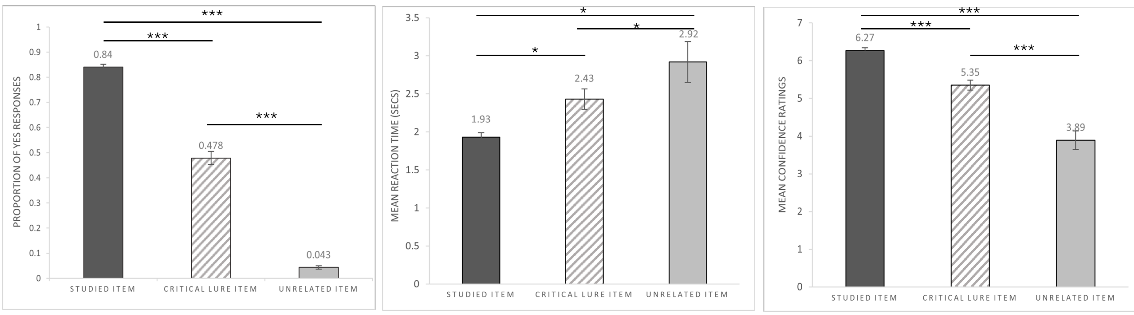
Lastly, participants completed the measure of individual differences in visual imagery, the VVIQ (Marks 1973). This questionnaire includes 16 items, and each of these items asks participants to create some mental images (e.g., *"The overall appearance of the shop from the opposite side of the road").* The participants then rate the vividness of these mental images between 1 and 5, 1 indicating "No image at all, you only "know" that you are thinking of an object", 5 indicating "Perfectly clear and as vivid as normal vision". The overall VVIQ score of a participant is calculated by summing all ratings. VVIQ has a Cronbach's α of .88 (McKelvie 1995).

2.2 Results

2.2.1 False Memory Effect for Objects (H1.1)

Repeated-measures ANOVA was performed to reveal the false memory effect in our paradigm and to test False Memory Effect for Objects. We evaluated the effect of Item Type (Studied, Critical lure, Unrelated) on the Recognition (i.e., proportion of yes responses) for objects as well as on mean Reaction Time and mean Confidence Rating. Figure 2.3 represents the results.

Figure 2.3 Proportion of yes responses, reaction times and confidence ratings for three item types[1]



2.2.1.1 Object recognition

Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 35.400$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .663$). The main effect was significant, $F(1.327, 67.668) = 737.406$, $p < .001$, partial $\eta^2 = .935$, indicating that there was a significant difference between at least two groups of item types.

Post-hoc pairwise comparisons with a Bonferroni correction revealed significant differences between all item types, all $ps < .001$. The recognition of studied items was significantly higher than of critical lure items and unrelated items. Moreover, the (false) recognition of critical lure items was significantly greater than unrelated items, which indicates the presence of false memory for critical lure items of objects, which we hypothesized.

[1] Note that the error bars show standard error, not standard deviation, unless stated otherwise.

2.2.1.2 Reaction time

While analyzing the reaction times, we first specified the outliers within our data by calculating the Z-scores. The ones that have a Z-score higher than 3.29 were replaced with values that were calculated by adding 2.5 standard deviations to the mean reaction time for the specific item (Mowbray, Fox-Wasylyshyn, and El-Masri 2019). A within-subjects ANOVA was performed to investigate the effect of item type on the reaction times for given yes responses. Results of the sphericity test showed that the assumptions were not met, $\chi^2 = 21.346$, $p < .001$; as a result, Greenhouse-Geisser correction ($\epsilon = .668$) was used to readjust degrees of freedom. The main effect of item type on reaction time was significant, $F(1.335, 42.732) = 12.821$, $p < .001$, partial $\eta^2 = .286$. This result indicates that reaction times significantly differed between at least the two groups of items.

To further investigate which item types significantly differed in terms of reaction times, post-hoc pairwise comparisons were made using Bonferroni correction. The analysis revealed significant differences between all item types in terms of reaction time, all $ps < .05$. Mean reaction time for studied items was significantly smaller than both for critical lures and for unrelated items. Moreover, mean reaction time for the recognition of critical lures was significantly lower than that of unrelated items. These results suggest that participants were the fastest when saying yes to studied items and were faster in (falsely) recognizing critical lures than unrelated items. Thus, reaction time data, consistently with recognition data, indicates the presence of false memory for critical lure objects.

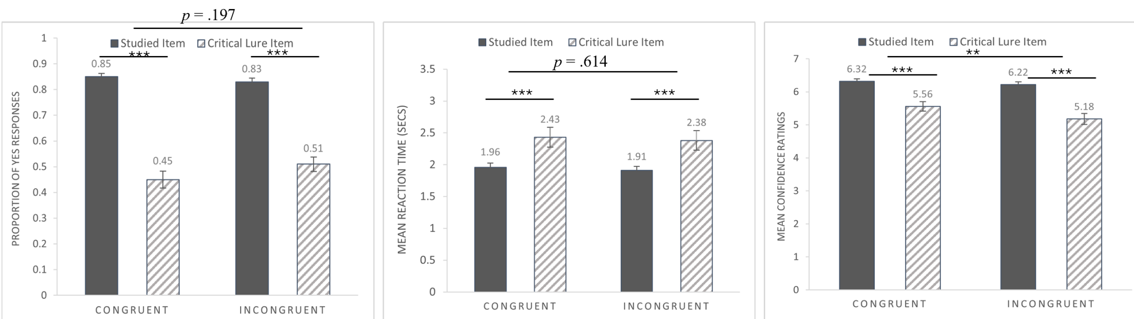
2.2.1.3 Confidence rating

Repeated measures ANOVA was conducted to compare mean confidence ratings of three item types (studied, critical lure, unrelated). The sphericity assumption was violated, $\chi^2 = 17.309$, $p < .001$; therefore Greenhouse-Geisser was used to correct degrees of freedom ($\epsilon = .700$). The main effect of item type on mean confidence ratings was significant, $F(1.401, 44.823) = 65.169$, $p < .001$, partial $\eta^2 = .671$. Post-hoc pairwise comparisons with a Bonferroni correction revealed significant differences between all item types, all $ps < .001$. Mean confidence rating for the recognition of studied items was significantly higher than that of critical lure items and unrelated items. Importantly, mean confidence ratings for critical lure items was significantly greater than mean confidence ratings for unrelated items. Thus, participants were more confident in their responses when falsely recognizing critical lure items, than when falsely recognizing unrelated items, again, supporting the false memory effect. Mean confidence rating for the recognition of studied items was significantly higher than that of critical lure items and unrelated items. Importantly, mean confidence ratings for critical lure items was significantly greater than mean confidence ratings for unrelated items. Thus, participants were more confident in their responses when falsely recognizing critical lure items, than when falsely recognizing unrelated items, again, supporting false memory effect.

2.2.2 The Effect of Conceptual Distinctiveness (H2)

To investigate the effect of conceptual distinctiveness on object Recognition, a 2 (congruent vs. incongruent) X 2 (studied vs. critical lure) repeated-measures ANOVA was conducted (Figure 2.4). The item type was included in the analyses to examine the possible differences of congruence effects for studied vs. critical lure items (Note that the unrelated items were not included in the analyses since they were not presented on congruent vs. incongruent backgrounds during the encoding phase).

Figure 2.4 Proportion of yes responses, reaction times and confidence ratings for congruent and incongruent conditions



2.2.2.1 Object recognition: item type x congruence

The results of the repeated-measures ANOVA revealed a significant effect of item type, $F(1,51) = 229.786$, $p < .001$, partial $\eta^2 = .818$, as the recognition of studied items were significantly higher than that of critical lure items. However, the effect of congruence was not significant, $F(1,51) = 1.712$, $p = .197$, partial $\eta^2 = .032$. The interaction between item type and congruence was significant, $F(1,51) = 6.036$, $p = .017$, partial $\eta^2 = .106$. Post-hoc pairwise comparisons revealed that recognition for the studied items was always higher than for the critical lure items, but more so for the congruent condition.

2.2.2.2 Reaction time: item type x congruence

To investigate the effect of item type and congruence on reaction time for object recognition, a 2 (studied, critical lure) by 2 (congruent, incongruent) repeated measures ANOVA was conducted on reaction times of given yes responses. The main effect of item type was significant, $F(1,49) = 23.59$, $p < .001$, partial $\eta^2 = .325$. The main effect of congruence or the interaction effect of item type and congruence was not significant, $ps = .614$, and .973, respectively.

Post-hoc pairwise comparisons using Bonferroni correction revealed that reaction time for the recognition of critical lures was significantly higher than that of studied items for both conditions. (Mdiff = .475, $SE = .098$), $p < .001$.

2.2.2.3 Confidence rating: item type x congruence

A 2x2 repeated measures ANOVA with item type (studied, critical lure) and congruence (congruent, incongruent) as the independent variables was conducted on the mean confidence ratings of given yes responses. The main effect of item type on mean confidence ratings was significant, $F(1,49) = 72.37$, $p < .001$, partial $\eta^2 = .596$. Mean confidence rating for the recognition of studied items was significantly greater than that of critical lures (Mdiff = .912, $SE = .107$), $p < .001$. Moreover, the main effect of congruence was significant, $F(1,49) = 7.15$, $p = .01$, partial $\eta^2 = .127$, meaning that the confidence ratings differed according to the levels of congruence, with mean confidence ratings being higher in congruent condition than in incongruent condition (Mdiff = .259, $SE = .097$), $p = .01$. However, the interaction between item type and congruence was only trending, $p = .055$.

2.2.2.4 Signal detection measures for object recognition

Signal detection indices (Green and Swets 1966) were calculated to additionally examine the use of encoding- or retrieval-based strategies, as it is a common approach in false memory literature (Gunter, Bodner, and Azad 2007; Huff and Bodner 2013; Huff, Bodner, and Fawcett 2015). Two types of indices are calculated in the scope of Signal Detection Theory, which are sensitivity (*d'*) and bias (c). Sensitivity index (also known as discriminability) is related to how much information was encoded in memory during the study phase. It is computed as a difference between the standardized rate of hits (signal present curve) and false alarms (signal absent curve). If the distance between these two distributions is higher, this indicates greater discriminability between hits and false alarms. In other words, if more information is encoded about studied items, the distributions of hits and false alarms would be more separated, leading participants to discriminate better. Therefore, sensitivity index can be used to gauge the encoding-based claims in terms of the effect of distinctiveness (Huff, Bodner, and Fawcett 2015; Wickens 2001).

The sensitivity index (*d'*) was calculated for both congruent and incongruent conditions using this formula:

$$d' = z(Hit\,Rate) - z(False\,Alarm\,Rate)$$

While applying this formula, Z scores are calculated by using the inverse of the standard normal cumulative distribution function (CDF; Macmillan and Creelman 2004). Using this calculation, if a participant has a hit rate or false alarm rate of 0 or 1, this would lead to infinite values. Therefore, 0 and 1 values of hit rates and false alarm rates were replaced with 1/2N and 1 - (1/2N) values, N being the trial number (Macmillan and Creelman 2004). Afterwards, the d' scores for each participant is calculated by subtracting Z scores of hit rates and false alarms for each participant.

