Confidence in Memory's Accuracy

Phyllis Margaret Windsor

University of Stirling

Submitted as a requirement for the degree of PhD

August 2022

Declaration.

This thesis has been composed by the candidate, the work is the candidate's own work, and has not been previously included in a thesis or dissertation submitted to this or any other institution for a degree, diploma, or other qualifications.

Acknowledgements.

Drafting a thesis, and designing and carrying out the laboratory experiments, is never an easy task, but one which was greatly helped by the encouragement and advice of my supervisors, Prof David Donaldson (now in St Andrews) and Dr Benjamin Dering; together with Dr Jamie Murray (now in Glasgow) who assisted me with setting up the programs for the experiments on E-prime. I could not have completed this work without the help and assistance of the undergraduates at Stirling University, who participated in the experiments and followed the instructions. Finally, I acknowledge the assistance provided by the University, who enabled me to present an early version of the work as a poster for the New Orleans Annual Meeting of the Psychonomic Society in October 2018. I also acknowledge the help provided by the University's Disability Support Team.

Dedication

This thesis is dedicated to my husband Stephen Windsor, without whose unfailing assistance, patience, and support I would not have made any attempt to resume my studies following a major post-operative stroke in January 2019, shortly after completing year one of my PhD.

Table of contents

List of Figures

List of Tables

Abstract.

Despite accounts of a consistent and strong association between confidence and memory accuracy, they can be shown to dissociate. That belief in the truth of a memory may be influenced more by our metacognitive judgments of the memory’s quality (vividness or distinctiveness) than its factual accuracy was tested using real-world colour natural scenes photographs as retrieval cues in a continuous-response source memory task. Importantly, the correlation between memory quality and confidence was very large. Reducing the visual vividness of the images did not affect this very large correlation. Displaying pictures in a homogenous list arrangement (where all pictures represented the same natural scene category) reduced performance and memory accuracy but did not disrupt the very large correlation between distinctiveness and confidence. The correlation between accuracy and confidence remained moderate to large throughout. The relationship between accuracy and confidence was moderated by the vividness or distinctiveness of the memory, increasing confidence in the memory for the same level of accuracy, causing them to dissociate. The association between memory quality and accuracy acted largely indirectly, mediated through confidence. A 2-AFC recognition memory task, also using natural scenes photographs, reproduced the very large correlation between vividness and confidence. Remembering moderated the relationship between vividness and confidence, remembering being associated with higher confidence and higher vividness. Increases in the proportion of correct responses associated with remembering was seen only at highest levels of confidence. The 2-AFC task revealed a very large correlation between familiarity and confidence. Recollection was of less importance for accuracy when familiarity was high and of more importance when familiarity was low. The quality of a memory is associated with confident recollection, even if the detail is incorrect. The findings have implications for theoretical accounts of eyewitness testimony, illusions of memory accuracy, false memories, and dual process theory.

Chapter 1: Introduction

Purpose of the thesis.

I seek to explain why we can be very confident that we remember an event well (i.e., accurately), even when the details are incorrect or lacking. Our autobiographical (episodic, or declarative) long-term memory consists of events personal to us and their associated details. Without this conscious awareness of our memories of the past, we would have no understanding of our place in the world or of our own personal history. But are our believed memories always accurate? That this study is important relates to the many theoretical views and opinions advocating a very strong and inviolate relationship between confidence and memory accuracy, which supports the unquestioning acceptance of highly confident eyewitness testimony and argues against the existence of false memories.

More broadly, our own understanding of our personal experiences imply that memory can be fallible. We may disagree with our partners about the exact details of a holiday we went on, or a film we watched together (when was it, what cinema?). Similarly, when asked to recollect words from presented lists under experimental conditions, it is well-known that participants commonly remember words as old that were not studied but were similar in category. That such memory illusions occur, both in the real world and in the laboratory, led Roediger & McDermott (1995) to the conclusion that "remembering events that never happened can occur quite readily" (p.812).

Research questions and hypotheses.

In asking what factors influence confidence in a memory's accuracy, and if we talk of a strong confidence-accuracy relationship, this implies that confidence should normally be unable to produce illusions of accurate remembering. and inaccurate memories could not normally be reported with high levels of confidence. When listening to the news on the radio or television, reading the newspapers in print or on your iPad, it is common to hear or read that people talk of remembering something vividly, presented as evidence that their memory is correct and cannot be challenged. In such cases, when we remember something and are confident that our memory of it is accurate, if confronted or shown proof that our memory is at fault, we may maintain that we are correct in what we remember, despite contradiction, for how can we be so readily deceived by our own memory? I was so sure… it seemed so vivid… the memory was so distinctive. My thesis aims to investigate why belief in our memory's accuracy can be misplaced. As is set out in detail below, the cognitive psychology literature is less than clear whether other factors (beyond accuracy) are important for confidence in remembering.

My thesis investigates one possible explanation, the hypothesis that we construct our confidence in memory not only on its accuracy, but also on its quality, the richness of the experience of remembering (defined, for example, in terms of vividness or distinctiveness). Alternatively, if my hypothesis is incorrect and memory quality has no association with confidence, then accuracy must drive confidence regardless of the experience of remembering (how rich, vivid, or distinctive the memory is).

Theoretical foundation.

One reason for the theory that memory confidence and accuracy are strongly linked lies in the way memory is conceptualised. For example, the lay person typically assumes that a memory of an event or experience will play back like a video clip and be a true record. Empirical data confirms that this assumption is widespread. In a survey of jurors who responded to whether memory was like a diary entry, a video camera, a library, rooms in a house, or a storehouse, Brewin, Li, Ntarantana, Unsworth & McNeilis, (2019) found no evidence of a difference in belief in these metaphors (on average scoring up to 5.3 on a 0-7

Likert scale[1]), i.e., people generally consider memory to be an unchanging and exact reconstruction of the event experienced. By contrast, however, memory theorists argue against the lay view. For example, Roediger and McDermott (2000) suggested “memories are not recordings, rather re-codings” (p.126).

Experimental findings also clearly show that memories are not an exact copy of an experience. For example, a museum tour paradigm demonstrated that reactivation of personal memories, triggered by viewing a photograph, could enhance or change them when reconsolidated, leading to memory alteration (St. Jacques & Schacter, 2013). Despite the lay assumption that memory and accuracy are tightly bound, it is clearly possible to be highly confident in the absence of memory accuracy. Also of relevance to my thesis is the literature on belief in memory’s accuracy. Almost 30 years ago Chandler (1994), commented that researchers seldom investigated the variables that might influence subjective confidence, and whether these variables also determined accuracy. This remains the case, with little apparent interest in trying to explore relationships between judgments of memory accuracy, i.e., confidence, and judgments of memory quality, e.g., vividness.

Nature of the study.

To begin to investigate the stability of the confidence-accuracy relationship in my thesis, I report a series of studies examining how the quality of memory influences judgments of both confidence and accuracy. Initially, I use a continuous-response source memory task[2] to compare participants making judgments of vividness using natural scenes colour photographs, either all in full colour or with half manipulated to make the images less visually vivid. The same task is used, in separate experiments, to compare participants

[1] A Likert scale uses arbitrary divisions from which respondents choose one option that best aligns with their view, i.e., participants rate their response on an ordinal scale. The 7-option answer has choices from 7 to 1, e.g., representing: very, neutral, or undecided (unsure), to not at all. However, there may be a lot of difference between someone who is undecided and someone who is neutral.

[2] See the literature review Chapter 2 (ii) Experimental tasks of memory retrieval, for full explanation

making judgments of distinctiveness regarding identical natural scenes colour photographs shown in two different contexts (by manipulation of the list arrangement in which the images are viewed). Across these experiments judgments of confidence and accuracy (specifically recollection or source memory) are examined in the context of variability in either vividness or distinctiveness.

In addition to the use of traditional quantitative measures of memory confidence and accuracy, in the present thesis I also employ qualitative measures to assess participants' experiences of remembering. The collection of qualitative data has been largely neglected within confidence-accuracy research. Here it is used to investigate what metrics participants base their judgments of memory quality on. This approach is important in the context of judgments of vividness and distinctiveness, because rather than the experimenter instructing participants what these judgments should mean, I seek to find out what is important to participants when making their judgments.

Significance of the study.

My research, using both a recollection and a recognition memory task, is designed to fill a gap in the literature concerning whether judgments of memory quality can influence the relationship between confidence and memory accuracy. The paucity of existing research in this area is clear from the results of literature searches. For example, entering the search terms vividness and accurate remembering (July 2022) yields no results for an exact match in PsychArticles, and six papers in Web of Science (2000 – 2019): one deals with working memory, and four with emotions and memory - one of which is an abstract. Google Scholar fares little better, reporting two papers from 2004 and 2011, both involving neuroimaging. Other quoted references mostly concern vivid visual imagery or emotional arousal and memory vividness, with a paper linking emotional events to lack of accuracy because of attentional narrowing (Heuer & Reisberg, 1990).

Entering the search terms confidence and vividness and memory accuracy into Web of Science in March 2022 produced a total of 10 published papers (three involved obsessive-compulsive disorders; two were based on flashbulb or emotionally charged memories; three referred to eyewitnesses or face recognition; one was a neuroimaging study and the other concerned football fans). This starkly highlights the gap in the literature, and it is apparent there has been little interest in trying to connect or explore relationships between these judgments.

Similarly, and disappointingly searching for articles on memory quality in PsychArticles only brought up four papers (concerning refugee hearing decisions, schizophrenia, childhood memories, and memories in "clever old people"). To be fair, in Web of Science (July 2022) my search revealed 292 review articles on the topic of memory quality, however, many of these investigate changes in memory with age, cognitive impairment, dementia, mental illness and brain injury. The potential effect of memory quality on confidence and accuracy using judgments of vividness and distinctiveness, rather than negative emotions, certainly merits further investigation.

Chapter 2: Literature Review

For the literature review, I searched Web of Science, PsychINFO, PsychArticles, Journal/Author Name Estimator (https://jane.biosemantics.org) and Google Scholar. From Web of Science in June 2021, the topic of memory in the research area of psychology, yielded 293,254 results from all databases, as far back as an article concerning judgment in discriminating noise – clangs and tones, in 1900 by Whipple, and "The illusion of false recognition. Contribution to the study of pathological conditions of the recognition of memories" by Arnaud in 1901, indicating that fallibility of memory was an issue concerning researchers at the beginning of the 20th century. In the decade 1900-1910, a total of 128 papers on memory are listed by Web of Science; 991 articles were published in the year 1975; rising exponentially to 12,570 in 2020. Refining the literature to include only highly cited papers takes the total down to a more manageable 1,267 articles (415 in the last five years), representing a disappointing 0.43% of the entire body of literature on the psychology of memory. These figures do not include rejected or unrevised papers, or research not submitted for publication because of negative findings or a null result.

Before evaluating the literature relating to my specific research topic, I outline the historical and current literature on memory and memory theory to provide some background to my thesis.

Historical perspective.

To understand the question asked, it is necessary to appraise what we consider memory to be. Zlotnik and Vansintjan (2019) observed that considering memory to be solely defined in old-fashioned terms, as the storage of experiences in the brain, is at odds with their proposed extended definition which acknowledges that "advances in information technology are pushing the understanding of memory into new directions. We now talk about memory on a hard drive, or as a chemical change between neurons. Yet, these different definitions of memory continue to co-exist" (p.1).

To start with old-fashioned terms, memory was first conceived of as unitary, i.e., all memories share the same system or process for storage and retrieval. Nonetheless the current perspective, which dates back to the 1980s, is based on compartmentalized memory storage and separate retrieval processes, with dissociation between these processes underpinning much of memory theory, as will be discussed later.

Division into short-term and long-term memory.

Baddeley and Warrington (1970) challenged the unitary theory of memory because of a suggested separation between short-term (working) memory and long-term memory when comparing amnesic patients to normal control participants. They initially studied one patient who had undergone unilateral temporal lobectomy (historically used for treatment of epilepsy), together with five other amnesic patients, comprising four patients with alcoholic Korsakoff's syndrome (a chronic neuropsychiatric syndrome associated with thiamine deficiency from malnutrition in the context of alcohol abuse), and a fifth patient with a presumed vascular cause for their symptoms.

These six patients, and six age-matched control participants were tested using five different experimental tasks: free recall of word lists of 10 nouns; short-term forgetting using sequences of 3 three-letter words (triads); a proactive interference experiment (the effect of previously learned materials on the acquisition and retrieval of newer materials); a paired-associate learning task (where a series of item pairs are shown and subsequently tested by showing only the first word of the pair, with the participant having to recall the second word); and a standard digit span test. To rule out dementia, participants were reportedly screened in order to select only those who were unimpaired on intellectual tasks. More contemporary views suggest the opposite approach, i.e., that older adults with intellectual impairment should be screened for dementia, (Burt, Aylward, et al., 2000)].

Baddeley and Warrington (1970) concluded that amnesic participants demonstrated normal short-term memory and impaired anterograde long-term memory performance on the free recall task**. A**fter a 30-second delay, percentage recall was reduced in amnesic participants for all but the final item in the list. Additionally, in a previously administered questionnaire (in a separate experiment) involving recall of recent events, performance was reduced in amnesic participants. Though, as patients with Korsakoff's psychosis have a history of significant, chronic alcohol abuse, it stretches credulity to believe they would in any case have paid much attention to recent events.

Baddeley and Warrington's (1970) results were used to support a theorised dichotomy implying separate encoding of short-term and long-term memory, with a "short-term memory store" or episodic buffer, to connect the two systems (Figure 1). The Baddeley and Hitch model of working memory comprises a control system or central executive (a scientific concept to separate the analysis of executive processes from the question of their anatomical location, i.e., it controls attentional processes rather than acting as a memory store), and two storage systems (the visuo-spatial sketchpad and the phonological loop), with the episodic buffer acting as an interface between the two sub-systems of working memory and long-term memory (Baddeley, 2003). As a further qualification, it was proposed that fluid systems involved reasoning ability and the ability to deal with new or complex information, whereas crystallised intelligence systems involved acquisition of knowledge, learning and skills.

We now know that Korsakoff's syndrome is not equivalent to temporal lobe damage, despite the key feature being impaired long-term memory performance, with a severe declarative anterograde amnesia (memories involving personal experiences), (El Haj, Nandrino, Kessels & Ndobo, 2021). Korsakoff's syndrome is also characterised by confabulations, i.e., incorrect verbal expressions to questions or situations in which an amnesic patient feels compelled to respond (Oudman, Rensen & Kessels, 2021). The issues following temporal lobectomy are also more complex, e.g., laterality affects visual memory, verbal memory, and recall (Morris, Abrahams, Baddeley & Polkey, 1995).

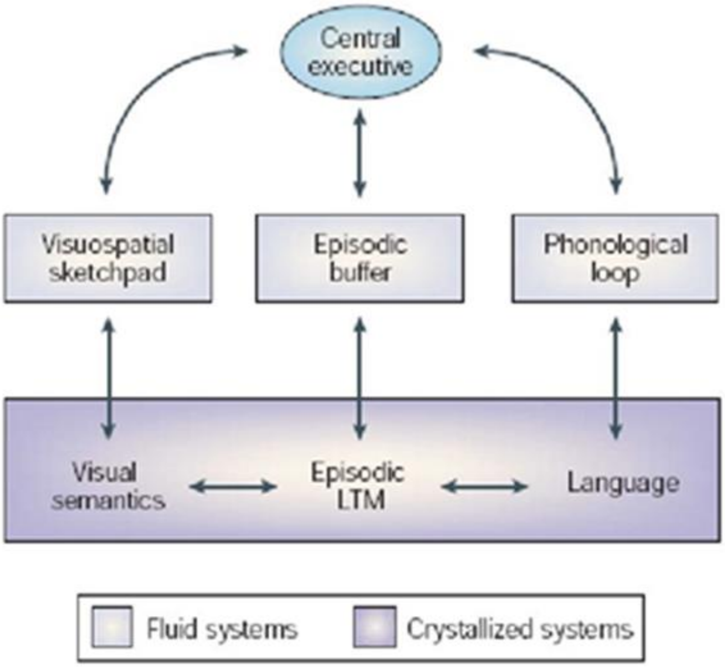

Figure 1. The Baddeley & Hitch model of working memory (taken from Baddeley, 2003)

Baddeley, Thomson, and Buchanan (1975) later proposed that short-term memory capacity is limited for auditory information (theorising an ‘articulatory rehearsal’ loop to explain the word-length effect. In a task using lists of between 4 and 8 words which were read aloud to participants, the percentage of correct word sequences was higher for lists comprising words of one syllable than those with five syllables. However, when the words were presented visually on flash cards the effect disappeared. Regardless, visual short-term (or working) memory capacity is thought to be similarly limited. Using differently orientated and differently coloured squares only four colours and orientations could be retained in visual working memory, i.e., the amount of visual information necessary to carry out a current ongoing task (Luck & Vogel, 1997; 2013). This may alternatively represent the amount of visual information we are able to pay attention to at any one time. Shipstead, Lindsey, Marshall, and Engle (2014) argue that individual differences in the capacity of working memory are due to multiple factors, with performance on complex span and visual array tasks (Luck & Vogel, 1997; 2013; Figure 2) related to reaction time differences on attention control tasks, e.g., Stroop[3] and Flanker[4] tasks.

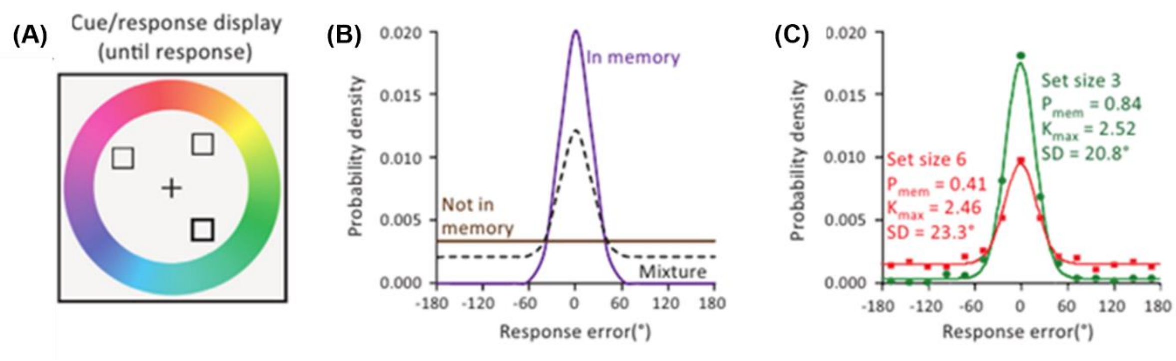

Figure 2. Illustrating (A) the colour wheel; (B) the hypothetical Von Mises distribution (used for circular dimensions such as hue), indicating that responses are normally distributed around the correct value. If the cued item is not remembered, then errors will be random, i.e., have a uniform distribution around the circle; (C) actual distribution of errors observed for set sizes of 3 and 6 coloured squares, demonstrating reduced short-term visual memory for 6 (red curve) compared to 3 (green curve) items (adapted from Luck & Vogel 2013).

Visual search theory, represented by the guided search model (Wolfe, 2020), suggests that when information in a scene is represented in the visual system, fixation will be on the item which attracts the most attention, and this item will be more richly represented than others. However, items further away which are still within the functional visual field around the fixation item will stimulate visual awareness but are less richly represented visually. It could be that the capacity of visual working memory is limited by the capacity of the functional visual field and that additional coloured squares not within the functional visual field are not well-represented or recalled.

---

[3] The **Stroop** task: in which participants quickly indicate the hue in which a word was printed (e.g., ink hue: red; word: ‘‘BLUE’’), where the hue and word are either congruent or incongruent.

[4] The arrow **Flanker** task: in which an array of five items is presented, the middle item is always an arrow. The participant indicates which direction this arrow is pointing. Flanking characters are congruent arrows, incongruent arrows, or neutral items.

So what has this to do with my research? As I use photographs of natural scenes in my studies, the amount of visual information we are able to pay attention to in the 600ms the cross and in the 2000 ms the photographs are displayed for is an important consideration, as is short-term memory for verification of the cross location (as detailed in Chapter 3: General Methods).

Long-term memory and its divisions.

Although widely accepted that memory is divided into short-term (working) memory and long-term memory, it could be argued that long-term memory is merely short-term memory encoded and consolidated because we have paid attention to it (in this case we do not have to conceptualize two different encoding systems for short-term or working memory, and for long-term memory). The idea that there were different types of long-term memory was based on Tulving's work (Tulving, 1983), conceptualizing a dissociation between particular types of long-term memory. There is some confusion in the literature as to the terminology.

Implicit and explicit memory.

Long-term memory may be referred to as either explicit or implicit [the latter defined as performing a task requiring no conscious recollection of previous experiences to facilitate it (Buchner & Wippich, 2000)]. Explicit memory, also known as declarative memory, typically refers to memories involving personal experiences, a memory that you have to consciously work to remember. Undoubtedly some memory processes are implicit (also termed procedural memory by some), as we do not require to consciously remember how to walk or ride a bicycle, pick up a glass of water or a knife and fork, use our mobile phone or a computer (unless we have suffered a significant brain injury). However, retrieved memories of personal experiences can arise spontaneously without conscious effort (Gardiner, Ramponi & Richardson-Klavehn, 1998), as can be the case with recognition without awareness.

Episodic and semantic memory.

We all have autobiographical or episodic memory (concerning personal past events), memories necessary to use a computer (procedural memory), and memory necessary for facts to sit an examination (semantic memory). Tulving, (2002) later described his concept regarding the division of long-term memory:

> “By the time I wrote Elements of Episodic Memory (Tulving, 1983), it had become possible to entertain the thought that the heuristic distinction was useful for the simple reason that it corresponded to biological reality. I proposed, therefore, that episodic and semantic memory represented two functionally separable memory systems” (p.3).

Tulving, like Baddeley, based his theory on neuropsychology, studying individuals suffering from amnesia following traumatic brain injury, notably Kent Cochrane (K.C.) born in 1951, who died in 2014 at the age of 62 years after a stroke. At the age of 30 years, he had a single vehicle motorcycle accident on his way home from work, resulting in brain injury. When he reached hospital, he was suffering from epileptic seizures and underwent surgical decompression of a left frontal subdural haematoma (where blood collects between the skull and the surface of the brain), remaining unconscious for 72 hours. Seven days following the accident, he was able to recognise his mother. Three weeks later, computerised tomography (CT) scanning showed chronic bilateral frontal subdural haematomas and a left occipital lobe infarction. He had right sided paralysis; a right homonymous hemianopia (a visual field defect involving the right half of the visual field of both eyes); retrograde amnesia with loss of personal (episodic) details, with intact semantic knowledge and anterograde amnesia for both types of memory, This dissociation, based on the case of K.C., was taken as confirmation of two functionally separable memory systems, i.e., division of long-term memory into episodic and semantic memory systems was proposed. A follow-up CT scan in

1982 showed severe medial temporal lobe injury and almost complete bilateral hippocampal loss.

It has been suggested that such “functional dissociations” have more to do with what explicit and implicit memory measures are applied in cognitive testing, than with separate memory processes or memory systems. In other words, results may be solely due to methodological artefact. Typically, amnesic individuals perform poorly on tests of explicit memory (conscious memories of previously experienced events), and normally on tests of implicit memory, e.g., when shown three-letter portions of words with the instruction to produce the first word that comes to mind in response, (Buchner & Wippich, 2000).

Following K.C.’s death, post-mortem magnetic resonance imaging (MRI) findings and pathology of his brain have been reported, together with a summary of detailed cognitive testing during his life, when functional magnetic resonance imaging (fMRI) had confirmed no detectable hippocampal activity (Gao, Keith, Gao, Black, Moscovitch & Rosenbaum, 2020). The hippocampus is now thought crucial to memory retrieval, with early life damage associated with impaired recall but not recognition, (Patai, Gadian, Cooper, Dzieciol & Mishkin, 2015). Gao et al., (2020) concluded that relationships between behavioural findings and brain pathology might need revising, as some areas of K.C.’s brain were more damaged, and some less damaged, than previously thought.

Regardless, Elements of Episodic Memory (Tulving, 1983), is still considered an essential treatise for students of cognitive psychology. It is taught knowledge that we rely on episodic memory to recall meaningful or significant events from our past experience (Prebble, Addis & Tippett, 2013), defined as a manipulation of knowledge in order to create spatial-temporal configurations of object and event concepts. Semantic memory is relied on to recall knowledge concerning the world, conceptual in nature and without reference to personal experience (Binder & Desai, 2011).

In day-to-day life episodic memory enables us to recognise people and places, and to remember who and what they are. As has been described for amnesic individuals, we mainly live in the present and spend little time reminiscing or reliving our past, apart from the elderly or isolated amongst us. Sometimes particular cues such as music, pictures, scenes, or smells, trigger spontaneous recollection of a personal memory that we may not have thought about for years. Otherwise, our daily lives are governed by what may be called the general knowledge of our daily circumstances or semantic memory (what we know), and implicit procedural memory, how to ride our bicycle to work or how to start up our computer, (Berry, Shanks & Henson, 2008; Tulving, 1985; Watkins & Tulving, 1975), as well as short-term memory to remember and write down messages, names, or phone numbers. These divisions merely represent memory for different things, whether facts or perceptual experiences, so why must memory be compartmentalised?

As an example of how the distinction between semantic and episodic memory can become blurred; a question such as, “What is the capital of France?” can cue retrieval of the answer Paris from semantic memory (knowledge of the world) but may also provoke the recall of rich and detailed memories of a personal trip to Paris (a particular and concrete experience located in personal time and space) from episodic memory, or memories of a film about Paris. In their review of semantic memory, Binder, and Desai (2011) acknowledge that episodic memory depends heavily on the retrieval of conceptual knowledge from semantic memory. Moreover, as memories in old age are commonly assumed to lack the rich detail of episodic memory, and to become more semantic in nature (Cabeza & St Jacques, 2007; Shing, Werkle-Bergner, Brehmer, Muller, Li & Lindenberger, 2010), this challenges the idea of two distinct and separate memory systems. Admittedly, Tulving (1985) accepted that both episodic and semantic memory could be involved in the recovery of knowledge about past events, proposing a model showing how one type of memory could compensate for lack of the other kind of memory (but still holding fast to the principle of two distinct and separate memory systems), as illustrated in Figure 3A (taken from Fang, Rüther, Bellebaum, Wiskott & Cheng, 2018).

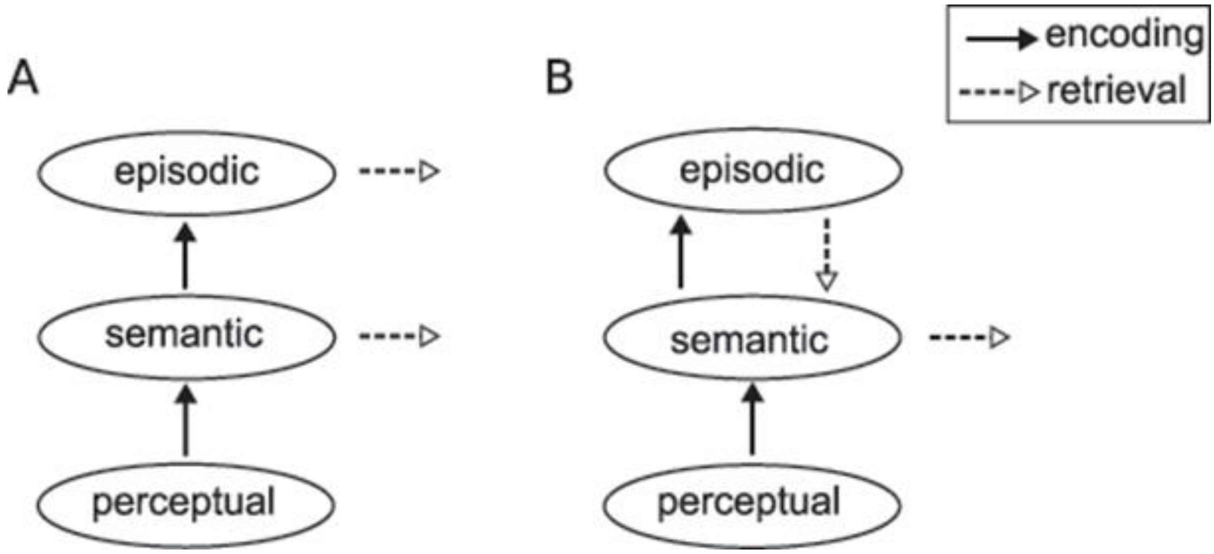

Figure 3. Schematics of the relationships between memory systems: (A) as suggested by Tulving (1985), with serial encoding, parallel storage, and independent retrieval; (B) the proposed model, where retrieval from the episodic system is not independent, but occurs through the semantic system (taken from Fang et al., 2018).

There have been challenges to the established view of memory. Even Baddeley (2020), questioned whether Tulving was right to theorise that semantic and episodic memory were two separate memory systems, based as it was on the study of amnesic patients (as indeed was Baddeley's own theory that short-term and long-term memory were two separate memory systems). A single-system approach to memory based on signal detection theory (SDT), asks if dissociations really do imply independent memory systems (Berry et al., 2006). In separate research, a computational model of semantic and episodic memory suggests interdependence, with reliance on semantic memory to represent spatial and temporal aspects of episodic memory, (Fang et al., 2018, Figure 3B). This is reinforced by work from Renoult, Irish, Moscovitch and Rugg (2019), who offer a review of behavioural, neuropsychological, and neuroimaging data supporting the interrelation of episodic and semantic memory, or the blurring of the distinction between them. Further evidence for interaction between episodic and semantic memory, linked to familiarity, is offered by Greve, van Rossum and Donaldson, (2007).

Encoding, Consolidation, Storage and Retrieval of Memory

Logically, memory requires an input system (termed encoding), a system for fixing the memory in neural networks (termed consolidation), storage somewhere in the brain via a memory trace (or engram), and output process or processes for retrieval of the engram (remembering or recollection). The accepted theories regarding memory division leave unanswered where and how does memory get stored in the brain?

Neuroimaging & neurobiology.

Although not using neuroimaging in my research, it is necessary to understand what this body of literature has to say about memory storage and retrieval, in order to recognise what may underlie behavioural research findings. Memory research has evolved over recent years with the advent of neuroimaging. See Poldrack and Yarkoni (2016) for a review of neuroscience research, explaining that the organisation of memory, i.e., cognitive function, arises from defined brain structures. The old concept of a schema is a pattern of thought or behaviour that organises categories of information and the relationships among them. Brought up to date by neuroscience, a schema may also refer to an associated system of neurons, forming a neural network, (a latent neurocognitive structure of strongly interconnected nodes potentially re-activated together). Ghosh and Gilboa (2014) view a schema as representing a neural network based on repetition, similarity, or commonality of events to characterise complex constructs, albeit comprising the gist and lacking unique detail for each event, which is of importance in relation to my conclusions. Information in a schema could be used to influence the encoding and retrieval of episodic memory and to guide context-specific behaviours.

Neurobiology is shedding light on the biological mechanisms such schema may rely on. Neural network models have been studied for some years. These models use vectors and probabilities e.g., to investigate recall or recognition (Chappell & Humphreys, 1994). Evidence from neurobiology research confirms that connectivity within neural networks relies on the strengthening of synaptic connections. Modifications to DNA by epigenetic changes that do not alter its sequence can alter gene expression to regulate neuro-steroid production within the brain. This allows synaptic connections between existing neural networks to become stronger (Colciago, Casati, Negri-Cesi, & Celotti, 2015; Colciago & Magnaghi, 2016). Changes in neurotransmitters at the level of the synapse facilitate transmission (i.e., retrieval) of memories (Mayford, Siegelbaum & Kandel, 2012). Yet, we are still far from understanding exactly how memories are captured at a cellular level.

In the recent past it was possible only to externally measure the electrical activity arising within the brain, using an electroencephalogram (EEG). This type of research uses electrodes attached to the scalp; similarly, event-related potentials (ERP) seek to investigate brain activity related to a presented stimulus in an experimental task, (cf. MacLeod & Donaldson, 2014; Murray, Howie & Donaldson, 2015; Murray, Ouyang & Donaldson, 2019). Magnetoencephalography (MEG) uses very sensitive magnetometers (housed in a helmet-type device and thus easier to apply), to record weak magnetic fields produced by the electrical currents occurring naturally in the brain e.g., (Boto, Hill, Rea, Holmes, Seedat, Leggett,... & Brookes, 2021). Although EEG and MEG both produce representations (maps) in real-time, localisation lacks fine detail.

However, nowadays cognitive psychology researchers have access to more expensive tools such as MRI scanners. Brain cell activity inferred by changes in blood flow forms the basis of functional magnetic resonance imaging (fMRI) to "pinpoint" which areas of the brain are activated when carrying out experimental tasks (e.g., Hassabis & Maguire, 2007;

Rissman, Chow, Reggente & Wagner, 2016; Zhao, Wang, Liu, Xiao, Jiang, Chen & Xue, 2015). While this gives better anatomical localisation of the processes studied, this method also has its drawbacks. There is a time lag of up to a few seconds for blood flow to change after brain activation, and degradation of images occurs due to fine movement resulting from respiration and heartbeat. The complexity and the accuracy of the method of analysis of the images is a source of debate that is open to question (Norman, Polyn, Detre & Haxby, 2006). In the same way as we know that different areas of the brain carry out different functions (on a basic level, the motor and the sensory cortex have been known about since the 1930s), we now consider that different areas of the brain are involved in various aspects of memory.

Of importance to my research, fMRI has been used to study the areas of the brain associated with episodic recollection and memory vividness, (Tibon, Fuhrmann, Levy, Simons & Henson, 2019; Zou & Kwok, 2022), both concluding that the angular gyrus is engaged in vivid recollection of episodic memory.

*We think that the world is a solid, vivid place, full of shape and colour and solid objects like this table and this microphone, and so on, but we actually create that in our heads out of the bits of information that hit the back of eyeballs or hit our eardrums or hit our tongues or whatever. (**Douglas Adams** 1952 – 2001).*

From a medical or physiological standpoint, perceptual experiences involve the interpretation of electrical impulses transmitted to the brain (the central nervous system) from the peripheral nervous system, i.e., from nerve endings in the skin, retinal cells in the organ of vision (the eye), cells in the organ of hearing and balance (the ear), and proprioceptive receptors in muscles and tendons (respectively relaying sensations of touch or pain, vision, hearing, balance and movement) specifically to the various parts of the brain responsible for their perception. Logically, the memory of those experiences would be stored in the same areas, making it more plausible that neurons in the part of the brain relating to the various perceptual inputs are involved in their memory, mediated by changes at the level of the synapse. This, however, is not an original concept, as described by Squire, (1986):

> "The view of memory that has emerged recently, although it still must be regarded as hypothesis, is that information storage is tied to the specific processing areas that are engaged during learning. Memory is stored as changes in the same neural systems that ordinarily participate in perception, analysis, and processing of the information to be learned. For example, in the visual system the inferotemporal cortex (area TE) is the last in a sequence of visual pattern-analysing mechanisms that begins in the striate cortex. Cortical area TE has been proposed to be not only a higher order visual processing region, but also a repository of the visual memories that result from this processing" (p.1612).

Visual perception and visual imagery have common representations in the visual cortex, (Dijkstra, Bosch & van Gerven, 2017). This could be interpreted as memories of visual stimuli being stored in the part of the brain where they were perceived, re-imagining being linked to the same anatomical region. Research into the Ebbinghaus illusion (where a stimulus surrounded by smaller or larger stimuli appears larger or smaller respectively), assumes that memory and perception are based on the same sensorimotor system (Rey, Riou & Versace, 2014).

Processing memory lies in connecting the disparate electrical impulses arising from an experience in order that the memory of it can be retrieved as a whole. Such holistic retrieval of memory is theorised to be based on pattern completion and to be localised in the hippocampus, e.g., based on fMRI analysis of a multi-element event paradigm (Grande, Berron, Horner, Bisby, Düzel & Burgess, 2019).

Sleep is thought to play an important role in allowing for the consolidation of memories, e.g., (Rasch & Born, 2013) or, and salient to my conclusions, active reorganization, when new memory traces are integrated within pre-existing long-term memories, (Straube, 2012). Once consolidated, memory is assumed to be stored as a memory trace or ‘engram’ (information imprinted in a physical substance, theorised to be the means by which memories are stored, i.e., the enduring offline physical and/or chemical changes that underlie the newly formed memory associations). If this is the case, then, experimental investigation of memory in the laboratory will only assess the discrete parts rather than the whole system of consolidated memory. However, neurobiology research points to more rapid encoding of memory than previously thought. As an example, Ison, Quiroga and Fried (2015), have shown that new associations can be formed even after a single stimulus, to facilitate rapid effortless retrieval of episodic memory. Their findings were concluded to be in keeping with continuous retrieval rather than an all-or-none process.

A current review on the engram (Josselyn & Tonegawa, 2020), suggests that engrams can be silent or active, according to how retrievable the memory is. It is also theorised that a given memory may be supported by functionally connected “engram cell ensembles,” each of which supports a component of the overall memory, but which may be dispersed across multiple brain regions. Although their studies are based on rodent brains, Josselyn and Tonegawa (2020) consider engram cells to exist in the hippocampus. As an aside and harking back to lay opinion of memory being like a storehouse, when waking up after my stroke I had the sensation of being in an empty warehouse when somebody had switched off all the lights. It felt like I was walking through deep wood ash, in the dark. Gradually, as I recovered, the lights started to come on again.

Such engram cells (activated by an experience which results in physical or chemical changes within the cell) are reactivated by subsequent presentation of the stimuli present at the original experience, which results in memory retrieval. Such reactivation or memory retrieval produced by the subsequent presentation of the stimuli present at the original experience has long been known as the ‘encoding-specificity principle’ (Tulving & Thompson, 1973), and is the principle on which many experimental memory paradigms are based, as retrieval from memory is commonly cued in cognitive psychology experiments, with the effectiveness of the cue dependent on how the memory was encoded, i.e., dependent on the context in which the information was stored. This process is crucial to the conclusions drawn from my research.

As previously referred to, once consolidated, a memory may be susceptible to change when retrieved and reinstated (St. Jacques & Schacter, 2013). It is suggested that the creation, storage, and retrieval of episodic memory depends on communication between a multitude of areas in the brain. Although agreeing that information is received and processed in sensory areas, such information is eventually relayed to the hippocampus, mediated by oscillatory firing patterns. Rodent brain experiments have demonstrated that certain frequencies in these firing patterns modulate long-term potentiation of memory, in other words representing the neural basis of memory encoding. An attempt to reproduce synchronisation between visual and auditory cortices in human participants (using video clips and EEG recordings of brain activity) produced a null result, under explanation that the theorised mechanism, or the stimuli used, were incorrect (Van der Plas, Wang, Brittain & Hanslmayr, 2020).

Although there are multiple experimental paradigms used to investigate memory, I will next provide an outline of the general groups of tasks that may be used to investigate episodic memory and accuracy judgments, rather than detailing their many variations.

(ii) Experimental Tasks of Memory Retrieval.

When assessing memory, it is considered impractical to test participant's personal memories. The recollection of real-life personal events, rich in meaning, is thought to produce memory errors because of the reconstructive processes involved in their retrieval, but the basic problem is that one can never be completely sure a memory is an accurate description of the original experience without corroboration from other sources (Koriat, Goldsmith & Pansky, 2000). It is considered too complex to test everything at once when examining memory, since it may be difficult to determine which memory process supports the results obtained. For this reason, researchers "break the continuity of memory into entities called memory tasks" (Sumrall, Sumrall & Doss, 2016, p.24). Even though memory tasks using word lists produce more replicable results in terms of accurate recollection (Schacter, Israel & Racine, 1999), such tasks can lose ecological validity compared to real-life memories.

To test memory, you need to decide which theorised memory process you wish to examine, which task is appropriate to test your hypothesis, what stimuli to employ, what the outcome measure is, how to calculate it, and, importantly, how to carry out statistical analysis in order to present your results in the best light. As most cognitive psychology (laboratory-based) research is carried out on psychology undergraduates, this eliminates the potential confound of changes in memory with age (e.g., Gallo, Cotel, Moore & Schacter, 2007), but also raises the issue that results might not generalise.

Characteristically, for experimental tasks of long-term (episodic) memory, at test (i.e., memory retrieval) participants are required either to recognise or remember (recollect) stimuli previously shown at study (encoding). Participants need to be given explicit instruction to make their responses, in order to standardise what process or processes they might be using to perform the task required. Fundamentally, the separation of recognition from recollection is based on the task the participant is asked to perform, [Do you recognise this word? How many words can you recall (remember) from the list you were shown?] Regardless, these two aspects of memory retrieval are theorised to be separate and distinct, recognition referring to the ability to "recognise" an event or stimulus as familiar, with recollection being the retrieval of related details from memory. However, both processes (i.e., familiarity and recollection) are theorised to contribute to recognition of a stimulus or event.

Testing recollection.

Recollection is the remembering (recall) of a memory trace, i.e., of information or events previously encoded and stored in the brain. Typically, to test recollection, participants are asked to study a list of items, historically lists of words, either written down or read aloud. They are then asked to recollect (i.e., retrieve from memory) the items that were presented, (e.g., Dobbins, Kroll, Yonelinas & Liu, 1998; Tulving & Watkins, 1973; Waddill & McDaniel, 1998). Without prompting, this is termed free recall. Recollection of words has been assumed to be retrieval without need for a search process (Gardiner 1988, p.447). However, recollection could also be a process of recognising whether words retrieved from memory were on the presented list, i.e., a type of recall to reject processing (Haist et al., 1992; Rotello & Heit, 2000).

Cued recall prompts recollection using variously, weakly associated words, categories, or the first letter of the to-be-remembered word (like filling in a crossword puzzle, the more letters there are, the easier it is to solve the clue). A recollection task sounds very simple to do in the laboratory, but what type of words, how many, and what time interval is best used before testing recollection, i.e., immediate, or delayed recall? If the time interval is too short, arguably you may not be testing long-term memory, if the time interval (or the word list) is too long, the recollection rate may be so low as to fail to provide any meaningful results. It is therefore common that researchers base their choices on reproducing or modifying published experimental methods. If the task has a previously reported record of success, then it should work when repeated. I freely admit that both of my experimental tasks are based on modifying earlier published methods. In either free or cued recall, the proportion of words recollected represents the accuracy of memory. In such a recall test, although the number of words that were not on the studied list can be accurately assessed, there is no way of identifying the number of guesses (without recollection) that resulted in a correct word.

Other factors to be aware of include primacy and recency effects, in that the first and last items in lists are better remembered (Tulving, 2008). This may be overcome by discarding results from these positions or buffering items (using dummy non-tested words before and after the tested words). The type of memory task employed may dictate the strategies used by participants, and thus influence the results, with concrete words such as dog, chair, table, or scarf being easier to visualise, i.e., the participant may visualise a dog wearing a scarf, sitting on a chair at a table to remember the words, so in effect the words are dual-encoded, corresponding with theoretical accounts of increased depth of processing and improved recall (Craik & Lockhart, 1972; Paivio, Walsh & Bons, 1994). Additional concerns relate to the Deese–Roediger–McDermott (DRM) false memory illusion (Deese, 1959; Roediger & McDermott, 1995), whereby presenting lists of semantically-related words can cue free recall of non-presented but associated words up to 40% of the time, e.g., a list of words including dream, pillow, bed, nightdress… might provoke recall of the non-presented but related word sleep. Although this occurrence does not change accuracy (the proportion of words correctly recollected) this has led to some researchers abandoning words and instead using nonsense syllables.

A further point for both recollection and recognition tasks is the use of rated and standardised stimulus sets, i.e., words (where standard lists of words and lures can be downloaded, including word lists that produce a DRM illusion), or pictures, such as the international affective pictures system (IPAS) or the natural scenes collection (comprising nature and campus scenes). The advantage is that results can be compared across different laboratories, (Snodgrass & Vanderwart, 1980), the disadvantage being that any inconsistencies continue to be reproduced.

Recognition tasks.

According to the American Psychological Association (APA) online dictionary https://dictionary.apa.org recognition memory "is the ability to identify information as having been encountered previously" and relates only to "declarative knowledge—factual material that is deliberately and consciously accessed—rather than to nondeclarative knowledge or other implicitly known information."

Recollection and recognition are posited to be functions of conscious (declarative, or episodic) memory. Haist, Shimamura & Squire (1992) used a forced-choice recognition task with lists of words, participants comprised 12 amnesic patients, six with Korsakoff's syndrome, and six with various other MRI-confirmed lesions (which takes us back down the well-trodden path of using participants with varying neuropathology or brain injuries), finding that recollection, recognition and confidence were all impaired in amnesic patients.

Recognition memory is classically tested using an old or new, yes or no recognition experimental task and underpins one of the dominant memory theories, signal detection theory (SDT). In a recognition memory task, a series of words or pictures (referred to as targets) are presented singly within a study phase. Participants are asked to remember them for a subsequent memory test. At test, the previously presented targets (designated 'old') are presented again singly, mixed with a set of 'new' pictures or words, not previously presented (non-targets, distractors, or lures). At test, participants are asked if they recognise each stimulus (i.e., yes, it is old) or not (i.e., no, it is new), in other words the task tests their ability to discriminate which items have been previously viewed in the context of the experiment.

Although the number of words used in a list may be limited for tests of recollection or free recall, it has been shown that participants can recognise many words and hundreds of pictures readily, the latter with recognition accuracy approaching a ceiling effect (Shepard, 1967). In effect we remember pictures better than words (Ally, 2012; Rajaram, 1996), i.e., a picture superiority effect.

Recognition is said to be easier than recall because you are not required to retrieve the right answer from memory, as the task is cued by the word or picture being presented again. You just have to say whether you recognise it. The 'purity' of the old or new task has been doubted because a recognition task is regarded as based on the two separate and distinct processes of recollection and familiarity; recognising the target as old could also mean that participants may remember (recollect) studying the target or they could just guess they had seen it before as the target looked familiar. If recognition can be due to recollection without involving familiarity, are recognition and recollection in this case the same? If recognition can be due to familiarity without involving recollection, is recognition in this case just another name for a feeling of familiarity? This "dual process theory" will be discussed below, but for a recent review, see Hockley, (2022). Mandler (2008) provides a detailed perspective on familiarity and its relationship to recognition memory. For a historical perspective on Mandler's 1980 article (or "the butcher on the bus" – as it later became known) see MacLeod, (2020). I will return to these topics later, when discussing familiarity.

Although my studies are based upon behavioural experiments, one cannot entirely ignore the contribution of neuroimaging to the topic of recognition memory (see Slotnick, 2013). Neuroimaging experiments using fMRI have studied hippocampal activity associated with familiarity. Without priming[5], familiarity may be difficult to replicate in the laboratory setting. MacKenzie and Donaldson, (2007) investigated facial recognition using ERP, finding

[5]priming is a technique in which the introduction of one stimulus influences how people respond to a subsequent stimulus. Priming works by activating an association or representation in memory just before another stimulus or task is introduced.

that familiarity and recollection dissociated in terms of anatomical location. Paller, Voss and Boehm (2007) debate the use of ERP showing familiarity on this basis (cf. Curran & Cleary, 2003), considering instead that there may be confusion between markers of priming (perceptual or conceptual) and familiarity.

The absurdity of testing memory is that using tasks with a binary outcome gives only a limited account of memory accuracy, which is either 0%, i.e., incorrect, or 100%, i.e., correct, or is based on an averaged proportion of correct responses for each participant (the pitfalls of this type of analysis will be discussed in Chapter 3, General Methods). In contrast to using the proportion of words correctly recollected to represent the accuracy of memory, for an old or new task, accuracy of memory is measured by assessing the ratio of the proportion of correct responses or hit rate (when an old item is correctly identified as old) to false alarms (when a new item is incorrectly identified as old), yielding a single performance outcome, frequently paired with a retrospective confidence judgment as a proxy for objective memory strength, and assessed using receiver operating characteristic (ROC) curves (discussed below in Metacognitive judgments of memory accuracy). The problem of guessing is an issue for all memory tasks, particularly recognition tasks; how do you tell what proportion of guesses the hit rate includes, which targets have been recognised, which were recollected (remembered), which guesses were made in the absence of recollection, and which were made due to a feeling of familiarity?

The remember or know question.

One way of addressing these issues has additionally depended on the answer to a Remember or Know question, by asking participants, after making their response, if they had either remembered the stimulus or had known they had seen it before in the context of the experiment. Tulving (1985), had suggested that memory judgments for both recollection and recognition depended either on remembering or knowing. The remember or know question reflects the extent to which “self-knowing consciousness” was involved in the recovery of knowledge about a past event (in other words, episodic memory).

This approach has been questioned. Forcing participants to make a binary remember or know judgment means that if they do not remember the item, they must answer ‘know,’ even when this does not reflect the reasons behind their decision. Gardiner et al., (1998) carried out an experiment with participants asked to answer simply yes or no to whether they recognised an item as old; if they answered yes, next to indicate whether their response was based on one of three options: remember, ‘know’, or guess. Participants were interviewed to collect qualitative data regarding what their responses were based on, with ‘know,’ based either on familiarity or on just ‘knowing’ the item was old. Importantly, confidence was higher when the target was remembered than when the target was known, i.e., not remembered (Gardiner, 2001; Tulving, 1985). Significantly, Dunn (2004) considered the remember or ‘know’ question to be a measure of confidence rather than of different memory processes such as recollection or familiarity. Further issues relate to instructions given to participants on what the ‘know’ question means, which have been shown to bias the manner in which they respond to the remember question (Williams & Lindsay, 2019). Furthermore, doubt has been cast on what lay participants and psychologists understand by remember, recollect, or know (Umanath & Coane, 2020).

Recognition is said to occur without associated recollection, described as implicit recognition or recognition without awareness (Voss & Paller, 2010), theorised to be based on fluency, i.e., because the item seems familiar. This type of recognition is also termed recognition by familiarity (Mandler, 1980), alternatively, familiarity without recollection (Onyper, Zhang & Howard, 2010). Jacoby’s (1991) two-factor theory of recognition memory relies on either automatic or intentional processes, the latter applying to recollection (which in a divided-attention task is impeded, whereas an automatic process based on familiarity for recognition memory judgments is unaffected). Such "process dissociation procedures" are claimed to differentiate between automatic (implicit) and intentional (explicit) memory processes.

Gardiner (1988), and others (Anderson & Bower, 1972; Evans & Wilding, 2012; Mandler, 1980; Tulving & Watkins, 1973), have debated whether recognition and recollection are truly independent processes, or that recognition and recollection may not be different processes, and recognition should be considered a single process, e.g., (Onyper et al., 2010; Slotnick, 2013). Some researchers maintain that recollection and familiarity are separate and distinct processes. Items can be recognised either if they are recollected or if they are sufficiently familiar, forming the basis of dual-process theory (Yonelinas, Dobbins, Szymanski, Dhaliwal & King, 1996). Dual-process theory considers the two processes to be separate, but whether they are additive is disputed (Brainerd, Gomes & Nakamura, 2015; Cha & Dobbins, 2021).

When we recognise someone in real life, instead of just thinking "I remember them" or "I know I have seen them before," we may spontaneously or effortfully retrieve associated details such as who they are, in what context we saw them, and when we last saw them (the who, what, and when of memory). Mandler (1980) observed that such knowledge is rarely assessed in a typical old or new recognition memory experiment. This is of relevance for those working in the eyewitness field, who commonly use recognition of a suspect in a task where pictures of faces are shown singly, or as one photograph in a set of photographs, i.e., an N-AFC (alternative forced-choice) task, where multiple images are shown together equivalent to an identity parade line-up. The relative merits of the two tasks have been debated (see Wixted & Mickes, 2015). However, once the eyewitness is in court, they may well be asked about the what, and when, of their memory.

The recollect or familiar question.

In recognition tasks for previously encountered stimuli, it is assumed that recognition can be based on recollection, when recognition of a particular item is accompanied by conscious memory of its occurrence on the study list. Despite remember responses being associated with high confidence (Tulving, 1985), the intuitive use of ‘know’ (as in I know the capital of France is Paris) is that these answers may also be given with high confidence. For that reason, Dobbins, Kroll, and Liu (1998), and others (Evans & Wilding, 2012; MacKenzie & Donaldson, 2009; Williams & Lindsay, 2019; Yonelinas, Aly, Wang & Koen, 2010), have instead relied on recollect or familiar. As Gardiner et al. (1998) demonstrated, familiarity can be associated either with a ‘know’ response, or with a guess. I will return to this debate later on in the thesis (i.e., familiarity, Experiment 6).

2-AFC recognition memory task.

I discuss this particular type of recognition memory task here, as it forms the basis for two of my experiments. The 2-AFC task differs from the standard old or new recognition memory task in that although old items are shown separately at study, the two alternatives (old and new, target and lure) are presented simultaneously at test, with the participant having to choose (discriminate) between the two (Tulving, 1981; Zawadzka, Higham & Hanczakowski, 2017). The relationship between confidence and accuracy is a relative one in that information from both stimuli (the target and the lure) is available for the participant to make the decision. By this means, bias is eliminated, and thus ROC curve analysis is unnecessary. However, as well as guessing, participants may correctly identify an old target solely because they do not recognise the new item or lure.

Monitoring Source memory tasks.

In asking exactly what source memory is meant to be, I consulted the APA online dictionary https://dictionary.apa.org, where source memory is defined as:

> "Memory for the origin of a memory or of knowledge; that is, memory for where or how one came to know what one now remembers. This construct has been expanded to encompass any aspects of context associated with an event, including spatial-temporal, perceptual, or affective attributes."

The expansion of the construct relates to the use of source memory tasks in experimental cognitive psychology. A source memory task traditionally consists of two pieces of information, the first representing recognition memory for an item and the second representing recollection of the context in which the information was learned. As such, source memory is considered an essential part of episodic memory (Johnson, 2005; Johnson, Hashtroudi & Lindsay, 1993).

In a standard source memory experiment, items shown with differing source contexts at encoding are again shown at test without the context, with participants required to recognise the item and to recollect their encoding context, or source memory. The contexts can be intrinsic, such as words in different font colours (e.g., Doerksen & Shimamura, 2001), or extrinsic to the item, where participants have variously studied lists of words with background colour change, (e.g., Siefke, Smith & Sederberg, 2019) and words presented in different voices or in different spatial locations (e.g., Jaeger, Queriroz, Selmeczy & Dobbins, 2020; Troyer, Winocur, Craik & Moscovitch, 1999), to represent contextual or source information, also represented by a natural scenes background on which faces or objects are projected (Chua et al., 2012; Richter, Cooper, Bays & Simons, 2016). At test, presentation of the items cue recollection of the contextual information (source memory) that the item was presented with at encoding. Quite how a blue background on which a word is presented or the colour in which a word is shown represents the origin of a memory may be hard to follow. A source memory task is more likely a contrived way of separating the processes involved in recognition memory, with recovery of source memory theorised to be based solely on recollection of the context, which is (probably) not reliant on recognition. However, this may depend on the source memory task used. Recognition is easier than recollection; with an old or new judgment incorporating many guesses and varying degrees of recollection. In contrast, source recollection is more effortful, and source memory recollection is not entirely equivalent to the recognition of, e.g., an old word or picture (which might also be based on familiarity).

Outside of memory theory, and in accordance with the first part of the APA definition of source memory, more intuitively, source memory is also taken to mean the origin of information added to a memory after the event by a different actor, i.e., from the police, other witnesses, and information from TV programmes, films, or books. If the source of the information is forgotten at later recollection, this can lead to issues with source memory confusion and post-event misinformation as a cause of false memories, (e.g., Loftus, 2003; 2005), rather than the quality of the retrieved memory, or the misremembering of similar events. Source monitoring (or reality monitoring) may be used to check the accuracy of our memories, Johnson, Foley, Suengas and Raye, (1988) asked participants to recall a recent perceptual (actual) personal event and a recent imagined personal event, finding that actual events held more detail. The authors acknowledged that recent events may have been rehearsed more often, (i.e., reminiscence) hence leading to more detailed recollection (Öner & Gülgöz, 2018; Squire & Wixted, 2011). Actual events were found to be corroborated by other retrieved memories more often than were imagined events. However, it was uncertain if the participants believed in their imagined memories and was the detail of the actual events accurate?