On the other hand, the index of bias is related to the tendency of participants to embark on a more stringent (conservative) or a more relaxed (liberal) criterion when accepting or rejecting an item during the test (Wickens 2001). However, as stated by Huff, Bodner, and Fawcett (2015), in studies where distinctiveness is manipulated, this might result in a mirror effect where hit rates are observed to increase while false alarm rates are observed to decrease. As a result, since the calculation of criterion c requires the use of hit rate and false alarm rate, these measures moving in the opposite directions might result in misleading findings.

$$c = -\frac{1}{2}[z(Hit\,Rate) + z(False\,Alarm\,Rate)]$$

Therefore, it is suggested to use a different measure of response criterion, which is the lambda value (λ; Huff, Bodner, and Fawcett 2015), since it is calculated only using the false alarm rate. Similar to criterion c, higher values indicate a more conservative decision criterion, which may be understood as better monitoring at retrieval (Huff, Bodner, and Fawcett 2015). Lambda value was calculated as an index of response criterion of the participants in congruent and incongruent conditions using the below formula:

$$\lambda = Z(1 - False\,Alarm\,Rate)$$

First, in order to compare the congruent and incongruent conditions in terms of the sensitivity index, a paired-samples t-test was conducted on d' values of each condition. The results of the t-test revealed a significant difference between the sensitivity indices of congruent ($M = 1.27$, $SD = .70$) and incongruent conditions ($M = 1.02$, $SD = .56$); [$t(51) = 2.38$, $p = .02$]. The effect size of the difference between the conditions was measured using Cohen's d, which resulted in an effect size of .33. This can be interpreted as a small-to-medium effect size.

Considering the sensitivity values of the congruent and incongruent condition; it can be said that the ability of participants to discriminate between studied items and critical lures was better in the congruent condition, since the d' value is significantly higher.

To compare the conditions in terms of response criterion, paired-samples t-test was conducted. The results revealed no significant difference between lambda values of congruent ($M = .153$, $SD = .693$) and incongruent conditions ($M = -.026$, $SD = .582$); [$t(51) = 1.87$, $p = .068$]. According to these results, there was no difference in the response criterion of participants in incongruent versus congruent conditions; such that they were not more liberal or more conservative during the test in one or the other condition.

2.2.2.5 Analysis of the most and the least congruent lists

Since the distinctiveness in our study was manipulated using congruent versus incongruent backgrounds, and we had participants' ratings of how well the objects matched their backgrounds, we decided to further analyze the data to assess the effectiveness of our manipulation. To do this, we focused on a subset of stimuli with the highest congruency ratings for the congruent condition and the lowest ratings for the incongruent condition.

For this objective, from the congruent condition, we picked categories which had mean congruency (matching) rating above 6. There were five categories with ratings above six which were presented in congruent condition; and these were *beverages, clothing, furniture, vegetables and vehicles.* Next, we picked five categories from the incongruent condition as well with the lowest Likert ratings, and these categories had ratings lower than 1.40. The categories that we selected from the incongruent condition were *animals, fish, fruits, vegetables and watercraft.* After selecting the categories with highest congruence and incongruence, we conducted the same analysis to compare the recognition performance across two conditions.

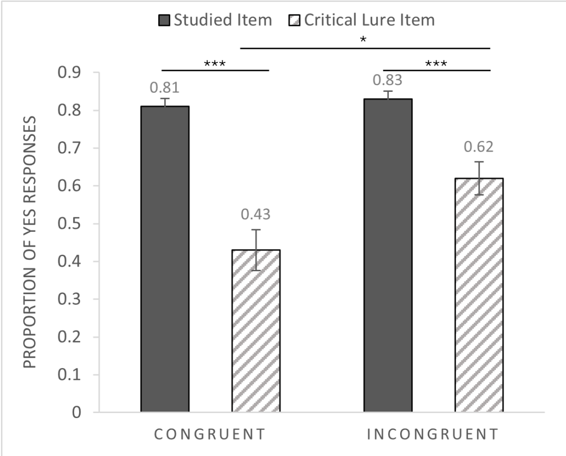
A 2 (item type) x 2 (congruence) repeated-measures ANOVA was performed on the proportion of yes responses. The main effect of item type was significant, $F(1,51) = 46.17$, $p < .001$, partial $\eta^2 = .475$. The recognition of studied items was higher than recognition of critical lure items, (Mdiff = .296, $SE = .043$), $p < .001$. The main effect of congruence was also significant, $F(1,51) = 7.10$, $p = .01$, partial $\eta^2 = .122$. This indicates that there was a significant difference between congruent and incongruent conditions, with incongruent condition having a higher proportion of yes responses (Mdiff = -.101, $SE = .037$), $p = .01$. The interaction between item type and congruence was also significant, $F(1,51) = 6.54$, $p = .014$, partial $\eta^2 = .114$, such that the difference between congruent and incongruent conditions was higher for critical lure items.

Post-hoc comparisons using a Bonferroni correction revealed that in congruent condition, recognition of studied items was higher than the recognition of critical lures, $p < .001$. The same applied for incongruent condition, with recognition of studied items being higher than that of critical lures, $p < .001$, which are results that are similar to our analyses made with the whole stimuli set.

Different from our previous analyses including all lists, post-hoc comparisons with a Bonferroni correction revealed a significant difference between the proportion of yes responses given to critical lure items in congruent condition and incongruent condition, p = .046. This result suggests that participants falsely recognized critical lures more in the incongruent condition, when a subset of stimuli with highest congruence and incongruence is used (Figure 2.5).

There were no differences in terms of reaction times and confidence ratings between congruent and incongruent conditions, $p = .667$ and $p = .868$, respectively.

Figure 2.5 Object recognition for the most and the least congruent lists



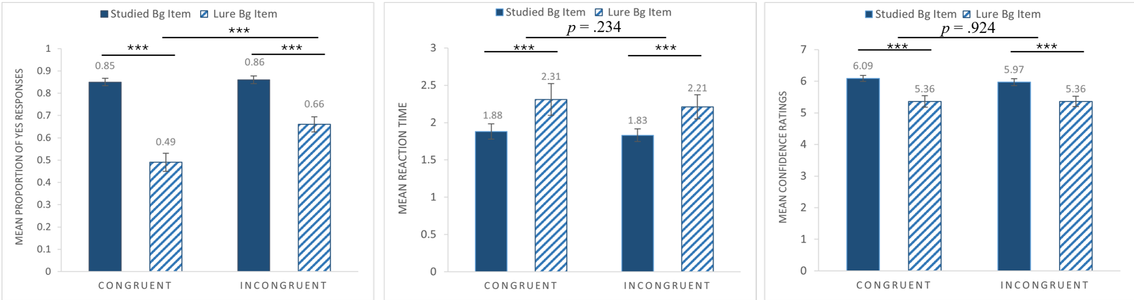
2.2.2.6 Signal detection indices for the most and the least congruent lists

Using the most and the least congruent lists, to compare the conditions in terms of sensitivity (d') and response criterion (λ), paired-samples t-test was conducted. For sensitivity, the results revealed a significant difference between d' values of congruent ($M = 1.2SD = 1.21$) and incongruent conditions ($M = .77$, $SD = 1.02$), ; [$t(51) = 2.38$, $p = .021$], indicating that more memory information was encoded and discriminability was higher in the congruent condition. For lambda, again, there was a significant difference between λ values of congruent ($M = .18$, $SD = 1$) and incongruent conditions ($M = -.30$, $SD = .82$), ; [$t(51) = 2.78$, $p = .008$]. This shows that participants set a stricter criterion for congruent items. These findings support the difference between false alarm rates for congruent and incongruent conditions, using the most and the least congruent lists.

2.2.3 False Memory Effect for Backgrounds (H1.2)

In order to additionally investigate false memory effects, i.e., whether participants were able to distinguish between studied, lure and unrelated backgrounds, we performed the same analyses as for objects. It should be noted that not all participants were presented with the same backgrounds, since participants saw particular categories with congruent or incongruent backgrounds due to within-subjects design and counterbalancing. Background recognition results are presented in Figure 2.6.

Figure 2.6 Proportion of yes responses, reaction times and confidence ratings for backgrounds



2.2.3.1 Background recognition

A 2 (item type: studied background, lure background) by 2 (congruence: congruent, incongruent) repeated-measures ANOVA was conducted on the proportion of yes responses given to background stimuli. Unrelated item type was not included in analyses since those items did not belong to either congruent or incongruent condition.

For the proportion of yes responses given to background items, the main effect of item type was found to be significant, $F(1,51) = 127.1$, $p < .001$, partial $\eta^2 = .714$, which designates that recognition of backgrounds differed significantly for studied vs. lure items ($M\text{diff} = .280$, $SE = .025$), $p < .001$. Moreover, the main effect of congruence was also significant, $F(1,51) = 20.9$, $p < .001$, partial $\eta^2 = .291$, indicating a difference in yes responses between congruent and incongruent background items ($M\text{diff} = \text{-}.093$, $SE = .02$), $p < .001$. Lastly, the interaction between item type and congruence was significant, $F(1,51) = 16.3$, $p < .001$, partial $\eta^2 = .242$. Pairwise comparisons revealed a significant difference between congruent and incongruent conditions for the lure background items, $p < .001$. Mean proportion of yes responses given to lure background items in incongruent condition was significantly higher than that of lure background items in congruent condition. There was no such difference for studied background items.

In terms of unrelated backgrounds, the false recognition of unrelated background items were significantly lower than the recognition of studied congruent backgrounds ($M\text{diff} = .796$, $SE = .017$) and studied incongruent backgrounds ($M\text{diff} = .806$, $SE = .016$), both $ps < .001$. Also, the false recognition of unrelated background items was significantly lower than the false recognition of critical lure congruent background items ($M\text{diff} = .433$, $SE = .038$) and critical lure incongruent background items ($M\text{diff} = .609$, $SE = .031$), both $ps < .001$. The mean proportion of yes responses given to unrelated background items was .056 ($SD = .063$).

2.2.3.2 Reaction time for background recognition: item type x congruence

A repeated measures ANOVA with item type (studied background, lure background) and congruence (congruent, incongruent) as independent variables was carried out on the mean reaction times for given yes responses to investigate the effect of item type and congruence on reaction time for background recognition. The main effect of item type on the reaction time for yes responses given to background items was significant, $F(1,45) = 11.102$, $p = .002$, partial $\eta^2 = .198$. This indicates that there was a significant difference between the reaction time of yes responses given to studied backgrounds and lure backgrounds. The main effect of congruence and the interaction of item type and congruence was nonsignificant, $p = .234$ and $p = .737$, respectively. In order to further understand the nature of the difference between studied and lure backgrounds, post-hoc pairwise comparisons were made with Bonferroni correction. The results revealed that mean reaction time for the recognition of studied backgrounds was significantly lower than (false) recognition of lure backgrounds (Mdiff = -.424, SE = .127), $p = .002$. This might indicate that participants spend more time when deciding to give a yes response to lure backgrounds, but they were faster when giving yes responses to studied backgrounds.

In terms of unrelated backgrounds, the mean reaction time for falsely recognizing unrelated backgrounds was significantly higher than the mean reaction time for recognizing studied congruent backgrounds (Mdiff = -1.12, SE = .268) and studied incongruent backgrounds, (Mdiff = -1.18, SE = .264), both $ps < .001$. Also, mean reaction time for falsely recognizing unrelated backgrounds was significantly higher than falsely recognizing critical lure backgrounds from both congruent (Mdiff = -.801, SE = .330) and incongruent conditions (Mdiff = -.85, SE = .327), $p = .021$ and $p = .014$, respectively. Mean reaction time for the false recognition of unrelated backgrounds was 2.82 ($SD = 1.74$).

2.2.3.3 Confidence ratings for background recognition: item type x congruence

Repeated-measures ANOVA was conducted on mean confidence ratings for given yes responses to backgrounds as item type and confidence ratings are independent variables. The main effect of item type was significant, $F(1, 45) = 48.87$, $p < .001$, partial $\eta^2 = .521$. This indicates that confidence ratings for recognition of studied backgrounds and (false) recognition of lure backgrounds significantly differed. The main effect of congruence and the interaction between item type and congruence was not significant, $p = .924$ and $p = .637$, respectively. The effect of item type was further examined using post-hoc pairwise comparisons with a Bonferroni correction. The results showed that mean confidence ratings for studied backgrounds was higher than mean confidence ratings for lure backgrounds, (Mdiff $= .695$, $SE = .10$), $p < .001$.

For unrelated backgrounds, mean confidence rating for their false recognition was significantly lower than mean confidence ratings for the recognition of congruent studied backgrounds (Mdiff $= 2.1$, $SE = .28$) and incongruent studied backgrounds (Mdiff $= 2.07$, $SE = .29$), both $ps < .001$. Also, mean confidence rating for falsely recognizing unrelated backgrounds was significantly lower than falsely recognizing critical lure backgrounds from both congruent (Mdiff $= 1.47$, $SE = .314$) and incongruent conditions (Mdiff $= 1.62$, $SE = .302$), both $ps < .001$. Mean confidence ratings for the recognition of unrelated backgrounds was 4.12 ($SD = 1.95$).

2.2.3.4 Signal detection measures for background recognition

We performed the same calculations for *d'* and lambda (λ) for the background recognition, using the formulas depicted in section 2.2.2.4. To compare the congruent and incongruent backgrounds in terms of the sensitivity index, a paired-samples t-test was performed using d' values of both conditions. There was a significant difference between the sensitivity values of congruent ($M = 1.22$, $SD = .96$) and incongruent conditions ($M = .67$, $SD = .81$); [$t(51) = 3.29$, $p = .002$]. This shows that the discrimination between studied and critical lure backgrounds was better in the congruent condition. Further, to compare the conditions in terms of response criterion, paired-samples t-test was conducted. There was a significant difference between lambda values of congruent ($M = .016$, $SD = 1.19$) and incongruent conditions ($M = -.61$, $SD = 1.00$), ; [$t(51) = 4.16$, $p < .001$]. This indicates that participants were more conservative for the congruent backgrounds in terms of their decision criterion during retrieval. That is, they expected stronger cues to say "yes" to a congruent background.

2.2.4 Individual Differences in Imagery and Object Recognition

To investigate whether the recognition performance was correlated with our individual differences in imagery measures (MRT and VVIQ), we conducted Pearsons' correlation analysis. The recognition of studied items from the congruent condition was significantly correlated with MRT scores, p = .039. In terms of reaction times, mean reaction time of false alarms for congruent objects were positively correlated with MRT scores, p = .029. For confidence ratings, MRT scores were positively correlated with mean confidence ratings of false alarms of critical lures from the incongruent condition, p = .006. In terms of VVIQ scores, they were positively correlated with confidence ratings of overall hit rates, p = .012; also, for both congruent and incongruent conditions, p = .040 and p = .006, respectively. Moreover, VVIQ scores were positively correlated with confidence ratings of false alarms of critical lures, p = .020, which was coming from the incongruent condition only, p = .018.

Table 2.1 Correlations between object recognition and individual differences in imagery

		MRT			VVIQ		
		Recognition	Reaction Time	Confidence Rating	Recognition	Reaction Time	Confidence Rating
Hit	Overall	.152	-.103	.040	.169	-.125	.346*
	Congruent	.287*	-.117	-.065	.193	-.150	.286*
	Incongruent	-.041	-.070	.124	.138	-.092	.377**
False Alarm (Critical Lure)	Overall	-.099	.086	.145	-.095	-.088	.321*
	Congruent	-.065	.308*	-.173	-.110	-.001	.266
	Incongruent	-.108	-.044	.377	-.043	-.134	.328*
False Alarm (Unr.)		.142	.213	.008	-.071	-.191	.300

Note: * p <. 05, ** p < .01

2.2.5 Strategy Questions for Object Recognition

Pearson's correlation analysis was carried out to examine whether recognition performance correlated with strategies used by participants. Questions 2 and 7 did not correlate with any of the recognition measures.

Question 1; "I indicated having seen an object (said yes) because I actually remembered exactly seeing it." was positively correlated with confidence ratings of the correct recognition of studied objects (hits), p = .009, both for congruent and incongruent conditions; p = .047 and p = .002, respectively.

Question 3; "I indicated having seen an object (said yes) because it felt familiar."

was positively correlated with hits, p = .005, both for congruent and incongruent conditions; p = .011 and p = .021, respectively. Moreover, Question 3 was correlated with false recognition of critical lure objects (critical lure false alarm), p = .041. For confidence ratings, Question 3 was correlated with confidence ratings of hits, p = .046, which was coming from the incongruent condition only, p = .030. Also, Question 3 was correlated with confidence ratings of critical lure false alarms from the incongruent condition, p = .030.

Question 4; "I paid more attention to the object than the background." was correlated positively with both hits and critical lure false alarms, p = .049; p = .050, respectively. It should be noted that these are marginally significant correlations. In terms of reaction times, Question 4 was correlated with reaction times of hits, p = .048, which was attributed to only incongruent condition, p = .035. For confidence ratings, Question 4 was negatively correlated with confidence ratings of critical lure false alarms, p = .016, which was coming from the incongruent condition, p = .004.

Question 5; "It was more difficult to memorize the object when the background was mismatching." showed a positive correlation with critical lure false alarms p = .046. This question also showed a positive correlation with lure false alarms from the congruent condition, p = .021; but not with critical lure false alarms from the incongruent condition, p = .326. In terms of reaction times, Question 5 only correlated with reaction times of unrelated false alarms, p = .037.

Question 6; "While memorizing the objects, I named the objects to myself." only showed a significant correlation with hits from the congruent condition, p = .037.

Correlation table for all strategy questions and object recognition measures can be found in Table 2.2. Tables for correlations between strategy questions and RT and confidence rating are presented in Appendix A.

Table 2.2 Correlations between object recognition and strategy questions

	Hit			False Alarm (Critical Lure)			False Alarm (Unrelated)
	Overall	Congruent	Incongruent	Overall	Congruent	Incongruent	
Q1	.180	.082	.238	.148	.141	.108	-.036
Q2	.116	.083	.077	.124	.087	.124	-.002
Q3	.387**	.350*	.319*	.285*	.269	.208	.065
Q4	.274*	.272	.163	.274*	.238	.228	-.103
Q5	.046	.079	-.040	.278*	.319*	.139	-.014
Q6	.209	.290*	.069	.095	.109	.049	-.087
Q7	.100	.081	.048	.148	.146	.099	-.054

Note: * $p <. 05$, ** $p < .01$

2.2.6 Individual Differences in Imagery and Background Recognition

The possible correlation between background recognition and individual difference measures (MRT and VVIQ) was also examined using Pearson's correlation. MRT scores did not correlate with any of the background recognition variables. In terms of VVIQ scores, they were found to be positively correlated with unrelated false alarms for backgrounds, p = .017. Moreover, VVIQ scores were correlated with confidence ratings of hits for incongruent backgrounds, p = .016; as well as with confidence ratings of false alarms for both congruent and incongruent backgrounds, p = .034 and p = .043. Correlation table for background recognition performance and individual difference measures can be found in Table 2.3.

Table 2.3 Correlations between background recognition and individual differences in imagery

		MRT			VVIQ		
		Recognition	Reaction Time	Confidence Rating	Recognition	Reaction Time	Confidence Rating
Hit	Congruent Background	.036	.032	.167	-.081	.000	.213
	Incongruent Background	.226	.015	-.032	.029	-.074	.333*
False Alarm	Congruent Lure Background	-.040	.063	.099	.001	.132	.309*
	Incongruent Lure Background	-.024	.104	.185	.012	.000	.285*
	Unrelated Background	.046	.075	.232	.329*	.087	.214

Note: * p <. 05, ** p < .01

2.2.7 Strategy Questions for Background Recognition

The correlation between background recognition performance and strategy use was examined using Pearson's correlation analysis. No significant correlations were found for Question 4, Question 6, and Question 7.

Question 1 was negatively correlated with reaction times of hits from the incongruent condition, p = .048.

Question 2 positively correlated with correct recognition of studied backgrounds from the congruent condition, p = .028; and with false recognition of lure backgrounds from the congruent condition, $p = .050$.

Question 3 positively correlated with correct recognition of studied backgrounds from the congruent condition, $p < .001$; and with false recognition of lure backgrounds from the congruent condition, $p = .003$.

Question 5 showed a positive correlation only with correct recognition of studied backgrounds from the congruent condition, $p = .033$.

Correlation table for all questions and background recognition can be found in Table 2.4. The tables of the correlation of strategy questions with RT and confidence ratings can be found in Appendix A.

Table 2.4 Correlations between background recognition and strategy questions

	Hit		False Alarm (Critical Lure)		False Alarm (Unrelated)
	Congruent Background	Incongruent Background	Congruent Lure Background	Incongruent Lure Background	
Q1	.156	.259	.127	.188	.125
Q2	.305*	.097	.274*	.144	.006
Q3	.457**	.136	.409**	.131	.075
Q4	.064	-.030	.197	-.152	-.090
Q5	.296*	.140	.193	.074	-.090
Q6	.137	.157	.013	.175	.030
Q7	.128	-.002	.045	.032	.108

Note: * $p <. 05$, ** $p < .01$

2.3 Conclusions

Study 1 results supported our hypothesis regarding false memory effect in object recognition (H1.1), since we observed that participants falsely recognized critical lure items more often than unrelated items. The findings related to confidence ratings and reaction times also showed that participants were the fastest and most confident when recognizing studied items, which is an expected finding. Notably, yes responses given to critical lures were also given faster and more confidently than yes responses given to unrelated items, which strengthens the evidence for false memory effect.

Consistently, in background recognition (H1.2), we observed a similar pattern, where participants recognized studied backgrounds more than critical lures and unrelated backgrounds. False recognition of critical lure backgrounds was higher than that of unrelated backgrounds, similarly to object recognition further supporting false memory effect. For reaction time and confidence ratings, yes responses given to studied backgrounds were given faster and more confidently than lure and unrelated backgrounds. Also, yes responses given to lure backgrounds were given faster and more confidently than yes responses given to unrelated backgrounds. Our results demonstrate that the false memory effect, comparable with those previously obtained effects in DRM and CAP paradigms, can be also observed when using visually similar pictures that belong to the same theme. In order to get a false memory effect, there is no requirement for a critical lure to be either the most representative item in a category or the highest semantic connectedness to the list items, but is it sufficient to have any item from the category as a critical lure. Our observation suggests that it's the semantic category of items or, possibly visual similarity of items, that matters for the false memory effect. Further research is needed to examine these factors in false memory formation, using visual and verbal stimuli.

Overall, participants were able to differentiate studied background items from lure background items, which strengthens the argument that participants paid attention to the backgrounds since the lure backgrounds were quite similar to the studied backgrounds and might have been difficult to distinguish. Our data also supports the possibility of forming false memories for simultaneously presented items, at a similar rate.

Our second hypothesis regarding the effect of conceptual distinctiveness on false object recognition (H2.1) was not supported. We observed no difference between congruent and incongruent conditions in terms of false alarm rates for critical lure items (for all object recognition data using all lists). In addition, when we analyzed the subset of data with the most and the least congruent lists, the differences were in the opposite to the predicted direction. That is, we observed higher false memory in the more distinct condition. This finding was also observed for background recognition data using all lists.

Even though there was no main effect of congruence on proportion of yes responses for the full set if items, the interaction between item type and congruence was found to be significant, so that the difference between the proportion of yes responses given to studied items and critical lure items appear to be larger for the congruent condition. This might indicate that participants showed better discriminability for the congruent condition. This inference aligns with the signal detection index, sensitivity *(d')*, since we found that participants had a higher sensitivity index *(d')* for congruent items than for incongruent items. Moreover, this finding was supported with data on confidence ratings, which were higher for the congruent condition than for the incongruent condition. Reaction time did not differ according to congruence.