There are different interpretations of results from experimental source memory tasks. Batchelder and Riefer (1990), suggest the ability to retain incidental features of items should be analysed using signal detection theory processes, (cf. Cooper et al., 2017; Dodson, Bawa

& Slotnick, 2007). Furthermore, it is suggested that source memory recollection can be specific or partial, depending on the amount of detail or type of detail that is recovered (Simons, Dodson, Bell & Schacter, 2004). Fox and Osth (2022) report that source memory may be retrieved for unrecognised items, based on whether source memory is encoded as part of the to-be-remembered item. The idea that source memory may be encoded as part of the to-be-remembered item is of importance for assumptions made in the source memory task employed in my thesis. It may, however, be unclear in some tasks exactly which is the item (object) and which the source (context), e.g.,

> “We have all had the experience of vividly remembering a conversation taking place at a certain place, time, and location. We may not, however, be able to recall the content of the conversation. As such, access to contextual information can occur in the absence of item information” (Cook, Marsh, and Hicks, 2006. p.835).

In their example, the context is the place, time, and location with the content of the conversation representing item memory. However, it would be equally plausible to suggest that the conversation is the item, with the content of the conversation being the contextual detail. However, might it not be more appropriate to suggest that all of these various contextual details represent the detail of the memory (and not the detail of the item or the detail of the source)?

(iii) Theories of Memory Retrieval

The testing of memory using experimental tasks is closely linked to memory theory, with the proviso that such experimental tasks are based on the separate entities tested rather than on the totality of memory. It is, however, necessary, to consider theories of memory retrieval, in order to fully interpret results from my experiments, and I give a brief resume here. There has been (and still is) debate in the field of cognitive psychology as to whether memory is retrieved as a continuous process or whether it is thresholded (all-or-none). When discussing continuous processes, it could be argued that recollection can be sudden and total (all-or-none), rather than a continuum of no recollection whatsoever, gradually increasing to full recollection. Threshold theories assume that either you do or do not remember the stimulus, consequent on a theorised threshold value for remembering. It could be the case that "sudden and total" recollection occurs at such speed we are unaware that it has been a continuous process.

Signal detection theory.

Every year I get my visual fields checked using a machine which generates points of light in various places within what remains of my field of vision. Each eye is tested in turn, and I need to click a button to indicate that I have detected a light point. SDT was originally an explanation for detection of radio signals on a background of static[6], and in psychology as the visual detection of light signals. For the latter, some threshold of neural activity has to be exceeded in order for light signal perception to occur (Tanner & Swets, 1954). The detection of visual signals was concluded equivalent to detecting a "signal in noise" ("noise" being background light intensity rather than radio static). The criterion of seeing (or detecting a visual signal) depended on psychological (an internal threshold or decision criterion), as well as physiological factors (an external threshold).

The perception of a light signal is due to the signal being compared to some internal criterion. In my annual visual field test, if the signal is above the criterion, I will click the button to indicate "signal present, light seen," and if the signal is below the criterion, I will not click the button hence indicating "signal absent, light unseen." However, I will sometimes decide to only click when I am absolutely certain that I have seen a point of light, other times

[6] Radio noise is a combination of natural electromagnetic atmospheric noise ("spherics", static) created by electrical processes in the atmosphere like lightning, manmade radio frequency interference (RFI) from other electrical devices picked up by the receiver's antenna, and thermal noise present in the receiver input

I will click even if unsure. By clicking, I am guessing there was a light point, i.e., a false alarm. Sometimes, because one eye is covered up (to test the eyes separately) my eye starts to water and my vision becomes blurred, making detection of the lights more difficult (a physiological factor). Another physiological factor relates to my bitemporal hemianopia[7] and left homonymous hemianopia[8]. Even if a light point is present in that part of my visual field, I am simply unable to detect it, which is not due either to a guess or to an internal criterion.

A number of mathematical assumptions made by Tanner and Swets (1954) can be overlooked in later accounts of SDT. The criterion of making a correct response to (or recognising) a light signal assumes that all false alarms are guesses. i.e., you claim there was a light signal when there was none to detect (ergo, you must have guessed one was there). The frequency of perceiving a light is assumed to be a function of light intensity, equated to the frequency of seeing a light as a function of neural activity. It is assumed that a linear relationship exists between light intensity and neural activity (the more intense the light signals the greater the average of the measure of the resultant neural activity). Figure 4 represents the mean of the distribution of responses if the light signal were repeated and its detection measured an infinite number of times.

The mean of the distribution is associated with the intensity level of the light signal (Tanner & Swets ignored signal detection variance with light signal duration or the size of the light signal, as “beyond their scope”). In my visual field test, some of the lights are larger and brighter, these are far easier to detect, others are small and dim, when it is more difficult to be certain that there has been a light point.

[7] A bitemporal upper quadrant visual defect can progress to a bitemporal hemianopia, most commonly the result of tumours located at the mid-optic chiasm, usually in the pituitary gland.
[8] A visual field defect involving the two left halves of the visual fields of both eyes.

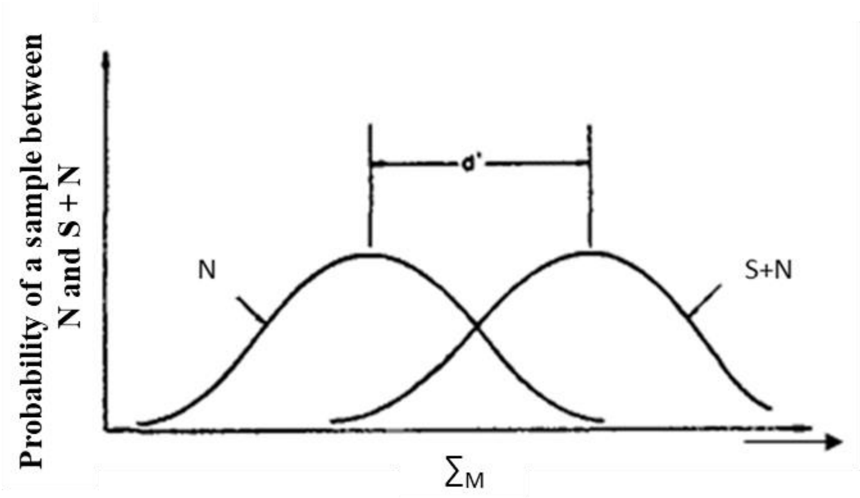

Figure 4. Hypothetical distributions of noise (N) and signal + noise (S+N); the greater the measure, ∑(M), the more likely it is that this sample represents a signal. (Taken from Tanner & Swets, 1954)

The two probability distributions shown in Figure 4 are (N): results obtained when no signal exists and noise or background light alone is sampled; and (S + N): when a signal is present, and signal plus noise is sampled. The distributions shown are Gaussian, with equal variance for all values of noise and signal + noise (with the proviso that this is not a true assumption). The observer bases their response on discriminating between one or the other, i.e., (N) or (S + N). In effect, light signal detection depends on the variance of the light signal on background light intensity and the greater this measure the more likely it is that the observation represents a signal. Others infer that a signal (S) represents memory strength, and that noise (N) equates to reliability (Zylberberg, Roelfsema & Sigman, 2014).

The work of Tanner and Swets was reviewed over 50 years ago by Banks (1970), focusing on how SDT had been used in psychology since the late 1950s and early 1960s. Weak perceptual signals may be ignored, either because they are not detected (the signal does not reach the threshold for detection), or because of the observer's response bias (decision criterion). In a recent history of SDT (Wixted, 2020), much of this has been re-visited using a commonly cited exemplar, relating SDT to how you can measure your perception of external stimuli (light, weight, sound) according to a Gaussian curve distribution (Figure 5).

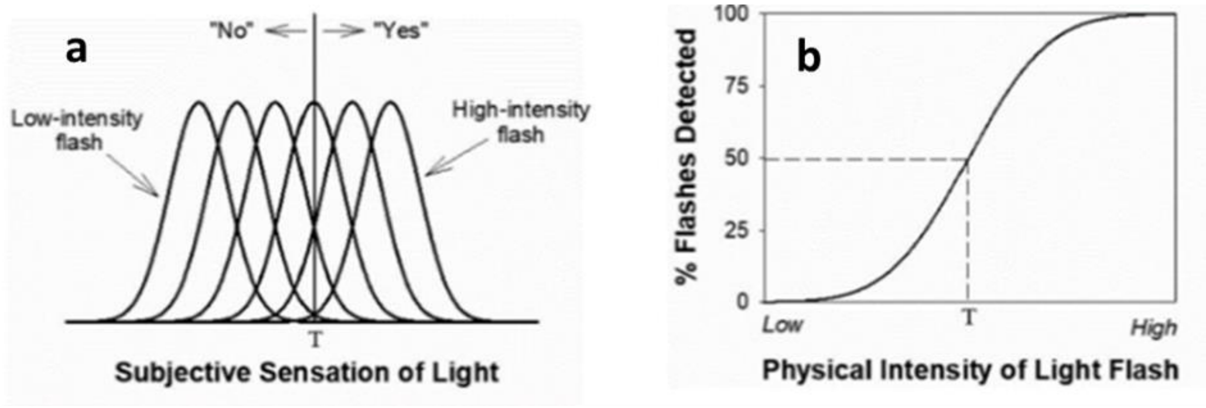

Figure 5. (a) SDT explained by means of flashes of light across differing levels of intensity (omitting stimulus-absent trials); (b) T is the threshold of conscious awareness relating the physical intensity of a flash to the percentage of flashes detected (taken from Wixted, 2020)

Plausible, but this conflates the use of SDT in memory signal identification with detection (perception) on a sensory continuum (Lockhart & Murdoch, 1970), meaning in effect that my visual field test is an allegory for how my memory works. SDT therefore is a convincing but simplistic way of explaining how a memory trace producing a discrete signal might be detected.

A false alarm rate (i.e., guesses, Tanner & Swets, 1954) versus hit rate graph was originally termed the memory operating characteristic (MOC). Each MOC is derived from a particular experiment, and the investigator needed to discover a distributional model to account for it. In most cases the MOC does not have a unique solution in terms of a single underlying model, not thought to be a great handicap, as Banks (1970) suggested that confidence ratings could be substituted (in many situations) to generate the same curve, now commonly used to create a receiver operating characteristic (ROC) curve.

The basic concept of SDT is the assumption of a gaussian (normal) distribution of the effects of old and new items on an abstract dimension known as the detectability index. The distance between the means of the two distributions scaled in z units (the number of standard deviations a given data point lies from the mean) is termed d-prime (d'), a theoretical construct of memory signal strength [as shown by Tanner & Swets (1954) in Figure 4]. The *S.D.* of the positive and negative distributions are assumed equal, essential for calculating d', i.e., a constant noise level across old and new items [assuming performance is dependent on a combination of memory strength and a decision criterion (Banks, 1970), Figure 6.

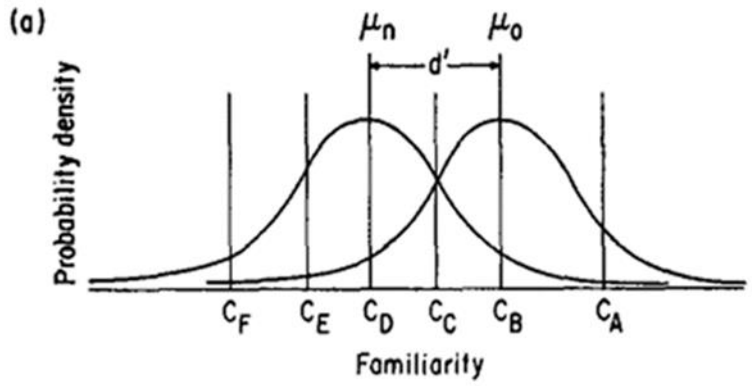

Figure 6. Normal distributions of new (N) and old (O) items along the familiarity (memory strength) axis assumed with the d' model, where σN = σO, (taken from Banks,1970)

In an old or new recognition task (N) is no longer noise but represents the strength of new items (or lures), and signal + noise (S + N) represents the strength of old items (O). The decision criteria, e.g., as shown in Figure 6 for successive trials, may become more lenient from C_B to C_F while d' remains constant, with the common variance as the measure. A decision criterion C_F that is too lenient to distinguish weak (below threshold) signals from "noise" leads to false alarms (indicating a signal when none was present – a guess). The axis of memory strength $\sum(M)$, Figure 4, is changed to familiarity in Figure 6.

The continuous curve of the mean memory signal [i.e., $\sum(M)$] reduces to zero, beyond which point there is no recollection. In a standard old or new recognition task, when no information (or signal) is detected, i.e., when recollection fails, you are tasked to indicate whether the item is old or new so perforce you must make a guess. The assumption is that false alarms (when a new item is falsely recognised as old) are always guesses. An alternative model of signal detection including a guess response criterion fails to fully account for false alarms not always being random guesses, (Gardiner, Ramponi & Richardson-Klavehn, 2002). Voss and Paller (2010) conceived of two different types of guessing: based either on the complete absence of retrieved information (when guessing is just a random, by chance guess), or based on relevant information present in neural networks, but unavailable for conscious awareness, the latter being more accurate, and based on unconscious familiarity, although confidence may be lower (Gardiner et al., 1998).

However, a lure may still evoke a memory signal due to similarity or familiarity effects. The “mirror effect” (Hilford, Glanzer, Kim & Maloney, 2019) is thought due to the strength of the items within the test affecting a participant’s decision criterion for different trials. In the strong condition, it is easier to identify new items as new and to identify old items as old than in the weak condition, (like dull and small points of light in my visual field test) hence the false alarm (guessing) rate varies accordingly, with false alarm rates and hit rates believed to vary in concert (as one rises, the other falls). The mirror effect may be produced by experimental variation, e.g., of familiarity, list length, or the similarity of targets and lures. In the mirror effect condition strong lures can evoke a strong memory signal for detection, such that the new item is falsely recognised as old. There is a body of literature confirming that some deceptive lures can be falsely recognised as old with varying degrees of confidence.

Unequal variance SDT (UVSD) model.

More recently, and in part to counter the proviso that the assumption made originally by Tanner and Swets (1954) is not a true assumption (of equal variance for all values of noise and signal + noise), it is suggested that an unequal variance SDT model should be regarded as the standard for decision-making in source memory recognition tasks. As Wixted, (2007) explains, “targets can be thought of as lures with memory strength added due to being on the study list, with this added strength differing across items” (p.153). False alarms are held to be guesses made to lures when recollection fails. The potential strength of lures is not acknowledged.

Starns and Ratcliff (2014) endorsed the unequal variance signal detection model for recognition memory, challenging the assumption that decisions as to whether the stimulus is old or new are based on a single overall strength value. Instead, they argue for fitting a “diffusion decision memory model” for accuracy and reaction time distributions in recognition memory, capable of explaining between-trial variability in memory evidence, i.e., memory is more variable for targets than lures. Starns (2021) further proposed that reaction time distributions help discriminate between decision-making models of recognition memory.

Threshold (discrete-state) theories of memory.

Threshold theories provided an earlier account of memory retrieval, in which the conscious experience of memory was divided by thresholds into fixed states, (Krantz, 1969; Paap, Chun & Vonnahme, 1999; Zhou, Osth, Lilburn & Smith, 2021). Below the internal threshold is the point at which the observer has no perception of a signal, i.e., a non-detection memory state, independent of an external threshold (where signal is either present or not present). Above the internal threshold is the point at which the observer perceives a signal, i.e., a detection memory state, based on the observer’s decision criterion. This has been largely superseded by SDT. However, I will also later discuss a thresholded rather than a threshold theory of recollection, which was originally derived employing the source memory task I propose to use in my thesis.

Dual-Process Theory.

Dual-process theory, previously referred to, suggests that a recognition task is based on two separate and independent processes, familiarity, and recollection (remembering). According to dual-process theory recollection is a threshold process, and familiarity an equal variance signal detection process (Yonelinas,1999; Yonelinas et al., 2010). This is the dual-process, signal-detection (DPSD) model of recognition memory (Anderson & Bower, 1972;

Onyper et al., 2010; Parks & Yonelinas, 2009; Yonelinas, 1994, 1999; Yonelinas et al., 1996; Yonelinas et al., 2010).

ROC curve analysis (discussed below) has been used to support discrimination between memory processes. Yonelinas (1999), speculated that to support a dual-process theory the ROC curve should be curvilinear and asymmetric (demonstrating that recollection and familiarity are separate processes). If recollection is expected to dominate performance and increase the high confidence hit rate on an experimental task it will result in a skewed ROC, i.e., not symmetrical. In his validation of the dual-process theory a source memory task (based on recollection and familiarity) was used; participants had to remember whether words were shown on the right-hand or left-hand side of the screen, using a 6-point confidence scale (Yonelinas, 1999). Results were compared to those from a standard recognition memory task, where familiarity also drives recognition. In the ROC curve analysis, source memory (based on recollection) gave a straight-line plot whereas recognition being based on either recollection or familiarity (but not both) produced a curved ROC plot (Figure 7).

Yonelinas (1999), reported that a linear function fitted the source memory data (apart from the lowest hit rate point), and a curvilinear line was a “better fit” for the recognition data. Notwithstanding, it is not best practice to present ROC curves for a 6-point scale as if this were a continuous variable (when smoothing of the curve is acceptable as there are multiple data points). Such limited ROC curves should rightly be presented as stepped, and should you draw a stepped curve for both tasks in Figure 7, you would be less likely to consider the processes to be distinct, making any difference between the two curves less obvious. This experiment, together with another three similar word-based experiments showed consistent results. In his discussion Yonelinas (1999), admitted that recollection might also be a signal-detection process, assuming that recollection and familiarity are independent processes contributing to recognition. Unfortunately, the demonstration of different memory processes may not be due to recruitment of the processes, rather, any differences could also be attributed to the fact that the tasks, i.e. source memory and recognition memory, differ in their level of difficulty.

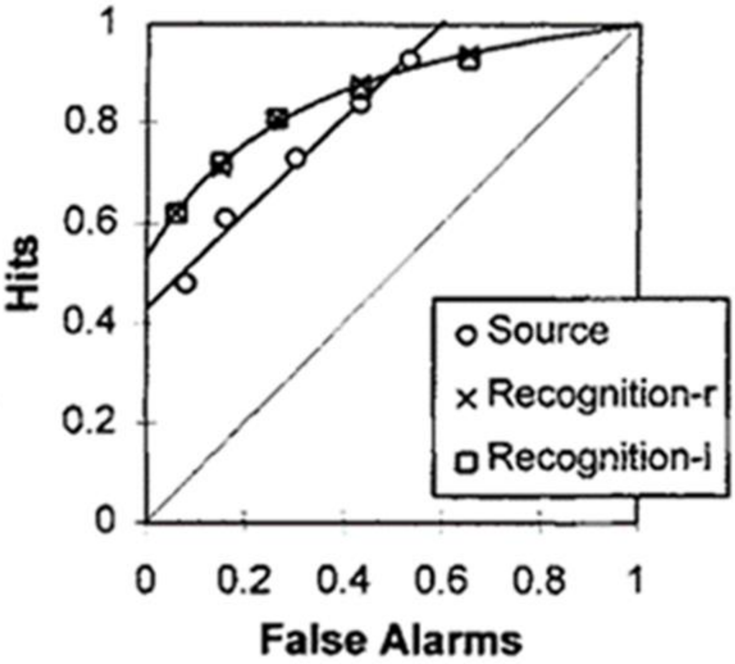

Figure 7. ROC for recognition and for source memory performance (taken from Yonelinas, 1999)

Gillund and Shiffrin (1984), provide a lengthy and complex retrieval model for recognition and recollection, relating to the use of search processes based on context, familiarity, and a dual-process procedure. The original dual-process theory proposed that familiarity and recollection were separate processes, where either one or the other contributed to recognition judgments. Yonelinas (2002), later theorised that recollection and familiarity were initiated in parallel at the time of memory retrieval and acted independently. He also conceded that source memory might not be process-pure, reiterating that recollection was thresholded and related to the detail of the memory, whereas familiarity was a continuous signal detection process, based on memory strength (because old items will be more familiar than new items), see Figure 6. However, is this assumption always true?

In “a critical review of dual-process theories of recognition,” Mandler (2008) cites his research presenting an outline of dual-process theory 40 years earlier (Mandler, Pearlstone & Koopmans, 1969). Source memory paradigms are used in experiments to validate a standard dual-process theory of recognition memory in order to isolate the process of recollection. Yonelinas et al., (2010) conceded that recollection could sometimes fail, therefore falling below a threshold, but that this would still be in keeping with the dual-process model.

Despite the literature originally proposing that recollection and familiarity contribute independently to recognition, others have considered the processes of recollection and familiarity to be combined, with participants deciding either which to rely on, or choosing the process which is stronger (see Cha & Dobbins, 2021). Regardless, exactly how the dual process theory of recognition memory works, continues to be debated. Slotnick (2013) provided a meta-analysis of fMRI studies into recognition memory, considering three main debates in the recognition memory literature: firstly, which part of the brain is associated with which process (familiarity or recollection, which is predicated on them being separate and independent); secondly, is recollection a continuous or threshold (all-or-none) process; finally, are recollection and familiarity separate processes or a single process? Slotnick’s (2013) conclusion is that the behavioural evidence points to recollection and familiarity being a single but continuous process, based on the hippocampus.

Wixted (2007) provides an alternative narrative of the dual process theory of recognition memory, based on an unequal variance signal detection model, supported by cognitive neuroscience research. That being said, it is admitted that the model does not effectively explain reaction time data, for which a diffusion decision model is more adequate, e.g., (Forstmann et al., 2016). According to a continuous dual process theory (CDP), recollection and familiarity are both continuous signal detection processes (Cha & Dobbins, 2021; Ingram, Mickes & Wixted, 2012; Wixted & Mickes, 2010). CDP is based on the remember, know, or guess question (cf. Gardiner et al.,1998), proposing that an old or new judgment in CDP is equivalent to an UVSDT model; if recollection fails, recognition is based on familiarity (if this exceeds a high-confidence know decision criterion), otherwise, the response will be a low-confidence guess. The circular debate results in remember equating to recollection and know equating to familiarity (but only if familiarity is strong). This reprises the argument against using such questions, as if you do not “recollect” the target you have to answer “familiar,” even if the target is only vaguely familiar and even if you are just randomly guessing (when a random guess may be even faster than a “true” familiar response).

Johns, Jones, and Mewhort (2021), conclude that recollection does not have to be considered a distinct process from recognition, but proposed that they rely upon different information sources. However, as the different information sources are represented by similarity (i.e., familiarity) and source-based information (i.e., recollection), this does not advance the argument. There is also a dual-recollection theory which further divides recollection into two processes, recollection of the target, and recollection of the context (Brainerd et al,, 2015; Nieznanski, 2020). Brainerd et al., (2015) used a source memory paradigm with a remember or know judgment and either a word frequency manipulation or the use of gist-based distractors. Nieznanski (2020), used word or picture triads to separate the two processes of context or target recollection from familiarity, utilizing a paradigm loosely related to the picture-similarity experiment of Tulving (1981). The conclusion was that level of processing impacted on context recollection for pictures but not on target recollection.

Of more relevance to my experiments is the finding that for pictures, recognition can be based on features within the picture or familiarity of the scene as a whole, and that recognition of features equates with recollection of the detail, (Reinitz, Peria, Seguin & Loftus, 2011). An old or new picture recognition paradigm was used, with colour photographs of natural scenes such as landscapes and seascapes (and including cityscapes). A pilot study had shown that recognition could be due either to recollection of features within the images or to a general feeling of familiarity. Sixteen participants viewed eight blocks of 12 pictures in the study phase (comprising 6 familiarity and 6 feature-based images), shown in low contrast and for varying durations. At test, study pictures were shown at random mixed with 12 new pictures (6 familiarity and 6 feature-based). After making an old or new decision, participants rated confidence on a 4-point Likert scale from 0 (guess) to 3 (sure). Test images remained on screen until the participant had completed all responses. This fact could, however, lead to pictures that were initially recognised due to their familiarity subsequently being recollected, i.e., a long deadline provides more time and thus can differentially affect recall, (e.g., Gillund & Shiffrin, 1984; McElree et al,, 1999). Reinitz et al., (2011) considered their results were akin to those for a remember/know (or recollect/familiar) response and took this to support a dual-process theory of recognition memory. Results were also consistent with a multidimensional model of the confidence-accuracy relationship, i.e., that confidence and accuracy arise from differing sources, with feature recollection affecting certainty, which drives confidence but not accuracy. However, according to their multidimensional model, feature recollection also affected memory strength, which itself drives both accuracy and confidence.

Some-or-None Thresholded Recollection.

To test whether recollection either fails below a threshold or is a continuous signal falling to zero (as in SDT), Harlow and Donaldson (2013) used a continuous-response spatial location memory task, where a series of locations of a cross based around a circle were shown, in relation to words. Presenting the related word cued retrieval of the position of the cross on the circle some of the time, with success dependent on recollection of related source information (the location of the cross) rather than on the participant recognising the related word (again based on source memory being process-pure and not reliant on familiarity). Retrieval success was measured by response accuracy, offering a continuous assessment of the accuracy of retrieval to provide a comprehensive account of trial-to-trial variability in source memory performance. Location error frequency plots from 0° to ±180° from the target location, summed across participants, allowed a visual representation of memory performance, (Harlow & Donaldson, 2013), Figure 8.

Zhou et al. (2021), considered that the additional information in response accuracy distributions might be more diagnostic of the underlying retrieval processes than that derived from ROC curves. It can be seen that performance plateaus from ±90° to 180°. In this thresholded some-or-none memory theory, all responses from ± 90° to 180° of the target are assumed to be guesses made in the absence of recollection, revealing a baseline of sub-threshold guessing that occurs equally at all locations around the circle. Location errors within ± 90° of the target involve both recollection and guessing, with the proportion of trials associated with recollection represented visually as an increase in responses above the baseline guessing rate.

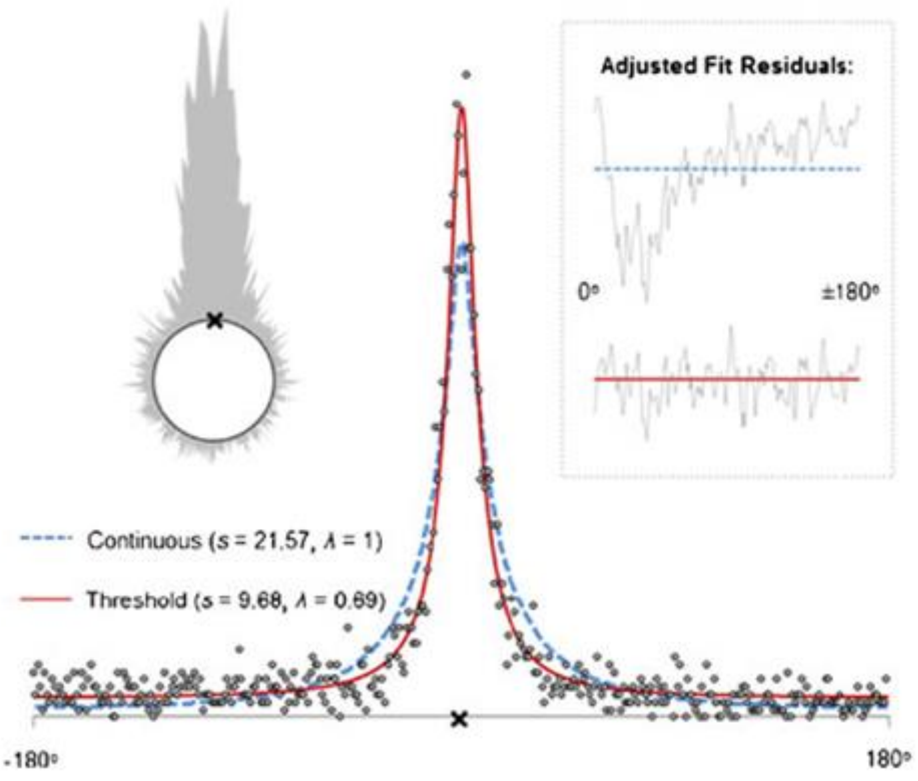

Figure 8. A plot of observed error distributions suggesting thresholded recollection, which consists of a mixture of successful recollections and subthreshold guesses that uniformly raise the distribution above zero. Responses are clustered around the target; there is significant subthreshold guessing when recollection fails. The adjusted fit residuals (inset) show that recollection is thresholded (solid red line) and not continuous (dashed blue line), as suggested by SDT (taken from Harlow & Donaldson, 2013)

Luck and Vogel (2013) considered that if the cued item was not remembered, then errors would be random, i.e., have a uniform distribution around the circle (cf. Figure 2). Murray et al., (2015) reported EEG data confirming a left parietal effect (measured using ERPs and studied as a neural correlate of recollection) associated with retrieval success but not with absence of recollection (a low accuracy or guess response), concluding that threshold memory models taking an all-or-none approach to recollection were inconsistent with these findings. It has been argued that source memory tasks, which assess the accuracy of recollection for contextual details linked to a retrieval cue, viewed as offering a theoretical thresholded some-or-none account of recollection (Zhou et al., 2021), are more appropriate than a continuous or signal-detection (UVSD model) approach. Others would disagree (e.g., Slotnick, Klein, Dodson & Shimamura, 2000).

Alternatively, Schurgin, Wixted and Brady (2020), investigating visual working memory strength, assume that the tail of responses around the target, e.g., in the colour wheel experiment (Figure 2) for short-term visual memory (Luck & Vogel, 2013), is not simply due to guessing, claiming instead that "noise" increases, and sensitivity decreases, the further the distance the displayed colour is from the target on the colour wheel, (Figure 9).

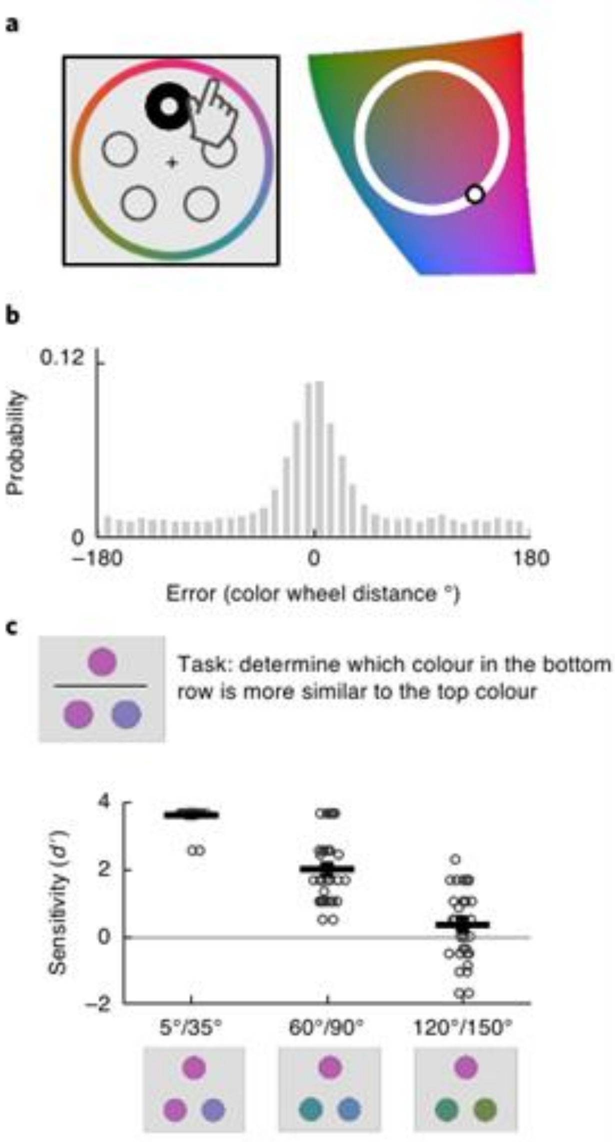

Figure 9. Illustrating: (a) a continuous response memory task using a colour wheel; (b) a histogram of results observed in such tasks, plotted as a function of distance in degrees of error from the target along the response wheel; (c) a triad scaling task where participants had to say which of two colours in the bottom row was more similar to the target colour (adapted from Schurgin et al. 2020)

However, when participants use a continuous response colour wheel, if they remember the colour was red, the participant could put their response anywhere within the red zone, leading to a substantial loss of precision (or decreased sensitivity). Schurgin et al. (2020) claim this is due to the encoded colour being most likely to generate the “maximum familiarity signal”, but that competition from other colours (especially similar colours) ensures that this will not always be the case, and the more “noise” that accumulates, the more likely a very dissimilar colour will be reported.

In their example (Figure 9c), it could be suggested that when neither colour in the task is similar to the target colour the participant will simply make a guess, unrelated to familiarity, visual memory strength or target confusability, accounting for loss of sensitivity, as can be seen in the distribution of responses in Figure 9c, or the fat tails on the error distribution graph, Figure 9b.

The data strongly concur with previous work (Harlow & Donaldson, 2013) in that responses more than 90° from the target are a result of guessing (Figure 8). Although their paper considers SDT in visual working memory, the argument of Schurgin et al. (2020) fails in this source memory paradigm, as guessing the location of a cross on the circle is due to failure of recollection and neither to a familiarity signal nor competition from other cross locations.

Zhou et al. (2021), considered that the heavy tails of responses in the location memory distribution represented errors for the recognised cues (claiming that when the cues were unrecognised participants did not search for location memory retrieval). However, the frequency of unrecognised cues was low, and their exclusion did not eliminate the heavy tails from the distribution of responses for recognised items. Assuming that participants only guess when recollection fails and they are unable to recognise the retrieval cue, they concluded that location memory retrieval was best characterised as a thresholded process (all-or-none or some-or-none), to be distinguished from threshold (discrete-state) theories of memory. Zhou et al. (2021), also analysed performance in the location memory task, finding that the best fit was when the proportion of guesses changed as a function of high or low confidence. As in SDT, their model has a decision criterion, resulting from the quantity of information retrieved to enable a response. Removing this perceptual "noise" (i.e., removing the tail of subthreshold guessing) produces a continuous trace falling to zero. This clashes with one of the original principles of SDT which bears repeating, i.e., there is a constant noise level across old and new items. In effect, guessing, or noise is expressed in a different way due to the nature of the task used. Across both theories, the basic curve remains the same whether drawn as a continuous curve falling to zero, or whether the same curve is drawn on top of a uniform baseline guessing rate. Notwithstanding, signal detection theory remains a hotly debated topic in some circles e.g., (Wilson, Harris & Wixted, 2020).

Other Memory Models.

Fuzzy trace theory.

Fuzzy trace theory has been championed by Brainerd and Reyna, cognitive psychology researchers in the USA. Based on an amalgam of dual process and false memory theories, in fuzzy trace theory (Reyna, Corbin, Weldon & Brainerd, 2016), it is hypothesized that memory is stored as either gist or verbatim memory, founded on laboratory experiments using sentences or word lists and also on everyday experiences, encompassing reality checking and source-monitoring (Reyna, 2000).

Fuzzy-trace theory proposes a theoretical division between precise (verbatim) memory and gist memory (based on inference or familiarity). As such, fuzzy-trace theory has been proffered as an explanation for false memories, the DRM memory illusion and eyewitness memory dissociations (Reyna et al., 2016). Consistent with this theory, fMRI data has demonstrated that different regions of the brain are responsible for high-confidence responses associated with true or false recognition, with correct recognition mediated by recollection regions of the brain and false recognition based on familiarity, (Kim & Cabeza, 2007).

Fuzzy-trace theory has not displaced SDT, for as Reyna et al. (2016) acknowledge, it does not offer a universal explanation across different stimuli and experimental paradigms. Questions remain as to how verbatim and gist memory can be stored and forgotten at different rates and that different retrieval cues are postulated for each type of memory trace. Arguments raised about the explanations from fuzzy-trace theory regarding source monitoring (information added to a memory by a third party, the source of which is forgotten at later recollection) are countered by Reyna (2000).

Global matching models.

Osth, Fox, McKague, Heathcote and Dennis, (2018) proposed this theory to explain source memory, i.e., the difference between the to-be-remembered item and the context in which it was learned. Many previous models have been based on comparing the shape of ROC curves, e.g., (Yonelinas, 1999). Words were used in the experimental task with source memory represented by font colour (red or green). In a global matching model, for tasks requiring item recognition, matching an item may receive its strongest contribution from its own representation in memory, while the other items on the list produce much smaller degrees of match, if they are unlike the cued item. However, in the particular source memory task used, half of the items in the list match the source cue (i.e., the font colour is either red or green), matching an item for source memory can resemble cases where half of the representations in memory bear a high similarity to the retrieval cue. Different list strengths were used and compared, using group averages and violin plots[9] assuming that there was no interference from prior memories. However, it is unclear whether this type of theory generalises to other tasks.

[9]Violin plots visualize the distribution of numerical data. Unlike a box plot, violin plots depict summary statistics by showing the shape of the distribution using a probability density function or density plot (smoothed histogram) for each variable, allowing comparison between groups.

Diffusion theory (diffusion decision model).

*Fast is fine, but accuracy is everything (**Wyatt Earp** 1848-1929).*

This theory explains decision-making processes, holding reaction time to be critical (Ratcliff, Smith, Brown & McKoon, 2016). Fast reaction times, e.g., when using online data collection, are linked to poor quality data, logically representing a trade-off between speed and accuracy in recognition tasks (Ratcliff et al., 2016; Ratcliff & Hendrickson, 2021).

Diffusion theory has been adapted to a circular diffusion model for continuous response source memory tasks by Zhou et al., (2021), who added an old or new recognition response (for the cue word) before the retrieval of source memory (the location of a cross on a circle). The frequency of unrecognised stimuli was low and guessing due to failure of recognition for the cue word (rather than recollection failure for source memory) did not fully account for the heavy tails of the error distribution (Figure 8). In other words, the participant can recognise the cue but not recollect source memory. The circular diffusion model assumes that precision (accuracy) is based on two factors, the quality of information retrieved, and the amount of evidence used to make the decision (decision criterion). However, see Onyper et al., (2010) for an alternative account based on a continuous SDT process for recollection and a dual-process theory of recognition memory.

Global memory models.

Similarly, other memory theories only deal with specific experiments and do not generalise. Global memory models have been criticised as being inconsistent (Ratcliff, Sheu & Gronlund, 1992), being based on recognition tasks. As an example, MINERVA2 has been used to explain the DRM false memory illusion and test-pair similarity effects i.e., Tulving's (1981) picture similarity experiment, (Arndt & Hirshman, 1998; Hintzman, 2001). This mathematical model hypothesises that retrieval cues activate several memory traces simultaneously, and that the sum of the values of the activated traces (rated according to their similarity to the cue) leads to an echo-intensity judgment, or familiarity/similarity judgment (with familiarity used as a surrogate for memory strength, as in SDT, see Figure 6). The choice in a 2-AFC task is made by selecting the item with the highest echo intensity. Ratcliff et al., (1992) tested global memory models of recognition tasks using ROC curve analysis, leading to the conclusion that existing global memory models needed revision.

(iv) Metacognition and Judgments of Memory Accuracy

Self-evaluation of memory is termed metacognition, referring to how we monitor and assess our own performance, i.e., being aware of our own memory processes. As the aim of my thesis is to address the question of what makes us consider our memory to be accurate, an understanding of metacognition is crucial.

In the model of Nelson and Narens (1990), the meta-level of monitoring and control is informed by, and can control, task performance. One important issue with metacognition is that many researchers regard it as predominantly relating to confidence in our performance in a memory task. There are alternative definitions of confidence, besides representing trust (belief) in the accuracy of our memory. Confidence may be described as our certainty that a chosen course of action is the right choice and that we are able to perform that action. In the latter case we might use our judgment of confidence prospectively, before attempting a task (Rhodes, 2019; Siedlecka, Paulewicz & Wierzchon, 2016). In other words, by means of our metacognitive processes (such as our confidence in carrying out a task) we can decide whether or not to attempt the task, exercising control by the meta-level of cognition (cf. McDonough, Enam, Kraemer, Eakin & Kim, 2021). In tasks where decisions can be withheld, metacognitive confidence judgments can be monitored, and the task response controlled. As an example, participants viewing a video of a mock crime were more likely to withhold low as compared to high confidence responses, and withheld responses were less likely to be correct (Evans & Fisher, 2011). As a further example that we have all experienced, we do not put our hand up in class to answer a question unless we are confident in our response, for fear of looking stupid (unless we have so much self-confidence we do not consider failure).

*All you need in this life is ignorance and confidence, and that success is sure. (***Mark Twain** *1835-1910***)**

Judgments of confidence in long-term memory apply equally to semantic memory as to episodic memory; although both self-confidence and metacognitive abilities may differ between individuals, they are considered stable within individuals (Brewer & Sampaio, 2012). For this reason, it is preferable to avoid between-subjects averaged comparisons when analysing confidence and other metacognitive judgments.

In experimental tasks this type of confidence may be termed a judgment of learning, i.e., how sure are you that you will be able to remember the word list correctly? Such confidence, or a feeling of knowing, rated prospectively has a weaker relationship with accuracy than when confidence is judged retrospectively, Siedlecka et al., (2016). This is only logical, “can you do it?” compared to “did you do it?” An individual’s confidence may depend more on individual differences such as their self-confidence rather than an individual’s prospective or retrospective metacognitive judgment of confidence.

More commonly, when referring to confidence, we mean a participant’s retrospective judgment of their performance in an experimental task or their assessment of the accuracy of a retrieved memory. When an individual is judging their confidence or trust in a memory, it is speculated they will assess memory strength and base their confidence rating on this (Busey, Tunnicliff, Loftus & Loftus, 2000). If confidence ratings are founded on memory strength, it comes as no surprise that confidence is considered to reflect accurate recollection.

Confidence.

By convention, confidence is reported as a retrospective metacognitive judgment of task accuracy and decision-making (Fleming & Dolan, 2012), to subjectively assess how well we think we performed in a task. Confidence may give the feeling of certainty that we have done well (e.g., we think we have performed well in an exam) or just knowing that something is true (e.g., we know we sat the exam). Confidence judgments might differ dependent on the experimental task, i.e., it has been suggested that in a 2-AFC task, ratings of confidence may be relative, based on the difference in memory strength of the two test items; on recognition of the target item as old; or on discrimination between the two items as to which is old, (Zawadzka et al., 2017). Relative confidence (e.g., we are confident we have done better in the exam than another person who did not revise for the exam), has been argued as not being the same as absolute confidence, i.e., in a standard old or new recognition task. Whether or not this is correct is debatable, it is still confidence, but confidence is not an all or none metacognitive judgment, it is continuous and varies in strength. However, relative confidence seems to be a straw man argument. A strong relationship (between confidence and accuracy) may also depend on how difficult (or easy) a task seems to participants. For example, Geraci, McCabe, and Guillory (2009) removed participants who were considered not to have understood their sure or unsure instructions

The confidence-accuracy relationship.

We trust our memories when we feel confident about them. My research questions, if our feelings of trust, belief or confidence do not correspond to the truth of the memory, what causes this dissociation? Whether the objective accuracy of memory is always strongly and positively related to how confident the individual is that their memory is correct, is a topic of debate in cognitive psychology, particularly in relation to eyewitness testimony (e.g., Brewer & Wells, 2006; Jaeger et al., 2020; Wixted & Mickes, 2015; Wixted & Wells, 2017), and false memory (e.g., Roediger & DeSoto, 2014). It has been acknowledged by some that eyewitnesses may be highly confident but inaccurate, and the use of predictive variables, e.g., time to decision and responses based on familiarity have been suggested (Grabman et al., 2019). However, others seek to disagree (Brewin, Andrews & Mickes, 2020; Semmler, Dunn, Mickes & Wixted, 2018), holding fast to the conviction that highly confident eyewitness recognition is highly accurate. Accounts of false memories and confidence-accuracy dissociation have been challenged, suggesting instead that confidence and accuracy remain tightly coupled, (e.g., Brewin et al., 2020; Roediger & Tekin, 2020; Wixted & Stretch, 2000).

Challenges to this view are claimed to be coloured by “methodological artefact and lack of external validity” (Brewin, Andrews & Mickes, 2020). The basic problem in this debate (apart from the narrow-minded opinion of some research groups) is the fact that our confidence is calibrated, not against objective accuracy (how accurate we were), but against subjective accuracy (how accurate we think we were). Those claiming the relationship between confidence and accuracy is weak, (Busey et al., 2000), find their conclusions dismissed as due to inappropriate methods of analysis, such as using point-biserial correlations to measure the strength and direction of the association that exists between a continuous and a dichotomous variable representing a yes or no decision in a recognition task (Brewer & Wells, 2006). However, the same argument can be used against those proponents of a strong link between confidence and accuracy (see below & Chapter 3).

Jaeger et al., (2020) using a source memory task, based in part on location (left or right side) found that confidence for correct responses was higher than that for errors, thus indicating a strong relationship between confidence and accuracy. Wais, Squire and Wixted, (2010) [commented on by Diana and Ranganath (2011)], used a word-based source memory task, asking participants to respond yes or no for each word to either how *common* the word was, or whether it described something you would *discuss* with a friend. Recognition was assessed three hours later, using thirty-two targets and eight lures arranged in six blocks of forty words each. The old or new decision was combined with a confidence judgment from 1 = definitely new, to 6 = definitely old. Source memory was judged on the same type of scale from 1 = definitely *discuss*, to 6 = definitely *common*, responses of 3 or 4 being considered as source memory guesses. Analysis was carried out on a by participant basis ($N = 16$). Wais et al., (2010) claimed that solely using high confidence old or new decisions with sub-groups representing correct or incorrect source memory decisions was justified to remove the "confound of memory strength". They found no significant difference between source correct or incorrect judgments for confidence ratings of 5 and 6 combined, however, when adding in confidence ratings of 4 (a source memory guess) the difference became significant ($p < .05$). This type of post-hoc analysis may lead to erroneous conclusions. In some research it can be reasonable to remove participants, e.g., Geraci, McCabe, and Guillory (2009) removed participants who were considered not to have understood their sure or unsure instructions when rating confidence.

One problem, highlighted by these examples, is that conviction that a certain theory is correct may lead one, knowingly or unwittingly, into ensuring the results or the conclusions fit your hypothesis; this is not objective scientific research. Whilst not entirely disputing a strong relationship between accuracy and confidence, we will all know from personal experience that we may not always be 100% correct when we are 100% confident and we may not always be 100% confident when we are 100% correct. We will have all confidently recounted a holiday anecdote, only to be told by our partner that we have misremembered the date, or the place. In other words memory for the detail is lacking or incorrect (i.e., source memory recollects the detail inaccurately, as compared to eyewitness recognition memory inaccurately recognising the wrong person). To claim that when we are highly confident we are most likely to be highly accurate does not answer the question: do we know when we are highly confident and when we are not highly accurate?

The direction of the relationship between confidence and accuracy is ambiguous (Roediger & DeSoto, 2014) and may depend on the experimental task used. Jaeger et al., (2020) maintained that dissociation of confidence and accuracy could depend on post-encoding manipulations, such as providing additional evidence in the form of true or false cues. Similarly, information sources out with task performance, representing the accumulation of evidence and information on error detection that occur once a decision has been made, might also be involved in metacognition (Siedlecka et al., 2016). This could lead to biases (i.e., heuristics) influencing participants' metacognitive judgments (Wilson & Brekke, 1994).

Jacoby, Woloshyn and Kelley (1989), concluded that confidence in memory occurs when any specific details are recollected (even if they do not apply to the target stimulus). Chandler (1994) also found that recollection of target information influenced confidence, asking if the same variables that influenced confidence also determined accuracy. He suggested that confidence and accuracy would dissociate if a variable were to influence accuracy without the participant being aware of the fact. However, this could be equally true if a variable were to influence confidence without the participant being aware of the fact.

Analysing the confidence-accuracy relationship.

Using a binary old or new recognition paradigm limits analysis of the relationship between confidence and accuracy, as it gives only two results for accuracy: yes or no, old or new. It is not possible to examine the relationships on a trial-by-trial basis, and accuracy results must therefore be averaged across groups. If confidence is assessed by asking a participant to use a 3- or 5-point Likert scale (Kensinger, Addis & Atapattu, 2011), this similarly reduces the amount of information for analysis. A Likert scale uses arbitrary divisions of confidence into low, medium, or high from which participants must choose the one option that best aligns with their view. The basic 5-option answer has choices from 5 to 1, e.g., representing: very confident, confident, neutral or undecided (unsure), not confident, not at all confident. There may be a lot of difference between someone who is undecided and someone who is neutral. Wu & Leung (2017), debate the analysis of Likert scales in social psychology, recommending the use of 11-point scales to deal with assumptions of normality and ease-of-use for participants. There is a paucity of data when using an ordinal format compared to continuous scales (Kandasamy, Kandasamy, Obbineni & Smarandache, 2020). Responses may depend on uncertain, inconsistent, imprecise, or indeterminate factors, and has been shown to be affected by providing feedback on correct or incorrect answers. Bornstein and Zickafoose (1999), studying eyewitness and general knowledge memory domains (using 4-AFC questions followed by a 5-point confidence scale, calculated an overconfidence score by subtracting the average confidence score from average accuracy for each participant.. It was admitted that very few responses were rated as 1 on the confidence scale. Overconfidence correlated between the two memory domains, in that feedback regarding overconfidence on general knowledge questions reduced average confidence scores for episodic memory questions. It should be pointed out that overconfidence might not be consistent across all trials for an individual participant. Averaging confidence and accuracy and subtracting one from the other assumes the overconfidence score is stable for an individual across all responses and may lead to erroneous inferences. It also raises another important question as to whether anyone should ever attempt to average responses to ordinal (Likert) scales. Unfortunately this is common practice, because you can readily calculate the average of 1,2, and 3, however, it is impossible to average “agree”, “disagree”, and “neutral. It is a moot point whether the averaged answer to a Likert scale provides a meaningful result, although admittedly may be used for comparison purposes.

Continuous 0 – 20, or 0 –100 response scales are commonly used to assess confidence, but one problem with such sliding scales is that of ‘anchoring’ (Delay & Wixted, 2020). In a quantitative estimation task, providing participants with a pre-determined anchor, (Jacowitz & Kahneman, 1995) could lead to some participants treating the anchor starting point (e.g., 0, 10, 20) as providing relevant information. In the same manner, previous answers could be used as anchors for subsequent responses see Epley & Gilovich, (2006) for further discussion of this problem].

Conclusions may depend not only on the task and how it is designed, but also on the way the results are analysed. For example, if you use a 0 – 100 continuous confidence rating scale in conjunction with a binary old or new recognition task, you will be plotting a continuous variable of confidence against a dichotomous variable of accuracy. Taking a mean value of confidence for correct or incorrect answers to enable a point-biserial correlation loses a lot of data, as Figure 10 demonstrates.

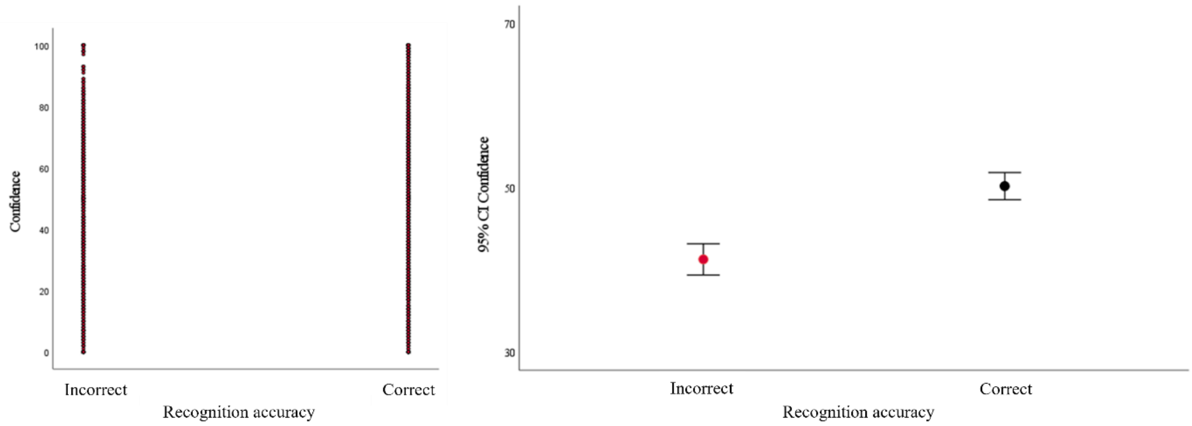

Figure 10. Plots for 2420 individual trials in a binary task, taken from Vividness, Experiment 3. (left): a bivariate scatter plot of a continuous vs. a dichotomous variable; (right): a point-biserial plot of the same data using mean values of the continuous variable (error bars = 95% C.I.). This tells you that mean confidence is significantly higher for correct responses, while concealing the fact that you can get high confidence incorrect and low confidence correct responses

Psychology experiments with high trial numbers may produce error data that is centred nicely around the mean, and therefore achieve statistically significant effects. However, the measure of spread is deceptive when the number of trials is high.

Receiver operating characteristic (ROC) curve analysis.

In situations such as that illustrated in Figure 10, receiver operating characteristic (ROC) curve analysis is commonly used. The ROC curve is a two-dimensional graph in which the false positive rate is plotted on the x-axis and the true positive rate is plotted on the y-axis, i.e., a graphical plot of how often false alarms (x-axis) occur versus how often hits (y-axis) occur for any level of sensitivity (or decision criterion), calculated for each retrospective confidence rating made by the participant on an arbitrary 3- or 5-point ordinal scale. Paralleling examples used in SDT (when comparing detection of complex memory traces to the detection of simple perceptual signals of light or sound), another example cited to justify the ubiquity of ROC curve analysis in psychology is their use for discrimination between medical diagnostic tests. Shaw, Pepe, Alonzo & Etzioni (2006) comment that “the value of ROC methodology in evaluating the accuracy of continuous markers is well recognised” (p.7). In this respect, false positive and false negative rates are crucial and dictate the more sensitive test, depending on how far the curve lies from the by chance straight line. Wixted, 2020) describes ROC curves as “nothing more than a plot of the hit rate versus the false alarm rate across different levels of response bias, holding discriminability constant” (p.213, Figure 11).

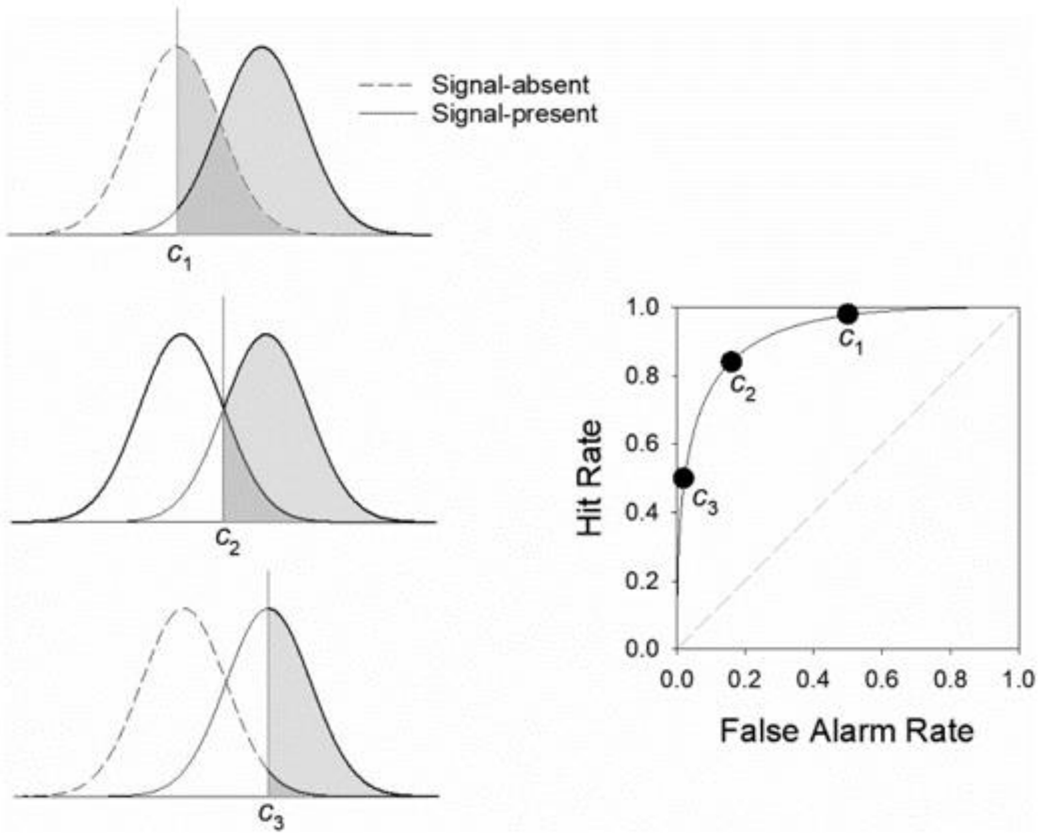

Figure 11. Signal detection interpretation of ROC data. As the criterion moves from liberal (C1) or low confidence to conservative (C3) or high confidence, both the hit rate and the false alarm rate decrease, i.e., the shaded regions to the right of the criterion (taken from Wixted, 2020)

Explaining ROC curves to surgeons, Carter, Pan, Rai and Galandiuk, (2016) suggest that when an ordinal predictor variable is used, multiple pairs of (x, y) are possible, leading to multiple points on the ROC curve, and giving a staircase (stepped) appearance. In contrast, a continuous predictor variable (such as serum marker levels in a test for cancer) results in a ROC curve with a smooth appearance. The smooth ROC curve shown in Figure 11, using a 3-point Likert scale for confidence should therefore accordingly be shown as a stepped curve. Furthermore, the ROC curve in Figure 11 informs me that high confidence (C_3) is associated with a lower hit rate, albeit with almost no false alarms, than low confidence (C_1). Although the low confidence judgment is associated with more false alarms, this interpretation could not have been the intention of the author, who counters this by claiming the critical factor is the ratio of hits to false alarms. In the C_1 condition, if the response “old” is made to every stimulus, all the old items will be correctly identified, but false alarms are high. In SDT (cf. Figure 6, Banks,1970), C_E is a lenient decision criterion, and it is not clearly explained how this is equated with low confidence C_1 in Figure 11.