Our hypothesis regarding the distinctiveness effect in background recognition (H2.2) was not supported. Instead, higher false recognition rates of incongruent backgrounds were observed. This finding was supported by signal detection indices; that is, significantly higher sensitivity *(d')* and higher criterion level (λ) was observed for the congruent condition. There was no difference between the congruent and incongruent background conditions in terms of reaction time and confidence ratings.

The analyses of individual differences measures in relation to memory measures showed that higher mental rotation scores were associated with higher hit rates and higher reaction times for false alarms of the congruent condition. Additionally, higher imagery vividness was associated with higher confidence in hits, in both congruent and incongruent conditions for objects, and for the incongruent condition for backgrounds. This is consistent with studies arguing that higher imagery abilities lead to better visual working memory performance (Beran et al. 2023; Keogh and Pearson 2011). However, we also observed that participants with higher MRT scores showed higher confidence ratings for false alarms from the incongruent condition. It was also shown that higher VVIQ scores were related to higher false alarms for unrelated backgrounds, and higher confidence in false alarms for critical lure backgrounds from both conditions. This suggests that higher imagery ability may also have the opposite effect on memory (Dobson and Markham 1993; Gonsalves et al. 2004) and may lead to higher confidence for inaccurate responses (Reisberg and Leak 1987). Correlations between recognition performance and strategy questions revealed that having the memory information of studied items (Q1) was related to higher confidence for hits. On the other hand, the sense of familiarity (Q3) was related to lower confidence for hits and false alarms to critical lures in the incongruent condition. The sense of familiarity (Q3) was also associated with higher hits, as well as with higher false alarms for critical lures. This might indicate that a sense of familiarity both enhances and diminishes the recognition performance (Huff, Bodner, and Fawcett 2015).

3. STUDY 2

Study 2 aimed to investigate the effect of perceptual distinctiveness on false memories in a pictorial Category Associates Procedure. The list items were presented on neutral versus colored backgrounds (within-subjects) and the colored backgrounds were homogeneous (all items within the same list had the same background color) or heterogeneous (each item within the same list had a unique background color) to manipulate perceptual distinctiveness. The newly created stimuli test was also re-tested for the false memory effect.

3.1 Method

3.1.1 Participants

Sixty participants took part in the experiment. The participants (18 Male, 42 Female) were recruited from Sabancı University SONA System (Systems n.d.), and all except one were Sabancı University students who received course credit or a meal coupon upon participation. The mean age of the participants was M = 22.3 (SD = 3.49). The ethical approval was received from the Sabancı University Research Ethics Council (SUREC) before the study.

3.1.2 Procedure

The experiment was conducted in a standard lab cubicle using the desktop computer. First, participants were seated in front of the computer, signed consent forms and were given instructions. Then, the experimenter left the room, and the participants completed the experiment by themselves. The experiment was carried out using PsychoPy software (Peirce et al. 2019). The duration of the experiment was approximately 30 min.

3.1.3 Materials

3.1.3.1 False visual memory task

Same as in the first study, we implemented Category Associates Procedure (Brainerd, Reyna, and Kneer 1995; Hintzman 1988) and used the same visuals representing objects from different categories. Unlike study 1, to investigate the effect of distinctiveness and encoding context on false memories, the backgrounds were manipulated not conceptually but perceptually, using solid colors as backgrounds. The task had a within-between mixed design. The within factor was neutral vs. color condition. Half of the 14 category lists had neutral background while the other half had colored background. The between factor was homogeneous vs. heterogeneous condition. In the homogeneous (non-distinctive) condition, all items of a category were presented on the same background color, while in the heterogeneous (distinctive) condition, each item of a category was presented on a unique background color. The uniqueness of a background may serve as a distinctive marker, as in (Howe 2008). In our experiment, a neutral baseline condition was also added with objects having the same background as the experiment setup background, which was gray.

3.1.4 Stimuli

Encoding phase

The same fourteen object categories from the first study were used for this study (Misirlisoy 2004; Van Overschelde, Rawson, and Dunlosky 2004). Each category included twelve items. For the heterogeneous condition, the twelve colors of the RGB color wheel were used as colors for the backgrounds. For the homogeneous condition, the seven colors out of twelve RGB colors were picked by considering the equal distribution of colors on the wheel. In this way, the picked colors were red, green, blue, yellow, cyan, magenta, and orange. These picked colors were used as solid color backgrounds for the objects, as the objects placed in the middle of the background. The dimensions of the stimuli were 960x540 pixels.

Recognition Phase

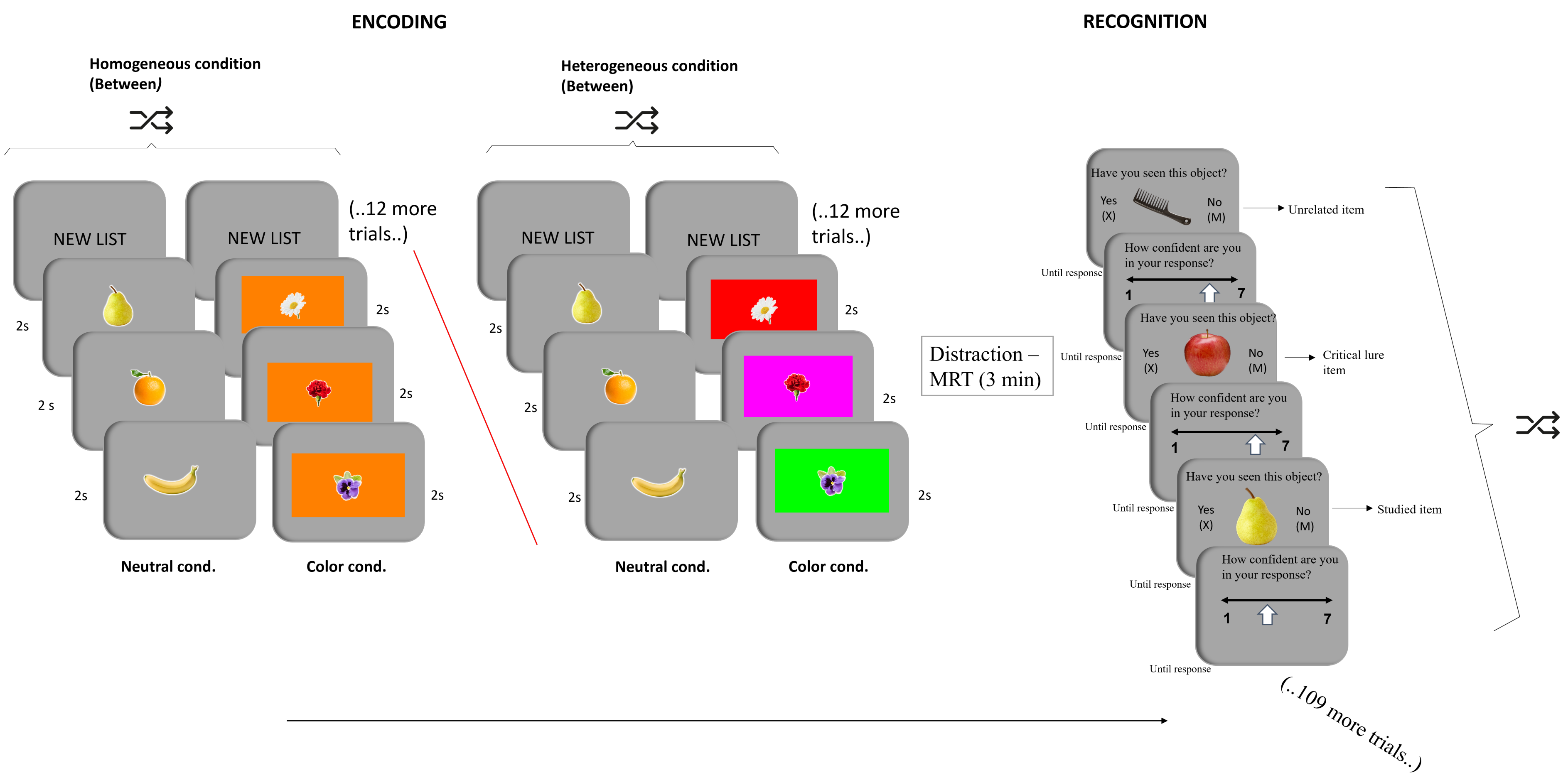
During the recognition phase, the memory for objects was tested. Similar to our first experiment, there were 112 test items; 56 of them were studied items, 14 of them were critical lure items and 42 of them were unrelated items. The size of the objects was standardized so that each of them had dimensions of 600x600 pixels. Unlike our first experiment, background recognition was not tested in this experiment. The stimuli set is shared at Open Science Framework (https://osf.io/nbfru/).

3.1.5 Task Design

The experiment included an encoding phase, a distractor phase and a recognition phase. Afterwards, participants were asked strategy questions. (See Figure 3.1). As in Study 1, in the *encoding* phase, participants were presented with 14 category lists visually. Between each category list, participants saw the "New List" title. The lists were counterbalanced so that each list both appeared with neutral and colored backgrounds, as well as homogeneous and heterogeneous backgrounds. Similar to our first study, during the *distraction* phase, participants performed the Mental Rotation Task (MRT; Peters et al. 1995; redrawn by Vandenberg and Kuse 1978) for a duration of 3 minutes. During the *recognition* phase, participants' memory was tested for objects. Each object appeared on screen and participants were asked "Have you seen this object?" and responded using keys X for "Yes" and M for "No". After each response, participants rated their confidence on a scale from 1 to 7 (1= "Not confident at all", 7= "Strongly confident").

Later, participants answered strategy questions similarly to the first experiment. The only difference was the question "It was more difficult to memorize the object when the background was mismatching." was omitted in the homogeneous condition and changed to "It was more difficult to memorize the object when the background was constantly changing" in the heterogeneous condition. As the last step, participants completed a vividness of imagery questionnaire, the VVIQ (Marks 1973).

Figure 3.1 Flow of Experiment 2



3.2 Results

3.2.1 False Memory Effect for Objects (H1)

3.2.1.1 Object recognition

A repeated-measures ANOVA was conducted to test our hypothesis regarding the false memory effect and to investigate the effect of Item Type (Studied, Critical lure, Unrelated) on Recognition (i.e., proportion of yes responses). Mauchly's test of sphericity revealed that the assumption of sphericity was violated, $\chi^2(2) = 15.176$, p = .001; therefore, degrees of freedom were corrected using Huynh-Feldt estimate of sphericity (ϵ = .833). Huynh-Feldt correction was used by considering the value, as suggested in literature (Blanca et al. 2023). The main effect of item type on recognition was significant, F (1.665, 98.249) = 394.500, p < .001, partial η^2 = .870. This indicated that at least two groups significantly differed from each other in terms of proportion of given yes responses.

Post-hoc pairwise comparisons were made with Bonferroni correction to identify which item types differed from each other. The results revealed a significant difference between all item types, all ps < .001. Studied items were recognized more than critical lures and unrelated items. The (false) recognition of critical lure items was higher than that of unrelated items. This result indicates that, consistent with Study 1, we observed the false memory effect.

3.2.1.2 Reaction time

A repeated-measures ANOVA was performed on reaction times of given yes responses to three item types (Studied, Critical lure, Unrelated). The assumption of sphericity was not met, χ^2 (2) = 12.081, p = .002. Therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity (ϵ = .826). The main effect of item type on reaction time was significant, F (1.652, 71.045) = 12.747, p < .001, partial η^2 = .229.