There are some issues with the use of ROC curve analysis in confidence-accuracy evaluation, with confidence ratings (or decision criteria) being placed along a strength-of-decision axis. Again, this means that retrospective metacognitive judgments of confidence are treated as equivalent to a rating of memory strength. Stretch and Wixted (1998), discuss how confidence criteria change when recognition accuracy is manipulated (such as by changing familiarity of the targets by allocating more time to the target in the study phase). By treating memory strength, accuracy, and confidence as the same thing you will produce a strong relationship, when it may not always be the case that confidence moves in a manner corresponding to a decision criterion, i.e., in lockstep. Accounts of ROC curve analysis are often far from clear. Explanations of how they are calculated may be lacking, otherwise too much detail is provided using complex mathematical equations, and their interpretation may go unchallenged.

Confidence accuracy characteristic (CAC) analysis

Confidence accuracy characteristic (CAC) analysis is proposed as an alternative to ROC curve analysis, to avoid the problems inherent in their interpretation, as certainly CAC analysis is more intuitive, by directly relating, e.g., identification accuracy to eyewitness confidence (Mickes, 2015). Absolute (rather than relative) accuracy is represented using calibration curves, calculated by dividing metacognitive confidence judgments into bins, i.e., 0-10,11-20… and determining the average accuracy for each bin. Plotting the results is equivalent to a calibration (Palmer, Brewer, Weber & Nagesh, 2013). This method of calibration equates with the confidence accuracy calculation (CAC) analysis proposed by Mickes (2015), assuming that participants can reliably rate their psychological experience on to the scale used, (Rhodes, 2019). In earlier research, Bornstein & Zickafoose (1999) also plotted calibration curves for average predicted confidence against accuracy for responses to both eyewitness and general knowledge questions.

Recently, and relevant to how my results from the recognition memory task used are analysed, Tekin, DeSoto, Wixted and Roediger, (2021) demonstrated how a CAC plot can be adjusted for use in a binary old or new discrimination task, as a direct alternative to ROC curve analysis, Figure 12. The authors used two different methods of calculation: either proportion correct = hits ÷ (hits + misses), or response-based accuracy, where proportion correct = hits ÷ (hits + false alarms) to create the plot, in either case finding that confidence was always highly related to accuracy, with either related or unrelated lures.

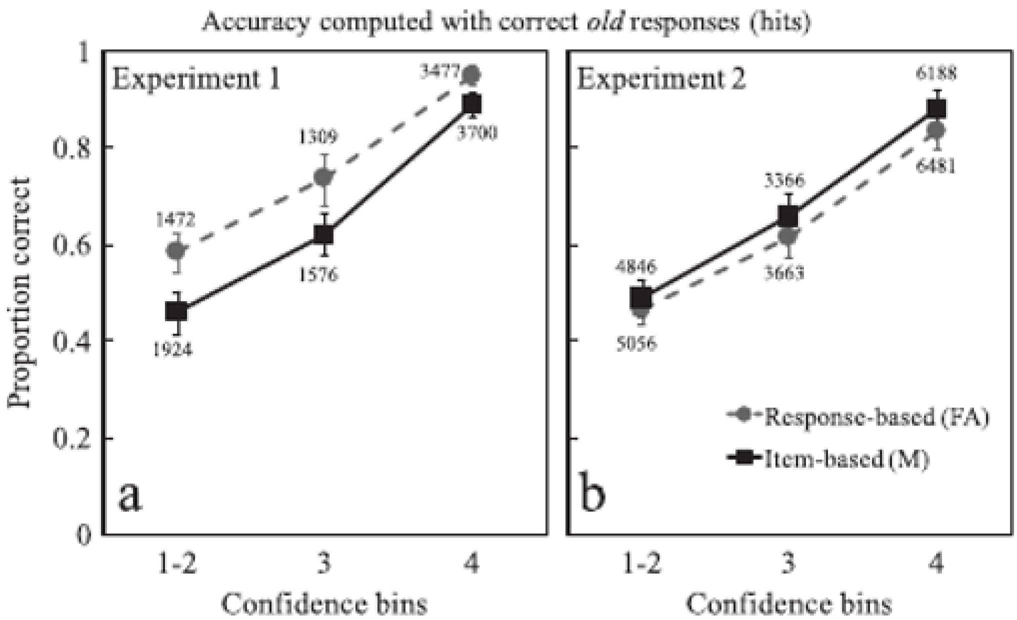

Figure 12. CAC plot of accuracy in an old or new recognition task against confidence; (a) with unrelated lures. (b) with related lures. Error bars indicate 95% confidence intervals. Response-based accuracy (dashed lines) was calculated as hits/ (hits+ false alarms), (taken from Tekin et al., 2021).

Illustration of the results of CAC analysis may be open to manipulation by simply adjusting the size of the confidence bins. Roediger and Tekin (2020) are of the opinion that a strong relationship between confidence and accuracy could be obscured by binning methods used in binary experimental tasks, at the same time pointing out that CAC plots (using 3 or 4 bins) show much higher correlations between confidence and accuracy. Tekin and Roediger (2017) reported no difference in CAC plots when binning data from either 100 or 20-point scales. This may reflect the binning method employed rather than the scales used for rating confidence. Otherwise, it could be assumed that participants are unable to separate 20 or 100 different degrees of confidence they may have, the result being that they, in effect, bin their own confidence ratings. Mickes (2015) used results from an earlier eye-witness experiment where participants had to identify a research assistant viewed at a distance of 10 metres who appeared for either 5 or 90 seconds. Confidence ratings were collected on a scale from 0 – 100 but binned into decile response options by Palmer et al. (2013), Figure 13. It was concluded that accuracy increases with confidence (particularly at the upper end of the confidence scale).

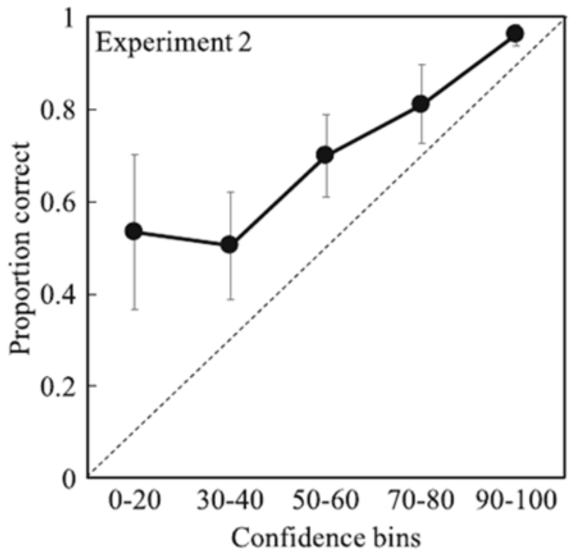

Figure 13. Calibration curves for choosers in the exposure duration conditions. Dotted lines denote perfect calibration. Error bars denote SEs (taken from Palmer et al., 2013)

In Mickes (2015), the same data are firstly illustrated with confidence in 5 bins, reproducing Figure 13. Next the data are plotted with confidence in 3 bins, Figure 14, under explanation that “noise” (i.e., unwanted, or inconvenient results) is reduced by combining the two lowest confidence bins. As the error bars overlap, there is no statistically-significant difference between the 0 – 20, 30 – 40, and 50 – 60 bins in Figure 13. In Figure 14, it now appears that for confidence ratings binned from 0 – 60, a difference may exist between the two conditions. This manipulation of confidence bins hides the previous finding that at low levels of confidence, accuracy increases in the 5s condition (one explanation being that participants are guessing. When data is collapsed into bins, it can disguise the 'true' picture.

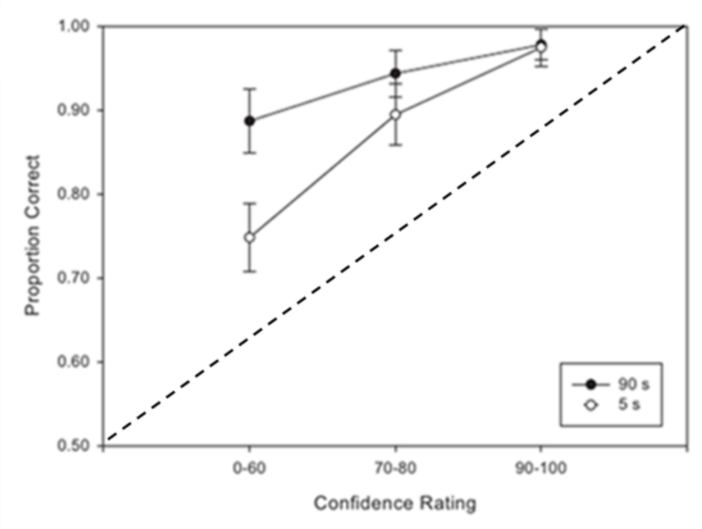

Figure 14. Confidence-accuracy correlation (CAC) curves for the exposure duration conditions with the lower confidence levels collapsed; the bars represent standard errors (taken from Mickes, 2015). The dashed line corresponds to the "calibration curve"

Chandler (1994) theorised that confidence ratings were not predicated by memory strength alone and could be affected by other (metacognitive) factors. The pitfalls in treating judgments of memory strength as equivalent to confidence (from an SDT standpoint) is further illustrated by Mickes, Wixted and Wais, (2007). Using a standard 150-word old or new recognition memory task, they asked 14 participants (one subsequently excluded) to enter a number ranging from 1 to 20 on a keypad to rate their memory strength and to indicate if the word was previously shown: 1 meaning that the word was definitely not on the list (thus they had no memory for it) and 20 meaning that the word was definitely on the list. Mickes et.al., (2007) gave participants additional verbal instruction, asking them to use 1 and 20 "only when they were 100% certain, one way or the other" (p. 859). Similar instructions have recently been used to investigate the relationship between familiarity, confidence, and memory accuracy (cf. Wais et al., 2010; Cha & Dobbins, 2021).

Response frequency data was plotted as proportion correct on the y-axis with memory strength or 'Rating' on the x-axis. This data, originally used to support an unequal variance SDT (adjusting SDT to explain the experimental results), is transformed over the course of several iterations (i.e., Mickes, Hwe, Wais & Wixted, 2011; Delay & Wixted, 2020; Wixted, 2020), subsequently represented as a line graph with confidence (rather than memory strength – assuming they are the same) on the abscissa, Figure 15.

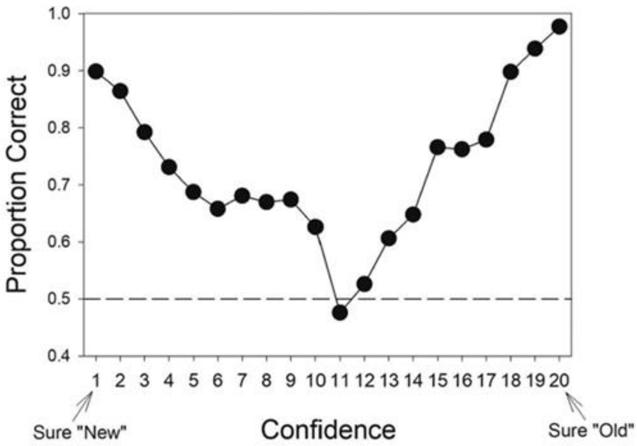

Figure 15. Accuracy (proportion correct) as a function of the confidence expressed in an old or new recognition decision (where 1 = sure new and 20 = sure old), (taken from Wixted, 2020)

The confidence scale is ambiguous, as a "sure new" judgment is rated as 1 on a confidence scale of 1-20; (originally rated on a memory strength scale, using 1 but only if the participant was 100% certain), now represented as a low confidence judgment with high accuracy. This serves to illustrate an important example of dissociation or inconsistency between confidence and memory strength ratings. If a "sure new" judgment means there is no memory strength for the old (target) item and participants used 1 "only when they were 100% certain (confident), how then does confidence equate with memory strength (unless you assume that no memory for the item can also mean high confidence)?

The condition under which participants are instructed to use confidence scales is further highlighted by Delay and Wixted (2020), using their example of a 20-point confidence scale, Figure 15. Even when there is no memory for a word, participants can be confident that it was definitely not on the list if it is not recognised, not familiar, and not recollected. To the uninitiated Figure 15 again demonstrates that low confidence judgments are accurate. This would be clarified if the two parts of the curve were illustrated separately, the left-hand side referring to lures and the right-hand side to targets; this way, it would illustrate appropriately that increasing accuracy exists with increasing (and not decreasing) confidence.

(v) Metacognition and Judgments of Memory Quality.

*Cognitive psychology tells us that the unaided human mind is vulnerable to many fallacies and editions because of its reliance on its memory for vivid anecdotes rather than systematic statistics. (**Steven Pinker**)*

So far, I have limited my discussion of metacognitive judgments of memory to judgments of memory quantity (represented by memory strength, judgments of confidence, and memory accuracy). Researchers in cognitive psychology often consider metacognition to be all about confidence, e.g., in how accurate we were in conducting a task or making decisions (Fleming & Dolan, 2012; Fleming, Weil, Nagy, Dolan & Rees, 2010). However, studies of human memory have shown that confidence in our memory is only loosely related to how accurately we remember. One factor in this may be the memory's quality.

Characteristically, metacognitive judgments of the quality of autobiographical (episodic) recollection involve the richness of detail remembered and the sense of reliving conferred by the memory (Rubin, Deffler & Umanath, 2019). When we judge our memories, we do not merely assess their accuracy by using metacognitive judgments of memory strength, i.e., confidence, but we also assess their quality (represented by remembered features or the detail of the memory). It is suggested that confidence may also be a judgment of memory quality (Jersakova, Allen, Booth, Souchay & O'Conner, 2017). In a cued-recall task, participants based their confidence in a memory's accuracy on the vividness of the memory, as well as the detail of recollection, i.e., on the quality as well as the quantity of their recollective experience (Robinson, Johnson & Robertson, 2000).

Individuals may use heuristics, i.e., fluency, and other concepts and beliefs, before making their confidence decision, suggesting one mechanism by which judgments of memory quality can influence the confidence-accuracy relationship. As an example, Busey et al., (2000) demonstrated that manipulating luminance between study and test showed that faces studied dim and tested bright confirmed participants' belief that a brighter test face produced better accuracy, thus causing a shift in confidence. In this case it follows that the confidence-accuracy relationship can be disrupted by confounding factors, i.e., metacognitive judgments of memory quality.

The quality of remembering is theorised as increasing belief in the memory, although, rarely, such memories can also be non-believed (Mazzoni, Scoboria & Harvey, 2010). Not only do we have to consider the effect of individual differences on the relationship between confidence and accuracy, but that confidence in memory's accuracy may depend more on its subjective and perceptual quality (Rhodes & Castel, 2008), than on its objective accuracy. Personal experiences (i.e., remembering in response to the remember or know question) that occur during a recognition task are thought to be a major factor in generating confidence judgments, in other words it is reasonable to assume that multiple memory processes may be responsible for judgments of memory confidence (Brewer & Sampaio, 2012).

Our judgments of memory accuracy may be based on factors which influence confidence but not accuracy, leading to dissociation, in effect creating an illusion of memory accuracy. My basic research question is whether quality judgments have a stronger influence on confidence in the memory's accuracy than actual accuracy itself. In my thesis, I speculate on whether memory quality is responsible for examples of dissociation between confidence and accuracy, and if memory quality is involved in the subjective experience of remembering.

It may be that we base our judgments of belief in a memory, not on its accuracy, but on how real it seems, by the richness or the quality of the experience of remembering, i.e., the vividness of the memory. Individuals may evaluate the memory's reliability (Zylberberg et al., 2014), which may in part be based on other factors, such as reality monitoring, using metacognitive judgments of quality such as vividness to aid retrieval monitoring and reduce errors (Gallo, 2013; Scimeca, McDonough & Gallo, 2011; Wong, Cramer & Gallo, 2012).

I will next introduce the two judgments of memory quality I intend to study, vividness and distinctiveness. Why these two? People readily describe their memories as vivid or distinctive (just listen to the news or read the papers and you will come across many examples of this); this being the case, it is more likely that participants in my experiments will understand what is meant by asking them to judge the vividness or distinctiveness of their memory.

Vividness.

When we consider our memories of our past personal events, how do we really know that the events happened just as we remember them? In my thesis I ask whether the vividness of our memories is related to confidence in their accuracy, and whether vividness is also strongly related to a task's factual accuracy. To the lay person, memory vividness signifies memory accuracy. However, in contrast to lay understanding of memory, it is claimed that the detail of vivid memories can be lacking (Fitzgerald, 1988; Heuer & Reisberg, 1990).

Neuroimaging research using fMRI demonstrates that vivid visual imagery or imagined events can lead to false remembering (Gonsalves, Reber, Gitelman, Parrish, Mesulam & Paller, 2004), i.e., was it a dream, did that really happen? Johnson, Hashtroudi & Lindsay (1993), presenting a framework of source monitoring, suggest that memory judgments can be based on heuristics, whereby the richness (associated perceptual detail) or quality of the memory can override organised processes. When asking if memory quality supports confidence more than it supports accuracy, current literature in cognitive psychology maintains a role for vividness and distinctiveness in memory accuracy (Hunt, 2013; Scimeca, McDonough & Gallo, 2011; Thakral, Wang & Rugg, 2015).

We commonly use vividness to describe a memory, basing our judgments of belief in the accuracy of a memory on how real it seems, referring to the richness or the subjective quality of the experience of remembering. When recalling a significant event from our past, if we refer to the memory as seeming vivid, it may feel as if we are reliving the experience (Buchsbaum, Lemire-Rodger, Fang & Abdi, 2012; Rubin, Deffler & Umanath, 2019). More recently, the role of the vividness of a memory and confidence in the memory's accuracy to provide the experience of recollection, has been investigated using fMRI procedures and source-based memory tasks, (Tibon, Fuhrmann, Levy, Simons & Henson, 2019; Zou & Kwok, 2022), concluding that the angular gyrus is involved in the vivid recall of episodic memory. As it is also considered that confidence produces fMRI activation in the hippocampal and perirhinal cortex (de Zubicaray, McMahon, Dennis & Dunn, 2011), one might speculate that the vividness of a retrieved memory trace or engram is associated with a rich conscious experience of remembering and thus enhances confidence in the accuracy of the memory (Richter et al., 2016). However, the angular gyrus is also thought related to the memory accuracy. What remains unanswered is whether or not confidence in the accuracy of the memory, bolstered by vividness, correlates with factual accuracy of the memory.

What makes a stimulus or memory vivid?

Vividness can have differing meanings; as well as referring to the external features of a stimulus, it refers to the subjective quality of a memory. Both may be of interest, as depending upon the task used and on what criteria vividness is judged, an object may be recollected vividly because it had a bright colour, or it activated recall of an experience, thus it may also be recollected vividly because it had personal significance or provoked an emotional reaction (Cooper, Kensinger & Ritchey, 2019; Ford & Kensinger, 2017; Habermas, Diel & Welzer, 2013; Kinugawa, Schumm, Pollina, Depre, Jungbluth… & Dere, 2013).

I have a vivid recollection of being in a class at primary school, on that day the teacher asked the pupils in turn to give the names of animals beginning with the letter C: cat, cow, chicken, camel... When it was my turn, I answered capybara, and was told I had made it up, despite my maintaining it was a large South American rodent. The teacher refused to believe me, and it was at that moment I first realised that teachers do not know everything. Although I have no certain recollection of detail, such as what gender the teacher was, or what time of year it was, and can remember nothing else about that lesson, I am confident that the memory is correct, and can re-experience my sense of frustration at being disbelieved. This type of vividness, the vivid remembering or reliving of personal memories, has been categorised as internal vividness by Ford and Kensinger (2016).

To further illustrate the point, I highlight the literature on flashbulb memories. These are detailed and highly vivid memories, creating a snapshot of the time and circumstances in which the individual heard about a public disaster or a traumatic world event that evoked an emotional response. Memories of such events are claimed to be detailed, long-lasting, and vivid by the person recollecting them (Brown & Kulik, 1977; Edery-Halpern & Nachson, 2004; Luminet & Spijkerman, 2017; Pillemer, Goldsmith, Panter & White, 1988). Regardless, the detail associated with flashbulb memories may be inaccurately remembered (Roediger & Gallo, 2004). Support for this view can also be found in a phenomenological study, Levine, Lench, Stark, Carlson, Carpenter… & Frithsen, (2020), who, examining judgments of an emotional experience, commented, “vividness was not a particularly reliable guide to the actual accuracy of predicted and remembered emotion” (p.136). Because flashbulb memories may be perceived as vivid, and confidently recollected, even if contextual details are incorrect or lacking (Coan,1997; Doss, Picart & Gallo, 2018;2020; Felstead & French, 2021; Talarico

& Rubin, 2003; 2007), it has been suggested that when our memories of these events are vivid it is the gist we recall and not the detail (Heuer & Reisberg, 1990). Newman and Lindsay, (2009) report:

> "people make a variety of autobiographical memory errors, such as misremembering what they used to be like, misremembering details of momentous events, mistakenly believing they experienced events that they did not, and, in some cases, developing full-blown false memories of events that never happened" (p.1106).

Not only does this suggest that the importance of more than one aspect of vividness in memory studies needs to be taken account of, but also provides evidence that vividness can affect confidence, and accuracy is not always associated with vivid memories, providing the basis for my thesis.

Vividness & accuracy.

It is accepted that the more vivid or distinctive a stimulus is, the more accurately it is recognised (Jacoby, Baker & Brooks, 1989; Wichmann et al., 2002). The current literature shows varying degrees of association between memory vividness and memory accuracy. Autobiographical studies have shown no relationship between vividness and recall of accurate details for personally significant memories (Habermas & Diel, 2013). Using a continuous recognition procedure, colour was found to enhance recognition by about 5% for natural scenes (Spence, Wong, Rusan & Rasteger, 2006). Richter et al. (2016), using fMRI and a continuous retrieval task based on a colour wheel (cf. Luck & Vogel, 2013; Schurgin et al., 2020) found pairwise correlation between vividness and accuracy to be moderately positive. In recognition tasks, performance could rely on familiarity or implicit memory processes. It is worth noting that the differences between the tasks in these studies may modulate the relationship reported between vividness and accuracy.

An important point regarding the relationship between vividness and accuracy is that judgments of vividness may also rely on the vividness of visual mental imagery (e.g., Cuperus, Laken, van Schie, Engelharde & van den Houte, 2019; Ford & Kensinger, 2016; 2017), by asking participants how vivid remembered pictures seem when cued by a label or description previously shown with the image at encoding, representing another form of vividness judgment. This approach is highly dependent on an individual's capacity for mental imagery, which at its most extreme is termed aphantasia, shown to impact memory (Jacobs, Schwarzkopf & Silvanto, 2018; Keogh, Pearson & Zeman, 2021).

Distinctiveness.

Why should I investigate distinctiveness in my studies? It has been shown that distinctiveness may be related to confidence, coinciding with my hypothesis of memory quality influencing confidence more than accuracy. Metzger (2006) established that participants rated confidence in recognition judgments significantly higher for distinctive faces, even after delays of up to 6 weeks (when memory performance had declined). This provides evidence that dissociation between confidence and accuracy can occur in the context of distinctiveness. I suggest that although I am using natural scenes pictures and not faces, the findings support my examination of distinctiveness, acting as another quality of memory which may impact on the relationship between accuracy and confidence.

Distinctiveness, found to increase recognition accuracy, may differ from vividness in some respects, but it also overlaps with vividness. The word salient has commonly been misused to signify distinctiveness. Hunt (1995) described distinctiveness as "a descriptive term for events that violate the prevailing context - that is, for events that are perceptually salient" (p.106). Later, Hunt (2006) wrote, "Indeed, salience and distinctiveness sometimes are used interchangeably in psychological literature, a practice that has been encouraged by the isolation paradigm[10]" (p.5); salience meaning that an item is most noticeable, prominent, eye-catching, dramatic, remarkable, interesting, or important; also explained as something of significance that merits the attention given to it.

Salience, in this respect does not simply refer to a word in a different colour or conceptual category, or a flower in a different colour that attracts attention just because it is unlike the other items presented with it. This, however, **is** distinctiveness, a quality or characteristic that makes something different from others in a way that is easy to notice. Synonyms for distinctiveness include distinctness, disparity, dissimilarity, diversity, unlikeness, variation, contrast, and difference.

It is likely that we will be more inclined to believe, trust, or be confident in a memory that is distinctive than one that is similar, does not stand out, or is vague. We may describe an event or occurrence (or their memory) as distinctive or remember a person with a distinctive face. A distinctiveness heuristic was proposed to reduce false memory because recognition of words associated with pictures resulted in lower levels of false alarms than that for words shown alone, (Schacter, Cendon, Dodson & Clifford, 2001). One might argue that this is merely a demonstration of the picture-superiority effect (e.g., Ensor, Bancroft & Hockley, 2019; Whitehouse, Maybery & Durkin, 2006). Does this mean that a memory rich in detail will always be accurate, and does this fact reduce false memories?

The importance of context.

In an isolation list paradigm, one item in a list may be different from the remaining items because it is unlike or dissimilar to them, e.g., in shape, colour, or category, but is the item itself intrinsically distinctive? Thus, distinctiveness can be a relative concept rather than a fixed property of an event or stimulus (Craik, 2006). Koch, Akpan and Coutanche, (2020),

[10] An isolation paradigm is an old or new recognition memory task when all the items in a list are similar apart from the isolate. Here, the experimenter is using the task to signpost which predetermined item the participant will select as being distinctive

suggested that image memorability, both for faces and for scenes, was an intrinsic property of the image – implying that distinctiveness may be an intrinsic property of a memorable image. Notwithstanding, it is proposed in distinctiveness research that if the context or circumstances of an item are changed, its distinctiveness may also shift, e.g., by varying the colour of the background on which words are presented (Siefke, Smith & Sederberg, 2019), in which case the distinctiveness of the word can be considered to be extrinsic. For example, if you saw a sheep eating grass in a field it would not be distinctive, but if you saw a sheep wandering around your college it would be distinctive, not intrinsically but because an unremarkable stimulus becomes distinctive purely because it is not in its expected context. In contrast, in his example of a purple car with lemon-yellow polka dots (Hunt, 2013), the car remains in isolation, i.e., is distinct from all others regardless of the context, it will be remembered as being intrinsically distinctive (unless every car is purple with lemon-yellow polka dots).

Experiments differ in how distinctiveness has been artificially produced, and in whether such distinctiveness is extrinsic or intrinsic.

The main body of literature on distinctiveness (in the same way as for vividness) relies heavily on recall or recognition memory tasks to test whether “distinctive” items are better recalled or better recognised than “non-distinctive” items. Additionally, the research may also depend upon an isolation effect, (e.g., Brunel, Oker, Riou & Versace, 2010; Schmidt, Schmidt & Wilson, 2021). When reproducing an isolation effect, the extrinsic distinctiveness of an item is artificially manipulated, for example, by using words in different list arrangements where one of the words is unusual or represents a different category (e.g., Schmidt, 1985); or using a picture in a list of words, (e.g., Parker, Kember & Dagnall, 2017); by using words in different colours (e.g., Schmidt et al., 2021); or words in different font sizes (e.g., Kelley & Nairne, 2001). Hunt and Lamb (2001) comment that an isolated item “has the advantage at retrieval of prior categorical and distinctive processing relative to any of the control items," (p.1361). Nevertheless, researchers continue to debate exactly how the isolation effect works (see Gretz & Huff, 2020; Schmidt et al., 2021).

What makes a stimulus distinctive?

*The world is wide. No two days are alike, nor even two hours, neither were there ever two leaves on a tree alike since the creation of the world; and the genuine productions of art, like those of nature, are all distinct from each other. (**John Constable** 1776-1837).*

Although it is well-known that we can remember many pictures, i.e., when presented in an old or new recognition task, how do we pick out those that are distinctive and remember them better? Some experimental manipulations treat distinctiveness as an independent variable, i.e., as for external vividness, being an intrinsic property of an item (Hunt,1995; 2006), e.g., the distinctiveness of features highlighted by colours aids picture recall (Suzuki & Takahashi, 1997), Images, likewise, may appear distinctive due to their context or because experiments have manipulated image distinctiveness by violating our expectation of what items should look like or where items should be. Examples include images featuring incongruent items (violating their category), or bizarreness (e.g., Hunt & Lamb, 2001; McDaniel, Dornburg & Guynn, 2005; Morita & Kambara, 2021; Santangelo et al., 2015; Doss et al., 2018). This is a form of distinctiveness pre-determined by the nature of the experimental design, as much as one might say for an isolation list arrangement. The concept of distinctiveness that is independent of such designs is in effect an individual's interpretation of their perception and comprehension of an item or event, and therefore subject to individual differences, e.g., idiosyncratic self-generated difference cues shows that information relating to individuals' prior experiences may be used in processing distinctiveness (Hunt & Smith, 1996).

The literature shows that distinctive features are a significant factor for image memorability (e.g., Bylinskii et al., 2015; Rust & Mehrpour, 2020). It is accepted that some images are simply more memorable than others (Goetschalckx, Moors & Wagemans, 2019); it has also been demonstrated that some images stand out from their category when shown in a similar context and will be rated higher in distinctiveness, although it is disputed that this occurs, Bainbridge (2020). Relating to the use of natural scenes images in my experiments, and of potential interest, Broers and Busch (2021), found that images with high memorability were likely to provoke personal memories (episodic associations). In the same way as idiosyncratic personal factors underpin internal vividness (Ford & Kensinger, 2016), stimuli could equally be perceived as distinctive purely because they have emotional significance for the participant or resonate with personal memories. This may be of importance for judgments of memory quality relating to confidence. In view of this information and particularly if not pre-determining certain images as being distinctive, it would be important to collect participants' comments on what aspects of distinctiveness they used in their judgments.

Distinctiveness processing.

There has been discussion of exactly when the judgment of distinctiveness is made, i.e., at encoding, when the stimuli are first experienced, or at retrieval, when the stimuli are remembered. This is not generally asked regarding vividness, as it is assumed that vividness of a stimulus remains the same, i.e., it is intrinsic to the item tested. However, this may not be entirely true, as vividness may also change due to the context in which the picture or item is shown.

If, when viewing items at study, the first item is a picture of a red poppy, and this is followed by pictures of green or brown fields you may remember it as being distinctive (in colour, or in category) in comparison with the remaining items viewed. It will also be judged as vivid. However, if the first item is a picture of a red poppy, and this is followed by pictures which are all of red poppies, you will no longer regard it as being distinctive (or likely, vivid) in comparison with the remaining items viewed. At test, when you have seen the whole list, you might rate all of them, or none of them as distinctive, or look for different criteria within each picture which makes them different (i.e., is a leaf shown, in which direction does the poppy stem point?).

Other investigators propose that distinctiveness is a psychological process (see Hunt, 1995), i.e., the mental sorting of our memories; as such, this must result from distinctiveness processing at retrieval, when perception and understanding of the task enables the reinstatement of earlier processing (Hunt, 2006). Brunel et al., (2010) reported that distinctiveness processing occurs at retrieval due to a global difference (the difference between items over a wide range of categories) between memory traces; alternatively, Koch, Akpan and Coutanche (2020), suggested that image memorability, both for faces and for scenes, was an intrinsic property of the image rather than its context.

When discussing whether distinctiveness is rated at encoding or at retrieval, if a distinctive item is the first in a list and cannot be recognised as distinctive or different at encoding, it is still recognised better at retrieval. Nairne (2006) hypothesized that relational processing could assist with recall of items from isolation lists. As an example, and possibly creating another confound, that of criterion shift, if participants are asked to assess each of a series of items for distinctiveness when first seen (i.e., at encoding), their distinctiveness judgment (or criterion) may well change over the course of the list item presentation. To return to the example previously used of a red poppy in a cornfield; when viewing items at study, if the first item is a picture of a red poppy and this is followed by pictures of green or brown fields you may then recognise it as being distinctive at retrieval (in colour, or in subject category) in comparison with the remaining items viewed (when you have seen all of the items in the list). At encoding, you may rate subsequent items as less distinctive in comparison (regardless of the rating given to the first item). This is an example of the effect of an isolation list. However, if the first item is a picture of a red poppy, and this is followed by pictures which are all of red poppies, you will no longer recognise a picture of a red poppy as being distinctive, in comparison with the remaining items viewed. To labour the point, the first poppy would be rated as distinctive, the second poppy might also be rated distinctive, but by the time you get to the third, fourth, or fifth poppy, your criteria for distinctiveness, when judged at encoding, would have shifted. This is an example of Nairne's (2006) relational processing.

If the same item contexts are replicated at retrieval you will recognise a picture of a red poppy in the first (isolation list) example better than the same picture of a red poppy in the second (homogenous list) example. The assumption has been made that a picture of a red poppy has been more richly encoded at study when shown as an isolate. These lengthy examples involving red poppies also relate to a list composition effect. Nguyen and McDaniel (2015), demonstrated superior recollection for complex pictures (line drawings of animals, filled in, and with background details) relative to simple pictures (line drawings of animals) in mixed rather than homogenous list arrangements.

If the difference of an item in the context of similar items helps us to recall it, according to depth of processing theory, items may be processed more deeply and hence better recalled because they are paid more attention to at encoding, (Craik & Lockhart, 1972; McDaniel & Geraci, 2006), it has been proposed that the different item will receive more attention and will be more richly encoded, making it better remembered. Gallo, Meadow, Johnson, and Foster, (2008) suggested that enhanced distinctiveness at encoding is responsible for encoding level-of-processing effects on reducing false recognition. Nonetheless, experiments specifically designed to test for deeper encoding will naturally produce results in keeping with distinctiveness processing at encoding, (e.g., Gallo et al., 2008; Weigl et al., 2020). This has been characterised as an "attentional boost effect" (Smith & Mulligan, 2018).

This explanation, however, has long been countered by the inferior-recall hypothesis. Von Restorff, as cited by Hunt (1995; 2006), suggested that the less distinctive items or events are recalled less well than the more distinctive items or events. It may be the case that distinctive items are not encoded better, but the mundane items are less well recalled. This chimes with the context-change approach for defining distinctiveness taken by Siefke et al., (2019), where distinctive items are those with features that are different from others in the current temporal context. Even if we pay more attention to the distinctive event, we may also pay less attention to the routine.

We readily forget the mundane but remember the unusual. That this is true, it is only necessary to examine our own memory, looking back on the events of the recent past we remember only moments rather than all the routine or mundane details in between such (so-called) distinctive events in our day-to-day existence. Tulving and Rosenbaum (2006) also endorsed an inferior-recall explanation for the effect of distinctiveness on long-term memory, in that memory for distinctive episodes is not any greater than for other episodes, just that recall for everyday events is inferior. Evans and Baddeley (2018) demonstrated when natural scenes were selected from a single category (and with distinctive detail minimised), memory primarily depended on the participant’s intention to remember, in other words this could be interpreted as remembering the mundane only if we wish to do so.

Does the inferior recall hypothesis hold true in the context of a laboratory-based experiment? Schmidt (1985), testing recognition memory for items immediately prior to and subsequent to the isolate, demonstrated that background items were more poorly recalled from lists which contained distinctive targets, suggesting some interference with encoding for the items in close proximity, rather than an encoding boost for the distinctive target.

As discussed above, this approach is supported by neuroscience (e.g., Ghosh & Gilboa, 2014) where a neural network may be based on repetition, similarity, or commonality of events, comprising only the gist, with a lack of unique detail for each event. In Hunt’s (2013) example, memory for a car with lemon yellow polka dots will be distinctive, but apart from considering if we remember the context, will we remember any of the other cars we may have seen at the same time? Findings of impaired recollection for unique detail is paralleled in the flashbulb memory literature, where the accuracy of recalling details of these unexpected, emotionally-charged events has been questioned, (e.g., what you were doing or where you were, when you heard about the terrorist attack on the Twin Towers). When participants were asked about “distinctive details” for both flashbulb and everyday memories, about 42% of their responses were inconsistent and therefore inaccurately remembered (Talarico & Rubin, 2003).

Some researchers have suggested that distinctiveness processing may be used both at encoding and at retrieval, the processing of difference in the context of similarity (Dobbins, Kroll & Liu, 1998; McDaniel et al; 2005; Rajaram, 1998). Other investigators, e.g., Schmidt et al. (2021), have concluded that encoding and retrieval processes interact to enhance memory for distinctive items. This discussion may not directly answer my research question but may be of importance in interpreting the results of my experiments.

(vi) Familiarity.

As discussed above, familiarity is regarded as a memory process involved in the task of recognition. Familiarity might additionally be considered a phenomenological experience of memory and has also been described as a metacognitive judgment of the recollective experience relating to whether or not the stimulus has been encountered previously, (Jacoby, Shimizy, Daniels & Rhodes, 2005; O'Connor & Moulin, 2010). Some cognitive psychologists consider familiarity to be a judgment of memory strength (e.g., Wixted & Squire, 2011); others, a judgment of image memorability, (Broers & Busch, 2021). Regardless, familiarity is another aspect of memory which might affect the relationship between confidence and accuracy, and therefore one which I also choose to study.

Broers and Busch (2021) claimed that image memorability differentially affected recollection rather than familiarity in an old or new recognition memory task. In their study, pictures were shown for two seconds in an encoding block, at test participants had to decide if the picture was old or new, how confident they were (on a 1-6 Likert scale), make a remember (recollection) or 'know' response for pictures considered old, then, for pictures considered new, indicate whether the scene was unfamiliar, or provide a detailed response regarding why they thought it was new. The authors concluded that low memorability images were recognised based on them appearing familiar, whereas high memorability images were recognised because they were recollected. One problem that I have with the findings is that feedback was provided for the recognition decision, but the authors did not regard this as a confound for the effect of image memorability on confidence.

Remember responses in an old or new recognition memory task are associated with high confidence (Tulving, 1985), familiarity can represent either a high-confidence 'know' response or a low-confidence guess. This illustrates that there can be varying degrees of familiarity. A strong feeling of familiarity makes the individual confident they have seen the stimulus (or person) before, even if no detail is recalled of who the person is, i.e., a high-confidence 'know' response. A vague feeling of familiarity does not make an individual confident that the stimulus has been seen before, i.e., a low-confidence guess (Gardiner et al.,1998). This, however, is exactly the definition of familiarity provided by the story of "the butcher on the bus."

Nevertheless, it is difficult to reconcile this with the view that familiarity correlates not only with confidence, but also with a judgment of recollection strength. Cha and Dobbins (2021) demonstrated that correlation between ratings of familiarity and recollection strength was large ($r = .6$, $p < .001$). In their study, confidence was rated on a 1 – 3 Likert scale (low, medium, high). Both recollection strength and familiarity were found to predict confidence, although coefficients were much larger for recollection strength. It appears logical that if familiarity is related to the strength of recollection, it will also be related to confidence. However, if recollection strength (memory strength) equates to confidence, results should be exactly the same for correlations between confidence, recollection strength, and familiarity.

"The butcher on the bus"

We may view a situation, an item, or a person as being familiar, that the stimulus (or person) has been encountered before, but without any recollection of specific detail. Much of what has been written on familiarity relates to a seminal paper by Mandler (1980), describing a situation where one sees an individual on a bus, recognises them because they look familiar, but cannot recall who they are, or where they have seen them before. After a while one may confidently recollect that the person seen on the bus is the local butcher.

This may have given rise to the idea that familiarity is fast, and recollection (remembering) is a slower process (e.g., Gronlund, Edwards & Ohrt, 1997; McElree, Dolan & Jacoby, 1999). Mandler (1980; 2008) considers that seeing something or someone that is familiar leads to a memory search to compare patterns of memory with the perceived experience or event, to obtain retrieval of the original memory and "full recognition" of the item (recognition of target and recollection of context). This bolsters the conclusion that a familiar or know response is quicker in an old or new task than a remember response and resonates with a cognitive neuroscience account of the diffusion decision model, (Forstmann, Ratcliff & Wagenmakers, 2016), where a cue is compared with each item in memory in parallel, so that the better the match, the faster the process moves to the positive boundary.

In this respect, context (as for distinctiveness) is important. Next, consider if the person we recognised as the butcher had been in the right context (i.e., in the butcher's shop), we would have immediately remembered who they were, so is recollection really slower than a feeling of familiarity? In this respect, the context the butcher is seen in could be regarded as source memory and therefore recollected rather than recognised; otherwise, you might relate it to the encoding specificity principle (Thomson & Tulving, 1970) in that the context of the retrieval cue needs to match that at encoding, i.e., you last saw the butcher in the butcher's shop, and if you see the butcher again in their shop you will retrieve the memory of who the person is more effectively than when seen in a different context, such as on a bus. As another example, if we pass our partner or a close friend in the street, we will immediately recognise them, whatever the context (and not think first that they look familiar and then remember who they are). As well as recollection being equally fast, it may occur without necessarily having an initial feeling of familiarity. It could be concluded that when recollection is strong, it is fast, and there is no need for familiarity. This is what I aim to investigate in the final of my six experiments.

Consequently, what does Mandler's example mean? This perhaps has negatively influenced research into familiarity and recognition as much as the isolation effect has in distinctiveness. The situation he describes is not that common, nevertheless, it has been seized upon to validate dual process theory. Investigators in cognitive psychology have considered that familiarity represents a metacognitive judgment of memory strength, albeit lacking contextual detail (Cha & Dobbins, 2021; Slotnick, 2013; Wais et al., 2010; Yonelinas, 2002).

Fluency.

Importantly, confidence may be influenced by factors which make memory retrieval appear easier and more fluent. Familiarity may be facilitated by an unconscious attribution (heuristic) associated with fluency, i.e., the speed of processing. This might be hard to test, as there are many factors which can affect the speed of processing (i.e., recollection, internal factors such as self-confidence, fatigue, boredom, and distraction, or external factors such as the perceived difficulty of the task). A relationship between familiarity and fluency acts as a heuristic for accurate and fast recollection. A fluency heuristic may mean that in a 2-AFC task, if both items are recognised, participants judge that the item retrieved more quickly from memory is more likely to be correct. Thus fluency (and familiarity), increase confidence but not accuracy. In laboratory experiments, a recognition task conducted under time pressure leads to a trade-off between speed and accuracy, where faster decisions can sacrifice accuracy and accurate decisions are carried out at the expense of speed (Jones, Jacoby & Gellis, 2001; Yonelinas, 2002). This speed and accuracy trade-off, i.e., represented by the diffusion decision model (Ratcliff et al., 2016), introduces a confound when using response times to measure fluency.

As an example, if an answer to a quiz question pops up instantly in our mind, without our even having to think about it, we may accept the answer purely because retrieval was rapid (i.e., fluent), involved no conscious effort on our part and made us confident. Of course, the answer may be incorrect, and had we thought about it instead of accepting it, we might have realised this. Ratcliff et al., (2016) use an example of when an answer can change over time, when participants responded to “A bird is a Robin” (fast answer yes; slower answer no).

Using fMRI, Wais et al., (2010) concluded that hippocampal activity increased with memory strength even when source memory (recollection) was absent, and that this absent recollection implied decisions were based on familiarity of the item (the assumption being that both processes cannot occur together). The fact that the same held true for source guesses was dismissed, being said to reflect task-irrelevant (i.e., idiosyncratic) recollection. Their conclusion, if you agree with it, is that strong memories are linked to familiarity-based activity in the hippocampus; otherwise, that the hippocampus supports recollection-based and familiarity-based memory for complex multi-attribute stimuli, such as the spatial relationship between different aspects of a scene, (Wixted & Squires, 2011). However, the exact relationship between familiarity, recollection, and recognition, remains opaque. Slotnick (2013) suggested that behavioural evidence points to recollection and familiarity being a single process, yet also based on the hippocampus.

Manipulating fluency & familiarity.

Repetition and visual clarity (both being perceptual manipulations), using verbalisation and varying degrees of masking of the target word at study, are shown experimentally to increase fluency and produce feelings of familiarity (Whittlesea,1993; Whittlesea, Jacoby & Girard, 1990). Recognition memory refers to the ability to identify as familiar, a stimulus or a situation previously encountered. Inferential theory suggests that spontaneous accessibility or fluency can be used as a cue for memory strength or as a proxy for accuracy, (Kurdi, Diaz, Wilmuth, Friedman & Banaji, 2018). Processing fluency (i.e., how easily an item is processed due to prior exposure) may lead to higher subjective confidence, even when the presented information is inaccurate (e.g., Whittlesea, 1993). This may be of importance for illusions of memory accuracy. In day-to-day life an event or action may seem familiar because it has been performed many times before and hence the details are remembered less well, paralleling an inferior-recall explanation for the effect of distinctiveness on long-term memory (Tulving & Rosenbaum, 2006). Ghosh and Gilboa, (2014), proposed that a neural network may be based on repetition, similarity, or commonality of events, with a lack of unique detail for each event. Does familiarity in the same way therefore represent the gist of the memory rather than the detail of recollection?

Chua, Hannula and Ranganath (2012), projected faces on to real-world images of natural scenes in relational encoding blocks. In recognition test blocks, three faces were superimposed on the scene just presented, for a 3-AFC recognition task. Confidence was rated on a 5-point Likert scale and fluency was estimated using eye tracking fixations on the faces. Greater earlier viewing of the correct face was associated with both confidence and accuracy. When the cue (the natural scene image) was familiar, confidence was higher. This would be in keeping with the importance of context but was only significant for incorrect cue-target associations. Notwithstanding, Chua et al. (2012) concluded that reliance on cue familiarity or fluency rather than recognition of the target could lead to inflated confidence. Also of relevance are claims of memory illusions arising from the use of photographs in tests of recognition memory for words. In a recognition memory task, it was found that new words, when presented with a related photograph, were more readily viewed as being old, theorised as due to effects of familiarity and fluency (Wilson & Westerman, 2018). These findings indicate that familiarity of images may play a role in confidence and accuracy dissociation, and merit further investigation.

Whereas fluency can provide an illusion of memory accuracy, familiarity is similarly considered to influence decisions in recognition memory tasks due to speed (fluency) of processing. At its most extreme, familiarity is said to provide a feeling of déjà vu, (a further memory illusion that one has the experience of visiting a place before), possibly based on dual process theory, and said to be when a memory is recollected with a strong feeling of familiarity. However, a correlation between familiarity and déjà vu remains difficult to prove conclusively (O’Connor & Moulin, 2013).

(viii) False memories & memory illusions.

*The difference between false memories and true ones is the same as for jewels: it is always the false ones that look the most real, the most brilliant. (**Salvador Dali** 1904-1989)*

Claims from the DRM illusion.

I have already considered the DRM memory illusion in terms of testing recollection memory. Presenting lists of semantically-related words can cue free recall of non-presented but associated words up to 40% of the time, leading to Roediger and McDermott (1995), interpreting the results as “remembering events that never happened can occur quite readily.” This was later defined as “remembering an episode from even the recent past may involve a blend of fiction and fact,” (Roediger & McDermott 2000). That this might be slightly exaggerated depends on whether you consider a word in a list as equating to an event, implying an actual life experience. Brewin et al. (2020), offer a different view of such strategies, suggesting more modest conclusions, while admitting that memory can be prone to manipulation. Whilst this is not the topic of my research, I do however, theorise that the quality of a memory may lead to it being believed even if the detail is lacking or inaccurate, as discussed in relation to flashbulb memories (Heuer &Reisberg 1990); in other words, an illusion of memory accuracy rather than an illusory memory.

Implanted false memories.

At this point I will briefly introduce the contentious literature concerning false memory, epitomised by Coan’s (1997) “Lost in a Shopping Mall: An Experience with Controversial Research”, involving implanted childhood memories, also referred to by Loftus (2003). Brewin et al. (2020), conclude that under normal circumstances, without influence from external sources, memory is essentially reliable. An alternative approach by Howe (2011) suggests that memory illusions and false memories should not always be viewed negatively.

These issues have been addressed with information from neuroimaging (Gonsalves & Paller, 2000; 2002). They concluded that false memory could be due to misremembering, i.e., by confusing imagined events with actual perceived events, which may also be provoked by visual imagery. This type of memory confusion has been termed an autobiographical memory error by others, i.e., misremembering details of events leading to mistaken belief in experiencing events that did not occur (Newman & Lindsay, 2009). Equally, one might hypothesise that earlier previously experienced actual events could also be confused, leading to mistaken belief in the detail of events that actually did occur, but have been mixed up. Source monitoring (Hicks & Marsh, 1999; Johnson et al., 1993), memory confusion and post event misinformation have already been mentioned, as has the finding that imagined memories have less detail and supporting memories than do actual memories (Johnson et al,, 1988). Reality monitoring (Gallo, 2013; Scimeca et al., 2011; Wong et al., 2012), may be one way of verifying our own memories, along with other measures (e.g., Anastasi, Rhodes & Burns, 2000). Johnson (2006), in “Memory and Reality” offers further insights into these debated aspects of memory.

(ix) Summary & conclusions.

I have attempted to present the literature outlining the theoretical framework for my research, relating to why memory can be inaccurate and not an exact reproduction of an event. The types of memory can be debated – whether memory is just memory or whether it is correctly represented by multiple types of memory, all different from the others. Memory theory is primarily based on the experimental task of recognition of a previously viewed (encoded) stimulus, usually a word, presented in a list at study. Theories provide a framework on which to base our interpretation of how metacognitive judgments of our memory accuracy and quality may be linked to the factual accuracy of a stimulus or event.

Looking back on the original literature behind the major divisions of memory leads to the conclusion that the foundations are not as strong as would be imagined. Although no doubt we do have short-term memory, it is difficult to imagine this having a separate process for encoding as compared to long-term memory (Figure 1). It appears to be the same, but as there is no intention to remember short-term memory, we do not give it attention at encoding. If, as I consider, memory is unitary, the accepted divisions into semantic and episodic memory may be little more than theoretical concepts, indicating how much perceptual detail is encoded or associated with each individual memory trace or engram cell ensemble. In the same way, the process whereby we retrieve our memories has been divided, with the divisions derived from tasks of memory retrieval. The role of familiarity in recognition memory cannot be ignored, but I am uncomfortable with the framework for dual-process theory, being predicated on Mandler's anecdote, and confirmed by questions to which the answer is predetermined (a stimulus cannot be familiar if it is recollected; you cannot 'know' the answer if you remember the answer). Even if both processes are involved, the participant is forced to choose between the two. When we recognise something, is it familiar, do we immediately also recollect (remember) the detail of the memory for it?

Furthermore, Signal detection theory (SDT), a dominant model of memory retrieval is founded on the perception of present or absent light flashes. It may not be an entirely appropriate model for recognition memory, when used in situations where it is not the case that stimuli are absent (as with the light flashes), but that the lure was not previously encountered (new not old). This is not the same as there being no stimulus present, and therefore no memory, either recollection or recognition, for the stimulus. A new stimulus can still trigger recognition, e.g., if the stimulus triggers a personal memory (an innocent party's face in an eyewitness line-up looks similar to the face of a neighbour or a former acquaintance), meaning that the eyewitness believes they accurately recognised the face, particularly if detail, i.e. the source, of the memory is lacking. Given the importance of the relationship between confidence and accuracy in real-world terms, and although it is convenient to suggest that memory confidence and memory accuracy are very strongly correlated, this may not always be the case. There is evidence from the literature that the quality of memory can also influence belief in a memory. Belief in an incorrect memory's accuracy is postulated as a cause of dissociation between accuracy and confidence. I have also tried to lay out the reasons behind my choice of vividness and distinctiveness as the two exemplars of memory quality studied in my thesis. These admittedly may not be that different, despite what the literature claims.

The aim of my research is to explore the strength of the relationship between confidence and the quality of the memory, and to see whether or not this impacts on the theorised very strong confidence-accuracy relationship. I use behavioural procedures as I am more interested in why we believe our memories to be true than where the theorised processes may take place in the brain. Regardless, neuro imaging research using fMRI and ERP methods have provided interesting hypotheses regarding neural networks and how engram cell ensembles may work. The testing of long-term memory is based on the conjectured (and debated) difference between recollection and recognition. In my thesis I compare the results of recollection of source memory with recognition memory – not to see if the two aspects of memory are the same, but to find out if results replicate across different tasks, when using the same set of stimuli. In the conventional sense my studies are confined to long-term (autobiographic, episodic, declarative, explicit) memory; nevertheless I refer to the generic term memory in the thesis to avoid confusion. As the tasks used to assess long-term memory may not allow sufficient time for consolidation of the memory, it might be argued exactly what type of memory is being tested. However, more recent work suggests that memory consolidation following encoding may be more rapid than previously thought. In addition to the type of task performed, how the data is collected, how it is analysed, and how it is presented, all have a bearing on interpretation of the relationship between accuracy and confidence.

Here, I use both a source memory task and a 2-AFC recognition memory task to examine whether metacognitive judgments of vividness and distinctiveness relating to the quality of memory (i.e., remembering) are associated with confidence in the memory more than with the accuracy of the memory.. In the next chapter, I will detail the methods used in the studies.

Chapter 3: General Methods.

In this chapter, I will present the tasks used to investigate my research questions, and, importantly, discuss further how analytical methods can affect the theoretical conclusions which may be drawn from results, seeking to justify my approach to data analysis. I employ two separate experimental tasks relating to recollection or recognition memory across six experiments. Qualitative data is collected and analysed to identify on what basis participants make their metacognitive judgments of memory quality, represented here by either vividness or distinctiveness.

(i) Source Memory Task.

Assumptions.

For my thesis, experiments were designed to examine the relationships between memory accuracy, confidence, and judgments of memory quality. In a standard recognition memory paradigm, accuracy is captured as a binary response. The source memory task used provides the opportunity of analysing fine-grained continuous data for the variables of interest (memory accuracy, confidence, and memory quality). A novel source memory task was adapted for use from procedures outlined in previous work (Harlow & Donaldson, 2013; Murray et al., 2015; Harlow & Yonelinas, 2016), Figure 16, initially devised to test theoretical accounts of the nature of recollection.

In a source memory paradigm, considered process-pure (based on recollection and not recognition), participants study a list of items presented in different source contexts that require recollection. In the task I adapted for my thesis, originally the list of items were words associated with the position of a cross on a circle, representing the source context.

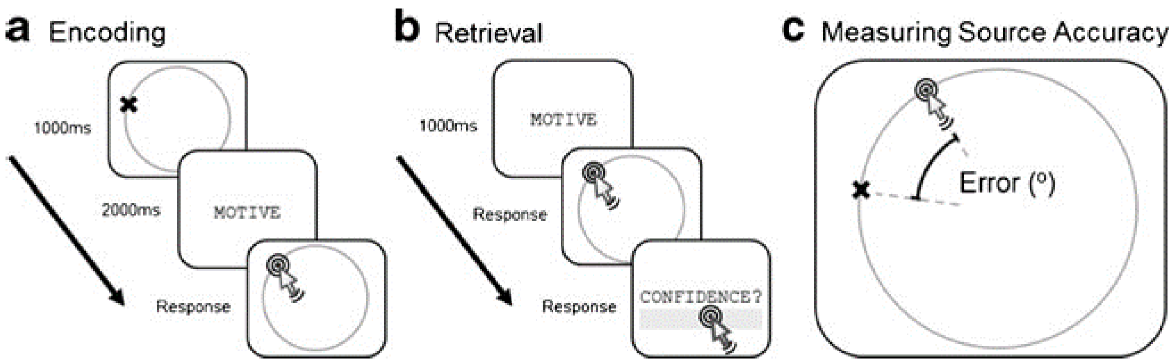

Figure 16. Continuous-response source memory task. (a) Encoding: participants memorised unique word/location pairs, indicating the location after each trial to confirm attention and provide a baseline measure of response error. (b) Retrieval: participants indicated the recollected location for each studied word and rated their confidence, using a mouse. (c) Location (source) accuracy was measured by calculating the arc length (in degrees) between the correct and recollected locations (taken from Harlow & Donaldson, 2013). Note: *93% of baseline trials were within 10° of the target location*

Participants must explicitly allocate their attention to the cross location and the related word to be able to perform the task. When the word is presented at retrieval to cue source memory, the position of the cross may be recollected because the word is recognised. If the participant does not recognise the word they may not search for the cross location (Zhou et al., 2021) and will rely on guessing to respond. Harlow and Donaldson (2013), and Murray et al., (2015) show the task can be completed with high degrees of accuracy, i.e., the binding of the retrieval cue and the cross location is not affected by successive presentation of these details.

Ethical Procedures.

Recruitment, Participation and Data collection.

The study was approved by the University of Stirling's General University Ethics Panel. Participants were undergraduate and postgraduate students at the University of Stirling, aged 18-37 years, with normal or corrected-to-normal vision and no history of abnormal colour perception.

Consent to participate.

Participants were provided with an information sheet prior to participation and written informed consent was obtained in person before the study procedure commenced. Participants volunteered their time; psychology undergraduates took part for mandatory course credits.

Confidentiality, privacy, and data protection.

All data collected, including basic demographic details (age, gender and educational level) was anonymised and identified by study number only.

Data collection.

The source memory experiments were conducted using E-Prime software, which also collected the data. Data was stored on and archived to a secure computer network within the university, password protected. Data was analysed on Excel spreadsheets and SPSS, also kept within the university's secure computer network. Consent forms and demographic information sheets were kept securely within a locked filing cabinet in the Psychology department.

Open practices statement.

The data and materials for all experiments are available at: University of Stirling in DataSTORRE: Stirling Online Repository for Research Data.

Experimental methods.

Stimuli.

Instead of using words as retrieval cues, real-world images of natural scenes were used for my adaptation of this source memory experiment. This manipulation fundamentally changes the nature of the paradigm, in that the picture and the related cross location are thought to be encoded together visually, as a single memory trace, i.e., according to the Item-Source Model, where source memory is a subset of item memory (Cooper, Greve & Henson, 2017; Fox & Osth, 2022). This, I feel, justifies using quality of memory for the natural scenes picture to correlate with confidence and accuracy for the recollected cross location. As compared to similar experimental procedures (i.e., Leshikar & Duarte, 2012; Richter et al., 2016), where the target object is shown overlying a background scene, in this task, the natural scenes image and the target (cross location) are shown sequentially. Participants are explicitly told to pay attention both to the cross location and the related picture to be able to perform the task.

When the picture is presented at test to cue recollection for the cross location, recognition of the previously studied cue picture may be influenced both by recollection and familiarity. Reinitz et al. (2011) proposed that an image could be recognised due to a distinctive feature being recollected, or due to general familiarity of the scene.

In order to minimise content that would potentially bias image memorability or recognition accuracy, images that included people or animals were avoided, e.g., (VanArsdall, Nairne, Pandeirada & Cogdill, 2017), as were those featuring incongruent or man-made objects, such as landmarks or famous buildings e.g., (Hunt, 2013; Santangelo, Di Francesco, Mastroberardino & Macaluso, 2015; Wichmann, Sharpe & Gegenfurtner, 2002). Natural scenes were found to cause more false alarms than man-made structures in a recognition memory task (Evans & Baddeley, 2018), also that features within natural images could be encoded rapidly and without intent. Nonetheless, to avoid similarity effects, pictures were included from four distinct categories of natural scenes: flowers, green landscapes and trees, mountains, or water. Previous evidence suggests that diverse types of natural scenes are associated with equivalent levels of recognition accuracy (Wichmann et al., 2002). Colours of the natural scenes were not manipulated, to avoid any bizarreness effects. As complex pictures of natural scenes may not always be easily represented verbally when performing the task, these pictures were chosen to mitigate any effect of dual encoding.

Suitable images were downloaded from two online psychology image databases, the McGill Calibrated Color Image Database (Olmos & Kingdom, 2004), and the U-Penn calibrated natural images of the Okavango Delta of Botswana (Tkačik, Garrigan, Ratliff, Milčinski, Klein, Seyfarth… & Balasubramanian, 2011). Emotional images were avoided. Durbin, Mitchell, and Johnson (2017) found that self-referential encoding detrimentally affected source memory, but only for neutral and negative pictures. However, others have shown that emotionally-valenced words can enhance source memory (Doerksen & Shimamura, 2001). In this experiment, all the pictures were considered to be neutral, although positive emotions could result from thinking the picture was attractive.

Because of the difficulty in downloading sufficient suitable images for the categories required, which did not include people, animals, or man-made structures, further photographs of natural scenes were selected from a personal database. All images were optimised for colour saturation and contrast (sharpness).

The natural scenes photographs used are now available to the psychology department for research purposes (there are no copyright issues as long as the downloaded online images are referenced in any scientific publications).

Procedure.

Experiments were conducted in person (i.e., with the researcher present) in a designated quiet psychology testing room using E-Prime (version 2.0 Professional, Psychology Software Tools Inc.), on a laptop PC (screen size: width 34 cm, height 18 cm). Photographs were presented in landscape orientation on a white background at a resolution of 640 × 425 pixels. All experiments involved one session lasting up to 40 minutes for each participant. Photograph presentation order was randomised by E-prime within each trial block during both study and test phases (to counterbalance for any effect of list order); additionally, trial block presentation order was randomised by E-prime for each participant within the experiment. Once participants had successfully carried out a short practice (using 8 different picture-location pairs), the experiment consisted of two phases (Figure 17). The participant could use their own strategy for the task.

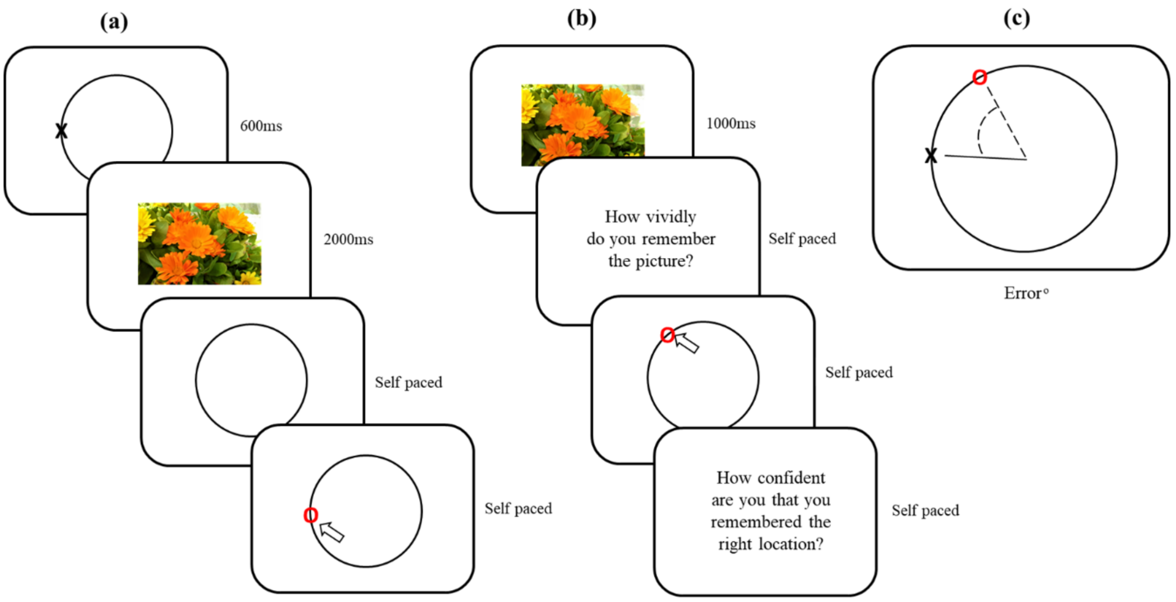

Figure 17. Protocol for experiments using judgments of quality at test: (a)relational encoding (study) block; (b)recollection (test) block; (c) measuring location (radial) error, where the radial error = distance in ± degrees between the participant's response (shown as a red circle) and the correct target location (shown as a cross on the circle).

Relational encoding (study) block.

Within the relational encoding blocks, trials commenced with the presentation of a black cross located on a grey circle (600 ms), followed by a photograph of a natural scene (2000 ms). Participants were explicitly asked to pay attention to both the photograph and its related cross location for a subsequent memory test.

Participants verified their attention by indicating the (now hidden) cross location using the mouse (self-paced). Responses within 6° of the correct location advanced participants to the next trial, otherwise the cross location was re-presented (250 ms) and the verification task repeated, as described previously (Harlow & Donaldson, 2013; Murray et al., 2015).

Recollection (test) block.

Each study block was followed by its corresponding test block; introductory instruction screens between study and test ensured that participants could not rely on working memory to perform the recollection task. The participant was shown each of the previously presented photographs (1000 ms) from the corresponding relational encoding block, i.e., all test images were old. Following each photograph, but after the judgment of memory quality had been made, the participant was presented with a circle and was asked to indicate the related cross location using the mouse (self-paced).

Behavioural data.

Judgments of memory quality.

Participants made their judgments of memory quality (either vividness or distinctiveness) for the natural scenes image at retrieval during the recollection test block presentation, immediately after the image had been re-presented to cue recollection of the cross location (but not while it was still on the screen), by answering, e.g., "How *vividly* do you remember the picture?" Response options ranged from not at all *vividly* to extremely *vividly* on a 0 –100 horizontal continuous scale.

Judgments of confidence.

Confidence was assessed at test after the memory quality judgment had been made, and after recollection of the cross location, Figure 17, with participants answering, "How confident are you that you remembered the right location?" Response options ranged from not at all confident to very confident, on a horizontal 0 –100 continuous scale, (cf. Harlow & Yonelinas, 2016).

Memory Performance.

Using this continuous-response source memory task enables memory performance to be assessed in two ways, (a) by using a calculated average recollection rate for each participant across all trials (equivalent to a standard binary task), and (b) by using the continuous variable of location accuracy [measured as degrees of error (0° to ±180°) from the target location and expressed as % accuracy for analysis.

(a) Recollection rate.

Location error frequency plots from 0° to ± 180° from the target location, summed across participants, allow a visual representation of memory performance, illustrated as an example in Figure 18a. Based on the assumption that all responses from ± 90° to 180° are guesses made in the absence of recollection, location errors within ± 90° of the target were used to calculate the proportion of trials associated with memory retrieval (1- twice the guessing rate from ± 90° to 180).

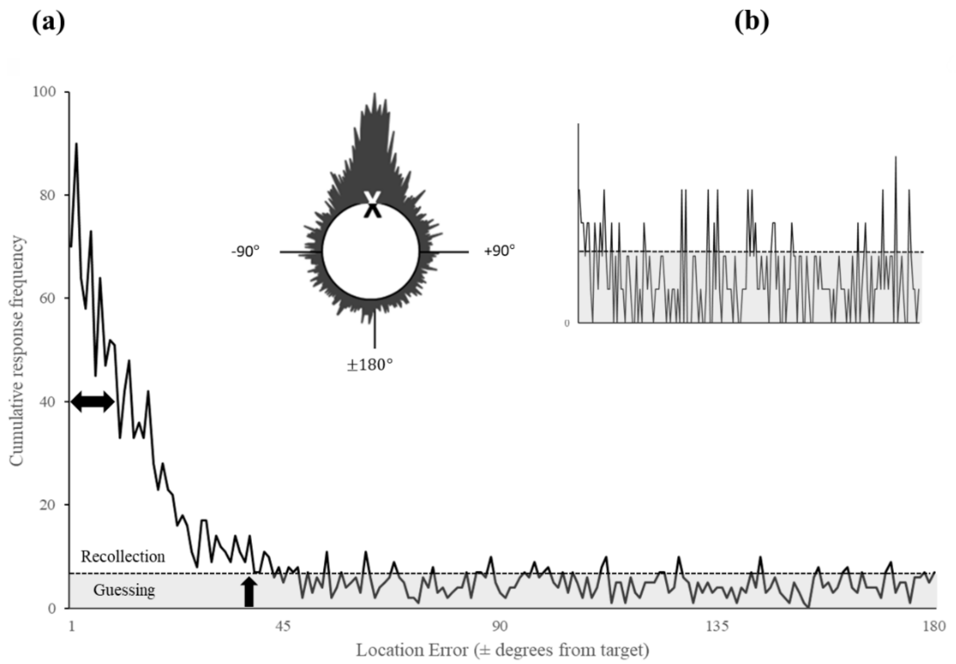

Figure 18. (a) A typical location error frequency plot illustrating source memory performance across all participants, i.e., recollection rate, guessing rate, scale parameter (precision – double headed arrow), and threshold (single headed arrow); recollected responses (cumulative response frequency) peak around the target location (0°); the guessing rate is illustrated. Inset is the wrapped distribution of recollected responses. As suggested by Harlow and Donaldson (2013), and Murray et al. (2015), recollected responses are clustered around the target location (0°) and illustrate significant subthreshold guessing; (b) A location error frequency plot illustrating an individual's location error plot when there is no recollection; recollected responses are flat (showing no peak) as the participant is guessing throughout the task.

The quality of the final average data sets can be monitored; if participants were unable to remember the locations and were guessing throughout the experiment, the location error frequency plot would have the appearance of that illustrated in Figure 18b; participants who could not or did not perform the task can be identified and their data excluded (cf. Zhou et al., 2021).

Precision.

The location error distributions were fitted to a mixed Cauchy plus guessing model (as per some-or-none thresholded accounts of recollection) using an online toolbox [Wessa P., (2013), Maximum-likelihood Cauchy Distribution Fitting (v1.0.3) in Free Statistics Software (v1.2.1), http://www.wessa.net/rwasp_fitdistrcauchy.wasp/Office for Research Development and Education]. The scale parameter of the Cauchy distribution represents half the width of the Probability Density Function at half the maximum height. Data from the model fit gives estimates of the shape of the distribution, providing a direct measure of memory precision[11], as depicted by the double-headed arrow in Figure 18a. A smaller scale parameter results in a taller and thinner curve (reflecting a high level of precision), whereas a larger scale parameter results in a shorter and wider curve (reflecting a low level of precision).

Threshold.

The recollection threshold, expressed in degrees from the target location, indicates the exact point at which recollection begins to contribute to responding, indicated by the single-headed arrow in Figure 18a, calculated using linear regression to determine when the slope of the distribution of errors changes from zero (i.e., is no longer flat because responding is not solely reliant on guessing), i.e., the point at which the maximum error occurs when

[11] Precision does not exactly mean accuracy. Precision refers to clustering of responses around the target, i.e., the 0° location, you can be precise (e.g., every response being precisely ± 4° or ± 10° away from the target). Accuracy, conversely, refers to how close each individual response is to the target

recollection is involved in responding. This assumes recollection to be a thresholded rather than a continuous signal-detection process, e.g., (Yonelinas et al., 2010; Zhou et al., 2021).

(b) Location accuracy

Rather than assessing accuracy based on a binary response to a question regarding source memory accuracy (yes or no, or what colour the word was presented in), here a different approach is taken. The accuracy of retrieving the correct location for the cross is calculated as degrees of error (0° to ±180°) from the target location and expressed as % accuracy for analysis, (i.e., if the participant puts the cross at test 180° from the target location, their accuracy is 0% and if it is ±90° from the target location, accuracy is 50%) this allows calculation of location accuracy as a continuous variable for each trial, permitting analysis on a by trial basis

Qualitative Data.

The distinction between the separate types of vividness (and also distinctiveness) referred to in the literature review, i.e., external, or internal, highlights the importance to researchers of assessing judgments of memory quality such as vividness both quantitatively, by rating how vivid a stimulus is on a scale, and qualitatively, to explore participants' perceptions of what was important to them when assessing vividness or distinctiveness. In essence, photographs or images may be recollected vividly because the pictures had the external property of vividness, or because they provoked self-referenced or personal memories. Procedures employed within the wider literature have given participants explicit instructions on judging visual vividness (Ford & Kensinger, 2016; Johnson, Kuhl, Mitchell, Ankudowich & Durbin, 2015; St. Laurent, Abdi & Buchsbaum, 2015).

In this source memory experiment, I explored the criteria used to judge memory quality by asking participants for free comments (once they had completed the experiment, and before they were debriefed) on, e.g., "What made the photographs vivid/distinctive?"

and, “What made the photographs *not-vivid/not distinctive*?” Comments were written down verbatim and later transcribed for thematic analysis, using methods outlined in Braun & Clarke (2006).

Data Analysis Plan

As referred to in the literature review, methods used for the analysis of results are of critical importance, particularly as regards their interpretation. The APA has issued guidelines for authors (Appelbaum, Cooper, Kline, Mayo-Wilson, Nezu & Rao, 2018), which include brief paragraphs on sample size, power, precision, psychometrics, and analytic strategy – to be described for “inferential statistics and protection against experiment-wise error for primary, secondary and exploratory hypotheses” (p.7).

Fundamental to my research is how the data are analysed, and here I will provide the rationale behind the methods used. As two different experimental tasks are employed, details of the analytical methods are outlined for each task separately.

In the continuous-response source memory task described in this section, indicators of memory performance [recollection rate, scale (precision), and threshold] were assessed using averaged values across the 120 trials for each participant, calculated at the end of each experiment as described. Results from the memory performance indicators informs as to whether participants were able to perform the task successfully. These results were not used to support my hypotheses.

All data from the variables of interest [location memory (source memory) accuracy, memory quality (vividness or distinctiveness), and confidence in source accuracy], were collected as continuous data on 0 –100 scales. It is customary for data to be averaged for each participant, and subsequently on a group basis (depending upon the task employed), which can lead to incorrect data interpretation. As a general point, averaging data has a place in tracking changes in a cohort over time. Although it is easier to present the data in the text or in tabular form if this approach is taken, much of the detail that can be seen in the full data set is lost. If data for the variables of interest are averaged for each participant, not only does it mean information loss, but simple regression correlations applied to aggregated data may produce biased results, (e.g., Bakdash & Marusich, 2017).

The statistical analyses used data from all individual trials, i.e., from each picture-cross location pair for correlation analysis of vividness or distinctiveness, accuracy, and confidence, to give a comprehensive picture of the distributions for each variable of interest. Although this may be criticised for ignoring between-participant and between-image variation, firstly, it avoids individual variation in confidence as a confounding factor; furthermore, image memorability scores have been shown consistent across participants and across experiments with varying contexts and setups (Bylinskii, Isola, Bainbridge, Torralba & Oliva, 2015), justifying this approach. Individual or image variability is not being studied, but rather how memory quality ratings interact with confidence and memory accuracy across the population of participants, on a trial-by-trial basis. Hidden covariates, e.g., subject or item differences, and subject-item interactions exist, (Hintzman, 1980). To ignore these potential effects and to simply average data across all 120 pictures for each participant, means that data is lost for individual pictures, i.e., each picture-location pair (or trial) in terms of memory quality. Examples of this approach exist in the published literature, using scatterplots of individual trial data, e.g., (Bylinskii et al., 2015); Harlow & Yonelinas, 2016).

Bar or column charts are ubiquitous in peer-reviewed published memory research, used to highlight differences between group means, with error bars representing ± *S.E.*, or 95% confidence intervals (± 2 *S.E.*). Examples of the use of bar charts in memory accuracy literature include (Iida, Itsukusima & Mah, 2020; Mickes et al., 2011; Simons, Peers, Mazuz, Berryhill & Olson, 2009; Sitzman, Rhodes & Kornell, 2016; Wong et al., 2012; Zhang &

Luck, 2011). However, as shown by Weissgerber, Milic, Winham, and Garovic (2015) this type of plot can mislead and hides the true data distribution.

In the analysis of results, univariate scatterplots, bivariate scatterplots, and variable density plots are used to allow for examination of the data distribution. As such the full data may suggest different conclusions from the averaged (summary) statistics, (Weissgerber et al., 2015). I will use vividness, Experiment 1 to illustrate the difference between averaged data and individual trial data for the continuous response source memory task. All participants ($N_{participants}$ = 16) completed the same number of trials (120), as a result, averaged data gives the same mean values, but with larger standard errors as the number of trials is larger (N_{trials} = 1920), for example, using the variables of interest *M* (*S.E.*): vividness 67.4 (2.62 vs. 0.58); accuracy 74.1 (2.73 vs. 0.64); confidence 54.9 (3.05 vs. 0.66). However, there is also significant disparity in the histograms showing the data distribution, Figure 19.

Significance levels.

A further consequence of using data from all trials is that the substantial number of individual trials analysed makes even tiny differences result in significant *p*-values. Here, I report the unstandardized *ß* coefficient (*B*)[12]. Interpretation of effect sizes were related to the R^2 change, taken as .01 (1%) = weak; .09 (9%) = moderate; and .25 (25%) = strong (Cohen, 1992; Taylor, 1990).

[12] Unstandardized coefficients are ‘raw’ coefficients produced by regression analysis when the analysis is performed on original, unstandardized variables. An unstandardized coefficient (B) represents the amount of change in a dependent variable Y due to a change of 1 unit of independent variable X

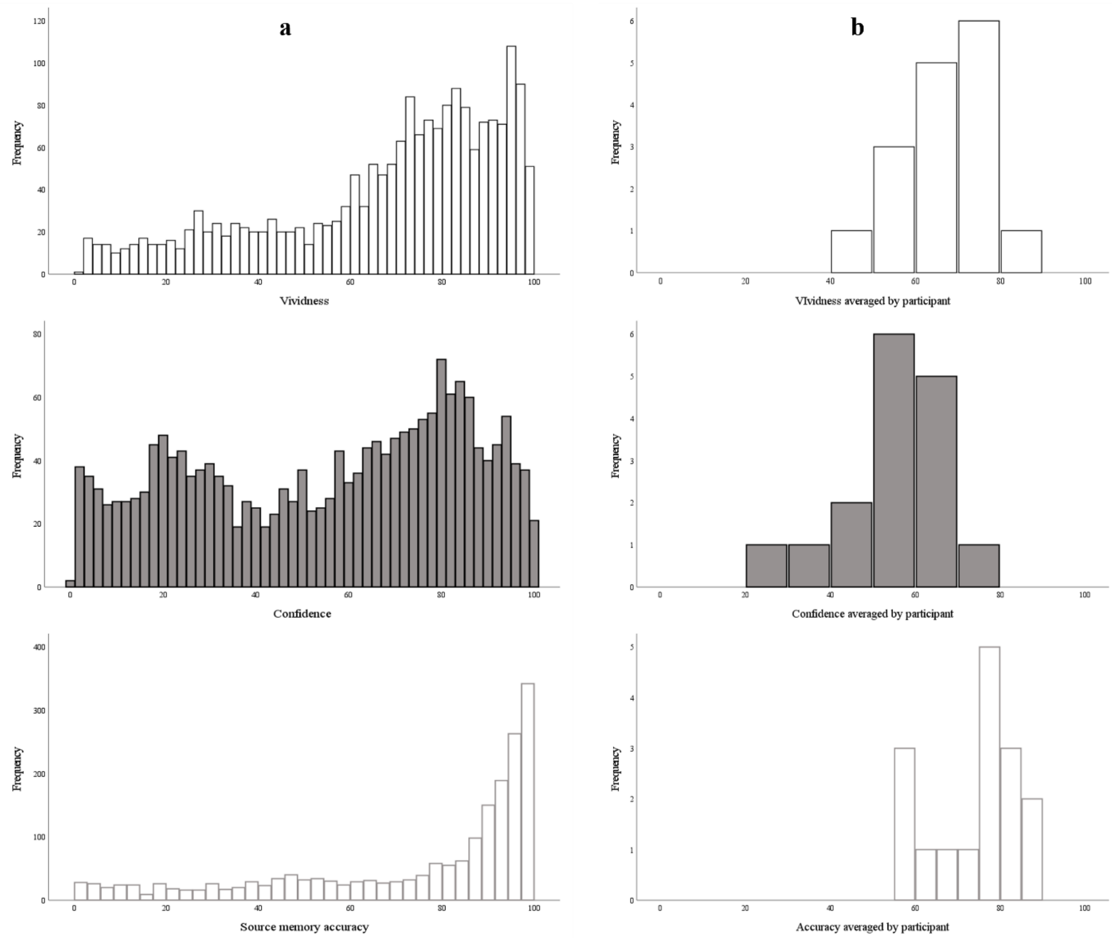

Figure 19. Vividness, Experiment 1: histograms of the frequency distributions of vividness, accuracy, and confidence, illustrating the difference between (a) all trial data and (b) averaged data by participant. Despite having the same mean values, the distributions are clearly not the same

Data outliers can skew the averaged data set, giving the erroneous impression that the data cluster around a point different to where they really cluster. In considering Figure 19 this is particularly apparent for the accuracy data, the histogram representing all trial data follows the pattern of the location error frequency curve seen in Figure 18a, however, the averaged data loses the longer spread of low frequency/low accuracy results, giving the false impression not only that the average participant will not be guessing, but also that they will only achieve 75-80% accuracy in their performance, with a fall off at higher levels of accuracy. The distribution of confidence between zero and 20% is lost and the peak is at less than 60% rather than around 80% as seen in the full data set. Similarly for vividness, the low-frequency/low vividness data is missing, and peak levels are lower than are apparent from the full data set.

A histogram aims to approximate the underlying probability density function that generated the data by binning and counting observations. Some of the data for correlational analysis is presented as a density plot (or smoothed histogram), e.g., Figures 28 and 29. A density plot is a useful non-parametric technique for visualizing the underlying distribution of a continuous variable. The scale on the x-axis is used to locate the bins. The scale on the y-axis is the probability density function of the variable, used to specify the probability of the random variable falling within a particular range of values, as opposed to taking on any one value. This probability is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range. The probability density function is non-negative everywhere, and its integral over the entire space is equal to 1. Key features of the data are easy to discern (central tendency, bimodality, skew), and they afford easy comparisons between subsets.

It is clear therefore that when averaging data, results can be deceptive. Group responses are useful if not much data has been collected for each participant, and such averaged responses may not capture all the characteristics of individual participants' data.

Correlational analysis

Correlation analysis was performed to assess the strength of relationships between the variables of interest (in this case vividness or distinctiveness, confidence, and accuracy). Roediger and DeSoto (2014), although for a recognition memory task rather than a source memory task, claimed that methods of analysis can produce differences in correlation results. Here the correlation analyses are initially assessed using data from all trials. Although this provides a large sample size for data analysis, it has also been suggested this can produce misleading results by overlooking specific relationships within groups which may differ from the overall picture, i.e., gender, educational performance, or age groups may differ in their responses. This is not considered a major issue in my studies, involving a restricted age group of participants having similar educational ability. If a trend exists in groups of data this might be reversed when the data are combined. This is termed Simpson's paradox, a statistical occurrence where an association between two variables when the population is divided into subgroups emerges, disappears or reverses when data is combined for the population as a whole. For a full discussion of the relevance of this effect in studies of memory retrieval (see Hintzman, 1980). This could result in unjustified causal inferences, hypotheses may depend upon which data analysis is used. To address this point, and to allow for comparison of the correlations across all trials, within-participant correlations are derived and averaged, these results are displayed alongside the Pearson's *r* values derived from all trial data. Detailed tables of individual correlations and scatterplots (with regression lines for individual participants) are illustrated in Appendix B, which gives the opportunity of assessing the data to see if there are groups of cases, e.g., with a negative correlation when the overall correlation is positive. Here, I employ a simple averaging of the within-participant Pearson's *r* correlations. It is accepted that this may lead to an underestimation of the correlation coefficient (Silver & Dunlap, 1987). The literature is divided as to whether Fisher's *z* transformation should be used (e.g., Corey, Dunlap & Burke, 1998; Silver & Dunlap, 1987). Although there may be bias in averaging correlations without transformation (resulting in higher or lower values for *r*) the general opinion is that the level of bias is very small. Strube (1988), advised that the bias between the two methods lessened as the set size increased. I also report the variance (R^2) for all correlations.

Another important issue is that regarding correlational analysis which uses averaged values of variables for each participant, rather than averaging within-participant correlation coefficients. The former might not provide an accurate estimation of correlation between the original variables. Furthermore, the correlation coefficient (*r*) can appear to be much stronger than it would be if all the data were used without averaging (i.e., Table 1 and Figure 20). Your theory may depend upon which data set you choose to present.

Table 1. Vividness, Experiment 1: illustrating a correlation matrix comparing the correlation coefficients obtained from data averaged for the 16 participants (shaded) with data averaged across all 1920 trials

Vividness, Experiment 1 (120-image set in full colour)			
I.V.	Vividness	Accuracy	Confidence
Vividness	(1920)	(16)	(16)
Pearson's *r*	1	.74	.72
Accuracy	(1920)	(16)	(16)
Pearson's *r*	.33	1	.78
Confidence	(1920)	(1920)	(16)
Pearson's *r*	.65	.46	1

Table 1 and Figure 20 clearly demonstrate how a fewer number of averaged data points can result in stronger correlations than many individual trial data points, (note that the exception is the large correlation between confidence and vividness, where the difference is less apparent). For a recent example of averaged data being presented as bar charts and correlation plots with sparse data, see an fMRI investigation of vividness and confidence (both assessed on 0-4 Likert scales), from Zou and Kwok (2022).

Comparing data from Table 1 and Figure 20 regarding correlation between confidence and accuracy, if you maintain that these are tightly coupled you would naturally present your results using averaged data, to support your hypothesis. In effect, correlating averaged values, although convenient, does not always give an accurate estimation of correlation between the original variables. We highlight results in Appendix A, (Table 8 & Figures 14 – 16) to illustrate this point. In our analysis of results, scatterplots and variable density plots are used to allow for examination of the data distribution[13]. One consequence of using data from all trials is that the substantial number of individual trials analysed makes even tiny differences result in significant *p*-values (see Sullivan, & Feinn, 2012).

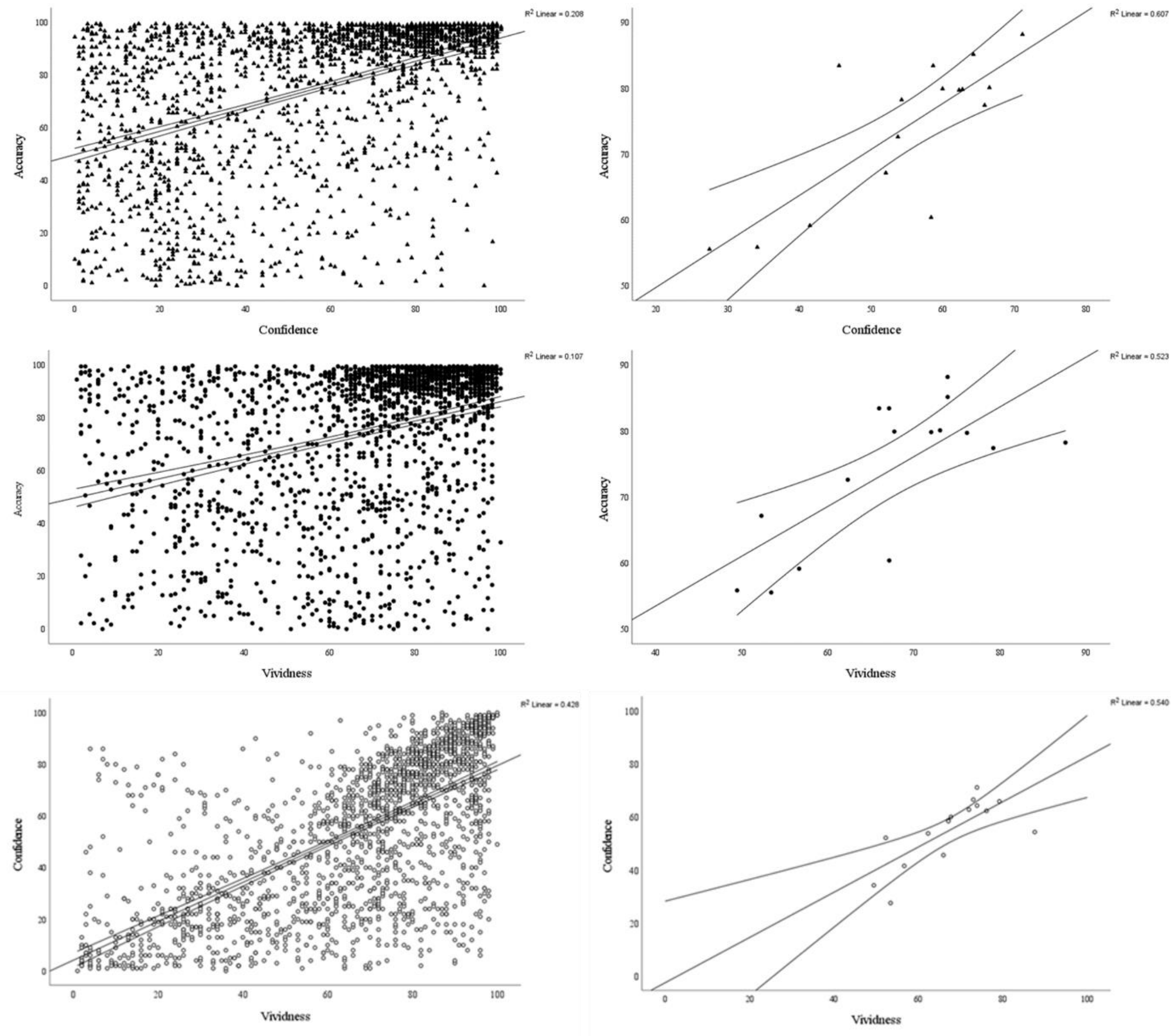

Figure 20. Vividness, Experiment 1: Illustrating bivariate scatterplots for the data shown in Table 1 to compare the correlation analysis for data averaged by participant (right) vs. all trial data (left). Linear regression lines are shown with 95% confidence intervals

[13]A density plot is a useful non-parametric technique for visualizing the underlying distribution of a continuous variable. The scale on the x-axis is used to define the size of the bins for aggregating the data. The scale on the y-axis is the probability density function of the variable, used to specify the probability of the random variable falling within a particular range of values, as opposed to taking on any one value. This probability is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range. The probability density function is non-negative everywhere, and its integral over the entire space is equal to 1. Key features of the data are easy to discern (central tendency, bimodality, skew), and they afford easy comparisons between the image sets.

Data were analysed using the Statistical Package for Social Scientists (SPSS Statistics for Windows, Versions 27.0 and 28.0, released 2020/2021. Armonk, NY: IBM Corp.), and Jamovi [(The jamovi project (2021). jamovi (Version 1.6, Computer Software). Retrieved from https://www.jamovi.org)].

As a further means of exploring the relationships between the variables of interest, the data were tested for moderation effects and for mediation. Both were carried out using the PROCESS Macro for SPSS (Hayes, Montoya & Rockwood, 2017), and results were presented according to APA recommendations.

Moderation Analysis

While correlational analysis looks at the effect of a predictor variable on a dependent or outcome variable, it cannot predict causality or the direction of the relationship. Here, I also seek to use moderation analysis to look at possible three-way interactions between variables. A moderation analysis investigates, for example, whether memory quality judgments moderate the relationship between accuracy and confidence. As for simple correlational analysis, such analyses do not determine causation or the direction of the relationships, but merely serve to confirm the associations, see Hayes and Rockwood, (2017). The analysis can be used with a continuous or a dichotomous moderator. A simple moderation analysis tests whether the regression (i.e., slope) coefficient (indicating the strength of the relationship between the predictor accuracy and the outcome confidence), differs significantly conditional on the moderator (e.g., memory quality). Testing for moderation effects estimates conditional effects of accuracy on confidence at different values of memory quality. For a continuous moderator, the conditional effects are estimated, and simple slopes tested for the sample mean *M* of the moderator and at ± 1 *S.D.*, illustrated using a simple slopes plot. Effect sizes are estimated using R^2 values.

Mediation Analysis

a further way of assessing interactions between variables uses mediation analysis. In a mediation model it is hypothesized that the effect of a predictor variable upon an outcome variable operates, either fully or in part, due to an intervening or mediator variable. The primary goal of mediation analysis is to estimate the pathways of influence. For the analysis, the traditional four-step method (e.g., Baron & Kenny, 1986) is discarded, and a simple mediation analysis using PROCESS (Hayes & Scharkow, 2013) is performed. Mediation analysis using PROCESS is conceptualized in a path-analytic framework (Hayes et al., 2017). The result is to estimate the two pathways of influence from X to Y, one direct and one indirect (e.g., Preacher & Hayes, 2004), rather than putting together a set of discrete hypotheses about the process based on a pattern of significant and non-significant effects (Hayes & Rockwood, 2017).

Mediation analysis can be used to test, e.g., whether the relationship between source memory quality and accuracy acts directly or is partly mediated through confidence. The pathways of influence from source memory quality to accuracy are firstly, the pathway that operates through the mediator, confidence (the indirect effect of memory quality on accuracy) and secondly, the pathway which bypasses confidence (the direct effect of memory quality on accuracy).

(ii) 2-AFC Recognition Memory Task

Rationale.

It could be argued that in the source memory task, any relationship demonstrated between memory quality, confidence, and accuracy is due to the fact that recollection was tested rather than recognition, therefore judgments of confidence and accuracy are somehow different from those obtained for eyewitness recognition memory. Examining recognition memory for a previously studied picture may be influenced both by recollection and familiarity. Judgments are made at test, when new items are shown singly, interspersed with old (previously studied) items, meaning that when there is no signal detected (the stimulus is new in the context of the experiment) the participant relies on guessing to respond. The guessing rate is often unknown or assumed and participants have a higher chance to guess successfully in comparison to a source memory task.

I decided on a 2-AFC recognition memory task utilising Tulving's (1981) picture similarity experiment, again using natural scenes colour photographs as stimuli. A 2-AFC task shows both old and new stimuli together at test and the participant is required to discriminate between the two as to which is old. My choice was to eliminate response bias and the use of ROC curve analysis. The task enables evaluation of performance (proportion of correct responses), a judgment of confidence, and a remember or 'know' or a recollect or familiar question, additionally a rating of memory quality, for the pictures recognised as old. However, for each individual trial, performance is either correct or incorrect. Before going on to describe the procedure, I will introduce the experimental paradigm used.

Similarity effects in recognition memory.

Illusions of memory accuracy may arise from the false recognition of deceptive lures (Roediger & McDermott, 1995; Tulving, 1981; Weigl, Pham, Mecklinger & Rosburg, 2020). Tulving's (1981) picture similarity experiment manipulated stimuli at test to make targets and lures similar, either perceptually or mnemonically, to produce a dissociation between confidence and accuracy. The picture similarity experiment was an old or new task with photographs, shown singly at study, but presented as picture pairs in the test phase with lures either perceptually similar to the targets (A-A'); perceptually similar to other (B) pictures, not used as targets, shown in the study phase (A-B'); or dissimilar either to targets or any pictures previously shown (A-X). For the paired images at test, participants were asked to discriminate between them to decide which one was old. A confidence judgment was also required. Tulving (1981) produced a confidence-accuracy inversion between picture pairs with perceptual similarity (A-A'), and mnemonic similarity (A-B'): participants judging (A-A') pairs showed higher accuracy with lower confidence than for (A-B') pairs. Although participants could better discriminate between targets and lures when the pictures were similar rather than dissimilar, i.e., they were more accurate, they were also less confident. Fandakova, Johnson, and Ghetti (2021) reported higher accuracy but lower recollection (remembering) in the A-A' condition., not remembering being associated with lower confidence. Participants in the A-A' condition were thought to base their judgment on identification of the most diagnostic feature rather than on subjective recollection of the A picture (Dobbins, Kroll & Liu, 1998; Hembacher & Ghetti, 2017). In the A-B' condition, the presentation of the two dissimilar pictures discouraged identification of the most diagnostic features but encouraged participants to assess their identification by relying on subjective recollection, repeating Tulving's explanation. In Fandakova et al. (2021), the experimental task included a betting manipulation and used pictures of everyday objects; there was no A-X condition.

Tulving (1981) explained the dissociation as representing two types of similarity relations in forced-choice recognition memory. When pictures were perceptually similar, the (A-A') condition, the retrieval process to determine which picture was old was more elaborate and resulted in higher accuracy than when the pictures were perceptually dissimilar, when memory of the picture similar to the lure was relied on to make the decision. In effect, the more rapid retrieval from memory of a familiar B' lure, and the lengthier process in the case of (A-A') pictures, led to a confidence-accuracy inversion. If the cue and the target were similar, participants were more accurate but less confident. Roediger and McDermott (2000), signpost the DRM memory illusion as an example of remembering with high confidence events that never occurred, thus explaining Tulving's result with (A-B') pairs for related but unstudied pictures, recognised with high confidence and low accuracy. Such deceptive lures are thought less likely to be remembered than 'known'.

Meanwhile, in an identity parade, and contradicting the results from Tulving's (A-A') condition, eyewitnesses are claimed to be less accurate (more prone to false recognition) if the suspect and fillers are similar, i.e., the more similar lures are to targets, or innocent people are to suspects in an identification parade, the more confident false recognition (false identification) is likely to occur, (DeSoto & Roediger, 2014; Grabman, Dobolyi, Berelovich & Dodson, 2019; Roediger & DeSoto, 2014). Regardless, Colloff and Wixted (2020), argue for similar lineups in an identification parade, as in Tulving's (A-A') condition, to "improve the accuracy of recognition". This makes no sense to me; the same cognitive psychologists hold fast to the account of highly confident eyewitnesses always being highly accurate whereas here we would have highly accurate but not confident eyewitnesses. However, ROC curve analysis was used to interpret the results and troublingly, a figure presented as a CAC plot almost exactly reproduces that shown here as Figure 15; illustrating that these highly accurate identifications are also highly confident, which was not as Tulving (and others) found.

Tulving's (1981) findings of a "confidence-accuracy inversion" have been replicated numerous times (Dobbins, Kroll & Liu, 1998; Fandakova et al., 2021; Heathcote, Freeman, Etherington, Tonkin & Bora, 2009; Hembacher & Ghetti, 2017). The importance of similarity effects in recognition memory cannot be overlooked as it seems uncertain whether this makes it easier, or harder, to discriminate between targets and lures. No doubt it would be easier in an identification parade for the suspect to be shown in a line-up if the fillers (non-perpetrators) were all similar, but dissimilar to the suspect, than if all were similar to the suspect. Similarity effects can be problematic in recognition tasks as well as in eyewitness identification.

The significance of a remember or know judgment.

This type of recognition paradigm has been associated with a remember or ‘know’ question (Gardiner, 1988; Rajaram, 1993; Tulving, 1985), this may also include a guess option (Gardiner et al.,1998). The question represents the extent to which self-knowing consciousness (i.e., episodic memory) is involved in the recognition response. Dunn (2004) considered that remember responses were based on confidence. This experiment provides the opportunity to investigate these questions further. I will point out that distinctiveness was not assessed in the 2-AFC task, because of possible confounds arising from picture-similarity conditions. However, familiarity was assessed, using a recollect or familiar question, e.g., (Dobbins et al., 1998).

Methods

Recruitment and consent.

Due to issues with Covid-19 lockdown restrictions these experimental tasks were carried out online using the Gorilla™ cloud-based psychology software testing platform (https://app/gorilla.sc). Recruitment was carried out via Psych Web (Stirling University Psychology Sign-Up System (sona-systems.com). To ensure data quality, potential participants were asked to view a short PowerPoint presentation to explain the studies (as in person discussion was not possible), before confirming sign up and were individually invited to participate on a specified date. Participants were also provided with a link to an approved information sheet prior to participation. When the participant logged in to the experiment site using their personal link the study information was repeated; a short demographic questionnaire asked for student status, age and gender identity and electronic informed consent was obtained before the study procedure could begin. To ensure complete confidentiality and data security, the software provided participants with a unique non-identifiable key which they could use to withdraw their data from the study for up to 14 days after completion. All participants were psychology undergraduates at the University of Stirling, aged 18-37 years with normal or corrected-to-normal vision and no history of abnormal colour perception, who took part for mandatory course credits. The studies were approved by the University of Stirling's General University Ethics Panel.

Data security.

Gorilla™ does not collect IP addresses automatically, in order to ensure participant anonymity. Performance data is stored against a Private ID (a random anonymous digit string). In compliance with BPS requirements, identifying data, demographic information and performance data are all stored separately. They were downloaded separately and joined together outside Gorilla using the Private IDs provided. By default, data from each participant only became accessible when the participant had completed the whole experiment within the allocated time (up to 2 hours from commencement).

Stimuli.

Although not stated explicitly, picture pairs in previous experiments (Dobbins, Kroll & Liu, 1998; Tulving, 1981) featured animals and buildings, which have been shown to influence recognition accuracy. As previously discussed, colour natural scenes images were chosen; these comprised equal numbers of the four different categories of natural scenes subject (green landscapes and trees, flowers, mountains, or water). All photographs were optimised for clarity (contrast) and colour saturation.

Procedure.

Study phase.

Procedures followed those documented in Tulving's (1981) original experiment. The experiment was carried out online using the participant's own tablet or laptop PC, (but not on a mobile phone), instead of participants sitting in a darkened room watching a projector slide-show, as in the early 1980s. A total of 196 photographs were used for the study, selected, and downloaded as described for the source memory experiments (indeed some images were used in both experiments). Of these, 160 images, were arranged into three blocks and shown sequentially in fixed order. Firstly, a buffer block of 56 photographs, which were not used for the test sequence; secondly, a target block of 48 photographs, consisting of 36 A targets and 12 B photographs (which were not shown again but were similar to the B' lures); finally, a second buffer block of 56 photographs, which were not used for the test sequence (Figure 21).

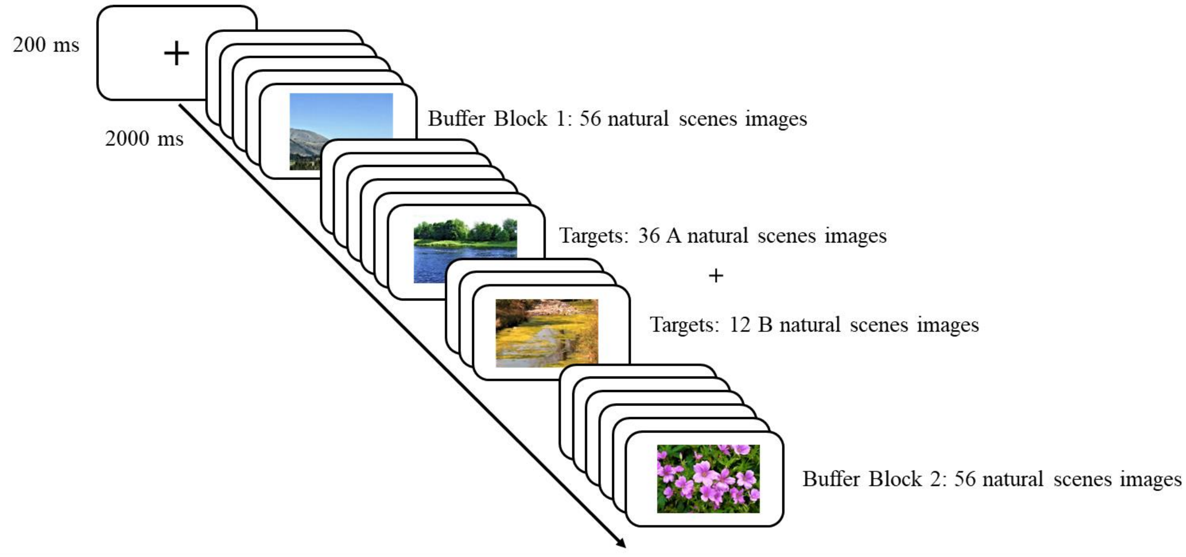

Figure 21. 2-AFC recognition memory task, Procedure for study image sequence:160 pictures were shown in random order within both buffer and target blocks, each image was displayed for 2 seconds following a fixation cross. Targets were not identified as such to participants, who were instructed to try to pay attention to all the images.

All photographs (which were counterbalanced for subject type) were shown in random order within each block. Each photograph was displayed singly for 2 seconds following a fixation cross. No indication was given to participants as to which pictures were targets; all 160 pictures were shown consecutively with no interval between blocks; participants were asked to pay attention to all of the pictures, some of which would be used for a subsequent recognition memory task. Pictures were screened to eliminate similar photographs from the test and buffer blocks. At the end of the presentation sequence participants were asked to look at a "Where's Wally" puzzle for 2 minutes to allow for a standard timed break between study and test sections.

Test sequence.

36 selected photographs designated as A targets were presented between two buffer blocks as described in the study sequence (Figure 21); 12 selected photographs designated as B pictures were also shown at study but were not shown again at test. At test, the 36 A targets were shown paired with a lure, (a new previously unseen picture), with 12 picture pairs shown in each of three conditions: (A-A′), the lure was perceptually similar to the A target; (A-B′), the lure was dissimilar to the target but was perceptually (mnemonically) similar to a B picture shown at presentation within the target block; and (A-X), the lure was dissimilar to the target or to any other picture shown at presentation. All test picture pairs were matched and counterbalanced for natural scenes subject category (3 of the 12 test pairs were taken from each natural scenes subject category in each of the three conditions), Figure 22.

All participants carried out the same task, viewing the same test picture pairs in each of the three conditions. Test picture pairs were displayed for 6 seconds, being shown in random sequence, i.e., the picture-pair conditions were randomly displayed. The target picture was pseudo-randomized to appear either on the left- or right-hand side of the screen, this did not change between participants.

Behavioural Responses.

Ratings for both vividness or familiarity and confidence applied only to the image selected as old and followed directly after the decision had been made. Details are supplied under each experiment; responses were self-paced and were made using sliders, on a scale from 0 – 100. Rather than a remember or ‘know’ question or a recollect or familiar question I used a remember, yes or no, and a recollect, yes or no question, answered by mouse click on a labelled response button. After the final response had been made, on pressing the continue button the screen moved on to the next test pair.

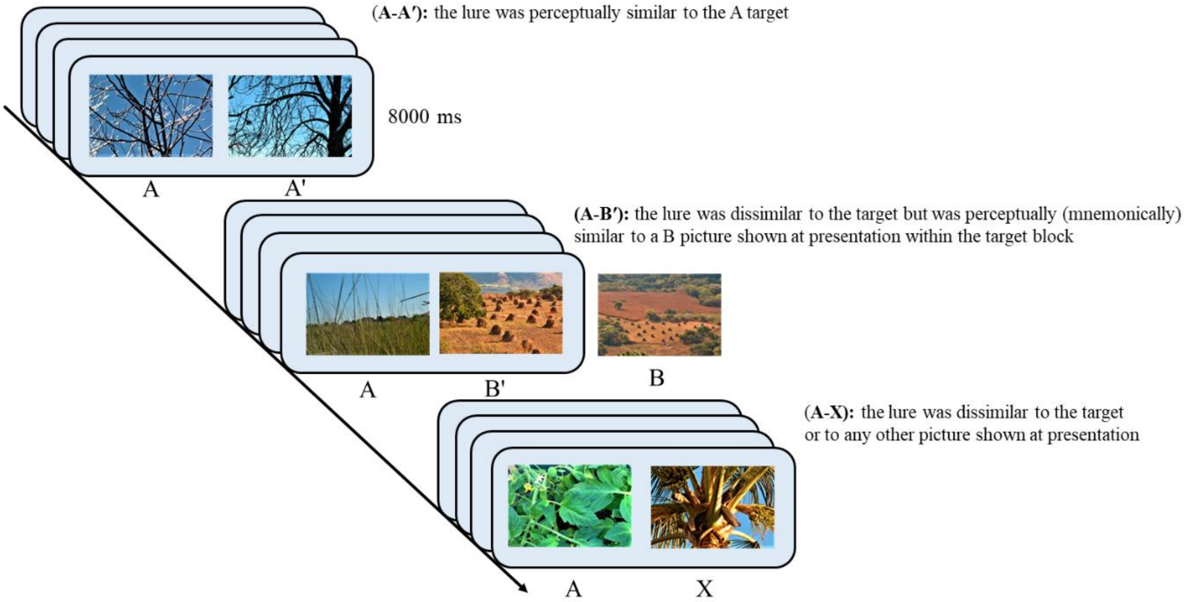

Figure 22. Vividness and Familiarity, Experiments 3 & 6, Example picture pairings used in the test phase (all pairs were shown in random order), for illustration purposes, the target A picture is on the left and the lure on the right. For comparison, the B picture shown in the target block is also illustrated (not shown at test).

Response time (milliseconds).

The software platform automatically collected response time data for the old or new decision at test. If the participant did not make the decision within the allocated time, although they could still go on to make behavioural responses, their data was excluded from the analysis (as they had not indicated an old or new decision).

Data Analysis plan.

The 2-AFC task represents recognition accuracy as a binary response. The question “if we are accurate are we confident” can be addressed in diverse ways. Pickering (2017) gives another interpretation on improving the clarity of data presentation in clinical situations. A bivariate confidence (a continuous variable) and accuracy (a binary variable) scatterplot will give the distribution shown in Figure 34, also termed a strip plot, (cf. Figure 10). More detail is shown in a univariate scatterplot, to allow for critical evaluation of the continuous data, as suggested by Weissgerber et al., (2015).

Confidence, and vividness or familiarity, were analysed on a by trial basis to compare continuous variables by the binary outcomes of performance (accuracy: correct or incorrect recognition) and remember/recollect (yes or no) judgments. There were differences in mean values averaged across all trials, compared to using averaged scores for each participant; not every participant completed 36 test trials (due to timed-out trials being excluded). Another way of approaching data description is to average confidence scores for each participant (with error bars representing 95% *C.I.*, and display the results by groups, (Figure 23).

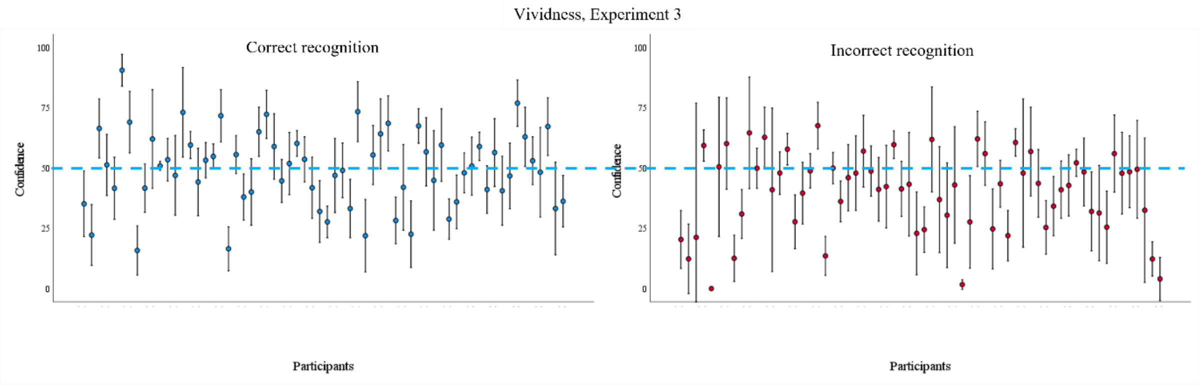

Figure 23. Vividness, Experiment 3: confidence data averaged for each participant and presented as an error chart (error bars represent 95% C.I.). The y-axes are identical and the dashed blue line is placed at confidence = 50 to aid comparison

There may be a lack of clarity when presenting data in this way. Demonstrably there is considerable variation in the averaged results; error bars are wider for incorrect responses, representing more variance in participant's confidence ratings when recognition is inaccurate. As many of the averaged results are close to confidence = 50, it is tempting to simply average the averaged data from participants. This will result in a point biserial plot, as illustrated in Figure 10. The abscissa scale can be altered to highlight any differences. The type of univariate (strip) plots illustrated in Figure 34 contain more information than in Figure 10.