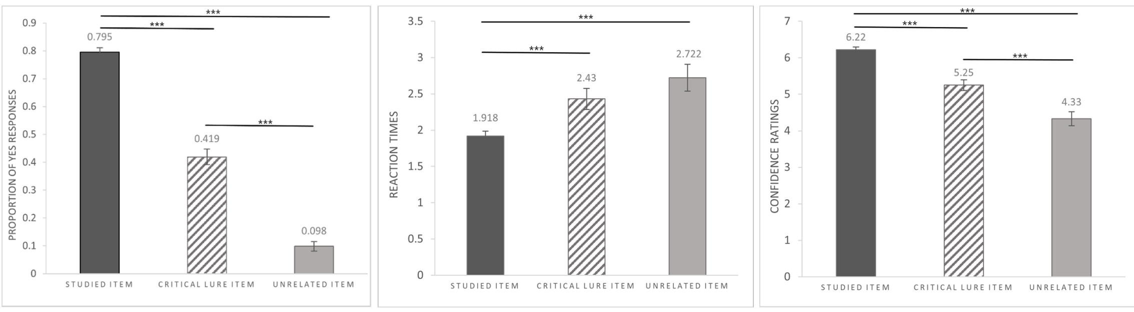
Post-hoc pairwise comparisons with Bonferroni correction showed that yes responses given to studied items were given faster than yes responses given to both critical lure items and unrelated items, both ps < .001. There was no significant difference between the reaction times for yes responses given to critical lure items and unrelated items, p = .416.

3.2.1.3 Confidence rating

To compare confidence ratings for given yes responses to three item types (Studied, Critical lure, Unrelated), a repeated-measures ANOVA was conducted. Mauchly's sphericity test revealed that the assumption of sphericity was violated, χ^2 (2) = 6.648, p = .036. Huynh-Feldt estimate of sphericity was used to correct degrees of freedom (ϵ = .906). The main effect of item type on confidence ratings of given yes responses was significant, F (1.812, 77.924) = 74.876, p < .001, partial η^2 = .635.

Post-hoc pairwise comparisons were made with Bonferroni correction to investigate which item types differed from each other in terms of confidence ratings. Results showed that, confidence ratings of yes responses given to three item types all differed from each other, all ps < .001. Yes responses were given to studied items more confidently than critical lure items and unrelated items. Additionally, confidence ratings of yes responses given to critical lure items were higher than that of unrelated items. Means and standard errors for recognition rates (proportion of yes responses), reaction times and confidence ratings can be seen in Figure 3.2.

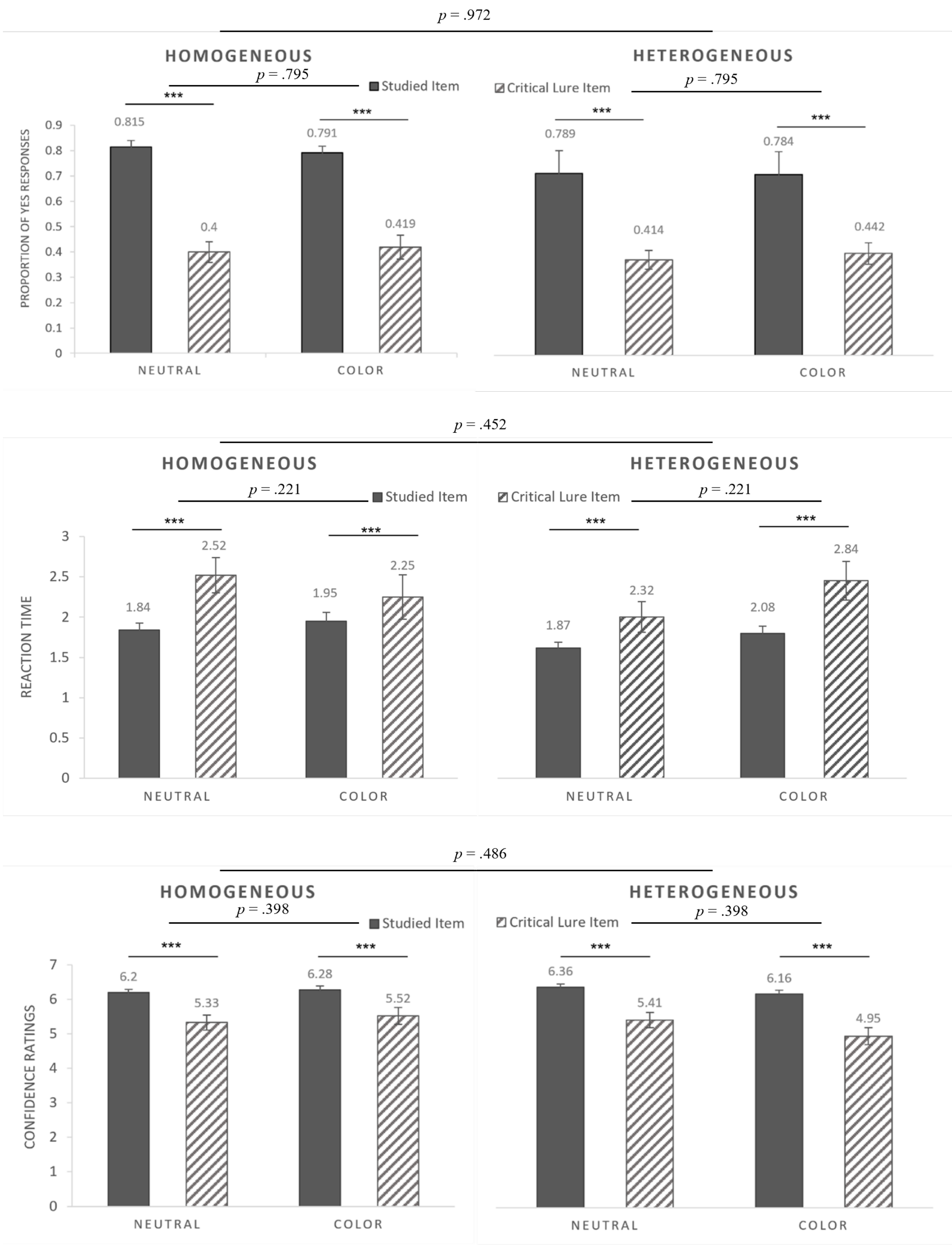
Figure 3.2 Proportion of yes responses, reaction times and confidence ratings for three item types



3.2.2 The Effect of Perceptual Distinctiveness (H3)

To investigate the effect of our background manipulation on recognition, reaction time and confidence rating, 2 item type (studied item vs. critical lure item) x 2 color (neutral vs. colored background) x 2 heterogeneity (homogeneous vs. heterogeneous background) mixed-design ANOVAs were conducted, where the first two factors were within-subject factors while the latter one is a between-subjects factor. (Note that, similar to Study 1, unrelated items were not included in this analysis since they were not presented during the study phase, thus were not subjected to our background manipulations). Means and standard errors for recognition rates, reaction times and confidence ratings of background conditions can be seen in Figure 3.3.

Figure 3.3 Proportion of yes responses, reaction times and confidence ratings for neutral vs. color and homogeneous vs. heterogeneous conditions



3.2.2.1 Item type x color x heterogeneity

The main effect of item type was significant, F (1,58) = 153.405, $p < .001$, partial $\eta^2 = .726$, with the proportion of yes responses given to studied items being higher than critical lure items. However, the main effect of the neutral vs. color condition or the homogeneous vs. heterogeneous condition was not significant, ps = .795 and .972. That is, recognition performance did not depend on background color or heterogeneity, suggesting no effect of perceptual distinctiveness, neither in terms of visual saliency or nor in terms of unpredictability.

3.2.2.2 Reaction time: item type x color x heterogeneity

The main effect of item type was significant, F (1, 55) = 25.194, $p < .001$, partial η^2 = .314, as yes responses given to studied items were given faster than that of critical lure items. The main effects of the color of the background (neutral vs. color) or the heterogeneity of the lists in terms of background (homogeneous vs. heterogeneous) were both insignificant, ps = .221 and .452, respectively.

3.2.2.3 Confidence rating: item type x color x heterogeneity

The main effect of item type was significant, F (1, 54) = 101.342, $p < .001$, partial η^2 = .652. Yes responses given to studied items were given more confidently than yes responses given to critical lure items. The main effects of neutral vs. color conditions or homogeneous vs. heterogeneous conditions were not significant, ps = .398 and .486, respectively.

3.2.2.4 Signal detection measures for object recognition

Sensitivity index *(d')*; which shows the amount of encoded memory information and the ability to distinguish studied items from unstudied ones, and the lambda value (λ), which shows the decision criterion (conservative or liberal) during retrieval, were calculated using the same procedure as in the first study.

To compare the sensitivity indices of the neutral vs. colored background conditions, a paired-samples t-test was performed. The results showed that there was no significant difference between the sensitivity indices of neutral (M = 1.22, SD = .80) and colored background conditions (M = 1.08, SD = .83); [t (59) = 1.63, p = .108]. This finding indicates that there was no difference between the neutral and colored background conditions in terms of the discrimination of studied and critical lure items, and encoded memory information.

To compare the neutral vs. colored background conditions in terms of lambda values, a paired-samples t-test was conducted. According to the results, it was observed that there was no difference between the lambda values of the neutral (M = .25, SD = .65) and colored background conditions, (M = .19, SD = .73); [t (59) = .822, p = .414]. This finding suggests that individuals did not differ in their decision criteria between the neutral and colored background conditions.

An independent-samples t-test was performed to compare the sensitivity indices of homogeneous vs. heterogeneous background conditions. The results revealed no difference between the sensitivity indices of the homogeneous (M = 1.19, SD = .81) and heterogeneous conditions (M = 1.07, SD = .73); [t (58) = .630, p = .531]. Participants did not differ in their ability to discriminate between studied items and critical lure items in homogeneous vs. heterogeneous conditions, and no differences were observed for the amount of memory information that is encoded in the two conditions.

Lastly, an independent-samples t-test revealed no differences between the lambda values of the homogeneous (M = .25, SD = .71) and heterogeneous conditions (M = .19, SD = .57); [t (58) = .378, p = .707]. That is, during retrieval, participants did not differ in their decision criteria between the homogeneous and heterogeneous background conditions.

3.2.3 Individual Differences in Imagery and Object Recognition

Pearson's correlation analysis was performed to explore the relationship between recognition performance and individual differences measures (MRT and VVIQ). The VVIQ correlated positively with recognition of studied items (Hits), p = .009. On further investigation, it was possible to observe that this correlation came from studied items in the colored background condition, since there was a positive correlation between VVIQ scores and Hits of colored background items, p = .004, but not between VVIQ and Hits of neutral background items, p = .09. VVIQ was also positively correlated with confidence ratings for Hits of colored background items, p = .047.

In terms of MRT, it was positively correlated with hits of items with neutral background, p = .043, while it was negatively correlated with false alarms of critical lure items from colored background lists, p = .034. For reaction times, MRT was negatively correlated with reaction times of hits, p = .015, for both neutral background items, p = .015 and colored background items, p = .026. In terms of confidence ratings, MRT correlated positively with confidence ratings of hits, p = .008, which can be attributed to hits of colored background items only, $p < .001$

Table 3.1 Correlations between object recognition and individual differences in imagery

		MRT			VVIQ		
		Recognition	Reaction Time	Confidence Rating	Recognition	Reaction Time	Confidence Rating
Hit	Overall	.198	-.313*	.337*	.334*	.076	.223
	Neutral	.262*	-.312*	.147	.221	.060	.157
	Color	.096	-.288*	.446*	.370**	.065	.257*
False Alarm (Critical Lure)	Overall	-.171	.164	.025	-.080	.172	.117
	Neutral	-.013	.020	.142	-.051	.021	.093
	Color	-.274*	.079	-.080	-.089	.143	.028
False Alarm (Unr.)		-.221	-.092	-.041	-.119	.020	.162

Note: * $p <. 05$, ** $p < .01$

3.2.4 Strategy Questions for Object Recognition

The relationship between the strategy questions and recognition performance was assessed using Pearson's correlational analysis. It should be noted that Question 5 ("It was more difficult to memorize the object when the background was constantly changing.") was only asked to participants in the heterogeneous condition, and it did not correlate with any of the recognition measures.