In a 2-AFC task, as two stimuli are shown together at test, participants can compare them to make their recognition decision, this eliminates response bias and the need for ROC curve analysis. A calibration curve or rather, CAC analysis is used as an alternative and more intuitive method of presenting old or new test results (Tekin et al., 2021). Unfortunately, this is also based on using binned and averaged data.

Chapter 4: Vividness.

Here, I use the continuous response source memory task described in Chapter 3. My aim is to explore whether memory quality, represented by vividness, influences our metacognitive judgments of confidence in the accuracy of recollection for the related detail of the memory (source memory). I ask if whether more generally, the detail of our vivid memories may be confidently, but inaccurately recollected.

(i) Vividness, Experiments 1 & 2.

Rationale

For these experiments, the continuous-response source memory task provides a measure of memory accuracy based on the participant at test recollecting information for the location of a cross on a circle (representing the detail of a memory), cued by a picture shown in a relational encoding block at study. The method provides fine-grained data for memory accuracy, confidence, and vividness. My focus is on comparing subjective confidence in the memory's accuracy with both the objective (factual) accuracy of recollection for the cross location and with the quality (vividness) of memory for the cue picture.

Here, based on the current literature, I hypothesise that vividness (as a surrogate for memory quality) colours our experience of remembering, i.e., gives us the illusion that we remember an event "well" or accurately, for general recall of our personal memories rather than specifically for eyewitness testimony. If we refer to a past memory as seeming vivid, it might feel like we are reliving the experience (Buchsbaum et al., 2012; Rubin et al., 2019), by recalling sensations such as smell or sound, as well as the emotions we may have experienced at the time. Vividness resulting from emotional memories may bias our judgments of belief in the accuracy of a memory (Doerksen & Shimamura, 2001; Durbin et al., 2017; Johnson et al.,1993), for this reason, neutral images were used.

We may base our assumption that externally vivid stimuli are recollected accurately on results of old or new recognition memory tasks showing, e.g., that accurate recognition of pictures in black-and-white (Wichmann et al., 2002), or grey-scale monochrome images of natural scenes (Spence et al., 2006) is inferior to that for colour pictures. In my experiments, I deliberately gave no instructions to participants concerning what was meant by vividness, so as not to influence the criteria they might use. However, qualitative data was collected at completion of the experiments to capture what participants were using to judge vividness of the images when remembered.

Results from two experiments are reported here, with picture sets either in full colour or manipulated to be less visually vivid. I hypothesise that memory quality such as vividness influences confidence in a memory more than vividness influences source memory accuracy.

Experiment 1

Methods.

Procedures were followed as described in Chapter 3: General Methods (i), using the continuous-response source memory task (Figure 17).

Power calculations.

Power calculations were performed using a freely available online psychology statistics tool to ensure that the planned number of participants was sufficient, [AI-Therapy statistics for psychologists (https://www.ai-therapy.com/psychology-statistics/power-calculator)]. A sample size of 32 participants provides sufficient statistical power (.8) to assess a between-groups correlation coefficient (Pearson's *r*) at 5% significance level (two-tailed).

Stimuli.

A 120-image set of natural scenes photographs was used in Experiment 1, which consisted of ten cross location-natural scene relational encoding blocks and ten cross location-natural scene recollection test blocks (each having twelve trials, counterbalanced for subject category, i.e., comprising three different images of flowers, green landscapes and trees, mountains, or water), as illustrated in Figure 24. No negative emotional pictures were used; participants were reassured that there would be no deception and that none of the images would be upsetting or distressing. A total of 120 different cross locations (out of a possible 360 locations 1° apart) were used, randomly paired with each of the photographs.

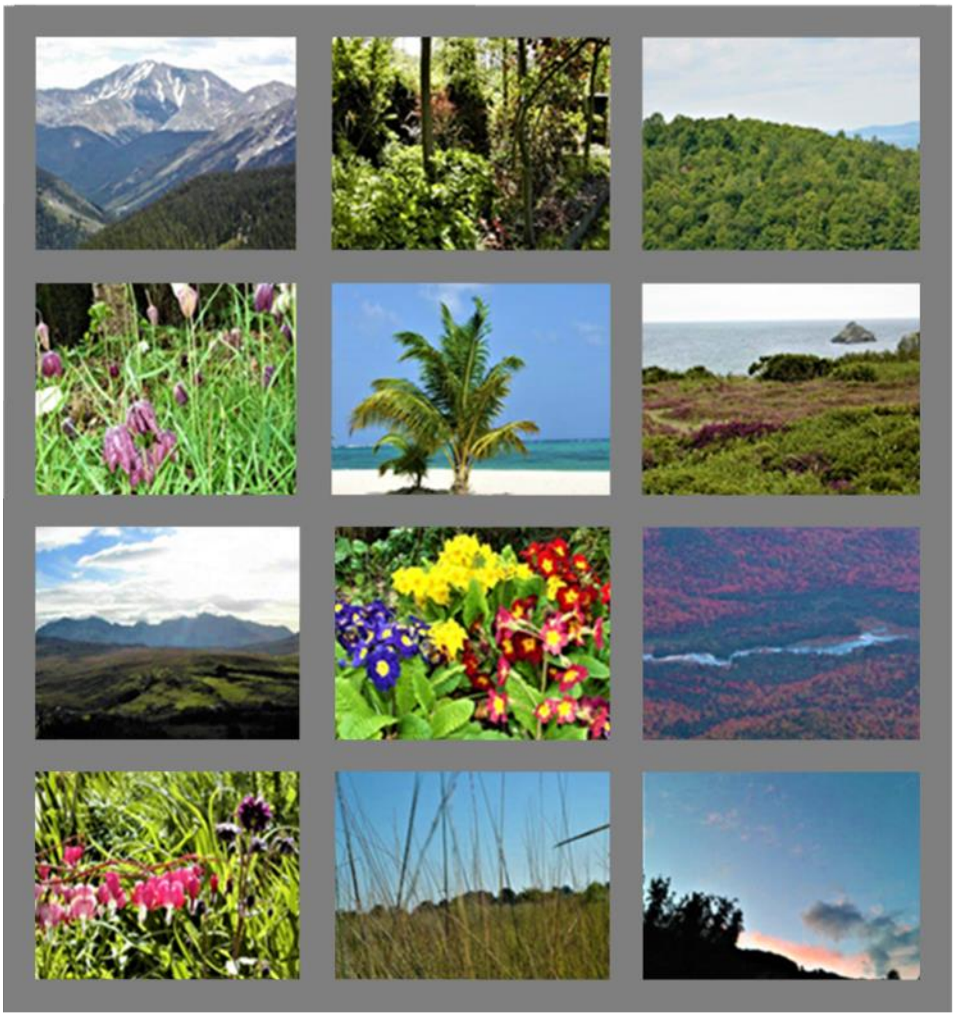

Figure 24. Vividness, Experiment 1: Example natural scenes photographs; 12 natural scenes colour photographs used for both study and test blocks (comprising three different pictures from each subject category: flowers, green landscapes and trees, mountains, or water)

Experiment 2

Rationale.

One criticism that could be made of the image set used in Experiment 1 is that all the images, being in full colour, were equally vivid in terms of physical properties (i.e., external vividness - sharp contrast, different hues), which might force participants to give subjective ratings of vividness on other dimensions. To counter this, in a second experiment, the 120-image set in full colour was augmented by the addition of 120 different photographs of natural scenes that had been manipulated to be less visually vivid. Rather than using black-and-white or grey-scale monochrome images, both of which manipulations are known to reduce recognition accuracy (Wichmann et al., 2002; Spence et al., 2006), the additional photographs were partially desaturated for colour and not adjusted for contrast (sharpness).

Methods.

Procedures were followed as described in Chapter 3: General Methods (i), using the continuous-response source memory task described (Figure 17).

Stimuli

A 240-image set of natural scenes photographs was used in Experiment 2, half of which were in full colour and half partially desaturated for colour. The experiment consisted of fifteen cross location-natural scene relational encoding blocks (each having sixteen trials, counterbalanced for subject category and colour saturation), and fifteen cross location-natural scene recollection test blocks, as illustrated in Figure 25. I point out here that it was not possible to use 12 trials in each block; with three photographs from each subject category, it would not be possible to counterbalance for colour saturation. The same procedures were otherwise followed as for Experiment 1.

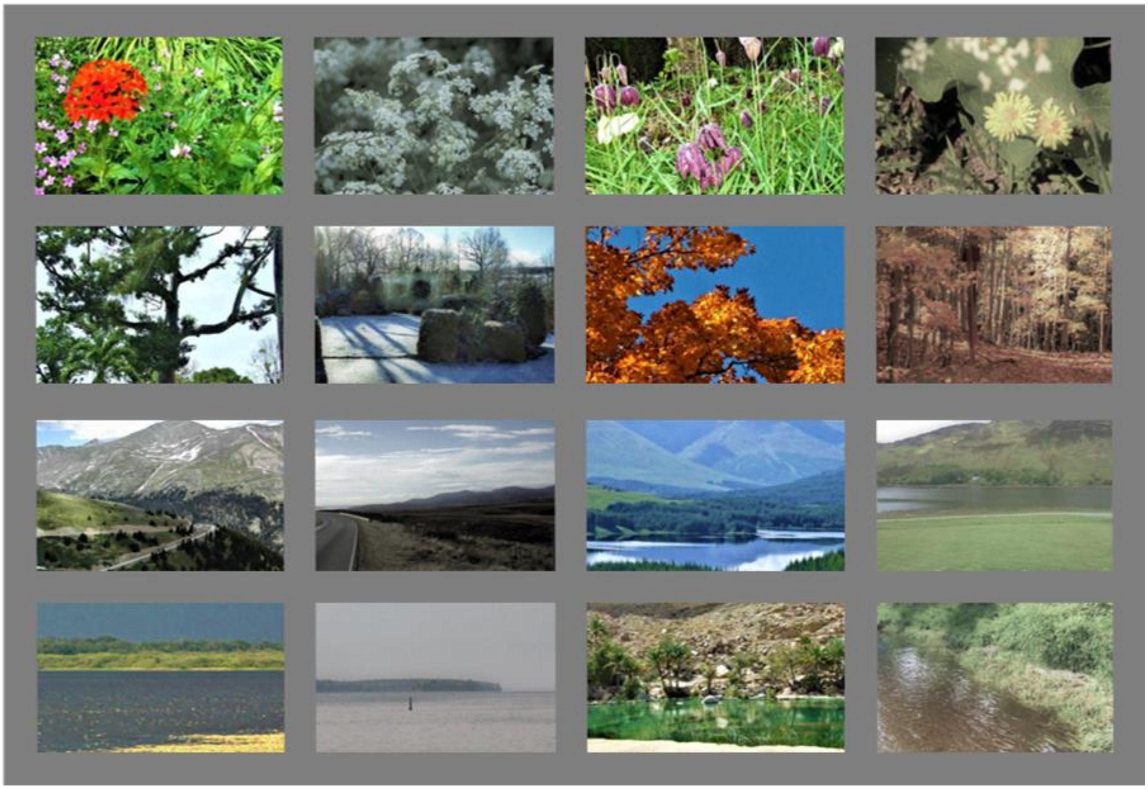

Figure 25. Vividness, Experiment 2: example natural scenes photographs. 16 natural scenes colour photographs used for both study and test blocks (comprising four different pictures from each subject category, counterbalanced for colour saturation: flowers, green landscapes and trees, mountains, or water)

Results

Participants.

There were sixteen participants in Experiment 1, mean age 29.5 years (*S.D.* 2.1 years); 14 identified as female. Data from all participants were included in the final analysis.

There were sixteen participants in Experiment 2, mean age 20.2 years (*S.D.* 3.6 years); 15 identified as female. Data from three participants were excluded from the analysis as their location error frequency plots showed no evidence of recollection, as illustrated in Figure 18b; data from thirteen participants were included in the final analysis.

Vividness rating by Experiment (picture set).

Results are compared for the two experiments (i.e., by the picture set used). In Experiment 2, with the addition of partially desaturated images, there is a more uniform distribution of vividness judgments throughout the range, Figure 26. In Experiment 1, individual trial results are clustered towards the upper end of the scale whereas in Experiment 2, although there is clustering at the upper end of the scale, results also cluster at the lowest end of the scale (corresponding to partially desaturated images), further confirmed by the variable density plots.

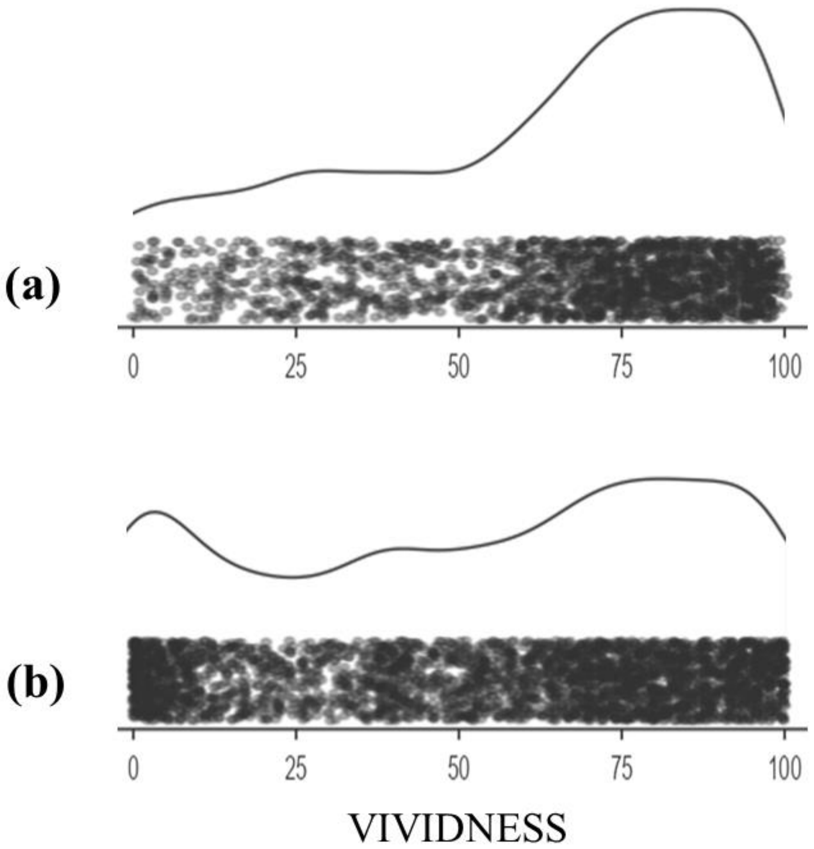

Figure 26. Vividness, (a)Experiment 1 (b) Experiment 2: univariate scatterplots overlaid by variable density plots illustrating vividness distributions using data from all trials. The difference in the number of trials in each experiment is apparent from the density of the scatterplots. Nevertheless, it is clear that the distributions are different and that in Experiment 2 a wider range of vividness ratings are made in comparison to Experiment 1, thus justifying the rationale behind Experiment 2

Data from either 120 trials (Experiment 1), or 240 trials (Experiment 2), allowed the source memory error to be independently characterised for all trials carried out by each participant and converted to provide a fine-grained measure of source memory accuracy as a percentage for each individual trial.

Memory Performance.

Despite the addition of partially desaturated images in Experiment 2 and the resulting list length discrepancy, the ability of assessable participants to carry out the task was unaffected. There were no significant differences in memory performance indicators (recollection rate, scale and threshold values) between the two image sets, i.e., Experiments 1 and 2, Table 2.

Table 2. Vividness, Experiments 1 & 2: memory performance [recollection rate, scale (precision) and threshold (degrees from target), M (SE)], results of independent t-tests to compare mean values for all assessable participants (N = 29) by experiment (image set)

Experiment 1: 120-image set, full colour
Experiment 2: 240-Image set, full colour/partially desaturated for colour

Image set	Variable	*n*	*M* (*SE*)	Independent samples *t*-test *t*(*d.f.*)	*p*-value 2-tailed
120-image set	Rate	16	0.56 (0.06)		
240-image set		13	0.45 (0.06)	1.27 (27)	.22
120-image set	Scale	16	10.07 (1.36)		
240-image set		13	12.07 (1.38)	0.92 (27)	.32
120-image set	Maximum Error	16	24.56° (2.88°)		
240-image set		13	25.69° (5.66°)	-0.19 (27)	.85

The similarity in overall performance is further illustrated by the location error frequency plots, (Figure 27). The difference in the number of trials carried out by each participant (twice as many in Experiment 2) is apparent from the response frequency scale on the ordinate, acknowledging the discrepancy in the number of participants between experiments.

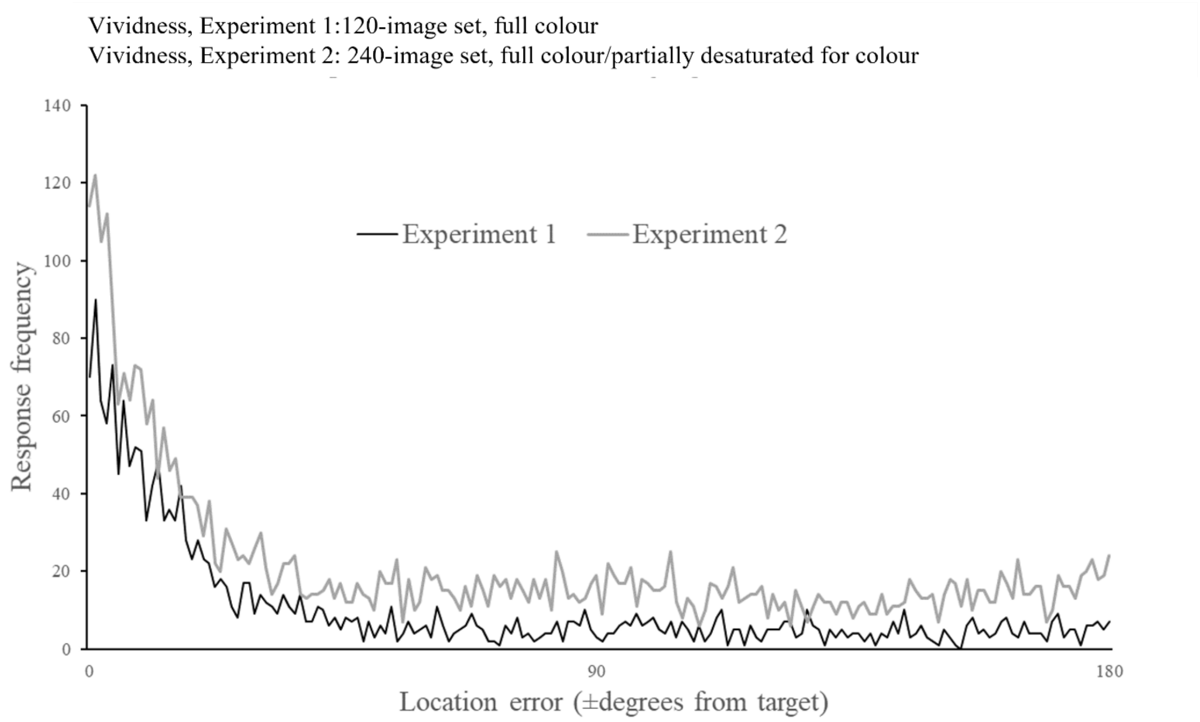

Figure 27. Vividness, Experiment 1 [(120-image set, full colour) N = 16]; Experiment 2 [(240-image set, full colour and partially desaturated for colour) N = 13]. Location error frequency plots showing response frequency, summed across all participants for each degree of error from the target (0° – ± 180°)

Variables of interest.

Variables of interest (vividness, source memory accuracy & confidence), averaged across all trials, are illustrated in Table 3, which shows that vividness, confidence, and source memory accuracy were all significantly higher in Experiment 1, when all images were in full colour. Due to the substantial number of trials the *p*-values cannot be relied upon; the effect size for accuracy was weak, with effect sizes for vividness and confidence above weak.

Table 3. Vividness, Experiments 1 & 2: Variables of Interest, averaged across all trials [M (SE)] by experiment, compared by R^2 change

Vividness, Experiment 1: 120-image set, full colour Vividness, Experiment 2: 240-Image set, full colour/partially desaturated for colour						
Experiment	Variable	*df*	*M (SE)*	Unstandardized *ß* coefficient (*SE*)	*p*-value[a]	R^2 change
120-image set	Vividness	1,5038	67.4 (0.6)			
240-image set			55.3 (0.4)	-12.31(0.86)	< .001	3.9%
120-image set	Accuracy	1,5038	74.1 (0.6)			
240-image set			67.6 (0.6)	-6.55 (0.87)	< .001	1.1%
120-image set	Confidence	1,5038	54.9 (0.7)			
240-image set			42.6 (0.6)	-12.3 (0.90)	< .001	3.6%

Note. [a]The p-value can appear significant because the number of trials is large; of more relevance is the effect size, denoted by R^2 change, which is documented only when the p-value is reported as significant [R^2 change: 1% - weak, 9% - moderate 25% - strong effect].

Correlations.

Correlations, using all trial data and averaged within-participant correlations are illustrated in Table 4. The averaged within-participant correlations result in a slight underestimation in comparison to those obtained from all trial data combined. However, this does not substantially affect the overall size of the correlations or the variance, further supporting the results obtained. For further clarity, Appendix B provides detailed individual correlations and *p*-values in Tables 23S and 24S, further illustrated by scatterplots showing individual linear correlation lines (Figures 67S-72S).

The correlation between vividness and source memory accuracy was moderate for both experiments. Variability in vividness of the pictures explained 10.7% of the variance in source memory accuracy in Experiment 1, and 15.9% of the variance in Experiment 2. The relationship between accuracy and confidence was moderate to large, and positive for both experiments, Table 4, regardless of the method of analysis; just over 20% of the variance in confidence can be explained by variability in source memory accuracy.

Critically, the correlation between confidence and vividness in both experiments was very large. In Experiment 1, variability in vividness of the pictures explained 42.3% of the variance in confidence, Table 4.

Table 4. Vividness, Experiments 1 & 2: correlation analysis by experiment (image-set), denoting relationships between variables of interest. Pearson's r-values are displayed using data from all trials and averaging within-participant correlations. It is evident that there is little difference between the two methods

Vividness, Experiment 1: 120-image set, full colour
Vividness, Experiment 2: 240-Image set, full colour/partially desaturated for colour

Experiment	I.V.	D.V.	*N*	Pearson's *r* *(all trials)*	*p*-value 2-tailed	R^2 variance	Pearson's *r* *(averaged)*	R^2 variance
120-image set	Vividness	Accuracy	1920	.33	< .001	10.7%	.29	8.4 %
240-image set	Vividness	Accuracy	3120	.399	< .001	15.9%	.37	13.7 %
120-image set	Vividness	Confidence	1920	.65	< .001	**42.3%**	.65	**42.3 %**
240-image set	Vividness	Confidence	3120	.76	< .001	**57.8%**	.69	**47.6 %**
120-image set	Accuracy	Confidence	1920	.456	< .001	20.8%	.41	16.8 %
240-image set	Accuracy	Confidence	3120	.48	< .001	23.3%	.43	18.5 %

Note. Cohen (1992) defined small, moderate (i.e., medium), and large *r* values as "about" .10, .30, and .50, respectively. R^2 represents the amount of variance in the dependent variable predicted by variability in the independent variable.

Results for the correlation analyses are further illustrated by the correlation matrices (scatterplots) and variable density plots (smoothed histograms), in Figures 28 and 29. From the correlation matrices, a very large positive relationship is illustrated between vividness and confidence for both image sets. Appendix B provides detailed individual correlations and *p*-values, further illustrated by scatterplots showing individual linear correlation lines.

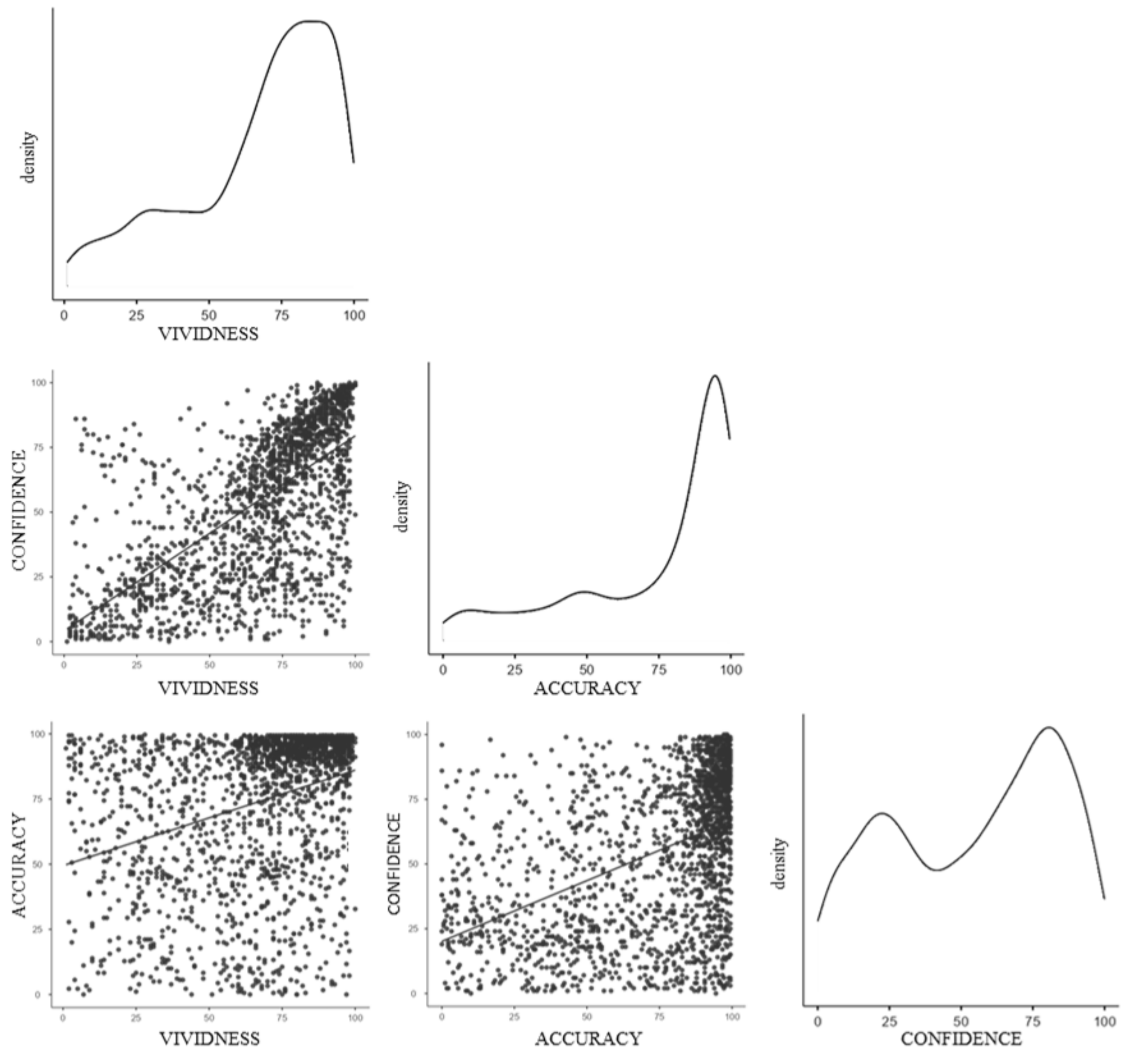

Figure 28. Vividness, Experiment 1, 120-image set in full colour): illustrating correlation matrices (scatterplots) and variable density plots (i.e., the smoothed histograms of the variable distributions) across all individual trials (N = 1920, where scales on the abscissa and the ordinate represent the data on a 0 –100 scale for the correlation matrices, and the scale on the ordinate represents the density function for the variable density plots

However, the variable density plots were not comparable, with a peak at low levels of vividness and of confidence for the 240-image set (Figure 29) assumed to correspond to desaturated or blurred images. Note that the variable density plots for accuracy are similar for both image sets and reproduce the location error plots illustrated in Figure 27.

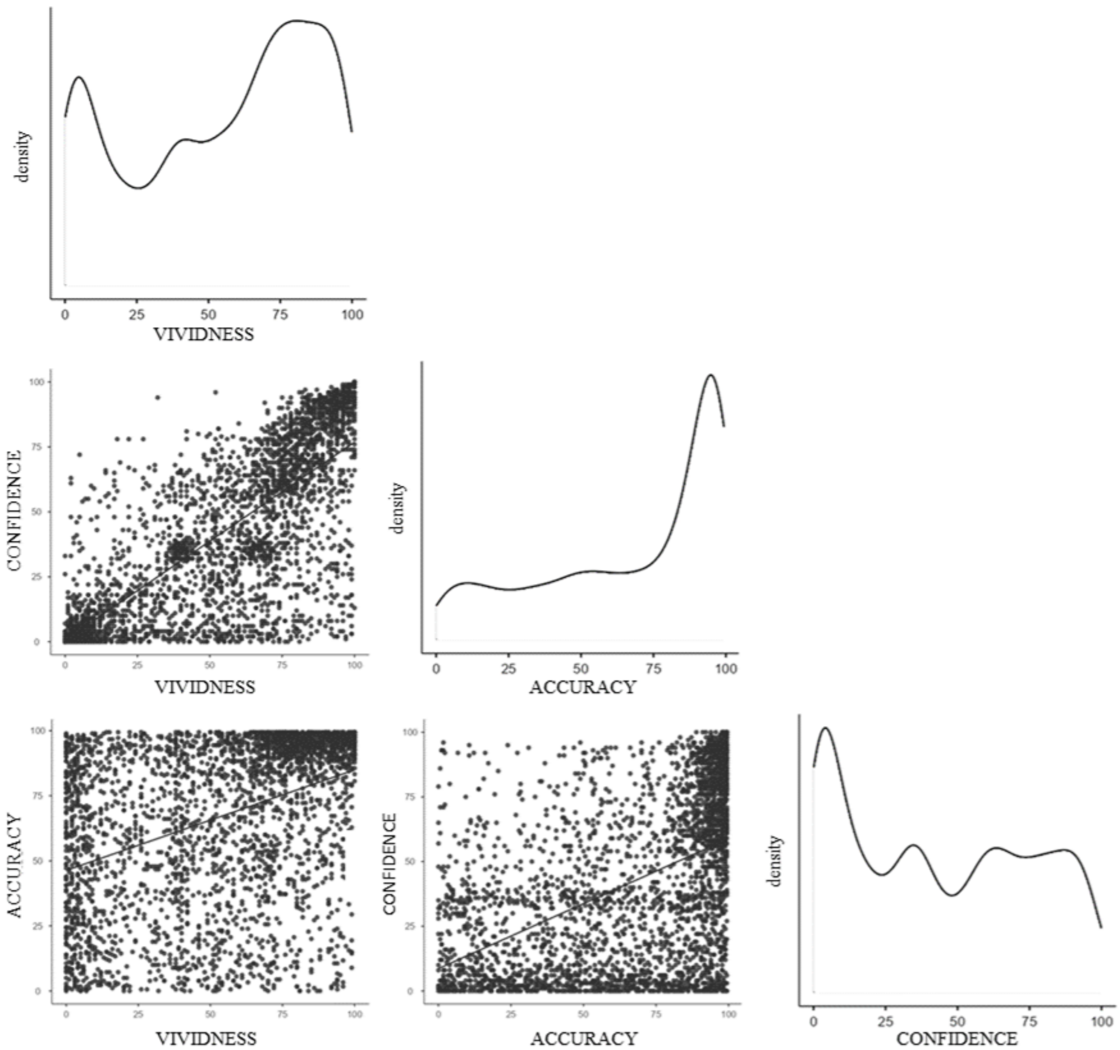

Figure 29. Vividness, Experiment 2, 240 image-set half in full colour and half partially-desaturated for colour: illustrating correlation matrices (scatterplots) and variable density plots, i.e., the smoothed histograms of the variable distributions across all individual trials (N = 3120), where scales on the abscissa and the ordinate represent the data on a 0 –100 scale for the correlation matrices, and the scale on the ordinate represents the density function for the variable density plots

Experiment 2: Manipulation of Vividness

Additionally, Experiment 2 affords the opportunity to compare results by stimulus type, i.e., comparing trials where the photographs were in full colour with trials where the images were partially desaturated for colour, to investigate how stimuli manipulated to be less visually vivid influence the results.

Memory performance.

The manipulated visual vividness of the retrieval cue images did not affect memory performance in this task. The similarity of memory performance across colour photographs and those partially desaturated for colour is evident from the location error frequency plot illustrated in Figure 30.

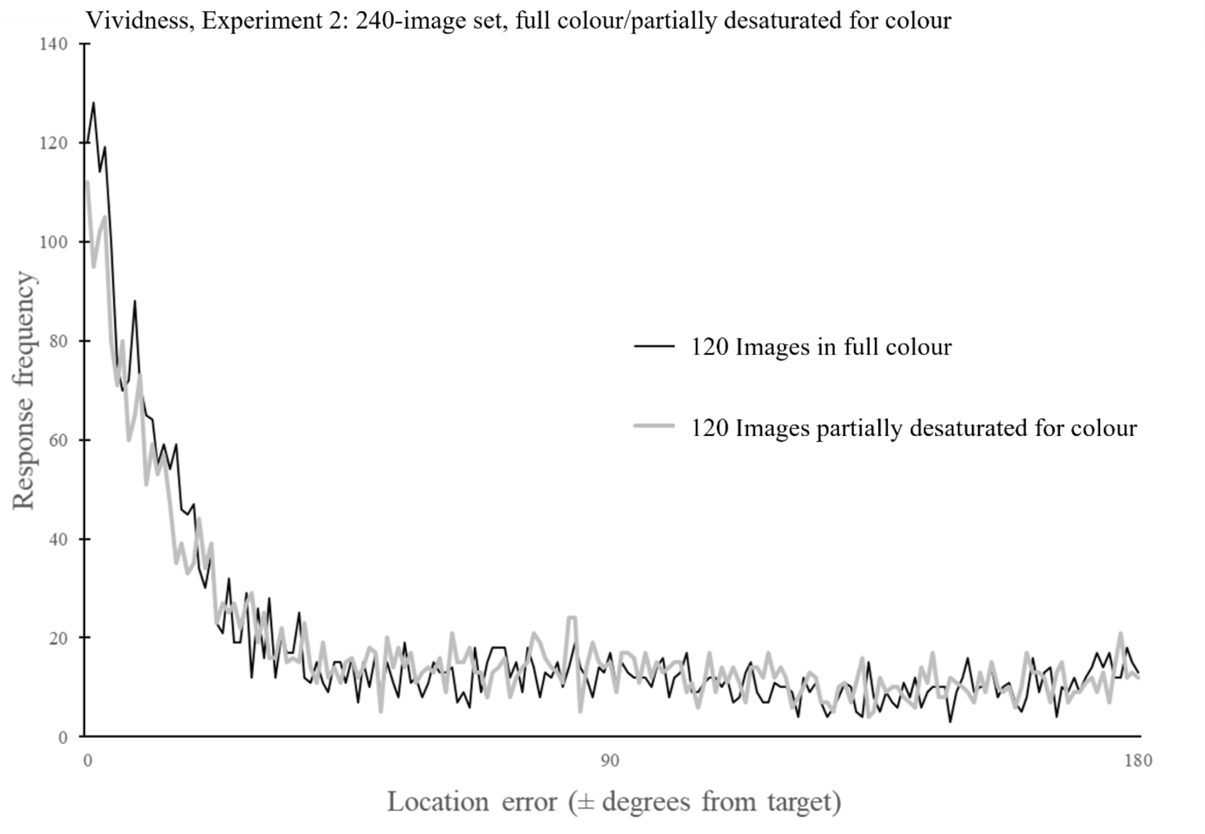

Figure 30. Vividness, Experiment 2 (240-image set): Location error frequency plots showing response frequency summed across participants for each degree of error from the target (0° – ± 180°) for photographs compared by image type,120 in full colour and 120 partially-desaturated for colour

Variables of interest (vividness, source memory accuracy & confidence).

For photographs in full colour, both vividness ratings and confidence were significantly higher compared to those partially desaturated for colour; the effect size was above weak for vividness, and weak for confidence (Table 5). Despite the significant differences in vividness and confidence, image manipulation had no significant effect on source memory accuracy, (*p* = .09), Table 5.

Table 5. Experiment 2: Variables of Interest (averaged across all trials) by stimulus type [M (SE)], compared by R^2 change

Vividness, Experiment 2: 240-Image set, variables by image colour (full colour vs. partially desaturated for colour)						
Stimulus type	Variable	*n*	*M (SE)*	Unstandardized *β* coefficient (*SE*)	*p*-value[a]	R^2 change
Full colour	Vividness	1, 3118	60.88 (0.78)			
Partially-desaturated			49.22 (0.82)	1.55 (1.13)	< .001	3.3%
Full colour	Accuracy	1, 3118	69.52 (0.78)			
Partially-desaturated			66.63 (0.78)	1.89 (1.11)	.09	-
Full colour	Confidence	1, 3118	46.81 (0.82)			
Partially-desaturated			38.3 (0.8)	8.51 (1.15)	< .001	1.7%

Note. [a]The *p*-value can appear significant because the number of trials is large; of more relevance is the effect size denoted by R^2 change, which is documented only when the *p*-value is reported as significant [R^2 change: 1% - weak, 9% - moderate 25% - strong effect].

Moderation analysis

Results from both experiments confirmed that vividness correlated moderately with accuracy but very strongly with confidence. To test whether the relationship between confidence and accuracy differed according to memory vividness for the retrieval cue picture, a simple moderation analysis was carried out. It is considered that, for this experimental task, as confidence is assessed after recollection of the cross position to yield location accuracy (Figure 17), that the continuous variable of accuracy acts as the independent (predictor) variable [X] for the analysis. The outcome variable [Y] for the analysis was the continuous variable of confidence. The moderator variable [W] for the analysis was the continuous variable of vividness.

Results confirmed a significant overall R^2 effect for the model, ($p < .001$); 60.5% of the variance in confidence is predicted by the model containing both vividness and accuracy. In terms of the whole model, the additional effect of accuracy on confidence was not significant ($p = .136$). The effect of vividness on confidence was significant ($p < .001$). The interaction effect (accuracy*vividness) on confidence was also significant ($p < .001$), with a small effect size, R^2 change = 1.9%[14]. At low levels of the moderator (vividness = 29.5), the [conditional effect of accuracy on confidence, *B* = 0.12, 95% *C.I.* (0.093, 0.144), $p < .001$]. At middle levels of the moderator (vividness = 59.7), the [conditional effect of accuracy on confidence, *B* = 0.27, 95% *C.I.* (0.25, 0.29), $p < .001$]. At high levels of the moderator (vividness = 89.9), the [conditional effect of accuracy on confidence, *B* = 0.42, 95% *C.I.* (0.39, 0.45), $p < .001$. A simple slopes plot for the moderation analysis, Figure 31, illustrates that the association between accuracy and confidence was conditional on memory vividness. As vividness increases, confidence increases for the same level of source memory accuracy; the effect is larger (steeper slope) at higher levels of vividness, i.e., the slopes diverge as vividness increases. These results identify memory vividness to be a positive moderator of the relationship between memory accuracy and confidence, the effect size is small (R^2 change = 1.9%).

[14] For the sake of completeness, the analysis was repeated with confidence as the predictor variable and accuracy as the outcome or dependent variable. The model was weaker although there was still a significant overall R^2 effect, ($p < .001$), reducing to 23.9% the amount of variance in accuracy predicted by the model containing both vividness and confidence. The individual effect of confidence on accuracy was significant ($p < .001$). In terms of the whole model, the additional effect of effect of vividness on accuracy was not significant ($p = .65$). The interaction effect (confidence*vividness) on accuracy was significant ($p < .001$), but with a trivial effect size, R^2 change = 0.4%. In this model, vividness did not moderate the relationship between accuracy and confidence.

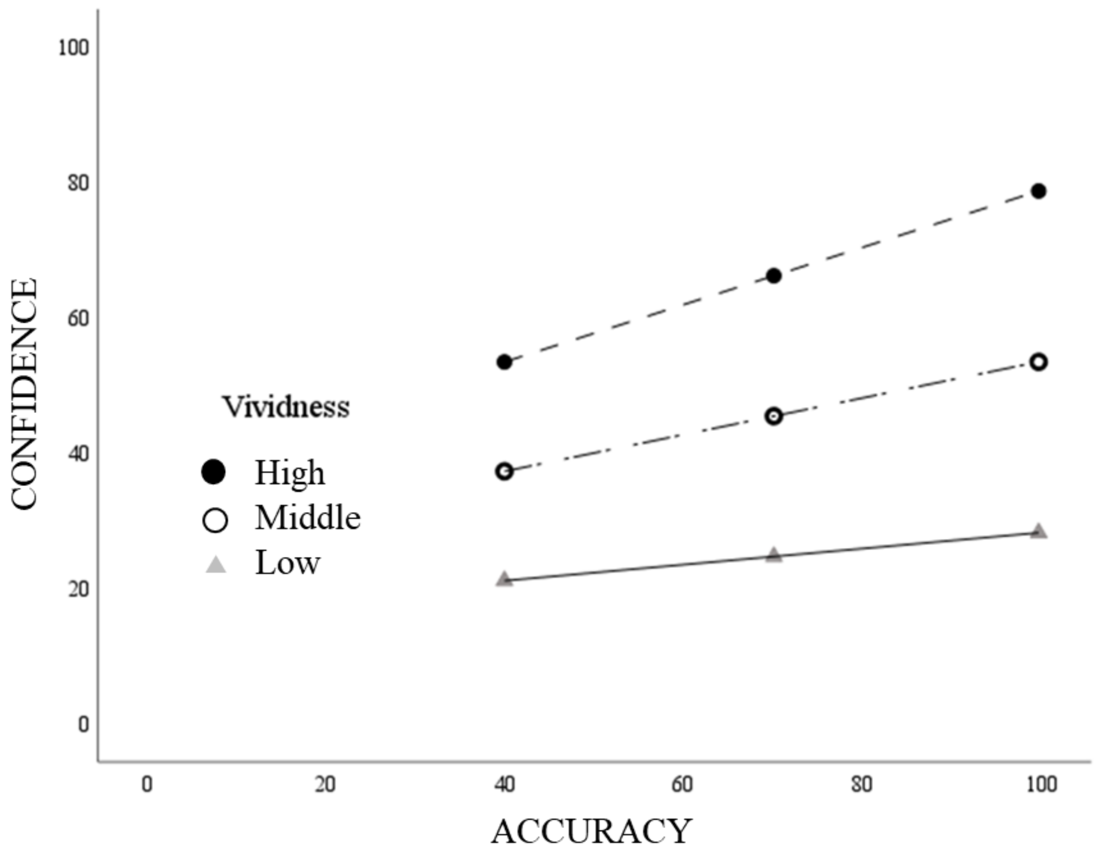

Figure 31. Vividness, Experiments 1 and 2: Simple slopes plots (of the interaction effect) for source memory accuracy at values (M ± 1 S.D.) with confidence by vividness, at high, middle, and low levels of the moderator

Mediation Analysis

There is an assumed strong relationship between vivid memories and accurate memories, however the correlation between vividness and accuracy was only moderate, compared to a very large correlation existing between vividness and confidence in memory's accuracy. Trials were analysed to investigate whether the association between vividness and accuracy acted directly or whether the association was indirect, carried through the mediator confidence. A simple mediation analysis was performed in SPSS using PROCESS (Hayes & Scharkow, 2013). The outcome variable [Y] for the analysis was the continuous variable of source memory accuracy, the predictor variable [X] was the continuous variable of vividness (vividness having been judged immediately after the cue natural scenes image was re-presented, and before the cross location had been indicated (Figure 17). The mediator variable

[M] was the continuous variable of confidence. To test for mediation, whether the indirect pathway was significantly different from zero was assessed. A bootstrapped (10,000 samples) estimate of the confidence interval was calculated for the indirect effect.

As illustrated in Figure 32, the path (a) from vividness to confidence was statistically significant ($B = 0.77$, $p < .001$). The path (b) between confidence and accuracy was statistically significant ($B = 0.41$, $p < .001$). The direct effect (c) of vividness on accuracy was statistically significant, but trivial ($B = 0.07$). For the indirect path (a*b), as the confidence interval [$B = 0.32$, 95% *C.I.* (0.29, 0.34)], did not include zero, it was concluded that the indirect effect of vividness on accuracy was significant. In the context of these experiments, the data did not support the conclusion that the vividness of a memory operated independently of this mechanism to affect accuracy.

I would reiterate that moderation and mediation analyses do not determine causation or the direction of the relationships, but merely serve to confirm the associations.

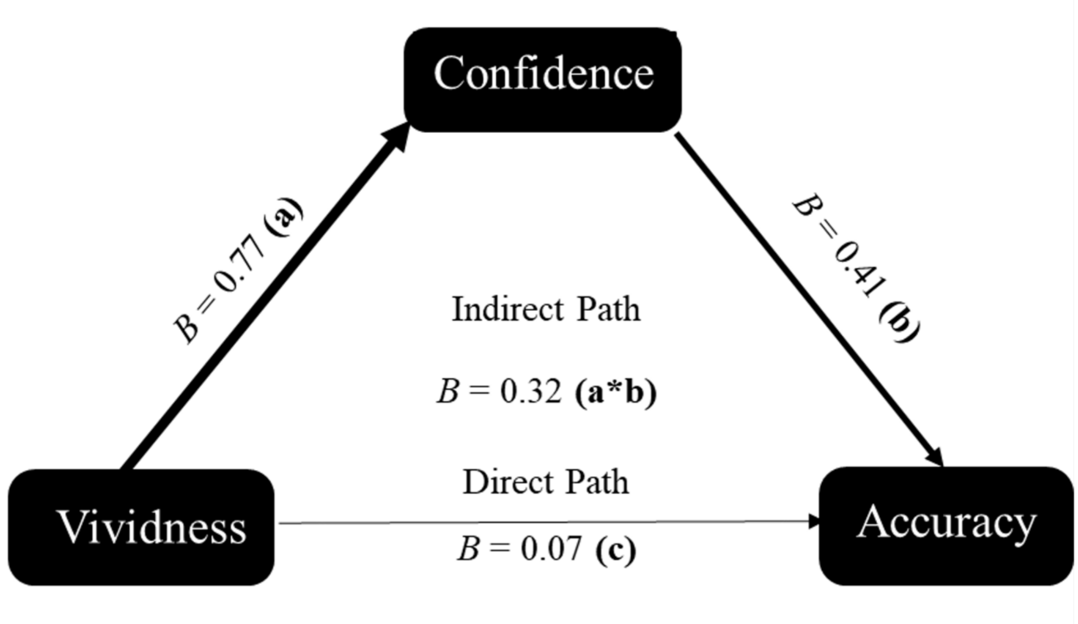

*Figure 32. Vividness, Experiments 1 & 2: The relationship between vividness and source memory (location) accuracy; (a) the path from vividness to confidence; (b) the path between confidence and accuracy; (c) the direct path from vividness to accuracy; (a*b) the indirect path from vividness to accuracy mediated through confidence (Hayes & Scharkow, 2013). Path (a) is larger than (b) and the direct path (c) is trivial; the weight of the arrows reflects the unstandardized beta (B) values*

Zou and Kwok (2022) also carried out a mediation analysis, hypothesizing that vividness partially mediated the relationship between confidence and accuracy. For the sake of completeness mediation analysis on the data from Experiments 1 and 2 was repeated using Zou and Kwok's (2022) hypothesis. The mediation analysis on this data set did not support vividness as a mediator of the relationship between confidence and accuracy; 82.1% of the relationship between confidence and source memory accuracy was carried through the direct path, i.e., was not mediated through vividness.

Qualitative Data Analysis

Criteria for judging vividness.

Qualitative analysis of what metrics participants based their judgments on was carried out by assessing free comments made at completion of the experiments. The data was then compared between experiments using thematic analysis. All 29 assessable participants commented on what made the images seem vivid to them; 27 of the participants also commented on what made the images *not-vivid*.

Data were initially collated into primary codes of external or internal vividness according to the literature (Ford & Kensinger, 2016; 2017). It became apparent that these two primary codes did not fully represent all the comments made. A third primary code was added to denote the participants' concepts of what vividness meant to them. These codes were sorted into themes, checking that data within each theme was coherent, ensuring finally to identify what aspect of the data each theme captured, to provide a consistent narrative. Some of the themes had additional sub-themes to give further information as to what participants were using to judge vividness.

Vivid images.

It was assumed that participants would focus primarily on the external (perceptual) vividness of the photographs. Despite the lack of instruction on what was meant by vividness, participants intuitively used an external primary code to judge images as vivid, external themes being colour (sub-themes: bright colours, or individual colours – red, yellow, orange, purple); subject (sub-themes: primarily flowers); and image quality (sub-themes: clear, bright or sharp images, and contrast). Internal themes referenced personal memories (sub-themes: personal associations, experiences, or connections, e.g., a photograph reminded them of where they used to live, a holiday, or they associated it with a book), and positive emotions, (sub-themes: liking or thinking the picture was attractive). Although the images were novel to participants, the use of an internal primary code was only ever cited when judging images to be vivid, and never to judge images to be *not-vivid*. This combination of external and internal themes fits well with previous findings within the wider literature. In a neuroimaging study, Ford and Kensinger (2016) reported differential patterns of connectivity between hippocampal and frontal brain regions associated with external and internal vividness. Importantly, the current findings add weight to the claim that internal and external sources of vividness are distinct – because internal features were only used to support vividness. Participants' use of concepts was divided into themes of features within the picture (sub-themes: a stand-out feature, a feature in the foreground of the photograph, or clear distinct lines within the image) and interest (subthemes: novelty and distinctiveness). The theme of concepts and its sub-themes highlights an overlap between the criteria participants used to judge vividness and those reported for judging both distinctiveness (e.g., Hunt, 2006) and image memorability (e.g., Rust & Mehrpour, 2020).

'Not-vivid' images.

Participants also used primary external codes to judge photographs as *not-vivid*; however, sub-themes were the opposite of those used for vivid images. Themes were colours (sub-themes: dull colours or individual colours – green, grey, brown); subject (sub-themes: primarily trees and green landscapes); and image quality (sub-themes: close-up, out of focus, fuzzy or blurred images). Importantly, self-referenced internal codes were never used to judge images as *not-vivid*.

Participants also used concepts to account for why a photograph was considered *not-vivid*, sub-themes again being the opposite of those used for vivid images. Themes were featureless images (sub-themes: no distinguishing features, large expanses of sea or sky, nothing stood out, lack of interest, not exciting) and similar pictures. Again, it is noted that participants are judging lack of vividness by comparable criteria as would be expected for similarity or lack of distinctiveness.

Data is illustrated by the percentage of participants using primary codes and themes to judge images as vivid or *not-vivid*, comparing Experiments 1 and 2, Figure 33.

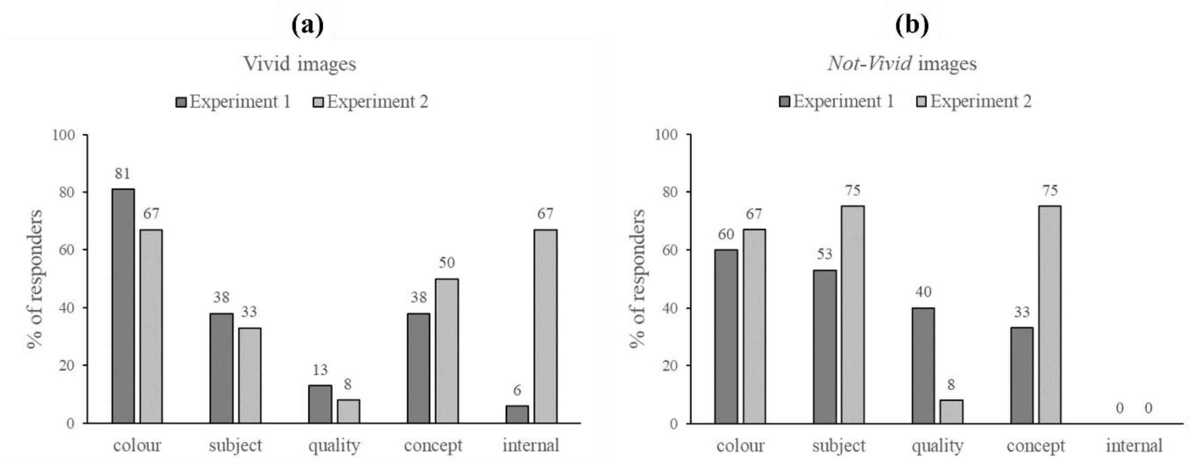

Figure 33. Vividness, Experiment 1 (images in full colour) & Experiment2 (images in full colour and partially desaturated for colour): illustrating themes used by participants to judge photographs as vivid or not-vivid. The numbers represent the percentage of responding participants who had mentioned the primary code or theme (noting that many participants mentioned more than one code and theme)

Comparison: judging images as vivid.

External themes of subject category and image quality were equally used in Experiments 1 and 2 for judging images as being vivid, the theme of colour being more frequently used in Experiment 1 (Figure 33). In Experiment 2, with half the images being of reduced visual vividness, in contrast to Experiment 1, concepts and internal codes were of greater importance for judging the images to be vivid.

Comparison: judging images as not-vivid.

The external theme of subject category was more commonly used in Experiment 2 as compared to Experiment 1 for judging images to be *not-vivid*. Image quality was not a factor of note in Experiment 2, being of more importance in Experiment 1 (when all 120 images were in full colour). To judge the images as *not-vivid*, when internal codes were never used, participants in Experiment 2 favoured the code of concepts (of what lack of vividness meant to them), Figure 33.

Summary.

Using a continuous-response source memory task across two image sets, one of which was manipulated to reduce visual vividness, with real-world natural scenes photographs as retrieval cues for recollection of the location of a cross on a circle, results were compared between the two image sets used (120-image set in full colour, Experiment 1 and 240-image set half in full colour and half partially desaturated for colour, Experiment 2). Memory performance did not differ significantly between the experiments. Full colour natural scenes pictures were associated with significantly higher confidence and vividness ratings, but there was no statistically significant difference in accuracy (when results for Experiment 2 were compared for trials using the 120 full-colour images and trials using the 120 images partially desaturated for colour, ratings of vividness and confidence were again significantly higher for full colour images, accuracy was also significantly higher, but the effect size was weak).

There was a moderate to large correlation between accuracy and confidence, a moderate correlation between vividness and accuracy, and critically, a very large correlation between vividness and confidence. This held true for both experiments (image sets), and for both methods of correlational analysis.

The association between accuracy and confidence was moderated by (conditional on) memory vividness; the moderation effect was larger at higher levels of vividness, Figure 31.

Mediation analysis showed unequivocally that the association between vividness and memory accuracy acted largely (82.1%) indirectly, mediated through confidence.

Qualitative analysis of the criteria participants used for judging images to be vivid or not-*vivid*, showed that the spontaneous recall of personal memories was only ever used to judge images as vivid. A new finding was that participants used their concepts of what vividness meant to them to judge images both to be vivid and not-*vivid*. These concepts overlapped with what would be expected for distinctiveness or similarity, and for image memorability judgments.

Discussion of results.

Confidence in the accuracy of a vivid memory is no guarantee that the memory is accurate. Vividness positively moderated the association between accuracy and confidence, changing ratings of confidence with no change in accuracy. Participants were more inclined to believe they had remembered the location of the cross accurately if the related retrieval cue was remembered vividly. Mediation analysis data did not support the conclusion that the association between vividness and memory accuracy operated directly, confirming that the association between vividness and accuracy was largely indirect, mediated through confidence.

The lack of a strong relationship between vividness and accuracy (with the mediation analysis indicating the relationship was largely indirect), compared to the very large correlation between vividness and confidence, fits well with findings from studies of both flashbulb memories and false memories (e.g., Talarico & Rubin, 2003; 2007).

Criticisms regarding basing memory quality upon the vividness of the natural scenes retrieval cue picture are acknowledged, it could rightly be argued that this is not source memory quality, however, it is a detail of the encoded memory. In the task, source memory is represented by the 120 or 240 individual locations for the cross around the circumference of a circle; if the picture is not recognised, the cross location cannot be recollected, leading to guessing occurring equally at all locations around the circle. My justification for using the quality of memory for the natural scenes picture as a measure of the quality of source memory is supported by participants' spontaneous comments that they related the position of the cross to a branch or specific feature within the picture, confirming that encoding the position of the cross onto the related picture is a feasible strategy to enable participants to carry out the task.

The self-report data highlights the importance of task instructions when studying vividness. If experimental participants are given explicit instructions to use external features of images such as colour, brightness, or contrast to judge vividness, there is a strong risk of losing the additional contribution to vividness made by both concepts and internal vividness, (the uncontrolled recollection of idiosyncratic personal memories and connections).

It might well be argued that the results may relate to the type of task used, particularly as eyewitness research is based on recognition memory and not on recollection of source memory. Whether results would replicate in a different task, i.e., an old or new recognition memory task is tested in Experiment 3, using the 2-AFC recognition memory task detailed in Chapter 3, General Methods (ii). Additionally this task provides the opportunity to investigate why remembering a target is associated with higher confidence, and whether this is also affected by memory quality.

(ii) Vividness, Experiment 3.

Historically the literature provides evidence that confidence and memory accuracy can dissociate. The 2-AFC task used is based on Tulving's (1981) picture similarity experiment, which demonstrated that participants can confidently recognise as old, new pictures that are similar to a picture seen at study, causing a confidence and accuracy to dissociate. In advocating that confidence and accuracy remain tightly coupled, (e.g., Brewin et al., 2020; Roediger & Tekin, 2020; Wixted & Stretch, 2000), such findings are disregarded.

Rationale

Experiments 1 and 2 confirmed a very large correlation between vividness and confidence in accurate recollection of source memory. Results of old or new recognition memory tasks have demonstrated that accurate recognition of pictures in black-and-white (Wichmann et al., 2002), or grey-scale monochrome images of natural scenes (Spence et al., 2006) is inferior to that for colour pictures. Establishing that the more vivid or distinctive a stimulus is, the more accurately it is recognised (Jacoby, Baker & Brooks, 1989), implies that such recognition will be confident. The additional information regarding whether remembering is associated with the recognition decision is provided by asking a remember or 'know' question. Importantly, remembering is also associated with high confidence (Dunn, 2004; Gardiner, 1988; Tulving, 1985), but is remembering always accurate? Gardiner (1988) called for further evidence to explain relationships between measures of conscious awareness (remembering) and confidence ratings, unresolved at that time. The correlation between remember and 'know' judgments and recognition accuracy was also queried.

In addition to investigating if the very large correlation between vividness and accuracy replicates in a separate memory task, I try to address the question of how remembering, confidence, and accurate recognition interact. In Experiment 3, participants are asked to rate their memory for the recognised picture for vividness; answer yes or no to whether they remembered the picture recognised (so as not to force participants to answer 'know' if a stimulus is not remembered); and rate their confidence in recognition accuracy.

Methods.

The experiment was carried out online. Procedures were followed as detailed in Chapter 3, General Methods, illustrated in Figures 21 and 22.

Power calculations.

Power calculations were performed using a freely available online psychology statistics tool to ensure that the planned number of participants was sufficient, [AI-Therapy statistics for psychologists (https://www.ai-therapy.com/psychology-statistics/power-calculator)]. A one-group sample size of 50 participants provides sufficient statistical power (.93) for an experiment designed to determine if two data sets (e.g., remember yes or no) are significantly different from each other at effect size = 0.5 and 5% significance level (two-tailed).

Stimuli

A total of 196 colour natural scenes images (including a majority of those used in Experiment 1) were selected. These comprised equal numbers of the four different categories of natural scenes subject (green landscapes and trees, flowers, mountains, or water). Details of the buffer and target blocks used at study, and the test picture-pair conditions used have been provided under General Methods. In keeping with Tulving's (1981) original study, participants were allocated 6000 ms to decide on their recognition response.

Test sequence.

The following on screen instructions were given to participants prior to them starting the test sequence (Figure 22):

> "You will see two images together. You will have 6 seconds to decide which picture you have seen before in the study phase. The buttons are labelled left and right, Press the button under the image you think is old. Once you press the button you cannot change your mind and the screen will advance to the vividness rating. A timer will count down the last 3 seconds. Try to remember the image selected for the following ratings. Press Continue when you are ready to start the testing phase."

Behavioural Responses.

Behavioural responses were collected in the following order, once the recognition decision had been made:

Vividness.

Participants rated vividness of the photograph selected as old (when it was no longer visible on the screen) on a 0 –100 sliding scale, being instructed: “Move the slider to rate the picture that you selected as old, for vividness, by answering the question “How vividly did you remember the picture?” There is no time limit. Press continue when you are happy with the slider position.” The slider starting position was set at 0; halfway through the experiment the vividness slider starting position was changed to 50 to mitigate anchoring effects (Epley & Gilovich, 2006; Jacowitz & Kahneman, 1995).

Confidence.

Participants rated confidence in their old or new decision on a 0 –100 sliding scale, being instructed: “Move the slider to rate your confidence in your old or new decision by answering the question “How confident are you that you remembered the old picture correctly?” Note if you were guessing, your confidence rating should be zero. There is no time limit. Press continue when you are happy with the slider position.”

The slider starting position was set at 0 for the duration of the experiment.

Remember yes or no.

Instructions for the remember response were broadly in line with those used by Rajaram, (1993). Participants were given the following onscreen instructions regarding the Remember question:

> “You need to interrogate your memory to answer the Remember question. Choose Remember "Yes", **Only** if: you have an experience of recollection for the picture you selected; you are consciously aware of some aspect of what was experienced at the

time the picture was presented in the study phase; you have a sense of yourself in the past and/or the picture brings back to mind a particular association, image, or thought you experienced at the time the picture was presented in the study phase. If there was another reason for your decision, just answer No. Think about the instructions before you answer. Once you have pressed the button Remember yes or Remember no, you cannot change your mind and the screen will advance to the next test pair."

Post study questionnaire.

As a quality check on how participants had interpreted the instructions, before leaving the experiment they were asked to complete a short online checkbox questionnaire based on how well they complied with the Remember yes instructions. "When you answered Yes to remembering the picture, on what did you base your judgment (check all that apply)"

1) I thought of a personal memory or association when I first saw the picture.
2) The picture reminded me of something.
3) I just thought the picture was attractive - I liked it!
4) I remembered that something happened in the room when I first saw the picture (I felt cold, or I coughed or something else interrupted me).
5) I remembered what I was thinking when I first saw the picture.

Participants were also asked to fill out a short checkbox questionnaire based on the reasons that may have prompted them to respond remember no: "When you answered No to remembering the picture, on what did you base your judgment (check all that apply)"

1) I just guessed.
2) The picture seemed familiar.
3) I just knew I had seen it before.
4) I did not recognise the other picture.

Participants were given the option of stating an alternative reason, however, only two participants used the option for remember yes, and no participant used this option for remember no.

Results.

Participants.

Experiment 3 included 70 participants of mean age 21.3 years (S.D. 5.4 years); the majority (56) identified as female. Each participant completed 36 trials. To ensure data quality (i.e., that all participants were attempting the task) response accuracy was analysed by participant. The mean response accuracy (proportion of correct responses) *M* = .59, (*S.E.* = .01)]. An outlier, with a response accuracy of .28 was removed from subsequent data analysis. To be clear, of the 36 trials carried out by the participant, 17 were timed out and the participant had answered remember yes to all 36 picture pairs. A further five participants who had also answered remember yes to all picture pairs were excluded from the analysis as it was considered they had either not understood or had not carried out the instructions. This data exclusion has a precedent, as instructions are crucial in determining how participants rate their responses, (Geraci et al., 2009). A further 110 trials (4.8%) across all participants were timed out and therefore omitted from the analysis (if timed out, participants could not make an old or new decision). The final data set for analysis comprised 64 participants, completing 2194 trials.

Performance.

Just over half of the images (53.7%) selected as old had been remembered, Table 6. Overall performance (proportion correct) was 0.62. Recognition was significantly more accurate if the stimulus was remembered (.66), than not (.57), [B = 0.092, $p < .001$], the effect size was trivial (R^2 change = 0.9%).

Table 6. Vividness, Experiment 3: performance (proportion correct) by remember response

Vividness, Experiment 3: performance by remember response		
Remember	Proportion correct	Total
no	.57	1015
yes	.66	1179
Total	(.62) 1360	2194

Variables of Interest.

Confidence.

Confidence was significantly higher *M* (*S.E.*) 62.0 (1.03) if the recognised image was remembered, than if not, 28.27 (1.08), ($p < .001$), the effect size was very strong (R^2 change = 32.8%). The relationship between confidence, vividness, recognition and remembering is illustrated in Figure 34. High confidence and high vividness decisions were both associated with correct recognition and remembering. Figure 34 clearly illustrates that the association was much larger for remembering than for correct recognition.

Vividness.

Vividness was significantly higher, 67.0 (0.69) if the recognised image was remembered, than if not, 32.4 (0.72), ($p < .001$), the effect size was very strong (R^2 change = 35.3%), Figure 34.

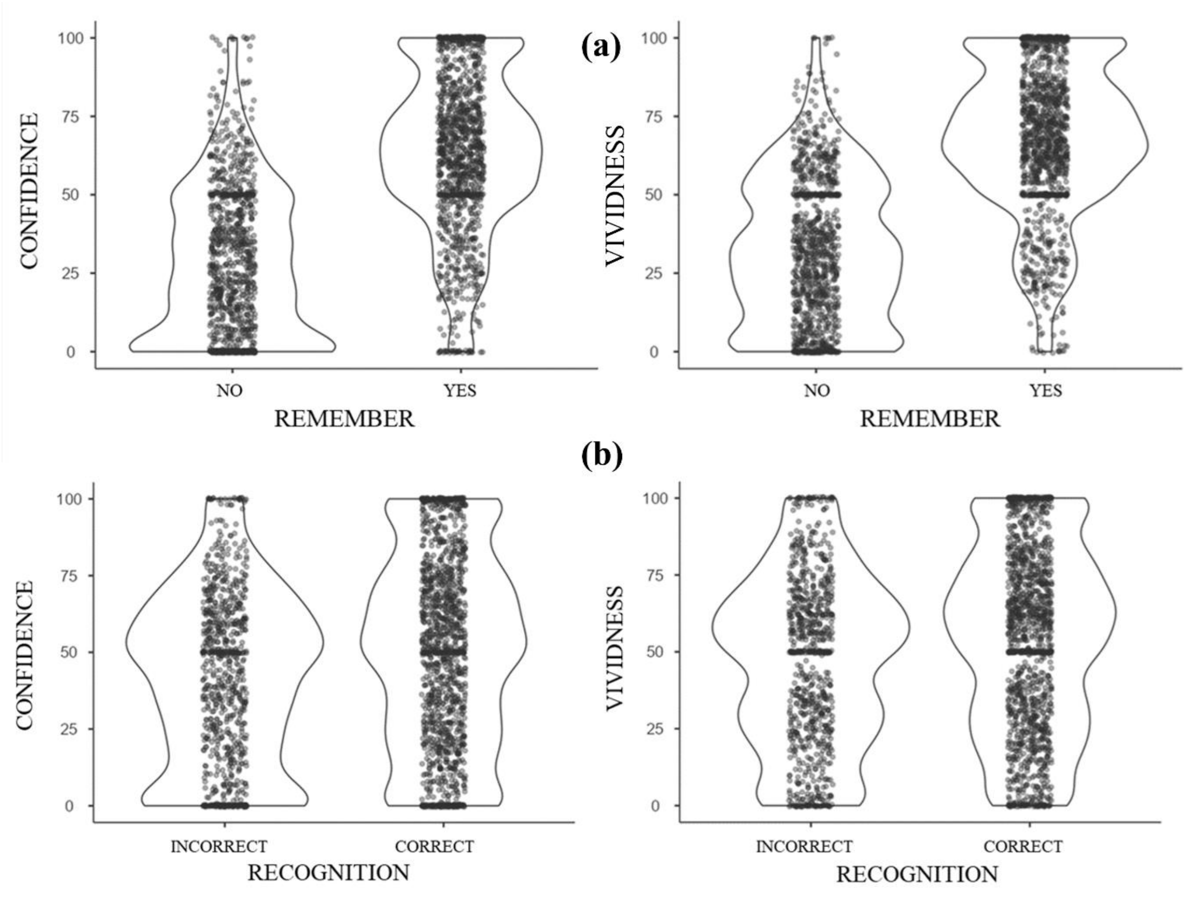

Figure 34. Vividness, Experiment 3: univariate scatter (strip) plots and overlaid violin plots[15] showing the distribution of confidence and vividness judgments for all trials (a) by remember response, and (b) by recognition accuracy (correct vs incorrect). There are more higher score trials when the stimulus was remembered, and more lower score trials when the stimulus was not remembered. Clustering of scores is seen at 100 only when the stimulus was remembered, and at 0 when the stimulus was not remembered

Correlations.

Confidence by vividness.

There was a very large correlation between confidence and vividness (r = .82), 67% of the variance in confidence (in recognition accuracy) can be explained by vividness of the recognised images. Within-participant averaged correlation was similarly very large (r = .79); all correlations were positive (Appendix B, Table 25S & Figure 73S). Anchoring was apparent but did not fundamentally change the overall pattern of the scatterplot, Figure 35.

[15] A violin plot is a hybrid of a box plot and a kernel density plot, which shows peaks in the data and the variable density profile

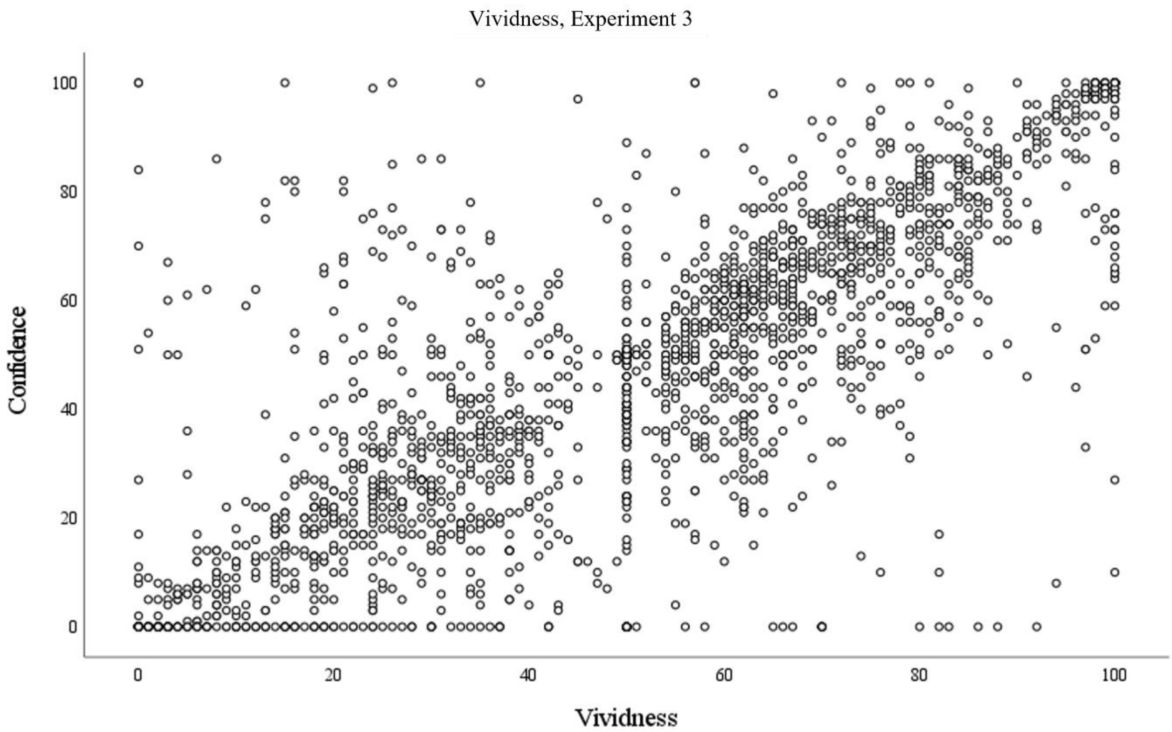

Figure 35. Vividness, Experiment 3: illustrating scatterplot of confidence by vividness across all individual trials (note anchoring at slider starting position for vividness = 50). Participants have been instructed to rate confidence = 0 if they were guessing (sliders started at 0 for both measures; for vividness, the starting position changed to 50 halfway through the trial)

Confidence by vividness, dependent on remembering.

Scatterplots show that trials associated with remembering the image recognised as old were predominantly distributed across the upper right quadrant of higher confidence and higher vividness trials, Figure 36. Conversely, if the image was not remembered trials were predominantly distributed across the lower left quadrant of low confidence, low vividness trials.

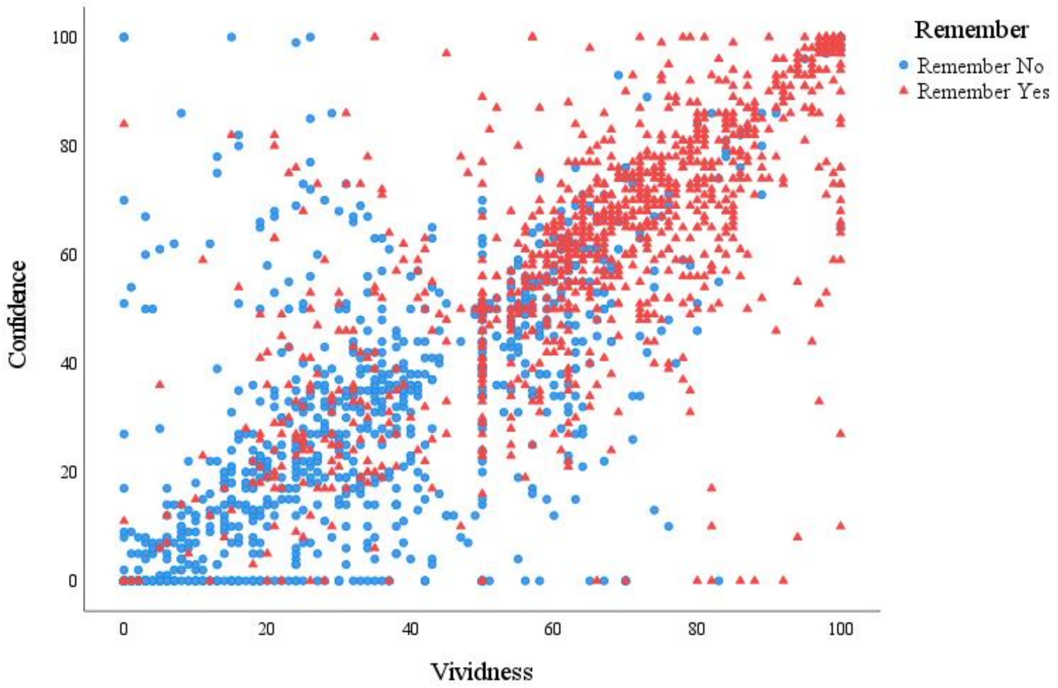

Figure 36. Vividness, Experiment 3: illustrating scatterplot of confidence by vividness across all individual trials by remember response (red triangles denote remember = yes)

Confidence by vividness, dependent on recognition accuracy (performance).

In contrast, whether recognition was correct or incorrect (i.e., high, or low accuracy trials), there was less of an obvious difference in how recognition accuracy (performance) was distributed between high and low confidence and high and low vividness trials, Figure 37.

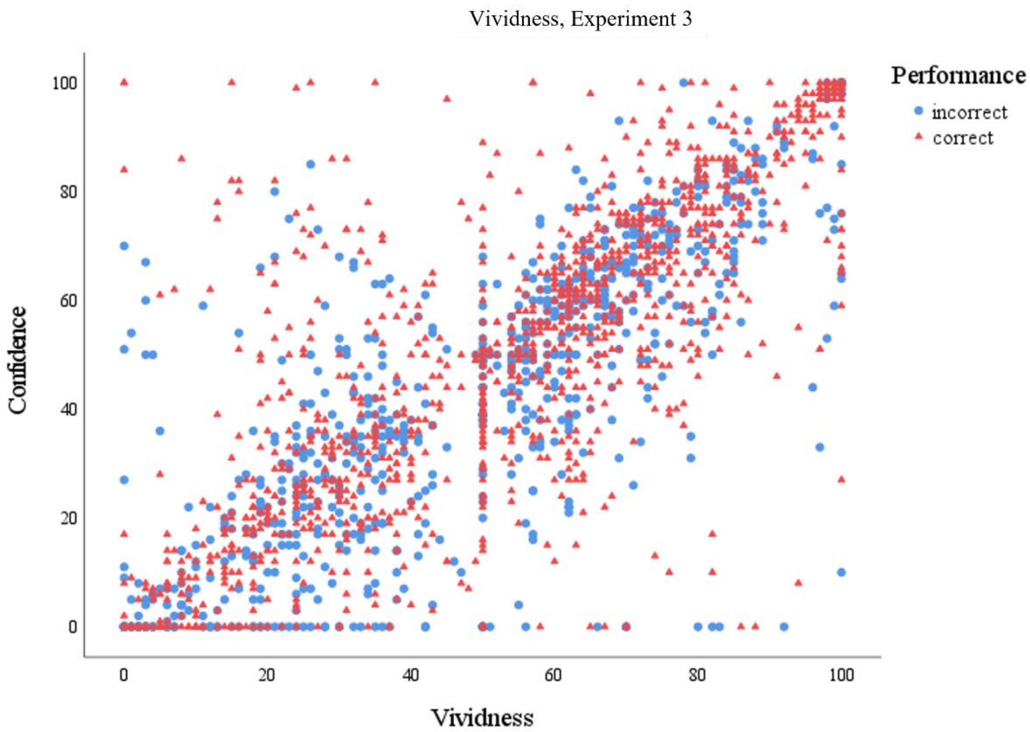

Figure 37. Vividness, Experiment 3: illustrating scatterplot of confidence by vividness across all individual trials by performance in the task (red triangles denote accurate recognition = correct)

Moderation analysis.

A simple moderation analysis was carried out in SPSS using PROCESS to investigate whether the relationship between vividness and confidence was influenced by whether or not the recognised picture was remembered (cf. Figures 35, 36 & 37). The predictor variable [X] for the analysis was the continuous variable of vividness. The outcome variable [Y] for the analysis was the continuous variable of confidence. The moderator variable [W] for the analysis was the dichotomous variable of remembering (yes or no) for the picture recognised as old, with the dichotomous variable of recognition accuracy used as a control variable for the model[16] (see path diagram, Figure 38).

[16] categorical control variables are variables that should be included in the statistical model which however are not of primary interest, i.e., the variable is suspected to be related with the outcome variable but not with the main independent variable of interest

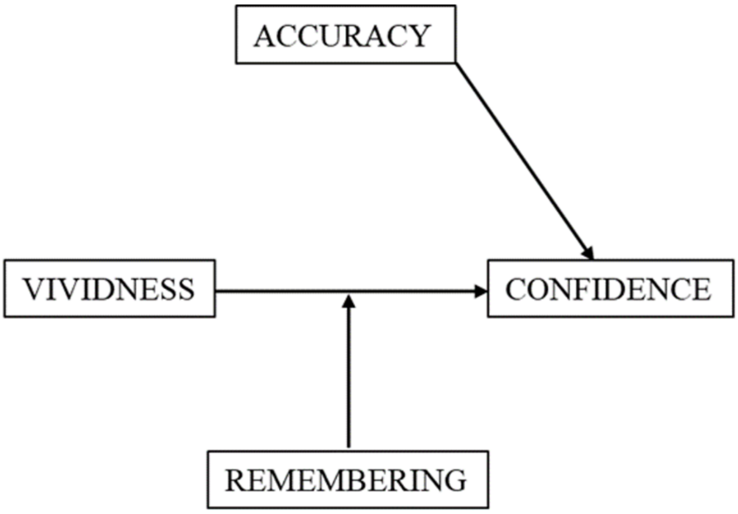

Figure 38. Vividness, Experiment 3: path diagram for the moderation model used to produce the simple slopes plot in Figure 39

Results confirmed a significant overall R^2 effect for the model, i.e., 68.9% of the variance in confidence is predicted by the model containing both vividness and remembering, with recognition accuracy as a control variable. Vividness had a significant effect on confidence [$B = 0.69$, 95% CI (0.65, 0.74); $p = .0001$]. In terms of the whole model, remembering the image did not have a significant additional effect on confidence [$B = 1.91$, 95% CI (-1.49, 5.53); $p = .27$)]. The interaction effect (vividness*remember) was significant ($p < .001$), but the effect size was trivial (R^2 change = 0.2%).

The effect of vividness on confidence (i.e., the slope, B) increased, conditional on whether the image was remembered. At low moderation (remember = no), the [conditional effect of vividness on confidence, $B = 0.7$, 95% *C.I.* (0.65, 0.74), $p < .001$]. At high moderation (remember = yes), the [conditional effect of vividness on confidence, $B = 0.82$, 95% *C.I.* (0.78, 0.86), $p < .001$]. The results identify remembering the image to be a positive moderator of the relationship between vividness and confidence, as illustrated by the simple slopes plot for the moderation effect, Figure 39.

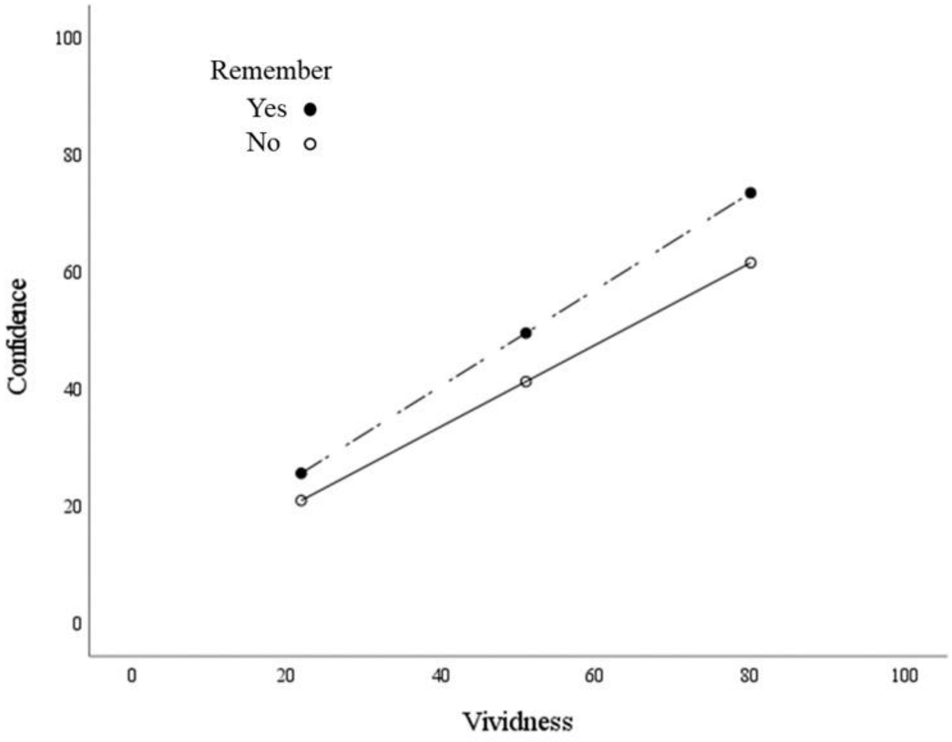

Figure 39. Vividness, Experiment 3: Simple slopes plots (of the moderation effect) for vividness with confidence by remember response

From the output, and illustrated by the plot, it can be seen that the effect of vividness on confidence in the recognition decision is buffered by whether the image was remembered. The positive effect of vividness on confidence becomes stronger when the image is remembered. The strength of the effect is larger with higher levels of vividness and the plots diverge as vividness increases (Figure 39). However, the effect size was trivial; at low levels of vividness there was only a small increase in confidence due to the additional effect of remembering the image.

Test picture-pair condition.

Trials in the mnemonically similar A-B' condition showed higher confidence and lower accuracy compared to the perceptually similar A-A' condition, Table 7.

Table 7. Vividness, Experiment 3: variables of interest (recognition accuracy, confidence, vividness, and reaction time), showing results by test picture pair condition

Vividness, Experiment 3: variables of interest *M (S.E.)* by test picture pair condition				
Test picture pair condition (n)	Proportion correct	Confidence	Vividness	Reaction time, ms
A-A′ (724)	.63 (0.02)	44.6 (1.09)	50.05 (1.07)	3178.3 (48.74)
A-B ′ (731)	.59 (0.02)	48.65 (1.13)	52.34 (1.08)	2974.9 (45.74)
A-X (739)	.65 (0.02)	47.21 (0.65)	50.58 (1.08)	2923.0 (44.28)

Differences in performance (proportion correct) by test picture-pair condition did not reach the level for statistical significance, ANOVA [F (2, 2191) 3.06; $p = .05$, $\eta^2 = .003$], Table 7. Confidence was lowest for perceptually similar A-A' test pairs. The overall difference reached significance, ANOVA [F (2, 2191) 3.36; $p = .04$, $\eta^2 = .003$], but the effect size was small, Table 7. There was no significant difference in vividness, ANOVA, F (2,2191) 1.24, $p = .29$, between the test picture-pair conditions.

Acknowledging the confound that may occur by setting a decision time of 6000 ms, participants took longer when the test picture-pairs were perceptually similar (the A-A' condition) compared to the other conditions. The overall difference reached significance, ANOVA [F (2, 2191) 8.48; $p < .001$, $\eta^2 = .008$], the effect size was small.

The correlation between vividness and confidence remained strong for all test picture-pair conditions, Figure 40.

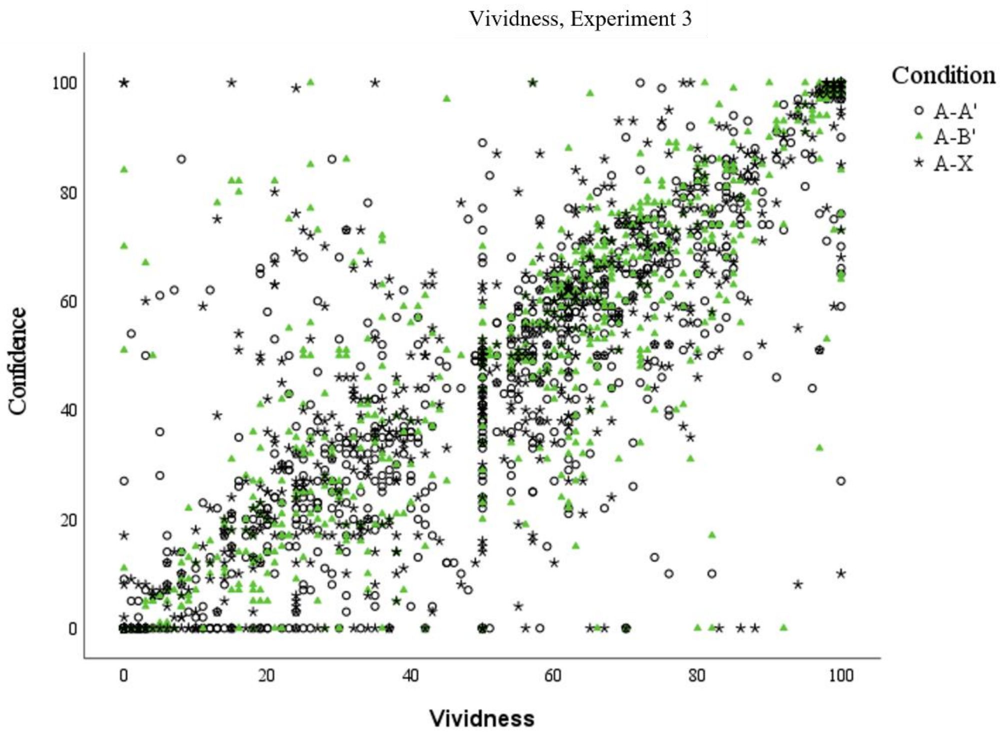

Figure 40. Vividness, Experiment 3: illustrating scatterplot of confidence by vividness across all individual trials by test picture-pair condition

Confidence Accuracy Calibration (CAC) Analysis.

Bivariate scatterplots between continuous and dichotomous variables cannot be carried out to illustrate the relationship between confidence and accuracy, and averaging performance and confidence data for each participant is problematic, as previously discussed. Confidence-accuracy characteristic (CAC) analysis was therefore performed. Analyses were carried out using a recently reported confidence-accuracy characteristic analysis devised for an old or new recognition memory task (Tekin et al., 2021). Admittedly, the report was for a standard old or new recognition task, rather than a 2-AFC task, however, results here were suitable for analysis based on performance: proportion correct = hits ÷ (hits + false alarms) to create the plot, with confidence binned into four levels (0–30, 31–50, 51–75, and 76–100). The cut points were chosen to allow for equivalent numbers (approximately 400 – 600 trials) in each bin. The plots allow comparison of the effect of remembering on performance for each level of binned confidence. Accuracy increased only at the highest level of confidence

(> 75), whether or not the recognised stimulus was also remembered (Figure 41). Notably, there was less effect of remembering on performance at lower levels of confidence.

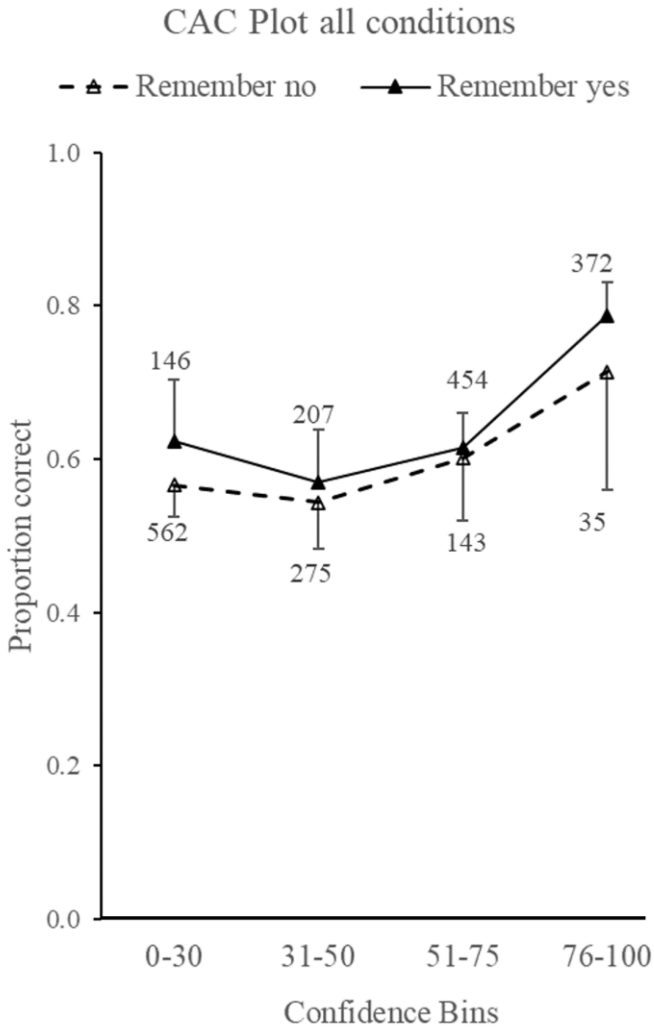

Figure 41. Vividness, Experiment 3: CAC plot of combined data for all test picture-pair conditions: with four bins of confidence illustrated, showing mean proportion correct by remember judgment (error bars represent 95% C.I. – only positive or negative bars are shown for clarity). The numbers represent trials assessed at each point

It could be pointed out that collapsing the lowest two or three bins of confidence would give a more impressive graph, if the aim were to associate confidence with accuracy (and noting that Figure 14 illustrates confidence binned from 0 – 60, 70 – 80, and 80 – 100). An additional point is how the ratio of trials changes between remembering yes or no as confidence increases, 10 times as many trials are remembered than not remembered when confidence is rated as > 75, confirming the well-known finding that remembering is associated with higher confidence than a ‘know’ response.

Analysis by test picture-pair condition was carried out to see if the pattern changed across similar, deceptive, or different lures, Figure 42. The results explain those in Table 7, showing increased accuracy with remembering in all but the A-B' test picture pair condition. Throughout the analyses, and as shown in Figure 41, it is clear there are more trials in low confidence bins when remembering is not involved in selection of the old item, most of the trials in the highest confidence bin (scoring > 75) are associated with remembering (as also illustrated in the scatterplot, Figure 36). Notably, it is only at this point that accuracy increases with confidence.

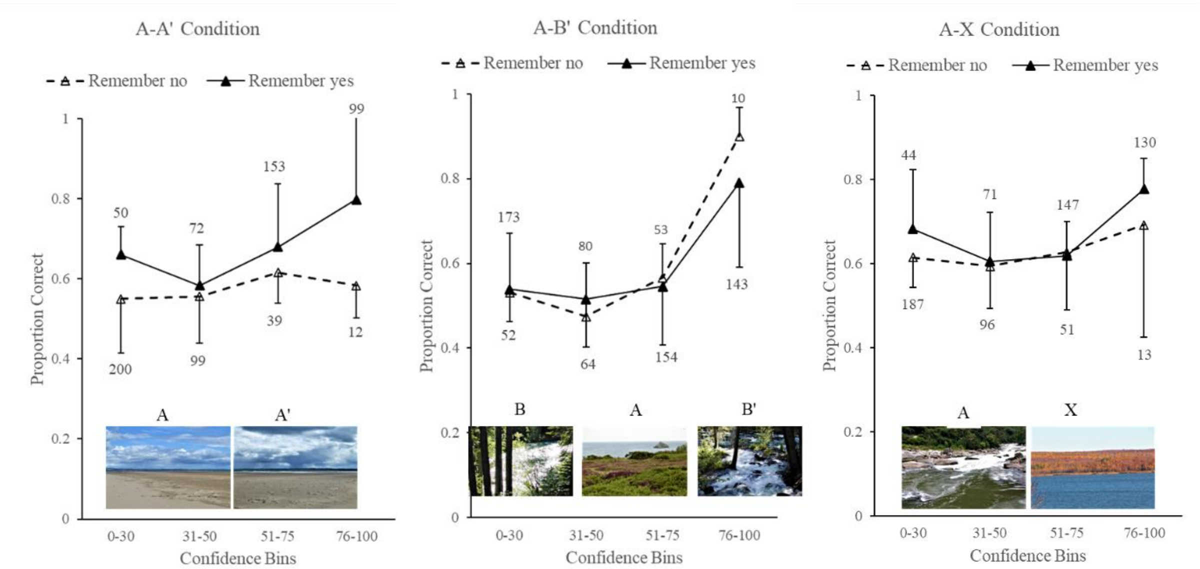

Figure 42. Vividness, Experiment 3: CAC plots by test picture-pair condition: showing mean proportion correct by remember judgment (error bars represent 95% C.I. – only positive or negative bars are shown for clarity. Example picture-pairs are illustrated for each condition; the B picture was shown during the target block in the study phase of the experiment, at test a similar B' image was shown as part of the test pair. Neither the A' nor the X images had been shown previously at study. The numbers represent trials assessed at each point

A-A′ condition.

To be clear, all of the A target pictures had been shown during the study phase. Remembering the target appeared to have the largest effect on performance in the A-A′ condition. When the two pictures were similar (Figure 42), accuracy increased with confidence ratings > 50 if the picture was remembered. If the picture was not remembered the plot was flat, with no change in accuracy with increasing confidence. The error bars separated only at the highest level of confidence (76 – 100).

A-B' condition.

When the paired B' picture was similar to (but not the same as) a B picture shown during the study phase, (Figure 42), in both remember conditions performance only improved with confidence ratings at the highest level (76 – 100); importantly, accuracy was higher when the recognised picture was not remembered. False remembering of a deceptive (mnemonically similar) lure can lead to an illusion of memory accuracy, i.e., participants were more confident but less accurate consistent with the confidence-accuracy dissociation found in Tulving’s (1981) picture similarity experiment (cf. Table 7). Of note, and as seen in the overall picture (Figure 41), there is a large imbalance in the ratio of remember yes and remember no judgments at the highest binned level of confidence. For this reason I have not presented a statistical *p* value and accept that the 95% *C.I.* bars overlap. However, I offer this as an interesting observation to explain the confidence and accuracy dissociation first demonstrated by Tulving.

A-X condition.

When the paired X picture was new and dissimilar to any pictures shown during the study phase, (Figure 42), remembering had less effect on performance. If the picture was also remembered there was a small performance advantage only at the highest level of confidence (76 – 100). If the picture was not remembered there was little change in performance with confidence.

Qualitative data: Post study questionnaire.

Participants were required to fill in a short post-experiment questionnaire to collect qualitative data on what metrics they had used from the instructions to respond remember yes, and also to find out on what basis they had answered remember no (other than none of the criteria applied).

Remember yes.

The results are expressed as a percentage of participants checking each answer (bearing in mind that they were instructed to check all responses that applied). The commonest reason cited for remembering the stimulus chosen (69%) was that the picture reminded the participant of something (i.e., triggered an idiosyncratic personal memory). The results clearly demonstrate that the majority of the responses were consistent with conscious awareness relating to a personal memory, with 42% having an emotional response, i.e., they liked the picture or thought it was attractive. Only 9% of participants used the fact that they remembered something happening in the room at the time of viewing the picture (Figure 43).

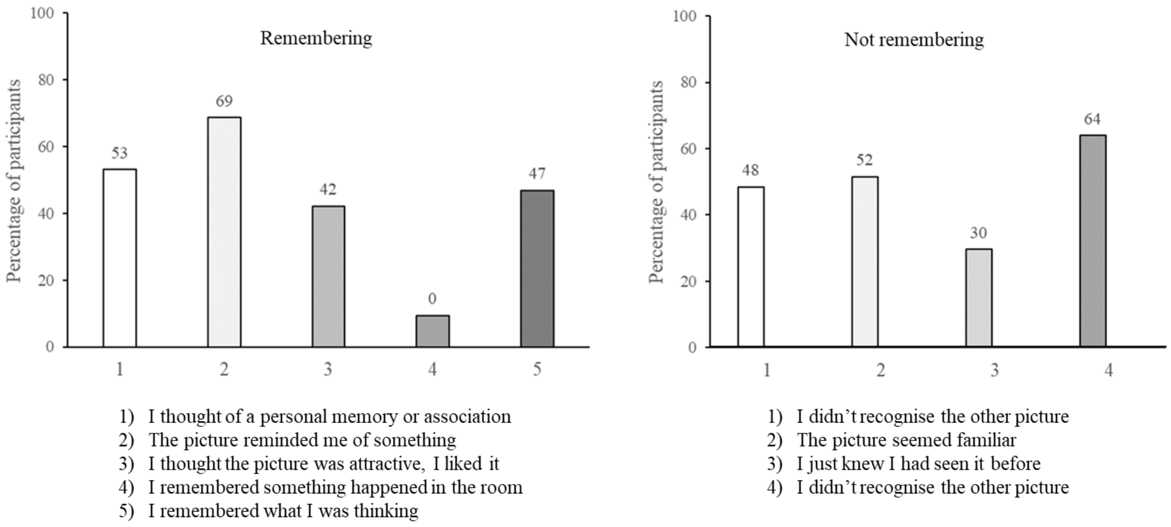

Figure 43. Vividness, Experiment 3: Qualitative data for remember = yes and remember = no analysed as a percentage of participants who responded to each question at the end of the experiment

Remember no.

The commonest reason cited for recognising but not remembering the stimulus (64%) was the fact that participants did not recognise the other picture (the picture not selected). There was little difference between guessing (48%) or thinking the picture was familiar (52%). However, the lowest percentage of responses were based on a 'know' response (a feeling of just knowing the picture was old), Figure 43. Gardiner, (2001), found that know responses outweighed guesses, particularly at longer response deadlines (to exclude a speed versus accuracy trade-off decision). Here, all the recognition decisions were made at a deadline of 6000ms, however participants' responses to the behavioural questions were all self-paced.

Summary.

Experiment 3 confirmed the very large correlation between vividness and confidence in recognition memory performance. This held true for both methods of correlational analysis. High vividness and high confidence were both associated with remembering the recognised image. Recognition accuracy (correct or incorrect trials) was not clearly associated with any level of vividness or confidence. The association between vividness and confidence was moderated by remembering; confidence increasing for the same level of vividness if the image was remembered.

The confidence and accuracy dissociation demonstrated by Tulving (1981) was replicated but although differences in confidence were significant, the effect size was small, and differences in performance (recognition accuracy) between the test picture-pair conditions did not reach statistical significance. Confidence accuracy characteristic (CAC) analysis illustrated for the mnemonically similar B-B′ condition, performance was impaired when the picture was remembered, but only for the highest bin of confidence (> 75). In the other test picture-pair conditions, higher performance (proportion correct) was associated with remembering the image, but again this was true only for the highest bin of confidence (> 75).

Discussion of results

Importantly, Experiment 3 replicated the findings of the source memory task, that there was a very large correlation between vividness and confidence; confirming the correlation held true for recognition memory as well as for recollection of source memory. Admittedly, both tasks used the same real-world natural scenes photographs. This was true for all three test picture-pair conditions. In asking why a remember response is associated with higher confidence. The results confirm that the vividness of the recognised image is strongly related to confident performance in the task, and that both vividness and confidence were strongly linked to remembering rather than the accuracy of the picture recognised. Trials with high confidence and high vividness were associated with remembering the recognised picture; there was no such association between confidence, vividness, and trials with correct target recognition, Figures 36 & 37.

The finding that remembering moderates the association between vividness and confidence can be taken to correspond to results from neuro imaging studies, where the angular gyrus is deemed responsible for assessing the conscious experience of vividness, but not confidence, of memory retrieval (e.g., Zou & Kwok, 2022). The pattern of results show that vividness improves the experience of remembering, and this channels the effect of vividness on confidence.

This reproduction of Tulving's (1981) picture similarity experiment did not show any statistically significant differences between the test picture-pair conditions in accuracy or vividness, the difference in confidence just reached significance, but the effect size was small. There was however a significant difference in reaction (response) time, Table 7. Participants took longer to make their response when the test picture-pairs were perceptually similar (the A-A' condition) compared to the other conditions. Fandakova et al., (2021) speculate that decision time is longer for perceptually similar test picture-pairs, due to a prolonged process of matching features in the image with memory, with accuracy being higher.

Tulving (1981), considered there were two kinds of similarity relations for mnemonically similar lures (the A-B′ condition), and perceptually similar lures (the A-A′

condition). If the cue and the target were similar, participants were more accurate but less confident; explaining the reasons for lower confidence for perceptually similar pairs (due to prolonged matching of features within the pictures) confirmed here by the significant increase in reaction time to reach a decision in this picture condition. Others claim that participants in the A-A' condition base their judgment on identification of the most diagnostic feature rather than on subjective recollection of the A picture (Dobbins, Kroll & Liu, 1998; Hembacher & Ghetti, 2017). Despite the accepted conclusion that this, in the A-A' condition, leads to higher accuracy and low confidence, here, Figure 42 shows that remembering the image is more common than not remembering in the A-A' condition, for decisions made with higher confidence.

In the A-B' condition, when the presentation of the two dissimilar pictures at test discouraged identification of the most diagnostic features, participants were then theorised to rely on remembering to make that judgment (Dobbins et al., 1998). However, as illustrated in the CAC plots (Figure 42), confident reliance on remembering the image may be associated with inaccurate recognition. Higher confidence decisions are more commonly associated with remembering the image and this does not result in improved performance, rather, when an image is not remembered and recognition is made with high confidence, the proportion of correct responses is higher. As Johns et al., (2021) comment: “In studies of false recognition, subjects not only endorse items that they have never seen, but they also make subjective judgments that they remember consciously experiencing them” (p.1). Given the results of Experiment 3, false recognition, a subjective judgment of consciously experiencing pictures the participants have never seen before, or an inaccurate memory is quite simply caused by confidently but mistakenly remembering an image that is similar, but not the same as, another picture shown at study. Such overlap between features of stored information may explain our common experience of misremembering personal events, also classed as “an error in reality monitoring" (cf. Gonsalves & Paller, 2002), or autobiographical memory errors, such as misremembering details of events leading to mistaken belief in experiencing events that did not occur (Newman & Lindsay, 2009).

Regarding a different facet of illusions of memory accuracy, Doss et al., (2018) found that context reinstatement both, as expected, increased correct recognitions, but was also found to increase incorrect recognition of similar objects, under explanation that conceptual information and perceptual information became confused to create this memory illusion. In their experiment, images of everyday objects were shown superimposed on a scene. Participants associated the item and the context so that they could recall the item when the context was re-presented in a 2-AFC choice. Perceptually similar lures, shown superimposed on the same context as the target object, were frequently and confidently recognised as old. The findings are similar to those in Experiment 3 for mnemonically similar pictures, however, in Experiment 3 this occurred in the absence of context confusion.

Next, I use the same source memory task and stimuli as for Experiment 1 with a memory quality judgment of distinctiveness.

Chapter 5: Distinctiveness.

We may think we know what is distinctive, for example, returning to my metaphor of a single red poppy in a cornfield, it will stand out and attract attention, it may be judged as vivid, but it will also be distinctive. But will it still be distinctive in a field of red poppies (where it still remains vivid) or even in a field of different flowers? As such, I considered it worthwhile to explore, as for vividness, whether the distinctiveness of our memory of a stimulus or event is similarly associated with confidence, and whether or not it will be associated with the accuracy of the memory's detail (source memory). When a memory is distinctive, are the associated details accurately remembered, or are we merely confident that we have remembered the details accurately? This is tested using the previously described continuous-response source memory paradigm. In the same way it is suggested that memory quality, represented by distinctiveness will also influence confidence in the memory's accuracy.

Rather than manipulating the visual vividness of the natural scenes colour pictures, one aspect of distinctiveness that requires attention is that of context, i.e., the list arrangement in which the photographs are displayed. Changing the context in which the photographs are displayed is theorised to affect their distinctiveness, whether or not this changes confidence in source memory accuracy, or the accuracy of remembering the detail of the memory can also be assessed using this paradigm. The criteria for judging distinctiveness will also be gathered following experiments and analysed by quantitative methods, according to the list arrangement used.

It is theorised that distinctiveness may increase the possibility of retrieval associated with conscious recollection of an item, i.e., remembering rather than knowing, or recollection rather than familiarity (Brunel et al., 2010; Rajaram, 1998). I thought it not possible to use the

2-AFC task from Experiment 3 for distinctiveness, as the similarity of pictures might bias judgments of distinctiveness. Admittedly, however, I could have used a different 2-AFC task.

(i) Distinctiveness, Experiments 4 & 5.

Rationale

The isolation effect is commonly used in tests of distinctiveness. However, it merely demonstrates that in a standard memory task, a pre-determined distinctive item (by experimental design or experimenter choice) is better recognised or remembered at test than other non-distinctive items, also pre-determined. So far we learn what the man in the street already knows, that a different item or event in the context of many similar items or events will attract our attention. Of course, a purple car with lemon-yellow polka dots (Hunt, 2013) is distinctive, it will capture our attention and thus it is said to be more richly encoded. We will be highly confident that if we saw the car again we would recognise having seen it before, however, can we remember the contextual detail, "the more exact knowledge of the event", (Mandler, 1980), i.e., what make the car was, where it was when we saw it, and when exactly did we see it.

Over two experiments, distinctiveness of an associated natural scenes colour photograph was judged when the image was re-shown to cue recollection of source memory for the location of a cross on a circle (as for vividness). Highlighting that none of the images shown were pre-determined as isolates or control items, the relationships between distinctiveness judgments, confidence, and source memory accuracy were assessed in two experiments using the same 120-image set of natural scenes photographs in full colour, split by the context (list arrangement) in which the pictures were presented. As shown in the literature, colour and distinctive features are significant factors which make an image memorable (e.g., Bylinskii et al., 2015; Rust & Mehrpour, 2020). It is accepted that some images are simply more memorable than others (Goetschalckx et al., 2019), therefore in order to avoid the possibility of this affecting results, the identical 120 natural scenes photographs were used across both experiments.

Two types of list arrangement (*list-types*) were employed: a mixed *list-type* where the 12 photographs in each relational encoding block at study and at test were made up of three photographs from each of the four different categories of natural scenes subjects (using the same list arrangement as used for testing vividness in Experiment 1), and a homogenous *list-type* where all 12 photographs in each relational encoding block at study and at test were different but from the same subject category. Neither *list-type* equates with an isolation list; although one might imagine that pictures of flowers and bright colours would stand out in a mixed *list-type* and be assumed distinctive, there is not just one picture of a flower, but either three or 12 pictures of flowers in each relational-encoding block. The order in which the pictures were shown was randomised by E-Prime, both at study and at test.

My choice of a homogenous list arrangement of photographs (all from the same category) was to determine if any are rated as distinctive, but also to determine if detail for the more similar items is less well recalled, with memory accuracy impaired by this *list-type*. Evans and Baddeley (2018) demonstrated that when natural scenes were selected from a single category (and with distinctive detail minimised), memory primarily depended on the participant's intention to remember. In my experiments, participants are explicitly instructed to pay attention both to the photographs and to the cross location, the instructions hopefully prompting their intention to remember, regardless of the *list-type* used in the experiment.

Importantly, participants were not prior informed as to the nature of the *list-type* to be used in their experimental task, nor even whether each block would have the same list arrangement. When distinctiveness is judged at test, participants will have seen all of the stimuli in the list at study and will understand the task that was presented. Regarding the question of exactly when distinctiveness is processed, it is acknowledged that here, distinctiveness is judged when the images are re-presented to cue recollection of the cross location. However, Brunel et al., (2010) considered that distinctiveness processing occurs at retrieval due to a global difference between memory traces (the difference between items over a wide range of categories).

To test the importance of item context, results for the two experiments are compared to answer the question whether the *list-type* (context) used to display identical images had any effect on their distinctiveness rating, on confidence, or on source memory accuracy for the recollected cross location. As earlier discussed, it is assumed that the natural scenes image and the location of the cross on the circle are encoded as one memory trace.

I hypothesized that if all the images in each relational encoding and test block are from the same natural scenes subject category (homogenous *list-type),* distinctiveness would be lower, as the images do not stand out from their class (category), assuming we pay less attention to the routine. However, the criteria participants used for judging distinctiveness of the images may shed more light on this.

Experiment 4.

Experiment 4 used a 120-image set of natural scenes colour photographs, displayed in a mixed *list-type*.

Methods

Procedures were followed as laid out in General Methods (i), illustrated in Figure 17, and as described for vividness, Experiment 1.

Stimuli.

Images were displayed in a mixed *list-type*, at study and at test, identical to the arrangement used for Vividness, Experiment 1. Experiment 4 consisted of 10 cross location-natural scene relational encoding blocks and 10 cross location-natural scene recollection test blocks, each block consisting of 12 trials, counterbalanced for subject category, comprising three different images of flowers, green landscapes, mountains, and water, Figure 44a.

Judgment of distinctiveness.

Participants made their judgment during the test block presentation immediately after the picture was re-presented to cue recollection of the cross location, and not while still on the screen, by answering "How distinctively do you remember the picture?" Response options ranged from "not at all distinctively" to "extremely distinctively," on a 0 –100 horizontal continuous scale. Participants indicated the cross location after this judgment (Figure 17).

Figure 44.(a) Distinctiveness, Experiment 4 (mixed list-type), 12 natural scenes colour photographs used for both study and test blocks (with 12 different pictures: 3 from each subject category (mountains, green landscapes and trees, flowers, or water); (b) Distinctiveness, Experiment 5 (homogenous list-type), 12 natural scenes colour photographs used for both study and test blocks (with all 12 pictures representing the same subject category; in this example flowers)

Experiment 5.

Experiment 5 used a 120-image set of natural scenes colour photographs, displayed in a homogenous *list-type*.

Methods

Procedures were followed as for Experiment 4, see also General Methods (i), Figure 17, and vividness, Experiment 1.

Stimuli.

Images were displayed in a homogenous *list-type*, at study and at test. To be clear the same 120-image set was used as for Experiment 4. Experiment 5 consisted of 10 cross location-natural scene relational encoding blocks and 10 cross location-natural scene recollection test blocks. Each block consisted of 12 trials including 12 different images, each from the same subject category (all were either flowers, green landscapes and trees, mountains, or water). In a presentation block where all the images are of flowers, each is different from the others in form and pseudo-randomised for colour, with no one picture pre-determined as an isolate (i.e., a red flower when all the others are white), Figure 44b.