Question 1; "I indicated having seen an object (said yes) because I actually remembered exactly seeing it." was negatively correlated with critical lure false alarms, $p < .001$, for both neutral background condition, p = .002 and colored background condition, $p < .001$. It was also negatively correlated with unrelated false alarms, p = .004. In terms of confidence ratings, Question 1 was correlated with confidence ratings of hits, p = .037.

Question 2; "I indicated having seen an object (said yes) because I remembered its category, but not the exact appearance." showed correlation with only confidence ratings among the recognition measures. It was negatively correlated with confidence ratings of hits, p = .028, which was observed due to the colored background condition, p = .012.

Question 3; "I indicated having seen an object (said yes) because it felt familiar." was correlated with critical lure false alarms, p = .022, which was an effect attributable to the colored background condition, p = .010. In terms of confidence ratings, Question 3 was negatively correlated with confidence ratings of hits, p = .026; only for neutral background items, p = .010.

Question 4; "I paid more attention to the object than the background." was negatively correlated with critical lure false alarms, p = .001; for both neutral background condition, p = .001 and colored background condition, p = .011. Moreover, this question was negatively correlated with unrelated false alarms, p = .003. For reaction times, Question 4 correlated with reaction times of critical lure false alarms, p = .022, only for the neutral background condition, p = .038.

Question 6; "While memorizing the objects, I named the objects to myself." was correlated with hits, p = .033, only for the colored background condition, p = .034. In terms of reaction times, Question 6 correlated positively with reaction times of critical lure false alarms from the colored background condition, p = .040.

Question 7; "While memorizing the objects, I paid attention to its visual details (color, shape, etc.)." was correlated with confidence ratings of unrelated false alarms, p = .011.

Correlations between strategy questions and recognition performance can be found in Table 3.2. Tables for correlations between strategy questions and RT and confidence rating are presented in Appendix B.

Table 3.2 Correlations between object recognition and strategy questions

	Hit			False Alarm (Critical Lure)			False Alarm (Unrelated)
	Overall	Neutral	Color	Overall	Neutral	Color	
Q1	-.051	-.010	-.079	-.543**	-.397**	-.557**	-.369**
Q2	-.056	-.044	-.056	.248	.242	.201	.092
Q3	-.169	-.173	-.130	.296*	.186	.330**	.104
Q4	-.132	-.194	-.044	-.417**	-.420**	-.326*	-.383**
Q5	-.248	-.203	-.237	-.060	-.021	-.079	-.218
Q6	.275*	.215	.274*	-.112	-.183	-.025	-.212
Q7	-.018	.021	-.051	-.069	-.088	.038	-.106

Note: * $p <. 05$, ** $p < .01$

3.3 Conclusions

In Study 2, we were able to replicate the false memory effect obtained in Study 1, thus supporting the hypothesis H1.1 Participants recognized studied items more than both critical lures and unrelated items, and they were faster and more confident in doing so. Also, false recognition of critical lure items was higher than unrelated items. Yes responses given to critical lure items had higher confidence ratings than that of unrelated items, but they were not given faster.

In terms of the effect of background color heterogeneity (i.e. perceptual distinctiveness), we were not able to support our hypothesis (H3.1), since there was no effect of the heterogeneity of the colored background on false memory rates. Mean reaction times or confidence ratings also did not differ between homogeneous and heterogeneous conditions. Also, there was no difference between the neutral versus color or homogeneous versus heterogeneous conditions in terms of signal detection indices *d'* and lambda (λ).

We compared the baseline condition (no background color) with colored background as a within-subject variable. We did not observe any effect of colored background on false memory rates (H3.2). No differences were observed for reaction times, confidence ratings and signal detection indices as well.

In Study 2, in terms of individual difference measures, we observed that higher imagery abilities were associated with better memory performance and faster RT for accurate responses. Unlike Study 1, they were not associated with false memory. In terms of strategy questions, similarly to Study 1, relying on encoded memory information (Q1) was related to better memory performance while relying on familiarity (Q3) was related to worse memory performance, as observed as higher false alarms to critical lures.

4. GENERAL DISCUSSION

4.1 Replication of the False Memory Effect

In both studies, we were able to observe the expected false memory effect using real-life objects within the CAP. The participants falsely recognized critical lure objects significantly more often than unrelated items. This finding was also observed for reaction time and confidence ratings. We observed false memory rates (48% in Study 1 and 42% in Study 2) that align with the existing literature where visual presentation modality was used, and false memory rates of 30-60% for DRM paradigm (Arndt and Reder 2003; Benmergui, McKelvie, and Standing 2017; Schacter, Israel, and Racine 1999) and 29-47% for CAP (Howe 2008; Intons-Peterson et al. 1999; Loprinzi 2023; Seamon et al. 2000) were observed. The false memory rates were significantly below the hit rates, which is common for studies that used visual stimuli (Seamon et al. 2000; Verma and Kashyap 2020).

Our results provided evidence of a false memory effect and extended previous findings using real pictures of objects. In the literature, when researchers aim to use real-life pictures, it is required that they create the stimuli set themselves, mostly by selecting items from thematic or categorical lists that can be presented pictorially (Wang et al. 2018). Our contribution is creating a stimuli set freely shared with the scientific community at OSF that can be further used in visual false memory studies.

Study 1 also showed a false memory effect using multiple versions of background pictures of different contexts (garden, living room, seaside etc) with rich pictorial details. This extends the finding of Koutstaal and Schacter (1997) where they used stimuli lists that include several pictures of the same objects (toys, cars, dogs etc). Differently, we observed the false memory effect using pictures of contexts with rich details, but not isolated objects.

4.2 The Effect of Conceptual Distinctiveness

In Study 1, we investigated the effect of conceptual distinctiveness on false memories by visually manipulating the semantic context that the studied items were encoded in.

Contrary to our expectations, we did not observe the reducing effect of conceptual distinctiveness on false memories, as there were no significant differences in false memory rates between the congruent and incongruent conditions when using the full set of the items. We also did not observe a difference in reaction times, but confidence ratings were higher for the congruent condition. Moreover, we observed a positive rather than negative effect of distinctiveness on false memory for background recognition performance, with higher false recognition for incongruent backgrounds. Overall, even though there was no significant difference between the false memory rates in two conditions, the pattern of results indicated that congruent stimuli can be processed more efficiently, reducing the false memory rate for congruent backgrounds and increasing the confidence for objects from the congruent condition.

No difference finding for object recognition was previously explained by Distinctiveness Heuristics. That is, when some items have higher distinctiveness in a study with within-subjects design, individuals may rely on a more stringent criterion applied to all test items during the retrieval (Schacter and Wiseman 2006; Schacter, Israel, and Racine 1999). This leads participants to reject any item that cannot pass through their decision criterion. In our case, the more conservative approach during the recognition test might have been applied to critical lure items from congruent condition as well, leading to similar rates of false recognition for congruent and incongruent lists (Roediger and Gallo 2017).

No difference in false memory rates between distinctive versus non-distinctive conditions due to within-subjects and distinctiveness heuristic could be explained by the retrieval-based theory. However, previous studies suggested that both encoding- and retrieval-based processes play a role in false memory formation (Hege and Dodson 2004; Shimane and Itoh 2022). We observed higher sensitivity index *(d')* for the congruent (non-distinctive) condition. This finding suggests that participants encoded more information and had higher discriminability in the congruent condition, when compared to incongruent condition. According to the impoverished relational encoding account (Hege and Dodson 2004), presenting items distinctively reduces the encoding of relational information that is related to a presented list. Overall, it can be concluded that both encoding-based and retrieval-based processes were effective.

Furthermore, when we used the most and least congruent lists for object recognition analysis, we observed an effect of distinctiveness in opposite to the expected direction. False memory rates were higher in the incongruent condition than in the congruent condition, and higher sensitivity *(d')* and higher lambda λ value was observed for the congruent condition. Here it should be noted that the same pattern of results was observed for background recognition using all lists. Additionally, higher sensitivity *(d')* and higher confidence for the congruent condition was observed when using all lists in the analysis. Considering all these findings, we conclude that the conceptual distinctiveness manipulation might have disrupted the memory instead of enhancing it, and led to higher false memory rates. The same pattern was observed by Frank et al. (2018) where they argued that incongruent items increase effort and attention during study, however this does not lead to enhanced recognition performance or lower false memories. Therefore, the authors concluded that semantic incongruence might have a detrimental effect on memory performance, contrary to our hypothesis and other studies that found distinctiveness reducing the effect of false memories (Arndt and Reder 2003; Huff and Aschenbrenner 2018). An argument that is similar to that of Frank et al. (2018) came from Ortiz-Tudela et al. (2017) where the authors suggested that the incongruency of the semantic context may disrupt remembering of the learned information. However, these studies did not manipulate conceptual distinctiveness by manipulating the congruency of backgrounds in a false memory paradigm. To our knowledge, this was only examined by Howe (2008) with children, and no effect of semantic context was found. Possibly, adults are more sensitive to semantic conflict and therefore more affected by incongruency.

4.3 The Effect of Perceptual Distinctiveness

In Study 2, we investigated the effect of perceptual distinctiveness on false memories by visually manipulating the background color. In terms of color heterogeneity of the backgrounds, we did not observe a difference in recognition performance between the homogeneous and heterogeneous conditions. Moreover, there were no differences between the two conditions in terms of sensitivity index *(d')* and lambda (λ). Thus, the finding of Howe (2008) that heterogeneity of background colors led to lower false memories (explained by the increased item-specific processing) was not replicated in our sample of young adults. Since Howe's study was conducted with children, it is possible that children were more sensitive to perceptual (e.g., color) rather than semantic properties of stimuli, and our null finding can be due to the age of our participants. This should be further examined by using other manipulations of visual saliency in different age groups. We did not observe the effect of backgrounds' color when compared to the baseline condition, as no differences in false memories were present between the two conditions. We also did not observe a difference in reaction times and confidence ratings. Additionally, we did not observe a difference for sensitivity index *(d')* and lambda (λ) between the neutral and colored background conditions, indicating that background color did not enhance the memory encoding nor it led to a change in the decision criterion of the individuals. This finding could be attributed to Distinctiveness Heuristics (Schacter and Wiseman 2006; Schacter, Israel, and Racine 1999) for Study 2 as well. However, different from Study 1, in Study 2 we did not observe a difference in sensitivity *(d')* indices of the two conditions. Therefore, we argue that perceptual distinctiveness did not lead to similar effects with conceptual distinctiveness in our case.

Taken together with Study 1, we argue that in our sample of young adults, perceptually manipulating distinctiveness was not effective, and did not disrupt the relational encoding process, while conceptual distinctiveness was able to do so, indicated by the sensitivity index *(d')*. However, this did not lead to a decrease in false memory rates, which we attributed to a possible detrimental effect of incongruency (Frank et al. 2018). This finding is the opposite of Howe (2008), which might be observed because our sample consisted of young adults and Howe (2008) used a sample of children. In the literature it has been suggested that adults are better able to draw out the gist of the meaning from the given context (Ghetti, Qin, and Goodman 2002), while children rely more on perceptual, verbatim information (Brainerd, Reyna, and Forrest 2002; Metzger et al. 2008). Thus, it can be the case that perceptual distinctiveness enhanced memory for children (Howe 2008) while conceptual distinctiveness was more effective in the memory encoding process for young adults in our study. This argument is also supported by Konkle et al. (2010) where they found that conceptual but not perceptual distinctiveness led to memory enhancement for visual information.