Results

Results were compared between Experiments 4 and 5, i.e., by *list-type*.

Participants.

Experiment 4 included 20 participants of mean age 20.8 years (*S.D.* 3.2 years); 16 participants identified as female. Data from all 20 participants were available for analysis.

Experiment 5 included 20 participants of mean age 20.6 years (*S.D.* 3.4 years); 15 participants identified as female. Two participants were excluded from the data analysis, their location error frequency plots indicated that they could not or did not perform the task and were guessing throughout (e.g., see Figure 18b).

Memory performance.

Memory performance was superior for images shown in a mixed *list-type*. The recollection rate was significantly higher, and the scale parameter significantly lower, (i.e., precision was significantly higher) in Experiment 4, with pictures in a mixed *list-type*. The difference in threshold values did not reach significance, Table 8.

Table 8. Distinctiveness, Experiments 4 and 5: Memory Performance Ratings, [Recollection Rate, Scale (Precision) and Threshold (Degrees from Target when Recollection Begins), [M (SE)], Results of Independent t-tests to Compare Mean Values for all Assessable Participants by List-Type (N = 38)

Distinctiveness, Experiment 4: mixed *list-type*
Distinctiveness, Experiment 5: homogenous *list-type*

List-type	Memory Performance Rating	*n*	*M* (*SE*)	Independent samples *t*-test *t(d.f.)*	*p*-value 2-tailed
Mixed	Rate	20	0.63 (0.05)		
Homogenous		18	0.35 (0.03)	- 4.5 (36)	< 001
Mixed	Scale	20	9.36 (1.01)		
Homogenous		18	14.42 (0.82)	-3.83 (36)	< .001
Mixed	Threshold	20	29.5° (2.8°)		
Homogenous		18	21.9° (3.3°)	1.76 (36)	.087

How memory performance changes, dependent on the context in which the same images are viewed, is further illustrated by the location error frequency plot in Figure 45. When pictures were shown in a homogenous *list-type*, recollection and precision were lower, resulting in a shorter and wider location error distribution, with a higher guessing rate suggesting reduced recognition of the pictures at test. Figure 45 also illustrates the difference in scale parameters (identified by the double headed arrows on the graph) and threshold values (broken vertical lines).

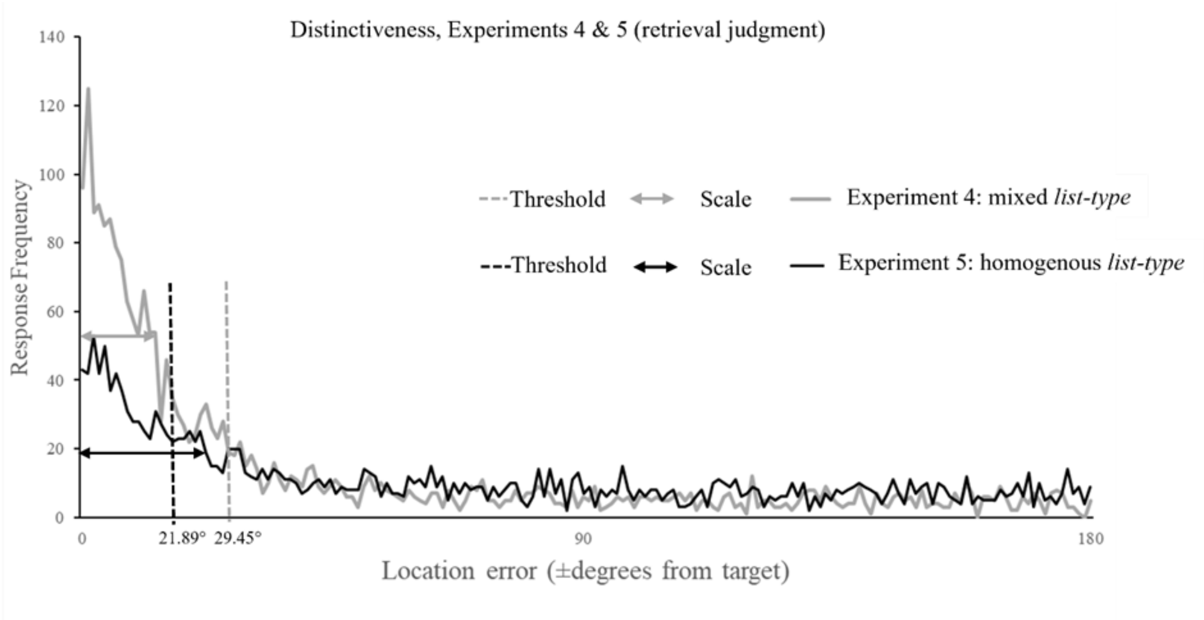

Figure 45. *Distinctiveness, Experiment 4 (mixed list-type) & Experiment 5 (homogenous list-type). Location error cumulative frequency plots. The scale parameter (indicating precision) is depicted for each list-type by the double headed arrows; threshold values are depicted for each list-type by broken vertical lines. The difference in recollection rate can be seen by the difference in height of the cumulative frequency plots close to the target location (0°). A higher guessing rate can also be seen for Experiment 5 (0.65) compared to Experiment 4 (0.37)*

Variables of interest.

Results, averaged across all trials, confirmed that distinctiveness, confidence, and source memory accuracy were all higher for images presented in a mixed *list-type*, compared to the identical images shown in a homogenous *list-type* (Table 9). Effect sizes were weak to moderate.

Table 9. Distinctiveness, Experiments 4 & 5: Variables of Interest, [M (SE)], Mean Values by List-type, (N = 4560) compared by R^2 change

Distinctiveness, Experiment 4: mixed *list-type*
Distinctiveness, Experiment 5: homogenous *list-type*

List-Type	Variable of Interest	*df*	*M (SE)*	Unstandardized *ß* coefficient (*SE*)	*p*-value[a]	R^2 change
Mixed	Distinctiveness	1,4558	59.3 (0.6)			
Homogenous			50.3 (0.6)	-8.99 (0.82)	< .001	2.6%
Mixed	Accuracy	1,4558	76.6 (0.6)			
Homogenous			63.9 (0.7)	-12.74 (0.86)	< .001	4.6%
Mixed	Confidence	1,4558	52.7 (0.6)			
Homogenous			40.3 (0.6)	-12.41 (0.80)	< .001	5.0%

Note. [a]the *p*-value may seem significant, but the number of trials is high. Of more relevance is the effect size, denoted by R^2 change: (1% = weak, 9% = moderate, 25% = strong effect).

Correlations.

Correlations between distinctiveness and source memory accuracy also differed between the experiments, being higher for images shown in a mixed *list-type*, Table 10. Table 10 also illustrates no appreciable difference between correlations made using all trial data compared with averaged within-participant correlations, in fact, some of the values are identical. For further clarity, Appendix B provides detailed individual correlations and *p*-values (Tables 25S & 26S), further illustrated by scatterplots showing individual linear correlation lines (Figures 74S-79S).

In summary, 14.3% of the variance in source memory accuracy was predicted by variability in distinctiveness of images viewed in a mixed *list-type*, compared to 5.1% when the same images were viewed in a homogenous *list-type*. The large correlation between source memory accuracy and confidence for the mixed *list-type* was only moderate for the same images shown in a homogenous *list-type*; 27% and 10% respectively of the variance in confidence was predicted by source memory accuracy.

Critically, and paralleling results for vividness, there was a very large correlation between distinctiveness and confidence in source memory accuracy (Table 10); this correlation was unaffected by the list arrangement of the photographs, i.e., there was no difference between experiments; 46% – 48% of the variance in confidence was predicted by variability in the distinctiveness of the images, despite the significant differences between the experiments (*list-types*) in distinctiveness, confidence, and accuracy, as illustrated in Table 9.

Table 10. Distinctiveness, Experiments 4 and 5: Correlation Analysis, split by list-type, denoting relationships between Variables of Interest. Pearson's r-values are displayed using data from all trials and compared with r-values and R^2 variances obtained by averaging within-participant correlations. It is evident that there is little difference between the two methods

Distinctiveness, Experiment 4: mixed *list-type*
Distinctiveness, Experiment 5: homogenous *list-type*

List-type	I.V.	D.V.	*N*	Pearson's *r* (*all trials*)	*p*-value 2-tailed	R^2 variance	Pearson's *r* (*averaged*)	R^2 variance
Mixed	Distinctiveness	Accuracy	2400	.38	< .01	14.3 %	.40	16 %
Homogenous	Distinctiveness	Accuracy	2160	.23	< .01	5.3 %	.24	5.8 %
Mixed	Distinctiveness	Confidence	2400	.68	< .01	46.2%	.68	46.2 %
Homogenous	Distinctiveness	Confidence	2160	.69	< .01	47.6 %	.67	44.9 %
Mixed	Accuracy	Confidence	2400	.52	< .01	27.5 %	.52	27.5 %
Homogenous	Accuracy	Confidence	2160	.32	< .01	10.2 %	.32	10.2 %

Note. Cohen (1992) defined small, moderate (i.e., medium), and large *r* values as "about" .10, .30, and .50, respectively. R^2 represents the amount of variance in the dependent variable predicted by variability in the independent variable.

Correlation matrices and variable density plots in Figures 46 and 47 illustrate the very large correlations between confidence and distinctiveness in both experiments. The variable density plot distributions for both distinctiveness and confidence differ between Experiments 4 and 5, with larger numbers of higher ratings of each variable shown for images displayed in a mixed *list-type*; both distributions show larger numbers of lower ratings of each variable shown for images displayed in a homogenous *list-type* (corresponding to the significant differences seen in the variable mean values, see Table 9.

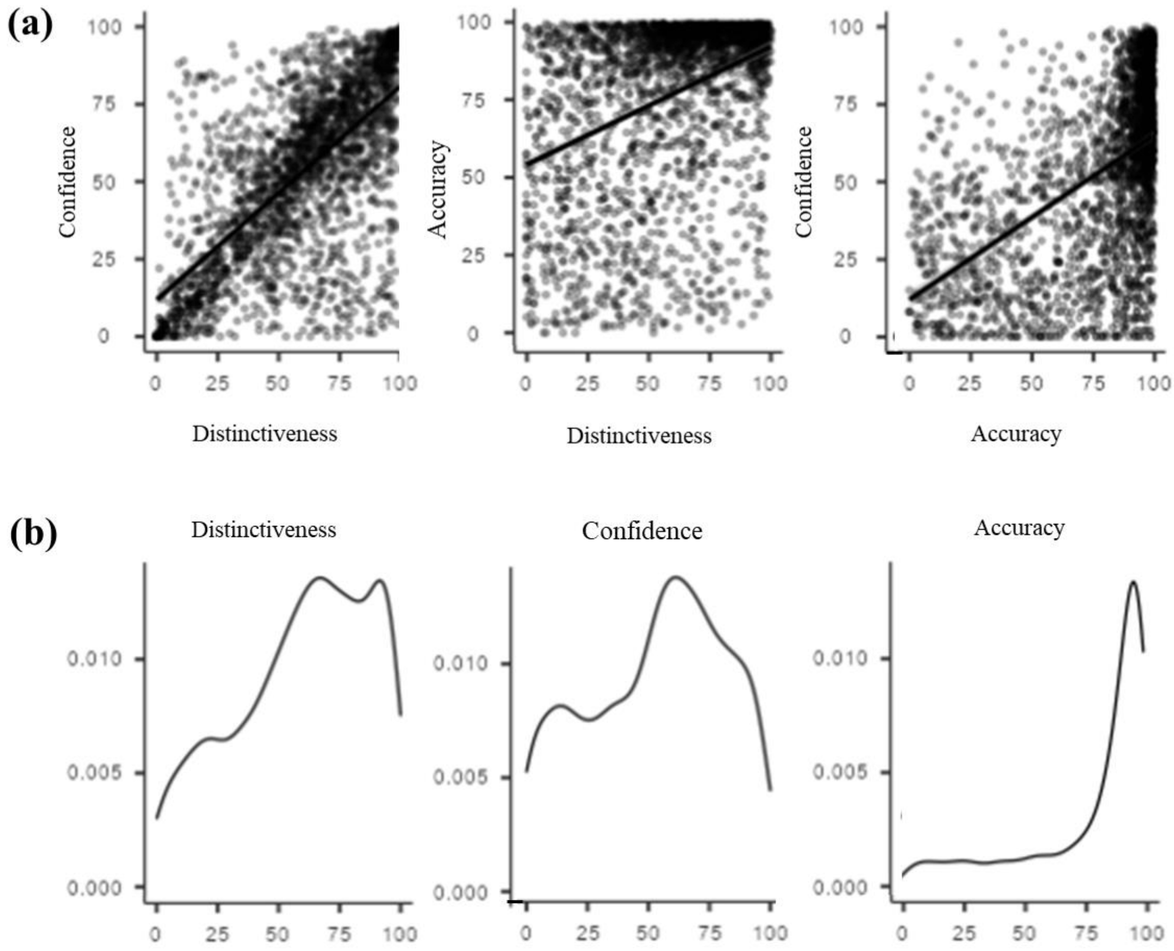

Figure 46. Distinctiveness, Experiment 4 (mixed list-type) showing (a)correlation matrices (scatterplots) and (b) variable density plots (i.e., the smoothed histograms of variable distribution) across all individual trials, where 0 –100 scales on the abscissa and the ordinate represent values for the correlation matrices, and the scale on the ordinate represents the density function for the variable density plots

Reduced correlation between accuracy and confidence in Experiment 5 might possibly relate to the generally lower confidence levels as illustrated in the variable density plot for confidence in Figure 47 associated with images displayed in a homogenous *list-type*.

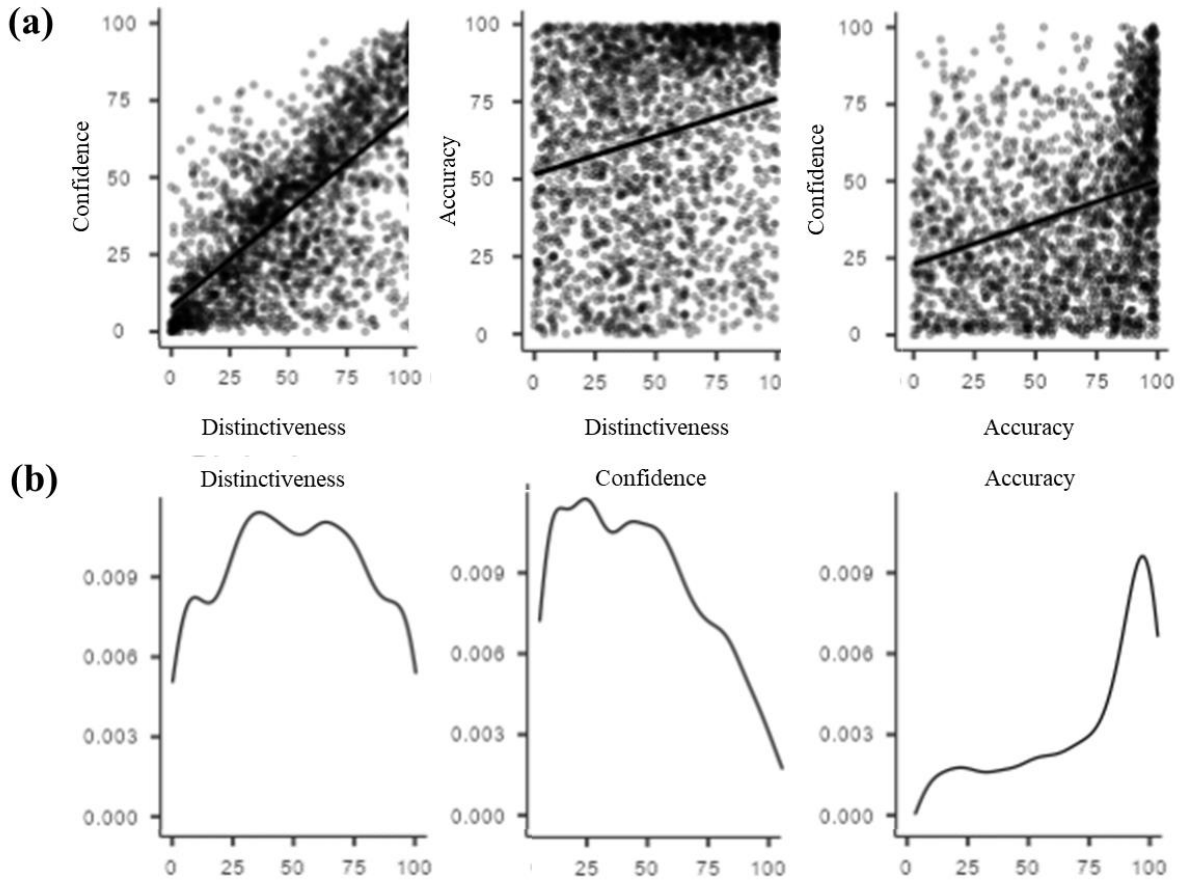

Figure 47. Distinctiveness, Experiment 5 (homogenous list-type) showing (a) correlation matrices (scatterplots) and (b) variable density plots (i.e., the smoothed histograms of variable distribution) across all individual trials, where scales on the abscissa and the ordinate represent the data on a 0 –100 scale for the correlation matrices, and the scale on the ordinate represents the density function for the variable density plots.

Moderation Analysis

Brigham (1990) found less distinctive faces to be associated with a lower confidence-accuracy relationship, suggesting that distinctiveness moderated the relationship between confidence and accuracy. A simple moderation analysis was performed in SPSS using PROCESS to test whether the strength of the relationship between the predictor source memory accuracy and the outcome confidence differs, conditional on the moderator of distinctiveness, see Chapter 3, General Methods for a full description. As the relationship between distinctiveness and confidence did not differ by *list-type*, data from both Experiments 4 and 5 were used for the analysis.

The outcome variable [Y] for the analysis was the continuous variable of confidence, the predictor variable [X] for the analysis was the continuous variable of accuracy, the moderator variable [W] was the continuous variable of distinctiveness. The dichotomous variable of *list-type* was used as a control variable for the model, as *list-type* was shown to affect the variables of interest, Table 9. The analysis used data from 4560 trials. The model was significant, R^2 = 54.9%, representing the amount of variance in confidence predicted by the model containing both distinctiveness and accuracy.

The additional effect of accuracy on confidence in the context of this model was significant but trivial (*B* = 0.07; *p* = < .001). The effect of distinctiveness on confidence was significant (*B* = 0.38; *p* < .001). The interaction effect (accuracy*distinctiveness) was significant, but trivial (*B* = 0.002; *p* < .001), R^2 change = 0.8%. The conditional effect of accuracy on confidence was calculated at high, middle, and low values of the moderator, distinctiveness (*S.D.* ± 1).

At low moderation (distinctiveness = 26.94), the [conditional effect of accuracy on confidence, *B* = 0.15, 95% *C.I.* (0.13, 0.18), *p* < .001]. At middle moderation (distinctiveness = 54.99), the [conditional effect of accuracy on confidence, *B* = 0.24, 95% *C.I.* (0.22, 0.26), *p* < .001]. At high moderation (distinctiveness = 83.04), the [conditional effect of accuracy on confidence, *B* = 0.33, 95% *C.I.* (0.3, 0.36), *p* < .001].

A simple slopes plot, calculated for the effect of source memory accuracy on confidence by distinctiveness at low, middle, and high moderation effects is illustrated in Figure 48. These results identify distinctiveness as a positive moderator of the relationship between source memory accuracy and confidence. From the output and the simple slopes plot it can be seen that as distinctiveness increases, the positive effect of accuracy on confidence becomes stronger. For the same level of accuracy confidence rises as memory distinctiveness increases; the effect is more marked at higher levels of distinctiveness, where the slopes tend to diverge. The effect size is small (R^2 change < 1%).

For the sake of completeness, the analysis was repeated, assuming confidence was the independent (predictor) variable and accuracy the dependent (outcome) variable[17]. For this model, the interaction effect between confidence and distinctiveness on source memory accuracy was trivial and negative.

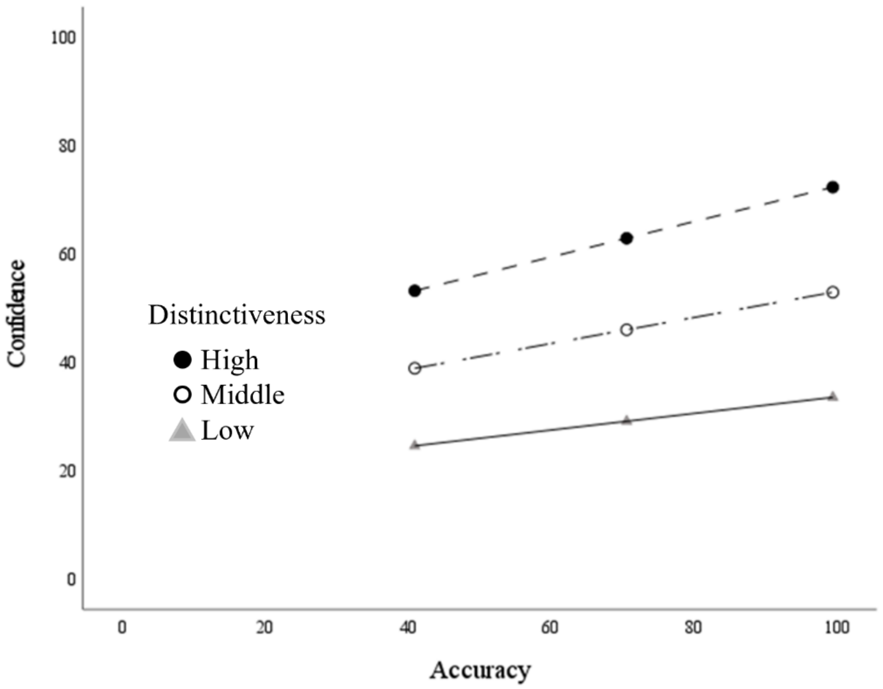

Figure 48. Distinctiveness, Experiments 4 & 5: simple slopes plot (of the interaction effect) for source memory accuracy on confidence by distinctiveness judged at retrieval at values of accuracy (M ± 1 SD), at high, middle, and low levels of the moderator

Mediation Analysis.

Next, it was asked, given the very strong correlation between distinctiveness and confidence for both *list-types* compared to that between distinctiveness and accuracy, whether

[17] Repeating the analysis with confidence as the predictor variable and accuracy the outcome or dependent variable, the model was weaker although there was still a significant overall R^2 effect, ($p < .001$), reducing to 21.7% the amount of variance in accuracy predicted by the model containing both distinctiveness and confidence. The effect of confidence on accuracy was significant (B = 0.48, $p < .001$). The effect of distinctiveness on accuracy was not significant (B = 0.05, $p = .07$). The interaction effect (confidence*distinctiveness) on accuracy was not significant, negative, and trivial (B = - 0.003, $p = .2$, R^2 change = 0.03%).

the relationship between distinctiveness and accuracy was mediated through confidence. Data from both Experiments 4 and 5 were again analysed, comprising 4560 trials. The outcome variable [Y] for the analysis was the continuous variable of accuracy. The predictor variable [X] was the continuous variable of distinctiveness. The mediator variable [M] was the continuous variable of confidence. Whether or not the indirect pathway a*b is significantly different from zero is assessed to test for mediation.

The path from distinctiveness to confidence (a) was significant [B = 0.69, 95% *C.I.* (0.67, 0.71), $p < .001$]. The relationship between confidence and accuracy (b) was significant [B = 0.46, 95% *C.I.* (0.42, 0.50), $p < .001$]. The direct path from distinctiveness to accuracy (c) was not significant [B = 0.03, 95% *C.I.* (-0.013, 0.063), $p = .19$]. The indirect path (a*b) was significant (B = 0.32), Figure 49.

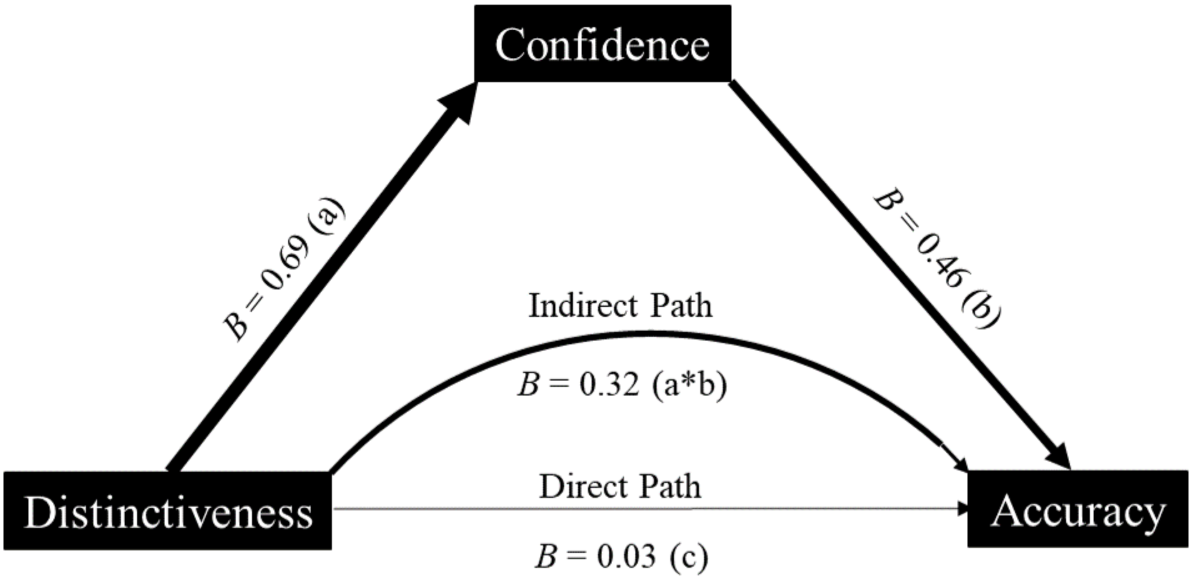

*Figure 49. Distinctiveness, Experiments 4 & 5: The relationship between distinctiveness and source memory accuracy; (a) the path from distinctiveness to confidence; (b) the path between confidence and accuracy; (c) the direct path from distinctiveness to accuracy; (a*b) the indirect path from distinctiveness to accuracy mediated through confidence (Hayes & Scharkow, 2013). It can be seen that path (a) is larger than (b); the direct path (c) is trivial; the weight of the arrows reflects the unstandardized beta (B) values*

For the indirect effect of distinctiveness on accuracy, the resulting confidence interval did not include zero, based on 10,000 bootstrap samples 95% *C.I.* (.29, .34.). The mediation analysis did not support the conclusion that distinctiveness operates independently from confidence in predicting accuracy; 91.4% of the relationship between distinctiveness and accuracy was carried through the indirect path, i.e., was mediated through confidence. In the context of these experiments, the data did not support the conclusion that the distinctiveness of a memory operated independently of this mechanism to affect accuracy. (Note that the analyses do not determine causation or the direction of the relationships, but merely confirm the associations).

Qualitative Data.

Qualitative analysis of what metrics participants based their judgments on was carried out by assessing free comments made at completion of the experiments. The data was then compared by experiment using thematic analysis. All 38 assessable participants commented on why they judged the images as distinctive, and 36 participants also commented on why they had judged the images as *not-distinctive*. The term *not-distinctive* was used to avoid the use of the term similar, as it was thought possible that this could bias participants’ answers. This data may help clarify what is important when distinctiveness is assessed in the absence of image manipulation or an isolation effect paradigm. When participants think an image is distinctive, precisely what criteria are they using?

Although no explicit instructions were given to participants as to what was meant by distinctiveness prior to the experiments, the starting assumption was that the use of natural scenes pictures would allow participants to focus primarily on external or fixed properties of the scene, such as colour and subject, or on their understanding or concept of what distinctiveness meant, i.e., relating to image memorability, where it is the difference of the photograph from the others shown, or different or distinct features within the photograph that are of importance (Rust & Mehrpour, 2020; Suzuki & Takahashi, 1997).

Following the existing literature (cf. Ensor et al., 2019; Hunt, 1995; 2006; Hunt & Lamb, 2001) data were initially collated into two primary codes: firstly, external (perceptual) features, subdivided into primary themes of colour, natural scenes subject category, and image factors (quality or contrast); secondly, concepts (i.e., conceptual distinctiveness – what participants understood distinctiveness to mean). A third primary code of internal factors (personal memories or associations spontaneously recollected on viewing the photographs), emerged on collating the data and was also added. The data were sorted into the codes and themes, checking each was coherent, ensuring that finally what aspect of the data each captured was identified to offer a consistent narrative. Sub-themes provided further information as to what participants believed they were using to rate distinctiveness. Notably, the same primary codes, themes and sub-themes were mentioned by participants regardless of the experimental *list-type*.

Distinctive images.

External primary codes were used by participants, themes being colour (sub-themes: bright colours, and purple, red, or yellow); subject (sub-themes: primarily flowers); and quality (sub-themes: bright or sharp images and contrast). The primary code of concepts was used, themes being interest (sub-themes: more going on, something significant), and a feature within the image (sub-themes: specific or standout features in the image, a feature pointing to or associated with the cross location). Internal primary codes were used to judge images as being distinctive in both experiments, (sub-themes: memories triggered by the picture, the picture reminded them of a place where they had lived or visited).

Not-distinctive images.

External primary codes were used by participants, themes being colour (sub-themes: all the same colour, dull colours, green or brown); subject (sub-themes: primarily forests and trees, grass, or sea); and image factors (sub-themes: blurry or unclear images). The primary code of concepts was used in both experiments, themes being lack of interest (sub-themes: no detail or features in the image, nothing going on) and similar pictures (sub-themes: all the same, landscapes). Internal primary codes of personal memories or associations were never used to judge an image as *not-distinctive*.

Figure 50 illustrates primary codes and themes used by participants to judge images as distinctive and *not-distinctive* for each experiment, expressed as a percentage of those responding.

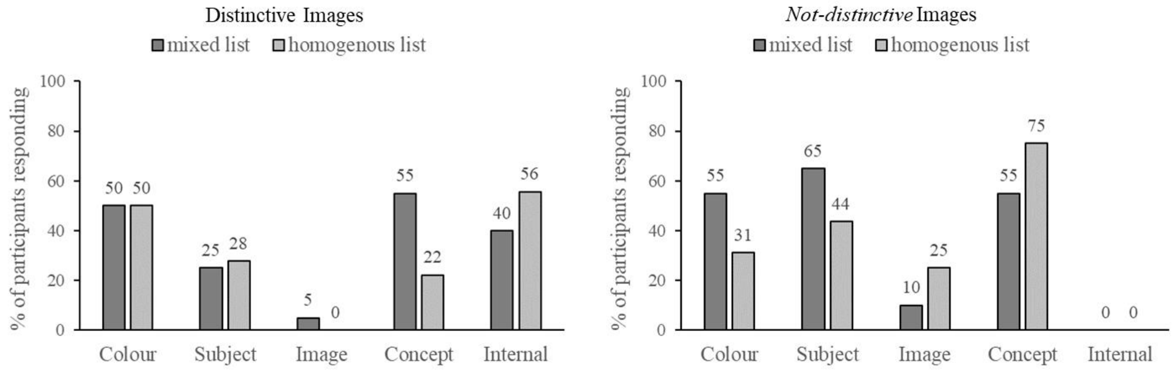

Figure 50. Distinctiveness, Experiment 4, mixed list-type & Experiment 5, homogenous list-type: showing codes and themes used by participants for judging images as distinctive or not-distinctive, split by list-type (experiment). The numbers represent the percentages of participants responding

Comparison: Judging images as distinctive.

Overall, and unsurprisingly (as the pictures were identical), there was little difference between the experiments in the use of external codes (colour or subject category) to rate distinctiveness. Concepts of distinctiveness were used more often for images shown in a mixed compared to a homogenous *list-type*, where participants based judgments of distinctiveness more often on internal factors.

Comparison: Judging images as not-distinctive.

External themes of colour and subject category were used more often for images viewed in a mixed *list-type* to rate images as *not-distinctive*. Images viewed in a homogenous *list-type* were primarily judged as *not-distinctive* using concepts of what lack of distinctiveness meant to participants.

Summary.

The results of Experiments 4 & 5 were broadly in line with those of vividness. However, unlike manipulating vividness by altering colour saturation, manipulating distinctiveness by altering the list arrangement of the same photographs was significant, in that memory performance was impaired, and ratings of distinctiveness and confidence were significantly lower, as was source memory accuracy, for images shown in a homogenous list arrangement (all from the same natural scenes subject category), Experiment 5. When the context in which the identical images were displayed was one of similarity in category, this significantly impaired ratings of image distinctiveness and the ability of participants to perform the task successfully. While no doubt some images were viewed as being distinctive despite their homogenous list arrangement (see appendix A4 more detail on this), it could be speculated that overall there was inferior recall for the non-distinctive images.

Importantly, a very large correlation was again shown between a judgment of memory quality, i.e., distinctiveness and confidence in source memory accuracy. Critically, the large correlation was unchanged, regardless of *list-type*. Correlations between distinctiveness and source memory accuracy were small to moderate. The relationship between accuracy and confidence was moderate to large and did not differ between the experiments, i.e., there was no effect of *list-type*. Distinctiveness was a positive moderator of the relationship between accuracy and confidence. Mediation analysis unequivocally revealed the relationship between distinctiveness and source memory accuracy to act almost entirely indirectly (91.4%), being mediated through confidence.

Qualitative data showed that vividness and distinctiveness were judged on broadly the same codes and themes as for vividness: as well as the external features of the pictures and the natural scenes subject category represented; their concepts of distinctiveness or lack of distinctiveness; participants also based their judgments of distinctiveness on the spontaneous recall of idiosyncratic personal memories triggered by viewing the images (i.e., internal factors).

Discussion of results.

Using a continuous-response task in two separate experiments where the same set of natural scenes photographs were presented in two different list arrangements to manipulate their distinctiveness, it is questioned not whether a distinctive item is better recognised, but if the distinctiveness of memory for the retrieval cue picture influences recollection of source memory for the associated details, i.e., the what, when, and where of memory. Relationships between the distinctiveness of the retrieval cue picture, source memory accuracy (retrieval of contextual detail for the cross location), and confidence in the memory's accuracy were investigated. The finding of a very large correlation between distinctiveness and confidence is supported by studies in face recognition (e.g., Metzger, 2006)

Even though the pictures were identical in both experiments, when images were shown in a context where they did not stand out in category, (i.e., a homogenous *list-type*) not only were they rated as less distinctive, but recall of associated source memory was impaired, confidence was lower, and memory performance impaired (Figure 45). That the very large correlation between distinctiveness and confidence was unaffected by manipulation of distinctiveness using different list arrangements is not so remarkable. If distinctiveness is rated low and confidence is rated low for a natural scenes image when shown in the context of a homogenous *list-type*, a correlation between distinctiveness and confidence should be unaffected.

Schmidt et al. (2021) proposed that the context in which images were viewed was of lesser importance than encoding and retrieval processes for judging distinctiveness. In my studies, images shown in a mixed *list-type* were judged on average as being significantly more distinctive. Nguyen and McDaniel (2015) similarly demonstrated superior recollection for complex relative to simple pictures in mixed rather than homogenous list arrangements.

Evidence from the qualitative analysis confirmed the overlap between factors important for image memorability, image distinctiveness, and vividness. Rust and Mehrpour (2019) considered that the concept of distinctiveness affected the memorability of images, which was not accounted for by perceptual (external) distinctiveness alone. In contrast Koch et al., (2020) suggested that memorability for scenes was an intrinsic property of the image, albeit in a recognition memory task, finding that memory performance was stable across differing individuals and that low memorability was associated with the similarity of images. Here, this might explain the results found for the identical images associated with similarity when shown in a homogenous *list-type*. As found for vividness, spontaneous recollection of personal memories triggered by the photographs, enhanced distinctiveness of memory for the picture (such recalled memories were only ever used for judging images as distinctive and never as *not-distinctive*). This coincides with reports from the literature, (cf. Hunt & Smith, 1996).

More generally, the present empirical findings also have implications for our recall of distinctive events. If a memory is perceived to be distinctive, as suggested for vividness, the differential effect on confidence may lead to illusions of accuracy for the related detail. Of course, we will distinctly remember a purple car with lemon yellow polka dots better than any other car we may have seen on that day (or have inferior recall for the other cars); we will confidently believe in our memory of where and when we saw it, or even what make it was. Awkwardly, given the results presented here, it is likely that our belief in the details may be greater than their recollected factual accuracy.

Chapter 6: Familiarity.

Next, I investigate familiarity. This is perhaps not so much of a discontinuity in the overall theme of my thesis, particularly regarding the discussion related to recognition memory accuracy (in which familiarity is said to be a memory process). While it is not the case that I imply familiarity is a type of memory quality that may affect the association between accuracy and confidence, it may still be the case that familiarity is a particular sort of memory quality. There again, possibly the familiar is viewed as not being distinctive:

*It is a function of art to review our perception. What we are familiar with we cease to see. The writer shakes up the familiar scene, and, as if by magic, we see a new meaning in it. (**Anais Nin** 1903-1977).*

Recognition is a form of memory retrieval which may be characterized by the phenomenological experience of familiarity, when something previously experienced is again encountered (O'Connor & Moulin, 2010). When a stimulus is presented in an old or new recognition memory task, a correct response can be provided in the absence of recollection, but a feeling of familiarity may not be reproduced in the absence of a previously encountered stimulus. Familiarity has been considered to represent a weak type of remembering, with lower confidence when associated with a guess rather than a 'know' response, (Gardiner et al., 1998), indicating that the quality of the memory is lacking. The 2-AFC recognition memory task outlined in General Methods and used in vividness, Experiment 3, provides an opportunity to test familiarity against recognition memory performance.

Familiarity: Experiment 6.

Rationale.

If the task of recognition is based on the processes of recollection and familiarity, both may not occur together. Recollection can be rapid and occur without any sense of familiarity being first experienced. A feeling of familiarity may not result in recollection, i.e., recollection can fail (Mandler, 2008; Mandler et al., 1969). In an old or new forced-choice task, the old target may be recognised because it appears familiar or because it is recollected. The new image (lure) may be falsely recognised as old (because it appears familiar, or it is mistakenly recollected) or correctly discriminated as being new (because it is not familiar, and it is not recollected). When a stimulus is not recognised, the participant may 'know' the answer, resulting in higher familiarity than if they have to guess to complete the task.

Using the same task as in Experiment 3, confidence should be higher if the recognised picture is also recollected (remembered). Reports from the literature suggest that recollection may be associated with lower familiarity. That familiarity is a weak type of remembering (lacking the detail of full recollection), with lower confidence, may be a result of how the question is asked. However, it is also proposed that familiarity is correlated to confidence, in which case familiarity should be high in association with recollection. In order to clarify these points, I explore how recognition of a picture as old is associated with familiarity, recollection, and confidence.

When using a recollect or familiar question, when responding, if the stimulus is not recollected the participant has to answer familiar (even if it is not), i.e., participants are forced to base their recognition judgment on the researcher's belief that the two processes are separate. Naturally, you obtain the answer you expect, that participants base their judgments either on familiarity or on recollection. Familiarity can vary in strength, so that in comparison with the term 'know,' it is a continuous rather than a binary variable. In asking a recollect or familiar question, the trouble with the familiar question is not no, it was not familiar or yes, it was familiar, but rather, <u>how</u> familiar it was.

Opinions in the published literature vary when considering how the two processes of recollection and familiarity act in a recognition memory task. According to dual-process theory, recollection is a threshold (single-state) process, and familiarity is an equal variance signal detection process (Yonelinas,1999; Yonelinas et al., 2010), arguing that they are independent processes, and that recognition is based on either familiarity or recollection, but not both (Yonelinas, 2002). Alternatively the processes are additive (Ingram et al., 2012; Wixted & Mickes, 2010), and both are continuous (unequal variance) signal detection processes. Another view is that participants are able to decide on which process they choose to base their recognition decision, by assessing the strength of their judgments for both recollection and familiarity (Cha & Dobbins, 2021).

Here, I attempt to establish the roles familiarity and recollection play in a recognition memory decision. In the 2-AFC task participants were asked to rate their confidence in recognition accuracy; whether they recollected the picture, yes or no; and rate how familiar it was, for each item recognised as old. Data was routinely collected on the response time for the old or new decision, which enables assessment of whether familiarity is fast (due to fluency), and recollection is, in fact, a slower process. Finally, it was asked whether the test picture pair condition would affect either familiarity, confidence, or accuracy. Consistent with Experiment 3, colour photographs of natural scenes were again used, in line with images employed in other tasks (Reinitz et al., 2012; Chua et al., 2012).

The source memory task previously used for testing metacognitive judgments of both vividness and distinctiveness was felt inappropriate for testing familiarity, recollection is understood to be a separate process from familiarity, i.e., a source memory task relies on recollection and not familiarity. Furthermore, the experimental protocol means that all pictures shown in the test blocks could be regarded as familiar, having just been shown in the relational encoding block.

Methods

The experiment was carried out online and recruitment methods, pre-test information and consent were as described in Chapter 3: General Methods. Detailed instruction was given to participants to explain familiarity, and brief instruction was given to explain recollection. In view of the number of timed-out trials in Experiment 3, participants were allocated 8 seconds to make their recognition decision, to try and reduce any speed accuracy trade-off due to time pressure. Otherwise, the same procedures were used, and stimuli comprised the same images as used in Experiment 3, (at study and for the test picture-pairs).

Test sequence.

Immediately preceding the test sequence, on screen instructions were given as follows:

> “You will see two images together, you will have 8 seconds to decide which picture you recognise as having seen before in the study phase (i.e., the picture is "old"). The buttons are labelled left and right. Press the button under the image you recognise as old. Once you press the button you cannot change your mind and the screen will advance to the rating screens, A timer will count down the last 4 seconds. Please try and keep within the allocated time. If you have no memory of the picture you can guess based on a feeling or on intuition (an informed guess), there is no problem with making this type of guess. It is important that you avoid making random guesses or simply press any button just to get through the experiment quickly. Try to remember the image that you recognised as old for the following ratings.”

Response time data in milliseconds for the recognition decision was automatically recorded by the Gorilla™ platform.

Behavioural responses.

These were requested in the following sequence, once the recognition decision had been made, and applied to the picture recognised as "old."

Confidence.

Participants rated confidence in their old or new decision on a 0 –100 sliding scale. On screen instructions were given as follows:

> "move the slider to rate your response in recognising the picture as old, by answering the question, How confident are you in your decision? There is no time limit. Press continue when you are happy with the slider position." (The slider position started at 0 for all trials).

Recollection, yes or no.

Participants were briefly instructed on screen how to respond to the recollect decision: "Recollect – you have a conscious awareness (of yourself in the past) when you saw the picture in the study phase." (Response buttons were labelled yes and no)

Familiarity.

Participants rated familiarity of the photograph selected as old on a 0 –100 sliding scale. On screen instructions were given as follows:

> "On viewing the scene, a feeling of familiarity raises the possibility that one has encountered the scene before, without recalling any details of the encounter. Familiarity may be a vague feeling, up to a strong feeling of familiarity (I am sure I have seen that before). Move the slider to rate how familiar the picture you recognised as old was by answering the question, how familiar did the picture seem? There is no time limit. Press continue when you are happy with the slider position." (The slider position started at 25 for all trials).

Results.

Participants.

Experiment 6 included 63 participants of mean age 20.3 years (*S.D.* 3.7 years); 56 identified as female. All participants completed 36 test trials. To ensure compliance with task instructions, and data quality (i.e., that all participants were attempting the task) the proportion of correct responses was analysed. The average proportion of correct responses [*M* (*S.E.*)] was .55 (0.01), there were no outliers. One participant, who had answered recollect yes to all test picture-pairs (apart from one timed-out trial), was excluded from the final analysis as it was considered they had either not understood or not carried out the instructions.

A total of 56 timed out trials (2.5%) across all participants were omitted from the analysis, i.e., it took the participant more than 8000ms to decide, and when timed out participants could not make a recognition decision. They were, however, able to make behavioural judgments, but these were excluded from the final analysis, which comprised 62 participants completing 2176 trials. Differences in the number of timed out trials between participants did not influence data analysis, as this was carried out using data averaged by trial for all test picture-pair conditions, rather than data averaged by participant.

Performance.

Just under half of the images (47.5%) selected as old were recollected, Table 11. Overall recognition accuracy (proportion correct) was .62, exactly matching that for Experiment 3. Recognition accuracy was higher if the stimulus was recollected (.74), than not (.58), [B = 0.092 (0.021), $p < .001$], the effect size was trivial (R^2 change = 0.9%).

Table 11. Familiarity, Experiment 6: performance (proportion correct) by recollection response

Familiarity, Experiment 6: performance by recollection response		
Recollected	Proportion correct	Total
no	.58	1142
yes	.74	1034
Total	1429 (.62)	2176

Variables of Interest.

The difference in the relationships between the variables of interest according to recollection (no vs. yes) and recognition (incorrect vs. correct) is apparent from Figures 51a and 51b, as is the similarity between confidence and familiarity distributions.

Confidence.

Confidence was significantly higher *M* (*S.E.*) 67.5 (0.81), if the image was recollected, than not, 27.3 (0.71), ($p < .001$), the effect size was very strong (R^2 change = 39.4%), Figure 51.

Familiarity.

Familiarity was significantly higher if the recognised image was also recollected, 67.4 (0.79), than not, 34.2 (10.72), ($p < .001$), the effect size was very strong (R^2 change = 30.8%), Figure 51.

Figure 51a demonstrates that confidence and familiarity ratings are concentrated in the upper part of the scale when recollection was involved in the recognition decision, compared to the lower half when not.

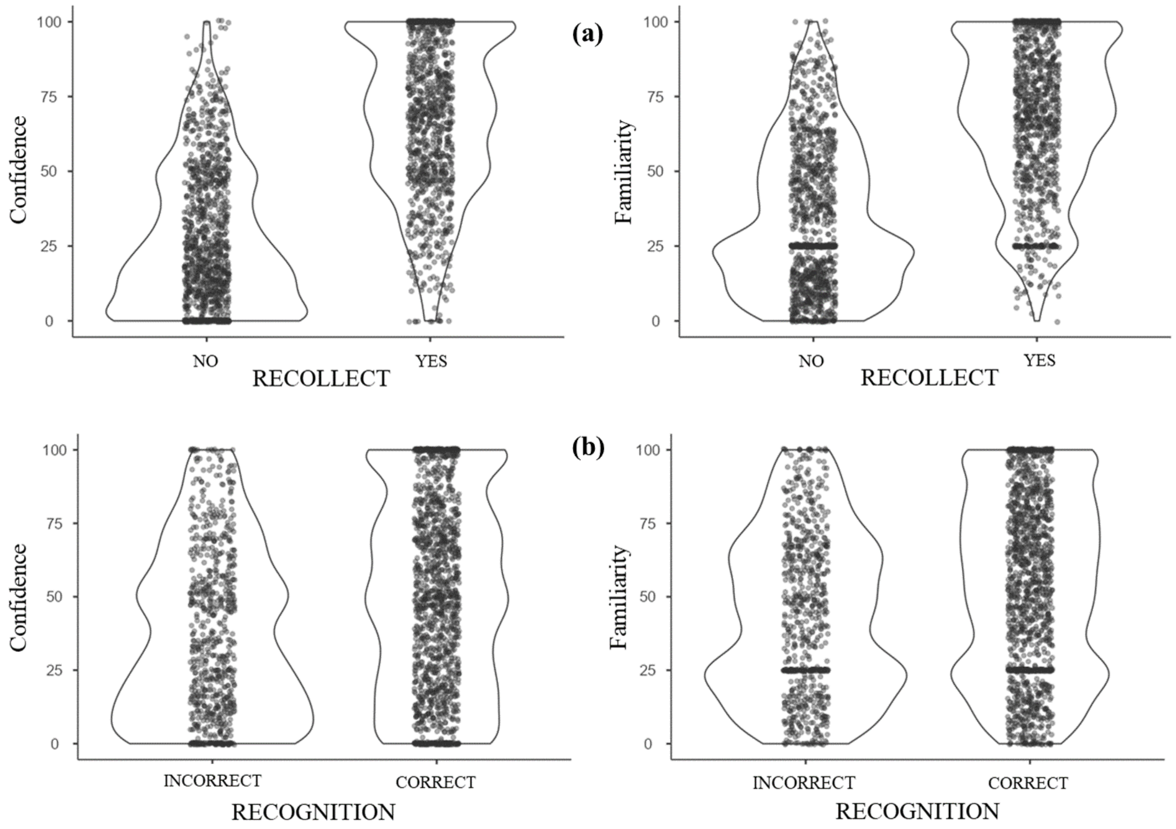

Figure 51. Familiarity, Experiment 6: univariate scatter (strip) plots and overlaid violin plots illustrating confidence and familiarity distributions for all trials: (a) by recollection (= no vs.= yes); (b) by recognition accuracy (correct vs. incorrect). Confidence ratings are concentrated in the upper half of the scale when the recognised image was recollected, compared to the lower half when not. The distribution of ratings is similar for familiarity. Note the effect of anchoring at the slider position = 25 for familiarity (regardless of recollection)

Reaction time.

Reaction time was significantly faster if the image was recollected, 3282.0 (42.59) ms than if not recollected, 4126.1 (52.60) ms, ($p < .001$), the effect size was moderate (R^2 change = 6.5%).

Correlations.

Confidence by familiarity.

The correlation between familiarity and confidence was found to be very large, (r = .77); 59% of the variance in confidence can be explained by the variability in familiarity of the images recognised as old. The averaged within-participants correlations confirmed the very large correlation between familiarity and confidence (*r* = .775); all correlations were positive (Appendix B, see Table 28S & Figure 80S).

Reaction time by familiarity.

There was a small negative correlation between familiarity and reaction time (*r* = -.22). As familiarity increased, reaction time was faster, but < 5% of the variance in reaction time was explained by familiarity. Individual differences, test picture-pair condition, and time pressure are suggested as affecting reaction time in this task.

A scatterplot of reaction time by familiarity was next explored by whether or not the image was recollected, Figure 52. The lack of a material difference in correlation between familiarity and reaction time dependent on recollection is clearly illustrated. However, variable density plots starkly highlight the difference in familiarity distribution consequent on recollection, also confirming confirmed reaction time to be faster when the recognised image was also recollected.

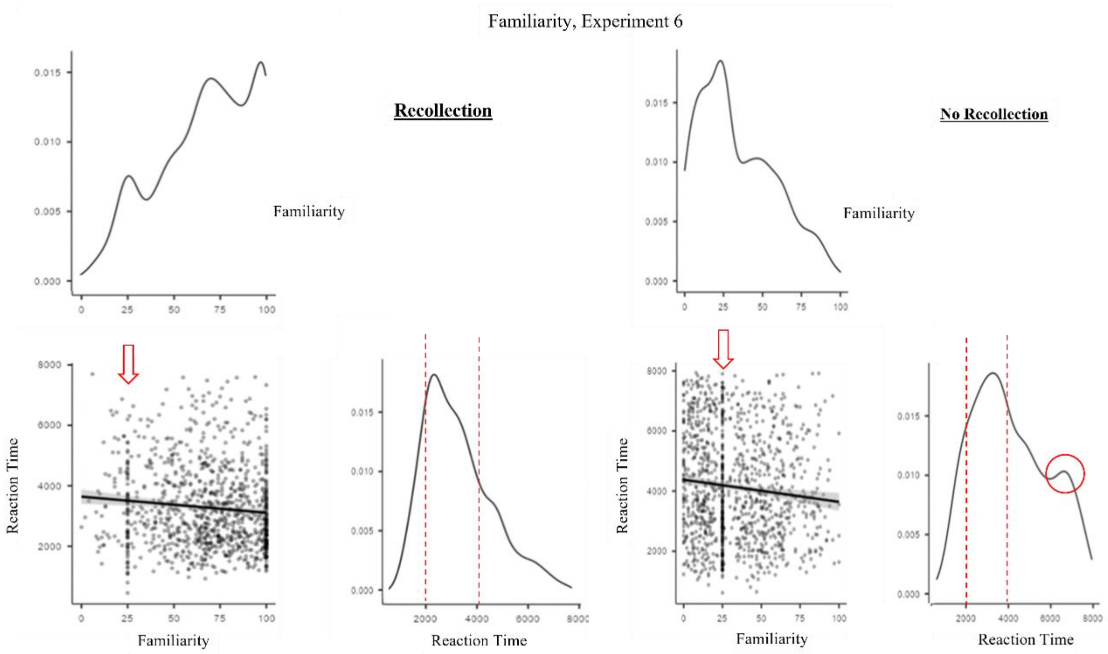

Figure 52. Familiarity, Experiment 6: illustrating variable density plots for reaction time and familiarity (i.e., the smoothed histograms of the variable distributions) and correlation matrices (scatterplots) of reaction time by familiarity, split by whether or not the recognised image was also recollected, where 0 –100 scales on the abscissa and the ordinate represent values for the correlation matrices, and the scale on the ordinate represents the density function for the variable density plots. Dashed lines are placed for comparison of reaction time density plots. Arrows highlight anchoring at familiarity = 25, the slider starting point. The second reaction time

peak when the image was not recollected, indicated by the red circle, is thought due to time pressure (participants had 8000 ms to respond).

A second small reaction time peak at 6-8000 ms, only seen when the image was not recollected, suggests decision time pressure (highlighted by a red circle). Anchoring at familiarity = 25 (red inset arrows) was more marked when the image selected was not recollected, when the familiarity response was more likely to be based on information provided by the slider starting point.

Confidence dependent on familiarity, by recollection.

Both confidence and familiarity were significantly influenced by whether or not the image selected as old had been recollected. A scatterplot of confidence by familiarity was next explored by whether or not the image was recollected, Figure 53. As is clearly illustrated, trials for images that were recollected were predominantly associated with higher confidence and higher familiarity judgments. Trials for images that were not recollected were predominantly associated with low confidence and low familiarity judgments.

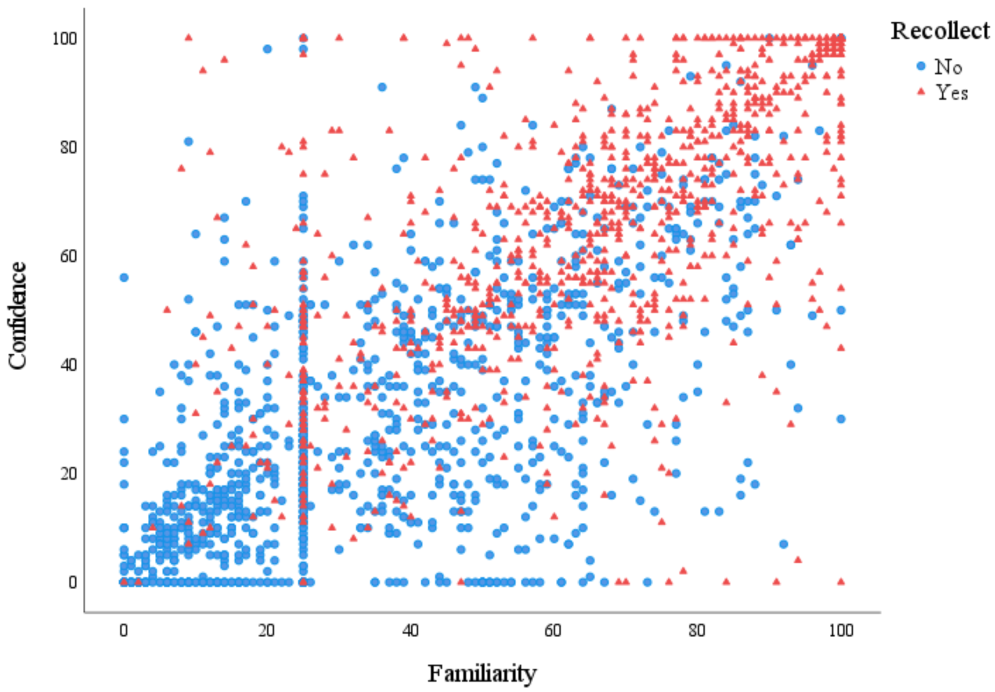

Figure 53. Familiarity, Experiment 6: bivariate scatterplot of confidence dependent on familiarity with markers indicating whether or not the stimulus was recollected. Although there is anchoring where the sliders started at familiarity = 25 and at confidence = 0. It is notable that the distribution of trials for images that were

recollected is predominantly associated with high confidence and high familiarity judgments. Images that were not recollected were predominantly associated with low confidence and low familiarity judgments

Confidence dependent on familiarity, by performance (recognition accuracy).