5. CONCLUDING REMARKS, LIMITATIONS AND FUTURE DIRECTIONS

Overall, in this study we provided evidence that prominent false memories can be observed using real-life objects and background pictures within a Category Associates Procedure. We also showed that although it did not have an effect on false memory rates, conceptual distinctiveness leads to changes in memory encoding while perceptual distinctiveness does not show such an effect in a sample of young adults. Considering the relevance of false memory to daily life, the importance of semantic context on the formation and reduction of false memories provides an essential direction for further research.

Our research has several limitations and discussing them might be necessary to provide further directions. Firstly, although we tested our newly created stimuli set in two studies and observed a false memory effect comparable with previous literature, while creating them we did not test whether the pictures we are using are completely depicting the intended objects and backgrounds. As we provide the stimuli set in OSF, we also provide a further direction to test the items in terms of compatibility. Moreover, as we observed high rates of correct recognition in both studies, it can be the case that we observed a ceiling effect due to the low task difficulty. This can be improved by adding new categories and increasing the item number during study and test.

For Study 1, we analyzed a subset of categories with the most and the least congruence ratings. However, it should be noted that this was an exploratory analysis, and these interpretations should be evaluated with caution. Analyzing a subset of lists might give incomplete results, since the items that were not included in this analysis still had an effect on items that were included.

Lastly, as it is a common problem in literature as we mentioned earlier, the conceptual definition of distinctiveness and whether it leads to changes in information processing is still not directly clear. Even though we performed the manipulations in both studies by relying on previous literature, we consider improving the conceptual and operational definition of distinctiveness to be a fruitful further direction for research. Further studies may benefit from neuroimaging methods to investigate whether distinctiveness manipulations actually lead to differences in attentional allocation and saliency and change the way the information is processed during encoding and retrieval.

BIBLIOGRAPHY

Alakbarova, Durna, Jason L. Hicks, and B. Hunter Ball. 2021. "The Influence of Semantic Context on False Memories." *Memory Cognition* 49(8): 1555–1567.

Arndt, Jason, and Lynne M. Reder. 2003. "The Effect of Distinctive Visual Information on False Recognition." *Journal of Memory and Language* 48(1): 1–15.

Battig, William F., and William E. Montague. 1969. "Category Norms of Verbal Items in 56 Categories: A Replication and Extension of the Connecticut Category Norms." *Journal of Experimental Psychology* 80(3, Pt.2): 1–46.

Beauchamp, Heather M. 2002. "Aural, Visual, and Pictorial Stimulus Formats in False Recall." *Psychological Reports* 91(3): 941–951.

Benmergui, Sarah R., Stuart J. McKelvie, and Lionel G. Standing. 2017. "Beneficial Effect of Pictures on False Memory in the DRMRS Procedure." *Current Psychology* 36(1): 136–146.

Beran, Michael J., Brielle T. James, Kristin French, Elizabeth L. Haseltine, and Heather M. Kleider-Offutt. 2023. "Assessing Aphantasia Prevalence and the Relation of Self-Reported Imagery Abilities and Memory Task Performance." *Consciousness and Cognition* 113(August): 103548.

Bjorklund, David F. 2005. *Children's Thinking: Cognitive Development and Individual Differences.* 4th ed. Australia; Belmont, CA: Thomson/Wadsworth.

Blanca, María J., Jaume Arnau, F. Javier García-Castro, Rafael Alarcón, and Roser Bono. 2023. "Repeated Measures ANOVA and Adjusted F-Tests When Sphericity Is Violated: Which Procedure Is Best?" *Frontiers in Psychology* 14(August): 1192453.

Brainerd, C. J., V. F. Reyna, and R. Kneer. 1995. "False-Recognition Reversal: When Similarity Is Distinctive." *Journal of Memory and Language* 34(2): 157–185.

Brainerd, C. J., V. F. Reyna, and T. J. Forrest. 2002. "Are Young Children Susceptible to the False–Memory Illusion?" *Child Development* 73(5): 1363–1377.

Coane, Jennifer H., Mark J. Huff, and Keith A. Hutchison. 2016. "The Ironic Effect of Guessing: Increased False Memory for Mediated Lists in Younger and Older Adults." *Aging, Neuropsychology, and Cognition* 23(3): 282–303.

Deese, James. 1959. "On the Prediction of Occurrence of Particular Verbal Intrusions in Immediate Recall." *Journal of Experimental Psychology* 58(1): 17–22.

Dewhurst, Stephen A., Rhian C. Pursglove, and Charlie Lewis. 2007. "Story Contexts Increase Susceptibility to the DRM Illusion in 5-year-olds." *Developmental Science* 10(3): 374–378.

Dobson, Matthew, and Roslyn Markham. 1993. "Imagery Ability and Source Monitoring: Implications for Eyewitness Memory." *British Journal of Psychology* 84(1): 111–118.

Einstein, Gilles O., and R. Reed Hunt. 1980. "Levels of Processing and Organization: Additive Effects of Individual-Item and Relational Processing." *Journal of Experimental Psychology: Human Learning and Memory* 6(5): 588–598.

Foster, Jeffrey L., Thomas Huthwaite, Julia A. Yesberg, Maryanne Garry, and Elizabeth F. Loftus. 2012. "Repetition, Not Number of Sources, Increases Both Susceptibility to Misinformation and Confidence in the Accuracy of Eyewitnesses." *Acta Psychologica* 139(2): 320–326.

Frank, Darya, Daniela Montaldi, Bianca Wittmann, and Deborah Talmi. 2018. "Beneficial and Detrimental Effects of Schema Incongruence on Memory for Contextual Events." *Learning Memory* 25(8): 352–360.

Gallo, David. 2013. *Associative Illusions of Memory.* 0 ed. Psychology Press.

Gallo, David A. 2010. "False Memories and Fantastic Beliefs: 15 Years of the DRM Illusion." *Memory Cognition* 38(7): 833–848.

Gallo, David A., and Henry L. III Roediger. 2002. "Variability among Word Lists in Eliciting Memory Illusions: Evidence for Associative Activation and Monitoring." *Journal of Memory and Language* 47(3): 469–497.

Gallo, David A., Jonathan A. Weiss, and Daniel L. Schacter. 2004. "Reducing False Recognition with Criterial Recollection Tests: Distinctiveness Heuristic versus Criterion Shifts." *Journal of Memory and Language* 51(3): 473–493.

Gallo, David A., Sivan C. Cotel, Christopher D. Moore, and Daniel L. Schacter. 2007. "Aging Can Spare Recollection-Based Retrieval Monitoring: The Importance of Event Distinctiveness." *Psychology and Aging* 22(1): 209–213.

Ghetti, Simona, Jianjian Qin, and Gail S. Goodman. 2002. "False Memories in Children and Adults: Age, Distinctiveness, and Subjective Experience." *Developmental Psychology* 38(5): 705–718.

Goldstein, B. E. 2014. *Cognitive Psychology: Connecting Mind, Research and Everyday Experience.* 4th ed. Cengage Learning.

Gonsalves, Brian, Paul J. Reber, Darren R. Gitelman, Todd B. Parrish, M.-Marsel Mesulam, and Ken A. Paller. 2004. "Neural Evidence That Vivid Imagining Can Lead to False Remembering." *Psychological Science* 15(10): 655–660.

Green, David M., and John A. Swets. 1966. *Signal Detection Theory and Psychophysics.* John Wiley.

Gunter, Raymond W., Glen E. Bodner, and Tanjeem Azad. 2007. "Generation and Mnemonic Encoding Induce a Mirror Effect in the DRM Paradigm." *Memory Cognition* 35(5): 1083–1092.

Hege, Amanda C. G., and Chad S. Dodson. 2004. "Why Distinctive Information Reduces False Memories: Evidence for Both Impoverished Relational-Encoding and Distinctiveness Heuristic Accounts." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(4): 787–795.

Hintzman, Douglas L. 1988. "Judgments of Frequency and Recognition Memory in a Multiple-Trace Memory Model." *Psychological Review* 95(4): 528–551.

Howe, Mark L. 2008. "Visual Distinctiveness and the Development of Children's False Memories." *Child Development* 79(1): 65–79.

Howe, Mark L., and Samantha Wilkinson. 2011. "Using Story Contexts to Bias Children's True and False Memories." *Journal of Experimental Child Psychology* 108(1): 77–95.

Huff, Mark J., and Andrew J. Aschenbrenner. 2018. "Item-Specific Processing Reduces False Recognition in Older and Younger Adults: Separating Encoding and Retrieval Using Signal Detection and the Diffusion Model." *Memory Cognition* 46(8): 1287–1301.

Huff, Mark J., and Glen E. Bodner. 2013. "When Does Memory Monitoring Succeed versus Fail? Comparing Item-Specific and Relational Encoding in the DRM Paradigm." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(4): 1246–1256.

Huff, Mark J., Glen E. Bodner, and Jonathan M. Fawcett. 2015. "Effects of Distinctive Encoding on Correct and False Memory: A Meta-Analytic Review of Costs and Benefits and Their Origins in the DRM Paradigm." *Psychonomic Bulletin Review* 22(2): 349–365.

Hunt, R. Reed, and James B. Worthen. 2006. *Distinctiveness and Memory.* Oxford University Press.

Intons-Peterson, M. J., Paola Rocchi, Tara West, Kimberly McLellan, and Amy Hackney. 1999. "Age, Testing at Preferred or Nonpreferred Times (Testing Optimality), and False Memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25(1): 23–40.

Israel, Lana, and Daniel L. Schacter. 1997. "Pictorial Encoding Reduces False Recognition of Semantic Associates." *Psychonomic Bulletin Review* 4(4): 577–581.

Kaplan, Robin L., Ilse Van Damme, Linda J. Levine, and Elizabeth F. Loftus. 2016. "Emotion and False Memory." *Emotion Review* 8(1): 8–13.

Keogh, Rebecca, and Joel Pearson. 2011. "Mental Imagery and Visual Working Memory." *PLoS ONE* 6(12): e29221.

Konkle, Talia, Timothy F. Brady, George A. Alvarez, and Aude Oliva. 2010. "Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects." *Journal of Experimental Psychology: General* 139(3): 558–578.

Koutstaal, Wilma, and Daniel L. Schacter. 1997. "Gist-Based False Recognition of Pictures in Older and Younger Adults." *Journal of Memory and Language* 37(4): 555–583.

Koutstaal, Wilma, Daniel L. Schacter, and Carolyn Brenner. 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.

Lampinen, James M., Jeffrey S. Neuschatz, and David G. Payne. 1999. "Source Attributions and False Memories: A Test of the Demand Characteristics Account." *Psychonomic Bulletin Review* 6(1): 130–135.

Loftus, Elizabeth F. 1997. "Creating False Memories." *Scientific American* 277(3): 70–75.

Loftus, Elizabeth F., and Jacqueline E. Pickrell. 1995. "The Formation of False Memories." *Psychiatric Annals* 25(12): 720–725.

Loprinzi, Paul D. 2023. "Effects of Pictorial and Imagery Encoding on False Memories.".

Macmillan, Neil A., and C. Douglas Creelman. 2004. *Detection Theory.* 0 ed. Psychology Press.

Marks, David F. 1973. "Visual Imagery Differences in the Recall of Pictures." *British Journal of Psychology* 64(1): 17–24.

Marsh, E.J., A.N. Eslick, and L.K. Fazio. 2008. "False Memories." In *Learning and Memory: A Comprehensive Reference*. Elsevier pp. 221–238.

McCabe, David P., and Anderson D. Smith. 2006. "The Distinctiveness Heuristic in False Recognition and False Recall." *Memory* 14(5): 570–583.

McKelvie, S. J. 1995. *Vividness of Visual Imagery: Measurement, Nature, Function Dynamics.* Brandon House.

Metzger, Richard L., Amye R. Warren, Jill T. Shelton, Jodi Price, Andrea W. Reed, and Danny Williams. 2008. "Do Children 'DRM' like Adults? False Memory Production in Children." *Developmental Psychology* 44(1): 169–181.

Michelon, Pascale, and Abraham Z. Snyder. 2006. "Neural Correlates of Incongruity." In *Distinctiveness and Memory*, ed. R. Reed Hunt, and James B. Worthen. Oxford University Press pp. 361–380.

Misirlisoy, Mine. 2004. Effects of Associative Processes on False Memory: Evidence From Converging Associates and Category Associates Procedures. Master's thesis.

Mowbray, Fabrice I., Susan M. Fox-Wasylyshyn, and Maher M. El-Masri. 2019. "Univariate Outliers: A Conceptual Overview for the Nurse Researcher." *Canadian Journal of Nursing Research* 51(1): 31–37.

Oliver, Merrin Creath, Rebecca Brooke Bays, and Karen M. Zabrucky. 2016. "False Memories and the DRM Paradigm: Effects of Imagery, List, and Test Type." *The Journal of General Psychology* 143(1): 33–48.

Ortiz-Tudela, Javier, Bruce Milliken, Fabiano Botta, Mitchell LaPointe, and Juan Lupiañez. 2017. "A Cow on the Prairie vs. a Cow on the Street: Long-Term Consequences of Semantic Conflict on Episodic Encoding." *Psychological Research* 81(6): 1264–1275.

Otgaar, Henry, Antonietta Curci, Ivan Mangiulli, Fabiana Battista, Elisa Rizzotti, and Giuseppe Sartori. 2022. "A Court Ruled Case on Therapy-induced False Memories." *Journal of Forensic Sciences* 67(5): 2122–2129.

Pardilla-Delgado, Enmanuelle, and Jessica D. Payne. 2017. "The Deese-Roediger-McDermott (DRM) Task: A Simple Cognitive Paradigm to Investigate False Memories in the Laboratory." *Journal of Visualized Experiments* (119): 54793.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. "PsychoPy2: Experiments in Behavior Made Easy." *Behavior Research Methods* 51(1): 195–203.

Peters, M., B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson. 1995. "A Redrawn Vandenberg and Kuse Mental Rotations Test - Different Versions and Factors That Affect Performance." *Brain and Cognition* 28(1): 39–58.

Reisberg, Daniel, and Sharon Leak. 1987. "Visual Imagery and Memory for Appearance: Does Clark Gable or George C. Scott Have Bushier Eyebrows?" *Canadian Journal of Psychology / Revue Canadienne de Psychologie* 41(4): 521–526.

Reyna, V.F., and C.J. Brainerd. 1995. "Fuzzy-Trace Theory: An Interim Synthesis." *Learning and Individual Differences* 7(1): 1–75.

Robin, Frédérique. 2010. "Imagery and Memory Illusions." *Phenomenology and the Cognitive Sciences* 9(2): 253–262.

Roebers, Claudia M., and Wolfgang Schneider. 2000. "The Impact of Misleading Questions on Eyewitness Memory in Children and Adults." *Applied Cognitive Psychology* 14(6): 509–526.

Roediger, Henry L. 1996. "Memory Illusions." *Journal of Memory and Language* 35(2): 76–100.

Roediger, Henry L., and David A. Gallo. 2017. *Associative Memory Illusions.* 3rd ed. Routledge.

Roediger, Henry L., and Kathleen B. McDermott. 1995. "Creating False Memories: Remembering Words Not Presented in Lists." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(4): 803–814.

Roediger, Henry L., Jason M. Watson, Kathleen B. McDermott, and David A. Gallo. 2001. "Factors That Determine False Recall: A Multiple Regression Analysis." *Psychonomic Bulletin Review* 8(3): 385–407.

Schacter, Daniel L. 2012. "Constructive Memory: Past and Future." *Dialogues in Clinical Neuroscience* 14(1): 7–18.

Schacter, Daniel L., and Amy L. Wiseman. 2006. "Reducing Memory Errors: The Distinctiveness Heuristic." In *Distinctiveness and Memory*, ed. R. Reed Hunt, and James B. Worthen. Oxford University Press p. 89–107.

Schacter, Daniel L., Daniel L. Cendan, Chad S. Dodson, and Erin R. Clifford. 2001. "Retrieval Conditions and False Recognition: Testing the Distinctiveness Heuristic." *Psychonomic Bulletin Review* 8(4): 827–833.

Schacter, Daniel L., Lana Israel, and Carrie Racine. 1999. "Suppressing False Recognition in Younger and Older Adults: The Distinctiveness Heuristic." *Journal of Memory and Language* 40(1): 1–24.

Schmidt, Stephen R. 1991. "Can We Have a Distinctive Theory of Memory?" *Memory Cognition* 19(6): 523–542.

Seamon, John G., Chun R. Luo, Sarah E. Schlegel, Sara E. Greene, and Audrey B. Goldenberg. 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(1): 120–146.

Shimane, Daisuke, and Yuji Itoh. 2022. "Picture Encoding Suppresses False Memory in the Deese-Roediger-McDermott Paradigm through Both Distinctiveness Heuristic and Impoverished Relational Encoding.".

Smith, Rebekah E., and R. Reed Hunt. 1998. "Presentation Modality Affects False Memory." *Psychonomic Bulletin Review* 5(4): 710–715.

Smith, Rebekah E., and R. Reed Hunt. 2020. "When Do Pictures Reduce False Memory?" *Memory Cognition* 48(4): 623–644.

Smith, Rebekah E., R. Reed Hunt, and Kathryn R. Dunlap. 2015. "Why Do Pictures, but Not Visual Words, Reduce Older Adults' False Memories?" *Psychology and Aging* 30(3): 647–655.

Stock, Adobe. n.d. "Stock Photos, Royalty-Free Images, Graphics, Vectors Videos.".

Stock images, Royalty-Free images, illustrations, vectors and videos. N.d. iStockPhoto.com.

Sugrue, Katrina, and Harlene Hayne. 2006. "False Memories Produced by Children and Adults in the DRM Paradigm." *Applied Cognitive Psychology* 20(5): 625–631.

Systems, Sona. n.d. "Sona Systems: Cloud-based Participant Management Software.".

Thomas, Ayanna K., and Mitchell S. Sommers. 2005. "Attention to Item-Specific Processing Eliminates Age Effects in False Memories." *Journal of Memory and Language* 52(1): 71–86.

Van Overschelde, James P., Katherine A. Rawson, and John Dunlosky. 2004. "Category Norms: An Updated and Expanded Version of the Battig and Montague (1969) Norms." *Journal of Memory and Language* 50(3): 289–335.

Vandenberg, Steven G., and Allan R. Kuse. 1978. “Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization.” *Perceptual and Motor Skills* 47(2): 599–604.

Verma, Kedarmal, and Naveen Kashyap. 2020. “False Memories for Semantic and Category Associates: Comparing Retrieval Strategies and Retention Interval.” *Psychological Thought* 13(2): 322–348.

Wang, Jianqin, Henry Otgaar, Mark L. Howe, Felix Lippe, and Tom Smeets. 2018. “The Nature and Consequences of False Memories for Visual Stimuli.” *Journal of Memory and Language* 101(August): 124–135.

Wickens, Thomas D. 2001. *Elementary Signal Detection Theory*. Oxford University Press.

APPENDIX A

ADDITIONAL TABLES FOR STUDY 1

Table A.1 Correlations between strategy questions and RT for object recognition

	Reaction Time						
	Hit			False Alarm (Critical Lure)			False Alarm (Unrelated)
	Overall	Congruent	Incongruent	Overall	Congruent	Incongruent	
Q1	-.078	-.041	-.110	.009	.004	.035	-.204
Q2	.161	.193	.108	.109	.056	.084	.058
Q3	.247	.268	.194	.074	-.001	.091	.188
Q4	.276*	.226	.294*	.121	-.020	.191	.112
Q5	.261	.256	.236	-.013	.017	-.001	.365*
Q6	-.121	-.071	-.161	.030	.016	.028	-.129
Q7	.125	.121	.108	-.119	.033	-.160	-.162

Note: * $p <. 05$, ** $p < .01$

Table A.2 Correlations between strategy questions and confidence ratings for object recognition

	Confidence Rating						
	Hit			False Alarm (Critical Lure)			False Alarm (Unrelated)
	Overall	Congruent	Incongruent	Overall	Congruent	Incongruent	
Q1	.360**	.277*	.412**	.065	.080	-.023	.168
Q2	-.030	.046	-.105	.087	.238	-.038	-.112
Q3	-.278*	-.230	-.302*	-.254	-.133	-.302*	-.016
Q4	-.039	-.019	-.051	-.332*	-.114	-.393**	-.025
Q5	-.109	-.095	-.117	-.003	-.092	-.028	-.077
Q6	.169	.091	.218	-.012	-.004	.012	-.137
Q7	.222	.187	.222	.038	.001	-.107	.016

Note: * $p <. 05$, ** $p < .01$

Table A.3 Correlations between strategy questions and RT for background recognition

	Reaction Time				
	Hit		False Alarm (Critical Lure)		False Alarm (Unrelated)
	Congruent Background	Incongruent Background	Congruent Lure Background	Incongruent Lure Background	
Q1	-.160	-.275*	-.096	-.098	.102
Q2	.151	.073	.055	.134	-.027
Q3	-.039	-.057	-.117	-.045	.088
Q4	.223	.150	.255	.141	.008
Q5	-.035	.012	.165	-.020	.300
Q6	-.032	-.136	.132	-.064	.065
Q7	.003	.001	-.027	-.077	.084

Note: * $p <. 05$, ** $p < .01$

Table A.4 Correlations between strategy questions and confidence ratings for background recognition

	Confidence Rating				
	Hit		False Alarm (Critical Lure)		False Alarm (Unrelated)
	Congruent Background	Incongruent Background	Congruent Lure Background	Incongruent Lure Background	
Q1	.270	.202	.075	-.013	-.155
Q2	.091	.039	-.035	.019	.187
Q3	-.156	-.156	-.107	-.136	-.030
Q4	-.140	-.007	-.197	-.211	-.085
Q5	-.094	-.020	-.067	.034	-.030
Q6	-.028	.060	-.075	-.083	.034
Q7	.267	.099	.115	.161	.051

Note: * $p <. 05$, ** $p < .01$

APPENDIX B

ADDITIONAL TABLES FOR STUDY 2

Table B.1 Correlations between strategy questions and RT for object recognition

Reaction Time							
	Hit			False Alarm (Critical Lure)			False Alarm (Unrelated)
	Overall	Neutral	Color	Overall	Neutral	Color	
Q1	-.048	-.044	-.027	.092	.074	-.010	.029
Q2	.092	-.012	.149	-.206	-.144	-.138	-.279
Q3	.196	.148	.204	-.227	-.071	-.241	-.017
Q4	.091	.096	.062	.295*	.271*	.081	.238
Q5	.048	-.041	.176	-.113	.172	-.246	-.150
Q6	.010	.064	-.073	.102	-.048	.271*	.044
Q7	.094	.038	.131	-.012	-.002	.035	.129

Note: * $p <. 05$, ** $p < .01$

Table B.2 Correlations between strategy questions and confidence ratings for object recognition

	Confidence Rating						
	Hit			False Alarm (Critical Lure)			False Alarm (Unrelated)
	Overall	Neutral	Color	Overall	Neutral	Color	
Q1	.270*	.223	.254	.040	.024	.072	.133
Q2	-.283*	-.168	-.322*	.059	-.018	.046	-.103
Q3	-.288*	-.330*	-.205	-.093	-.080	-.113	.006
Q4	.024	.006	.043	-.110	-.152	-.034	-.051
Q5	-.023	.030	-.087	-.058	-.098	-.007	.054
Q6	.154	.153	.126	.045	.068	-.034	-.010
Q7	.123	.155	.071	.162	-.030	.195	.381*

Note: * $p <. 05$, ** $p < .01$