A scatterplot of confidence by familiarity was further explored by recognition accuracy, representing performance in the task (correct or incorrect recognition), Figure 54. The distribution of correct or incorrect responses showed no clear association with ratings of confidence or familiarity.

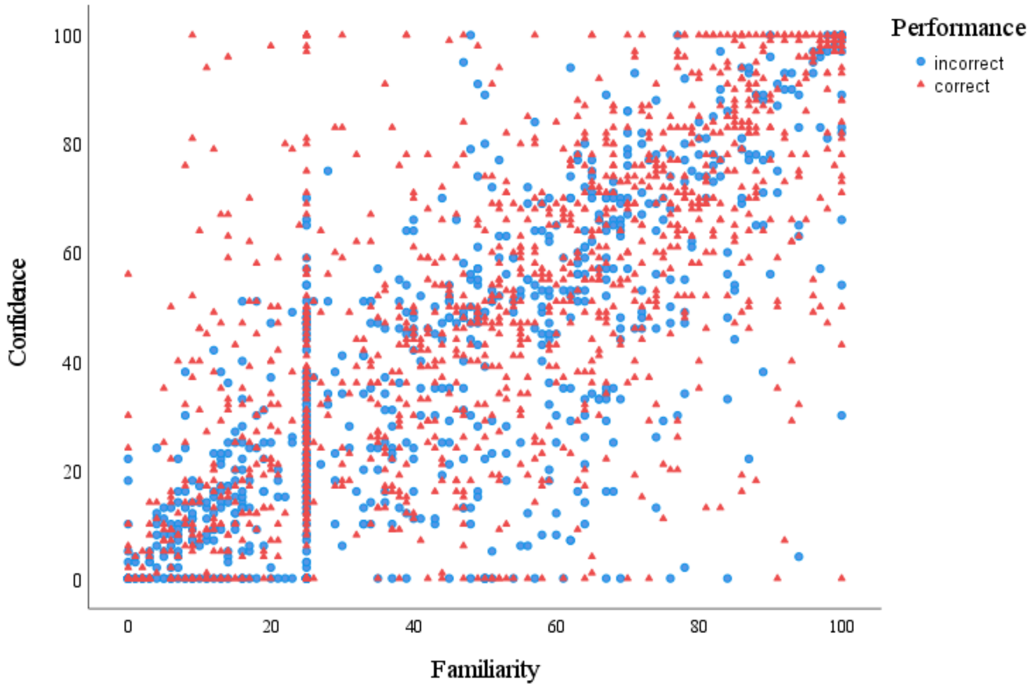

Figure 54. Familiarity, Experiment 6: bivariate scatterplot of confidence dependent on familiarity with markers indicating performance. Although there is anchoring where the sliders started at familiarity = 25 and at confidence = 0, the correlation remains strong. Unlike recollection however, there is no definite pattern from the distribution of correct or incorrect responses

Notably, in both Figures 53 and 54, there is a cluster of trials associated both with high confidence and with high familiarity ratings, which were both recollected and recognised accurately (correct). Note in each case these are illustrated by the small red triangles at the top right hand corner of the scatterplots. (This pattern was not so evident for vividness, Experiment 3, neither with remembering nor in association with accurate recognition).

Test picture-pair condition.

Variables of interest.

Performance (proportion correct) was lower for the mnemonically similar (A-B') test picture-pair condition, but there was no significant between-conditions difference, ANOVA [F (2, 2173) 0.42; p = .66]. Likewise, there were no significant between-conditions differences in confidence, ANOVA [F (2, 2173) 2.06; p = .13], or in familiarity, ANOVA [F (2, 2173) 1.92; p = .15], Table 12.

Familiarity, Experiment 6: variables of interest *M (S.E.)* by test picture pair condition

Test picture pair condition (n)	Proportion correct	Confidence	Familiarity	Reaction time, ms
A-A′ (721)	.63 (0.02)	44.6 (1.16)	50.02 (1.11)	3885.9 (62.83)
A-B ′ (731)	.59 (0.02)	48.04 (1.19)	51.48 (1.10)	3656.0 (59.79)
A-X (724)	.65 (0.02)	46.42 (1.20)	48.40 (1.12)	3634.3 (60.96)

A significant between-condition difference was demonstrated for reaction time (time to making a recognition decision). As in Experiment 3, reaction time was slower when the picture pairs were perceptually similar, (A-A'): ANOVA [F (2, 2173) 5.17; p = .006, η^2 .008]; however, in Experiment 6 when an extra 2000 ms were available to make the response, the effect size (η^2) was small, Table 12.

Moderation analysis.

To further explore the results a simple moderation analysis was carried out in SPSS using PROCESS, as described in General Methods. The analysis asks whether recollection of the recognised image influences the association between familiarity and confidence. The outcome variable [Y] for the analysis was the continuous variable of confidence. The predictor variable [X] for the analysis was the continuous variable of familiarity. The moderator variable [W] for the analysis was the dichotomous variable of whether the image was recollected (yes or no). The dichotomous variable of accuracy, (whether the image was recognised correctly, yes or no) was used as a control variable. Figure 55 illustrates the path diagram for the moderation model.

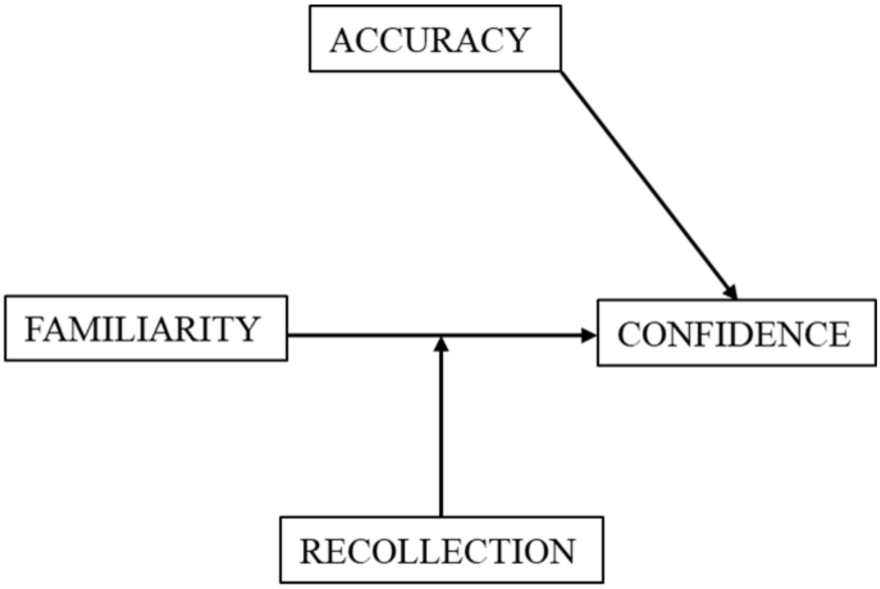

Figure 55. Familiarity, Experiment 6: path diagram for the moderation model used to produce the simple slopes plot in Figure 56

Results confirmed a significant overall R^2 effect for the model, i.e., 66 % of the variance in confidence was predicted by the model containing familiarity, recollection, and accuracy. As previously shown in the results, familiarity had a significant effect on confidence [B = 0.61, 95% *C.I.* (0.57, 0.66), $p < .001$], as did recollection [*B* = 13.9, 95% *C.I.* (10.2, 17.65), $p < .001$]. The interaction effect (familiarity*recollection) was significant [*B* = 0.08, 95% *C.I.* (0.014, 0.14), $p = .02$], but trivial (R^2 change= 0.09%).

At low moderation (no recollection), the [conditional effect of familiarity on confidence, *B* = 0.61, 95% *C.I.* (0.57, 0.66), $p < .001$]. At high moderation (recollection), the [conditional effect of familiarity on confidence, *B* = 0.69, 95% *C.I.* (0.64, 0.73), $p < .001$]. A simple slopes plot which estimates and tests at the mean of familiarity ± 1 *S.D.*, illustrates that the effect of familiarity on confidence is conditional on (moderated by) recollection, Figure 56. As seen from the *B* values, the moderation effect does not change materially with increased familiarity, i.e., the moderation effect is constant across the range of familiarity strengths.

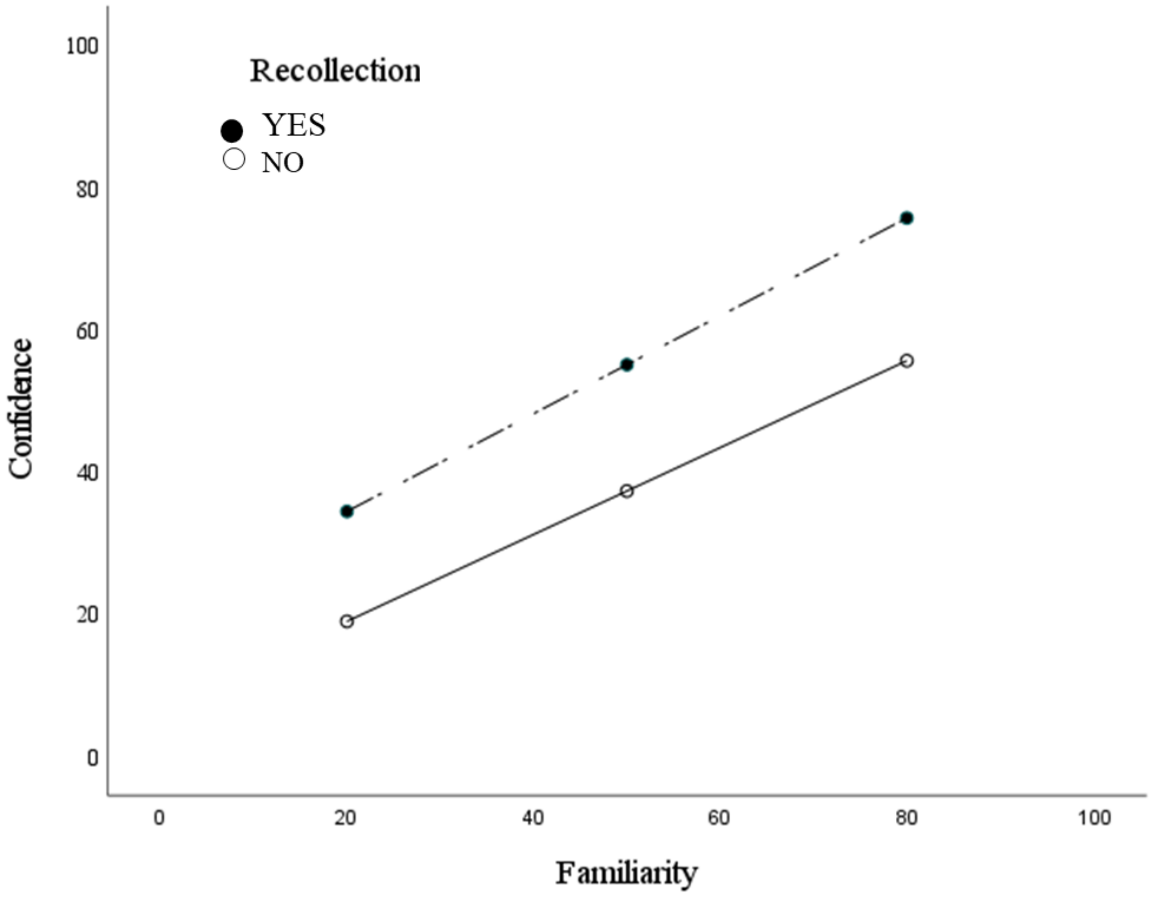

Figure 56. Familiarity, Experiment 6: Simple slopes plots (of the interaction effect): of familiarity with confidence by recollection, shown at levels of familiarity at mean ± 1 S.D.

Owing to the nature of the experiment, the binary variable of recognition accuracy is unsuitable for mediation analysis (i.e., to directly analyse whether or not the association of familiarity or confidence with performance acts directly, or indirectly through a mediator).

Calibration curve analysis.

As shown for Experiment 3, a further way of testing the relationship between familiarity, recollection, and recognition accuracy uses a calibration curve. This was carried out as for confidence accuracy characteristic (CAC) analysis but substituting familiarity for confidence. It is assumed that participants can reliably rate their psychological experience on to the scale used, (Rhodes, 2019), and a calibration curve is calculated by dividing metacognitive judgments into bins, i.e., of confidence (here used for a judgment of familiarity) and determining the average accuracy for each bin. Plotting the results is equivalent to a calibration (Palmer et al., 2013).

Figure 57 illustrates results using data from all trials. As familiarity increases so does accuracy of recognition. Importantly, when familiarity is low, recollection has a greater effect on recognition accuracy than when familiarity is high. Ingram et al., (2012) only used the highest ratings of confidence for participants to make remember and familiar judgments on the same scale, advising that high values for recollection and for familiarity both indicated high values of confidence (that the target was old). The ratio between the number of trials in each condition further confirm that not recollecting the image is predominantly associated with low values of familiarity, whereas recollection is predominantly associated with higher values of familiarity.

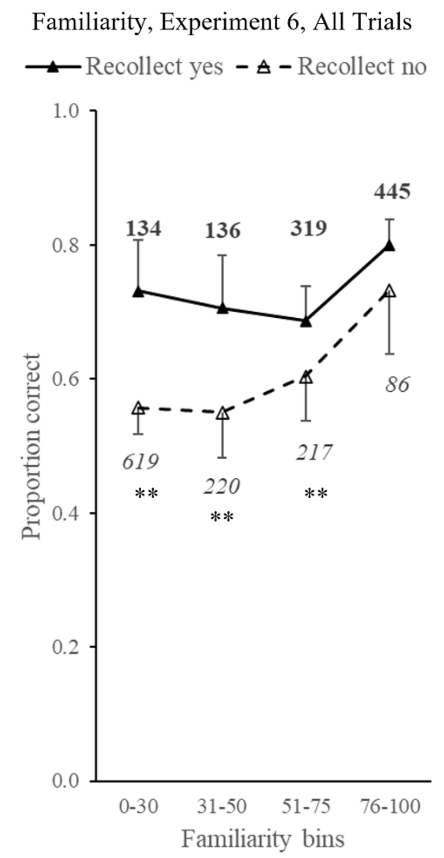

*Figure 57. Familiarity, Experiment 6: Data from all test pair conditions. The numbers represent trials assessed at each bin of familiarity; [bold type denotes number of trials associated with recollection, italics denote number of trials not associated with recollection; **indicates a significant difference in means; error bars represent 95% C.I. (only positive or negative bars are shown for clarity)]*

The same pattern applies when calibration curves are calculated for each test picture-pair condition, Figure 58. It is apparent that when familiarity is low, judgments relying on recollection are more accurate. When familiarity is high, the difference in performance (proportion correct) between targets recollected or not recollected is smaller. From Figures 58 and 59, it is apparent there are more trials in each condition associated with recollection when familiarity is high, i.e., in all but the lowest bins of familiarity (familiarity > 50).

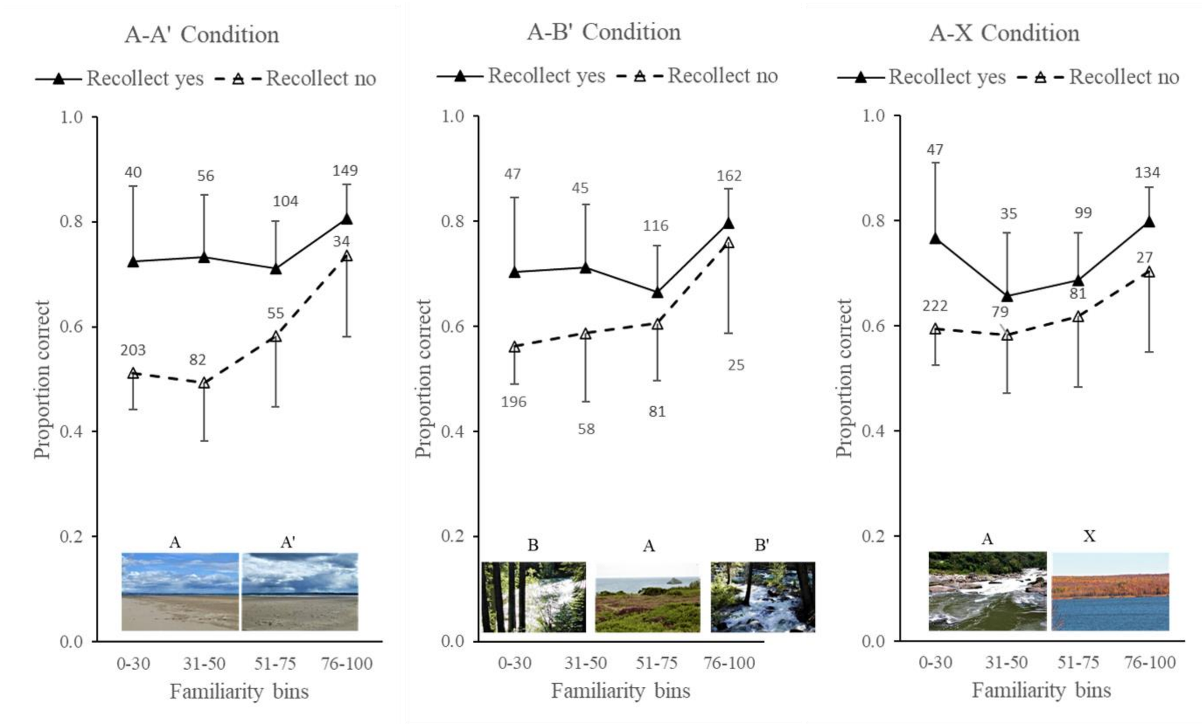

*Figure 58. Familiarity, Experiment 6: Calibration plots for each test pair condition illustrating mean recognition accuracy (proportion correct) by binned familiarity; [bold type denotes number of trials associated with recollection, italics denote number of trials not associated with recollection; **indicates a significant difference in means; error bars represent 95% C.I. (only positive or negative bars are shown for clarity)]. Example test pairs are shown for each condition; to be clear, all the A pictures are old, having been shown at study in the target block. The B picture illustrated was shown during the target block in the study phase of the experiment but was not shown again at test. Neither the A' (perceptually similar lure), B' (mnemonically similar lure) nor the X (dissimilar lure) had been shown previously at study (in these examples, all test pairs represent the natural scenes category of water)*

Summary.

In Experiment 6, the averaged proportion of correct responses (.62), exactly matched performance in Experiment 3. This suggests that the population of students were equivalent, and that the judgments required (vividness or familiarity; remember or recollect, yes or no) had no impact on performance in the task. Notably, the increase in decision time from 6000 to 8000 ms also had no effect on performance, i.e., there did not appear to be any change in a potential speed accuracy trade-off due to time pressure by allowing a further 2000 ms in which participants could make their decision. However, there was a significant reduction in the number of timed out trials, reducing to 2.5% compared to 4.8% in Experiment 3. As the same set of images were displayed in both experiments, the similarity in performance also suggests that the influence of the images on recognition was stable.

The correlation between familiarity and confidence was very large. In trials where confidence and familiarity were both high the recognised image was more likely to be recollected; if both confidence and familiarity were low, the recognised image was less likely to be recollected. No such pattern was seen for trials with either correct or incorrect recognition. The association between familiarity and confidence was conditional on (moderated by) recollection. This did not change across all levels of familiarity ($M \pm 1S.D.$).

Calibration curves provide a more intuitive visual interpretation of the results, demonstrating that as familiarity increased, the positive effect of information obtained from recollection on accuracy decreased, i.e., when the picture was rated as highly familiar (≥ 75), information provided by recollection of the picture was less important for performance in the task. When the picture was rated as low in familiarity, information provided by recollection of the picture was of more importance.

There were no significant differences between test picture-pair conditions in respect of confidence, accuracy, or familiarity, despite expecting familiarity to be higher for perceptually or mnemonically similar test picture-pairs. The decision response time was significantly slower for perceptually similar (A-A′) pairs. Furthermore, the response time for decisions made when the picture was recollected were significantly faster than those made in the absence of recollection.

Discussion of results.

Familiarity and confidence were closely coupled; from these results I conclude that assessing how familiar the recognised image was is of importance. Familiarity may be independent of recollection but acted in concert with recollection to make a recognition decision. Information provided by recollection of the picture was less important for improved performance in the task when the picture was highly familiar, and of more importance when the picture was not familiar. Results provided by the moderation analysis suggest the processes of familiarity and recollection are complementary in their effect on confidence.

Asking a recollect or familiar question is like asking what do you drink: coffee or tea? You may want to answer both. How do you decide which to opt for when forced to answer one or the other? When participants were allowed to make both judgments after an old or new recognition decision, when the image was recollected, familiarity was high. Despite the suggestion from Cha and Dobbins (2021), it is doubtful whether participants make a conscious decision to use recollection of the target if it was not familiar. Rather, I would theorise this represents an unconscious decision or heuristic, based on information from both recollection and familiarity, possibly related to fluency and response times. Rather than familiarity being associated with low confidence decisions, this experiment demonstrates that both familiarity and confidence can be high in association with recollection. When the image is recognised, but neither familiarity nor confidence are high, participants are less likely to have recollected the target.

At this point, I contrast my results with those of Cha and Dobbins (2021), who used a word-based old or new recognition task, theorising a continuous dual-process (CDP) model of recognition memory, which assumes that recollection and familiarity are equally weighted during a recognition judgment. For words recognised as old, ratings of both familiarity and recollection strength were required on a scale of 1 – 10 (in whichever order participants wished, the scales being presented as two halves of a circle). Ratings of familiarity and recollection strength were moderately correlated ($r = .6$, $p < .001$). Confidence was also rated, but on a 1 – 3 Likert scale (low, medium, high). Both recollection and familiarity predicted confidence, although coefficients were much larger for recollection. Nevertheless, much of the data was averaged (even though it may not be mathematically correct to take the average of low and high as medium); furthermore, when correlating a 1 – 3 Likert scale with a 1 – 10 recollection strength scale, you should rightly use a Spearman rank correlation, ρ). Recognition accuracy was lower when participants relied on familiarity rather than recollection strength to recognise the stimulus, i.e., when compared to recollection, familiarity was a low accuracy judgment. My results corresponded, in that accuracy based on familiarity alone (no recollection), was increased when the stimulus was recollected. However, the difference was greater for low ratings of familiarity compared to high ratings of familiarity. In my study, incorrect recognition could be associated with any level of familiarity. Confidence showed the same association with recollection as familiarity, with confidence and familiarity being highly correlated. Although you might argue that I am assessing relative confidence and relative familiarity in this 2-AFC task, hence the difference in results, their results might have been different had Cha and Dobbins (2021) used an equivalent 1 – 10 scale for confidence.

In "Recollection can be weak and familiarity can be strong", Ingram et al., (2012) also used a word-based old or new recognition memory paradigm, combining it with a source memory task (based on the location of the word or the colour of the screen it was presented on). Participants were required to rate confidence in their recognition decision on the same 1 – 20 scale as in prior experiments from their group (i.e., confident new = 1 to confident old = 20; cf. Mickes et al., 2007) but the researchers only considered the highest ratings of confidence (16 – 20), for participants to make remember and familiar judgments on a similar scale, advising that "responses of 20R and 20F should be equal in strength, they should both indicate 100% certainty that the word was on the list. Point bi-serial plots were used to illustrate changes in accuracy (proportion correct) comparing recollection and familiarity responses from the highest part of the scale (16 – 19 vs. 20), leading to the conclusion that recognition accuracy (but not source memory accuracy) improved with increasing levels of familiarity (this is broadly in line with the data seen in Figure 57), adding that both recognition and source memory accuracy improved with increasing levels of recollection (which is what one might expect, given that source memory is based on recollection and not familiarity).

Slotnick’s (2013) suggestion of recollection and familiarity being a single process bears consideration here. However, my results do not confirm this is the case. Familiarity and recollection do appear to be separate processes in the recognition memory task but (unlike dual-process theory) it is not the case of one or the other being used, neither does the participant consciously choose one or the other to rely on. What I found is that the processes act in concert, with information from recollection being of more importance in the recognition decision when familiarity is low, and of less importance when familiarity is high. Importantly, response times are faster when the stimulus is recollected.

Chapter 7: Metacognitive Judgments at Encoding

A potential problem with my research regarding the continuous response source memory task, is whether participants were reporting the quality of their memory for the picture at test, or simply reporting how vivid or how distinctive the picture seemed to them when viewed again. In the continuous-response task, as all the test images are old, unlike in a recognition memory task, this affords the opportunity to ask for judgments of memory quality at study, when the images are first presented in the relational encoding blocks. Our own perception of vividness may differ between when experiencing and when remembering a stimulus, e.g., flashbulb memories are said to be vivid because the event was significant when experienced (i.e., at encoding). To counter any comment that results might be affected because asking for a judgment of memory quality at encoding would make the pictures better (more deeply) encoded, might be important for directly comparing the two judgment conditions. However, the intention here is to compare results for the two image sets, which are both encoded in the same manner.

(i) Vividness at encoding, Experiments 1 & 2.

To answer this question the source memory tasks were repeated with additional participants who were asked to judge image quality at encoding. Results could then be compared to make sure that the judgments of memory quality reported were not simply judgments of the quality of the picture.

Rationale.

Cooper et al., (2019) found that images were recollected as less “visually salient” when remembered than when experienced. The natural scenes photographs might be better recognised at test, if it is assumed that asking participants to rate them for vividness means they will be more deeply encoded, but it is questioned whether this will translate to improved source memory recollection, i.e., the location of the related cross.

Procedure.

Both Experiments 1 & 2 were repeated to ensure a wide range of vividness judgments. Procedures were followed as laid out in General Methods, and as reported for Chapter 4. Participants were instructed to make their vividness judgments at study, immediately after the natural scenes picture was first presented (but not when still visible on the screen), Figure 59.

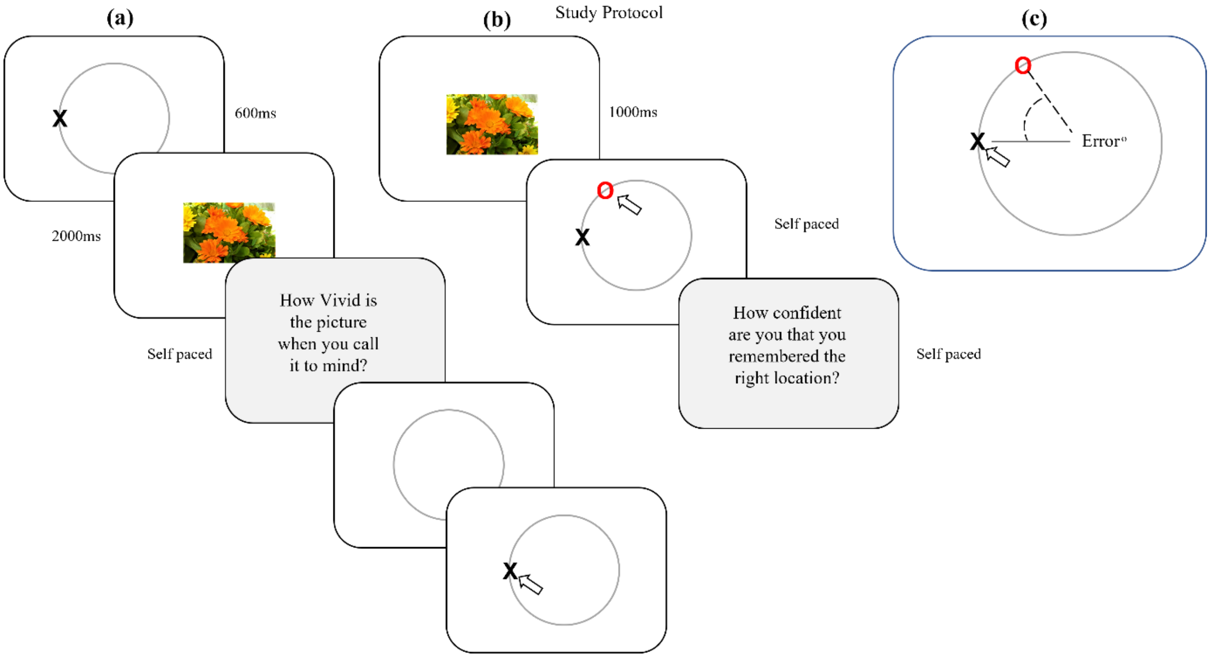

Figure 59. Vividness, encoding judgment, Experiments 1 & 2: Protocol for experiments using judgments of image vividness at encoding, during the study block. (a) Study (relational encoding) block; (b) Test (retrieval) block; (c) Measuring location (radial) error: the distance in ± degrees between the participant's response (shown as a red circle) and the correct target location (shown as a cross on the circle)

Stimuli.

The identical picture sets were used as shown in Figures 24 and 25, and the relational encoding blocks comprised the same picture-cross location pairs as were used previously, however presentation order of the images within each block (as well as the order of study and test blocks) was randomised for each participant by E-prime.

Results.

Participants.

There were 16 participants in Experiment 1, of mean age 27.3 years (*S.D.* 5.0 years); 10 identified as female, each carried out 120 trials; there were 16 participants in Experiment 2, of mean age 22.1 years (*S.D.* 4.3 years); 15 identified as female, each carried out 240 trials. The analysis included data from all 32 participants. In keeping with the reported experiments where vividness was judged during the test blocks, comparisons were made between Experiments 1 and 2.

Memory performance.

There were no significant differences in memory performance between the experiments. How well participants performed in the task was unaffected by the vividness manipulation of the natural scenes photographs in the two experiments. Making the judgment at study had no 'depth of encoding' effect on participants' performance in the task (Table 13, cf. Table 2).

The similarity is further illustrated in the location error frequency plot, Figure 61. Note the difference in response frequency (twice as many trials were carried out by the 16 participants in Experiment 2).

Table 13. Vividness, encoding judgment, Experiments 1 and 2: Memory Performance [Recollection Rate, Scale (precision) and Threshold (Degrees from target), M (SE)], Mean Values Averaged for all Participants (N = 32)

Vividness, Encoding judgment: (120-image set, full colour)
Vividness, Encoding judgment: (240-image set. full & partially desaturated for colour)

Experiment	Variable	*n*	*M* (*SE*)	Independent samples *t*-test *t*(*d.f.*)	*p*-value 2-tailed
120-image set	Rate	16	0.52 (0.18)		
240-image set		16	0.41 (0.28)	1.32 (30)	.2
120-image set	Scale	16	11.67 (1.26)		
240-image set		16	12.74 (1.39)	-0.57 (30)	.57
120-image set	Threshold	16	27.88° (2.46°)		
240-image set		16	33.06° (5.8°)	-0.82 (30)	.42

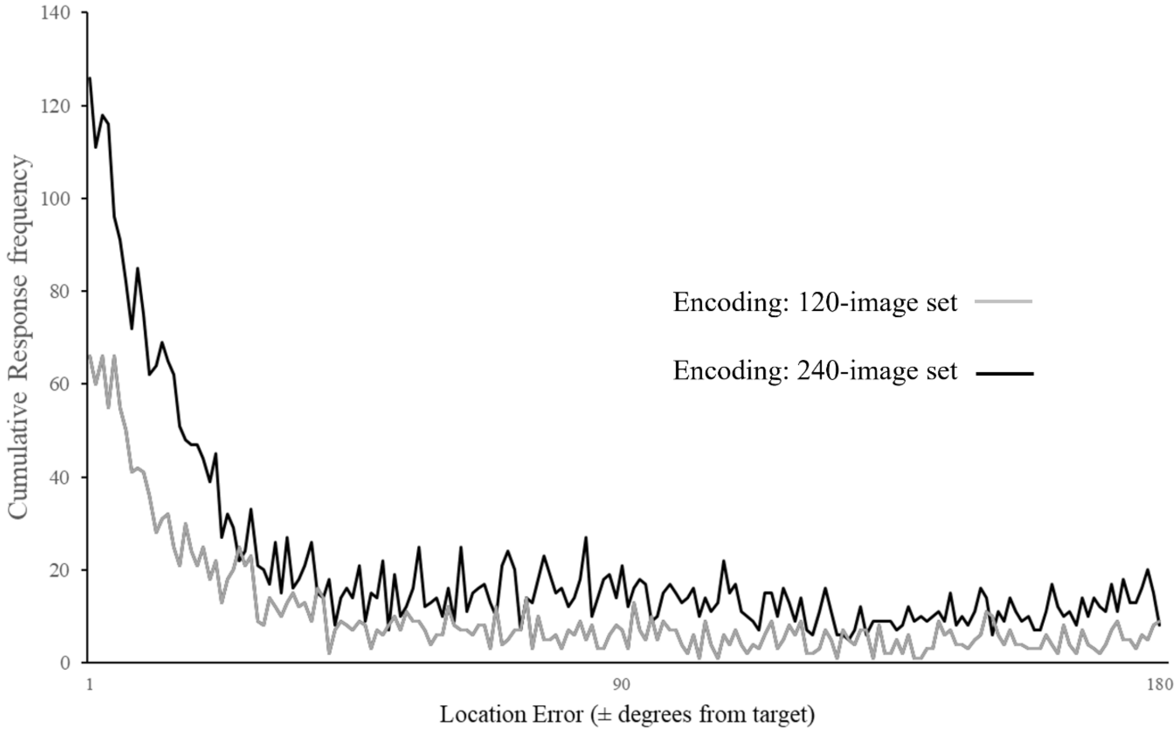

Figure 60.Vividness, encoding judgment, Experiments 1 (120 images in full colour) and 2 (240 images in colour and partially desaturated for colour): location error frequency plot showing response frequency, summed across all participants for each degree of error from the target (0° – 180°)

Variables of interest.

Vividness, confidence, and accuracy were all significantly higher for full colour images in Experiment 1, however, the effect sizes were trivial (R^2 change ≤ 0.2%), Table 14. In effect, when analysis was carried out on a by trial basis, as the *p*-values cannot be relied upon, when vividness of the image was judged at study, there were no meaningful differences in the variables of interest between images in full colour, or when half were partially desaturated for colour.

Table 14. Vividness, encoding judgment, Experiments 1 & 2: Variables of interest (averaged across all trials) by image set, [M (SE)], compared by R^2 change

Vividness, Encoding judgment: (120-image set, full colour)
Vividness, Encoding judgment: (240-image set. full & partially desaturated for colour)

Image set	Variable	*df*	*M (SE)*	Unstandardized *ß* coefficient (*SE*)	*p*-value[a]	R^2 change
120-image set	Vividness	1,5758	53.1 (0.5)			
240-image set			55.3 (0.4)	2.19 (0.63)	< .001	0.2%
120-image set	Accuracy	1,5758	71.3 (0.7)			
240-image set			68.2 (0.49)	-3.1 (0.83)	< .001	0.2%
120-image set	Confidence	1,5758	44.9 (0.7)			
240-image set			43.2 (0.4)	-2.56(0.83)	.002	0.1%

Note. [a]The p-value can appear significant because the number of trials is large; of more relevance is the effect size denoted by R^2 change, which is documented only when the p-value is reported as significant [R^2 change: 1% - weak, 9% - moderate, 25% - strong effect].

Correlations.

Correlations between vividness and source memory accuracy were small, Table 15. Variability in vividness of the photographs judged at encoding explained < 1% of the variance in source memory accuracy for the 120-colour image set and 1.4% for the 240-image set (half in full colour and half partially desaturated for colour).

In stark contrast to judging vividness of memory for the pictures, the strength of correlation between vividness at encoding and confidence was small to moderate for both experiments (Table 15); ≤ 6.3% of the variance in subjective confidence can be explained by variability in the perceived vividness of the stimulus at encoding. The relationship between accuracy and confidence was unaffected by judging vividness at encoding and remained moderate to strong for both experiments, (Table 15), accounting for ≤ 20% of the variability in confidence.

Results are confirmed by averaged within-participant correlations, Table 14. Appendix B details all individual results (Tables 29S & 30S, Figures 81S-86S).

Table 15. Vividness, encoding judgment, Experiments 1 & 2: Correlation analysis by experiment (picture set), denoting relationships between variables of interest Pearson's r-values are displayed using data from all trials and compared with r-values and R^2 variances obtained by averaging within-participant correlations. It is evident that there is little difference between the two methods

Vividness, Encoding judgment: (120-image set, full colour)
Vividness, Encoding judgment: (240-image set. full & partially desaturated for colour)

Image set	I.V.	D.V.	*N*	Pearson's *r* *(all trials)*	*p*-value 2-tailed	R^2 variance	Pearson's *r* *(averaged)*	R^2 variance
120-image set	Vividness	Accuracy	1920	.07	< .01	0.5 %	.06	0.4 %
240-image set	Vividness	Accuracy	3840	.12	< .01	1.4 %	.12	1.4 %
120-image set	Vividness	Confidence	1920	.23	< .01	6.9 %	.18	3.2 %
240-image set	Vividness	Confidence	3840	.25	< .01	5.3 %	.27	7.3%
120-image set	Accuracy	Confidence	1920	.45	< .01	20.3 %	.46	21.2 %
240-image set	Accuracy	Confidence	3840	.39	< .01	15.2 %	.38	14.4 %

Note. Cohen (1992) defined small, moderate (i.e., medium), and large *r* values as "about" .10, .30, and .50, respectively. R^2 represents the amount of variance in the dependent variable predicted by variability in the independent variable.

Qualitative data.

Analysis followed the procedures as described in Chapter 4, Vividness, Experiments 1 & 2. Qualitative data was judged on the same basis as described previously and sorted into primary codes and themes used by participants to judge images at encoding as either vivid or *not-vivid* for each experiment. There were no differences in the criteria used; primary codes, themes, and sub-themes were identical to those used by participants judging vividness of memory for the images. Results are expressed as a percentage of those responding, 32 participants commented on what made the images vivid; 24 participants also commented on what made the images *not-vivid*.

Vivid images.

When half of the images were less visually vivid (Experiment 2), the external theme of subject was more important, with image quality being used more often to judge the pictures in Experiment 1 as vivid (Figure 61a).

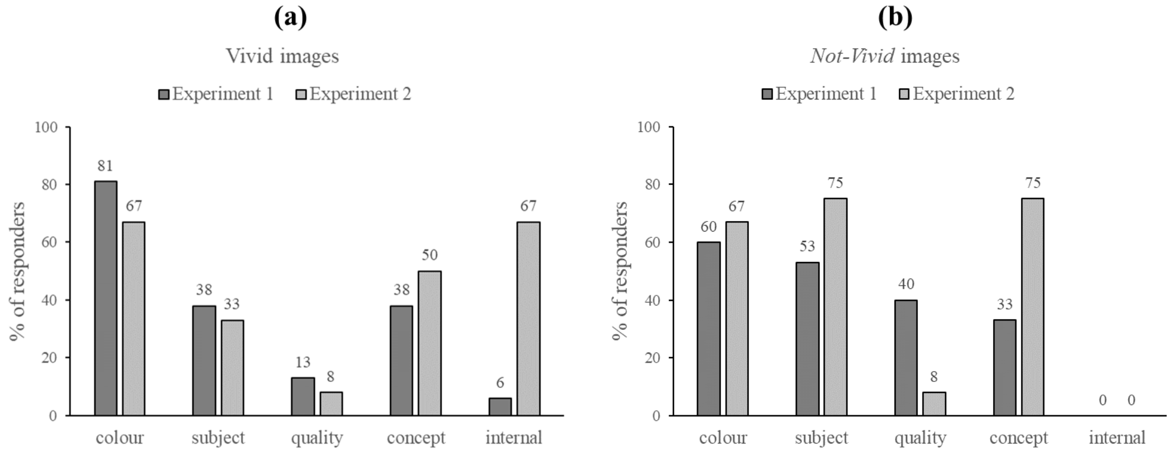

Figure 61.Vividness, encoding judgment, Experiments 1 & 2: illustrating themes used by participants to judge photographs as (a) vivid or (b) not-vivid. The numbers represent the percentage of responding participants who had mentioned the primary code or theme

Not-vivid images.

External themes of colour, subject, and image quality were all more commonly used for judging pictures as *not-vivid* when all the images were in full colour (Experiment 1), (Figure 61b). As found previously, internal codes were never used to judge an image as *not-vivid*.

Summary.

No differences were found between the experiments (picture sets), in terms of memory performance, or correlations. Despite vividness being higher on average for photographs when using a 240-image set with half partially desaturated for colour compared to the 120-image set all in full colour, both accuracy and confidence were lower. Correlations differed from those seen when vividness of memory for the pictures was assessed, in particular, the relationship between vividness and confidence was small to moderate, and that between vividness and source memory accuracy small or trivial.

(ii) Distinctiveness at encoding, Experiments 4 & 5.

Rationale.

Next, distinctiveness, Experiments 4 and 5 were repeated with the judgment of distinctiveness made at encoding, rather than at retrieval. This was also to confirm that distinctiveness of memory, and not of the image, had been judged in the previously reported experiments. The two separate list arrangements were used as before, and results compared between them. Additionally, these experiments might be useful to address the question concerning when distinctiveness is processed.

Procedure.

Procedures were followed as in Chapter 3: General Methods, and Chapter 5. Participants were asked to rate distinctiveness of the photographs during the encoding study block presentation (to assess distinctiveness when experiencing the stimulus), immediately after the photograph was first shown, Figure 62 (but not while still on the screen), by answering “How distinctive is the picture when you call it to mind?”

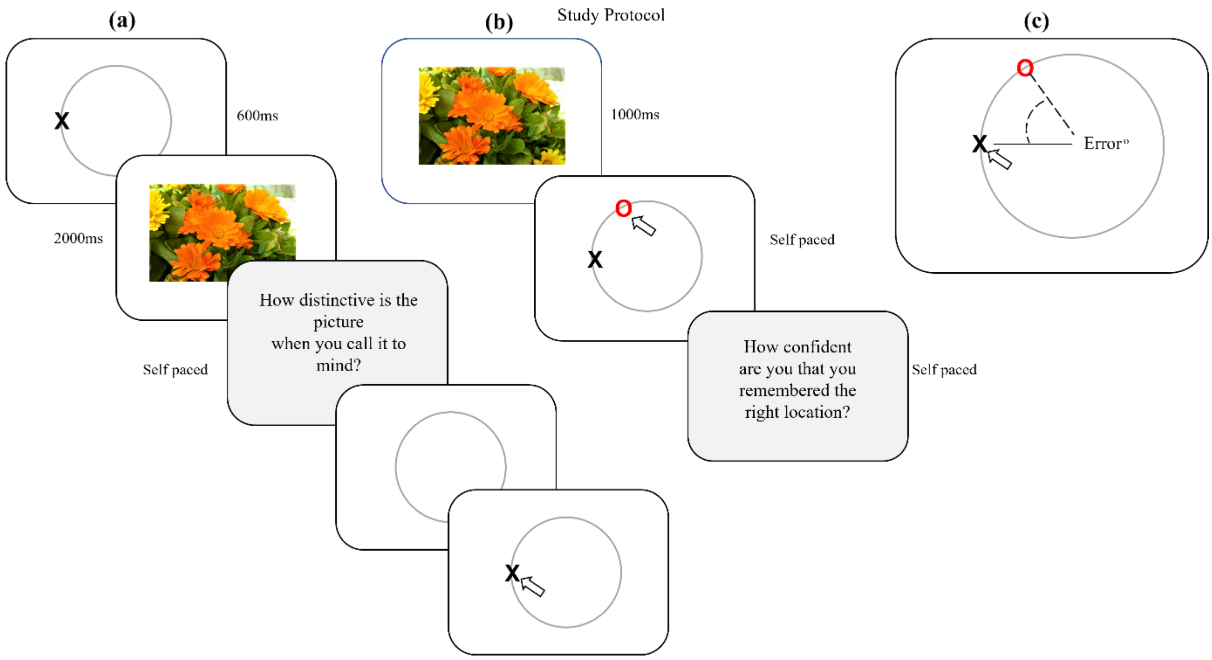

Figure 62. Distinctiveness, encoding judgment, Experiments 4 & 5: (a) Study (relational encoding) block; (b) Test (retrieval) block; (c) Measuring location (radial) error: the distance in ± degrees between the participant's response (shown as a red circle) and the correct target location (shown as a cross on the circle)

Results.

Participants

There were 20 participants in Experiment 4, of mean age 20.3years (*S.D.* 4.2 years);16 identified as female, all viewed photographs arranged in a mixed *list-type*, (Figure 44a). There were 20 participants in Experiment 5, of mean age 22.5 years (*S.D.* 5.4 years); 14 identified as female, all viewed photographs arranged in a homogenous *list-type* (Figure 44b). One participant was excluded from the data analysis as their location error frequency plots indicated that they could not or did not perform the task and were guessing throughout (see Figure 18b). Data from 39 participants were available (4680 trials). Analysis of results was carried out by comparing the two *list-types* (experiments).

Memory performance.

Unlike when distinctiveness of memory was judged during the test block, there were no significant differences in memory performance between the two experiments when distinctiveness of the images was judged at encoding, Table 16. To further illustrate the similarity in memory performance between the *list-types*, compare the location error plots in Figure 63, (cf. Figure 46).

Table 16. Distinctiveness, encoding judgment, Experiments 4 and 5: memory performance [recollection rate, scale (precision) and threshold (degrees from target), M (SE)], results of independent t-tests to compare mean values for all assessable participants (N = 39) by experiment (list-type)

Distinctiveness, Encoding judgment: (mixed *list-type*)
Distinctiveness, Encoding judgment: (homogenous *list-type*)

List-type	Memory Performance Rating	*n*	*M* (*SE*)	Independent samples *t*-test *t(d.f.)*	*p*-value 2-tailed
Mixed	Rate	20	0.49 (0.05)		
Homogenous		19	0.41 (0.05)	-1.15 (37)	.26
Mixed	Scale	20	12.1 (1.2)		
Homogenous		19	12.9 (1.1)	-0.54 (37)	.3
Mixed	Threshold	20	25.2° (2.8°)		
Homogenous		19	21.6° (2.8°)	0.9 (37)	.37

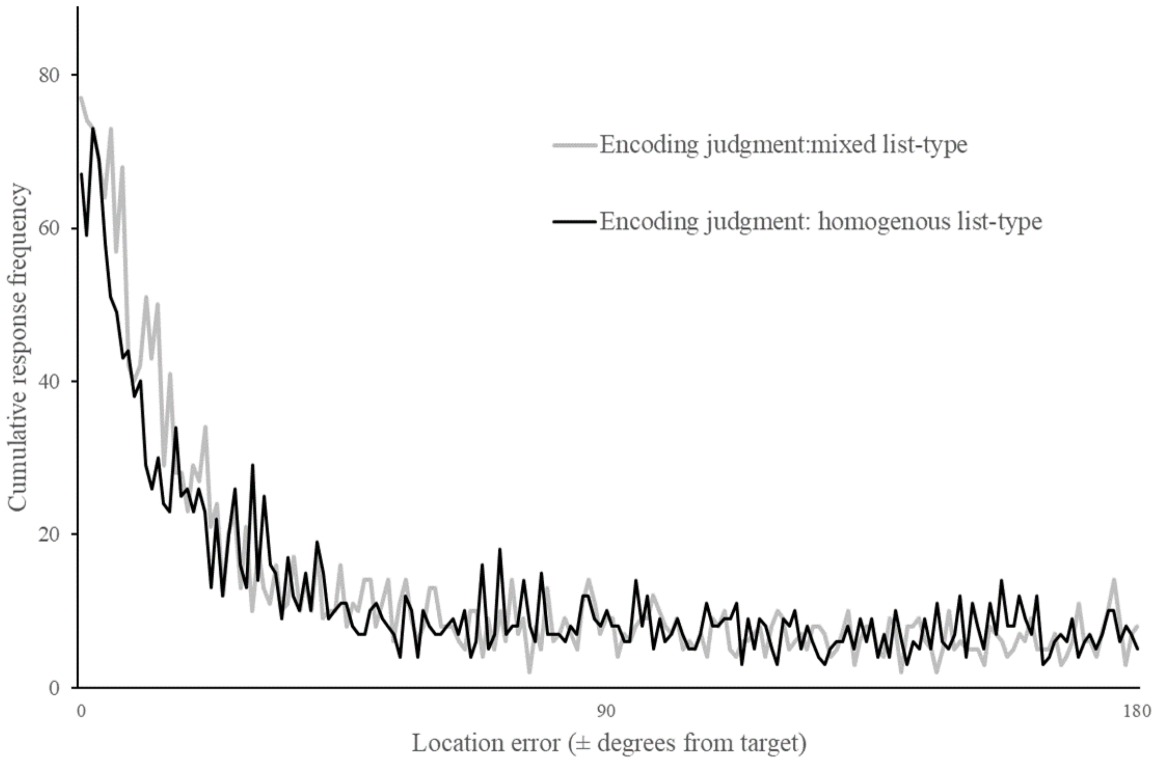

Figure 63. Distinctiveness, encoding judgment, Experiment 4 (mixed list-type) and Experiment 5 (homogenous list-type): location error cumulative frequency plots. In comparison with Figure 45 there are no apparent differences in recollection rate, threshold, scale (precision) or guessing rates between the two list-types

Variables of interest.

Distinctiveness was higher for images viewed in a homogenous list-type (Experiment 5), $p < .001$; the effect size was weak, (R^2 change = 1.5%), Table 17. (The same was seen for vividness, Experiment 2 using the manipulated set of images: although the higher vividness was significant, the effect size was trivial, see Table 16). Memory accuracy was higher for images presented in a mixed *list-type* (Experiment 4) compared to a homogenous *list-type* (Experiment 5); the effect size was trivial, (R^2 change = 0.3%). In contrast to judgments of memory distinctiveness made at test, there was no significant difference in confidence between the two *list-types*, (Table 17).

Table 17. Distinctiveness, encoding judgment, Experiments 4 & 5: Variables of Interest [M (SE)] by List-type, (N = 4680), compared by R^2 change

Distinctiveness, Encoding judgment: (mixed *list-type*)
Distinctiveness, Encoding judgment: (homogenous *list-type*)

List-Type	Variable of Interest	*df*	*M (SE)*	Unstandardized *ß* coefficient (*SE*)	*p*-value[a]	R^2 change
Mixed	Distinctiveness	1,4678	44.1 (0.4)			
Homogenous			49.8 (0.5)	5.75 (0.69)	< .001	1.5%
Mixed	Accuracy	1,4678	70.36 (0.6)			
Homogenous			67.2 (0.6)	-3.13 (0.88)	< .001	0.3%
Mixed	Confidence	1,4678	45.2 (0.6)			
Homogenous			44.3 (0.6)	-0.99 (0.86)	.25	-

Note. [a]The *p*-value may seem significant, but the number of trials is high. Of more relevance is its effect size, denoted by R^2 change: (1% = weak, 9% = moderate, 25% = strong effect), which is documented when the *p*-value is significant.

Correlations.

Correlations between distinctiveness and source memory accuracy were small for both experiments (*list-types*), Table 18. Correlations between accuracy and confidence were moderate to large, as had been the case throughout. In contrast to the very large correlations seen between distinctiveness of memory for the images and confidence, only moderate correlations were seen, ≤ 13.7% of the variance in confidence was accounted for by variability in the distinctiveness of the images when first experienced (at study). Results are confirmed by averaged within-participant correlations, Table 14. Appendix B details all individual results (Table 31S & 32S, Figures 87S-92S).

Table 18. Distinctiveness, encoding judgment, Experiments 4 & 5: Correlation Analysis, split by list-type, denoting relationships between Variables of Interest. Pearson's r-values are displayed using data from all trials and compared with r-values and R^2 variances obtained by averaging within-participant correlations. It is evident that there is little difference between the two methods

Distinctiveness, Encoding judgment: (mixed *list-type*)
Distinctiveness, Encoding judgment: (homogenous *list-type*)

List-type	I.V.	D.V.	*N*	Pearson's *r* (*all trials*)	*p*-value[a] 2-tailed	R^2 variance	Pearson's *r* (*averaged*)	R^2 variance
Mixed	Distinctiveness	Accuracy	2400	.11	< .01	1.2%	.14	2.0 %
Homogenous	Distinctiveness	Accuracy	2280	.08	< .01	0.6%	.11	1.2 %
Mixed	Distinctiveness	Confidence	2400	.37	< .01	13.7%	.36	13.0 %
Homogenous	Distinctiveness	Confidence	2280	.26	< .01	6.8%	.27	7.3 %
Mixed	Accuracy	Confidence	2400	.42	< .01	17.6%	.40	16.0 %
Homogenous	Accuracy	Confidence	2280	.37	< .01	13.7%	.36	13.0 %

Note. Cohen (1992) defined small, moderate (i.e., medium), and large *r* values as "about" .10, .30, and .50, respectively. R^2 represents the amount of variance in the dependent variable predicted by variability in the independent variable.

Qualitative data.

All 39 assessable participants commented on what made the images distinctive, with 35 also commenting on what made the images *not-distinctive*. Primary codes and themes used by participants to judge images at encoding as either distinctive or *not-distinctive* were identical to those used to rate memory distinctiveness at test (retrieval). Results are expressed as a percentage of those responding, Figure 64.

Distinctive images.

Overall, there was little difference between the experiments in the use of external codes to rate distinctiveness, noting that the identical images are shown in both list arrangements; image factors and concepts of distinctiveness were used more often to rate pictures in a homogenous *list-type*; and internal codes were used more often to rate images shown in a mixed *list-type*, Figure 64.

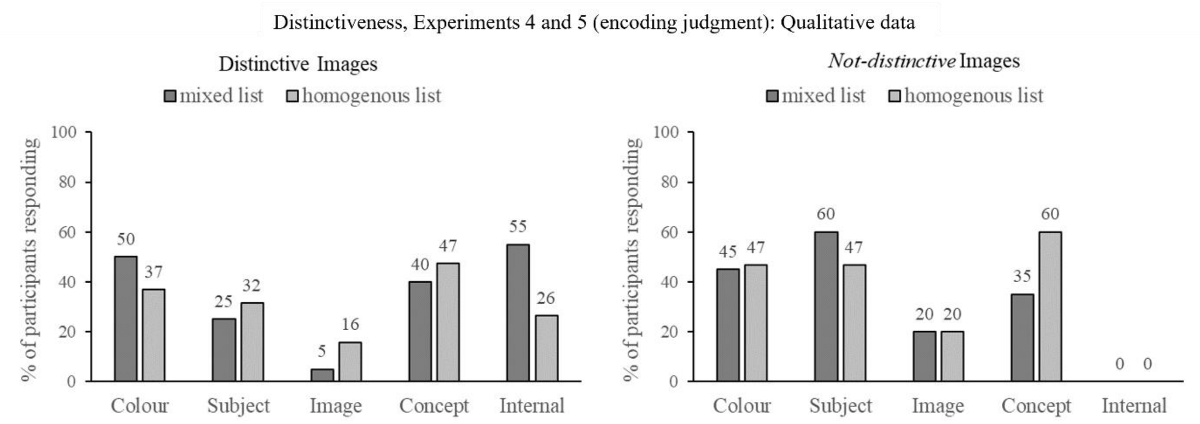

Figure 64. Distinctiveness, encoding judgment, Experiment 4 (mixed list-type) & Experiment 5 (homogenous list-type): qualitative data analysis, column charts showing codes and themes used by participants for judging images as distinctive or not-distinctive. The numbers represent percentages of the participants who responded

Not-distinctive images.

There was little difference between experiments in the use of external codes to rate images as *not-distinctive* (the external theme of subject was used more commonly for images shown in a mixed *list-type*); the participant's concept of what was *not-distinctive* was again used more often for images shown in a homogenous *list-type* to rate the images, Figure 64. Internal codes were never used.

Summary.

In these experiments, where distinctiveness was judged when the picture was first experienced (at study, encoding), in contrast to when judgments were made at test, there were no differences in memory performance between the two list arrangements for pictures. However, images viewed in a homogenous *list-type* produced inferior results for source memory accuracy, and despite average distinctiveness being significantly higher for the identical photographs shown in a homogenous *list-type* as compared to a mixed *list-type* there was no significant difference in confidence between the two arrangements. As seen for vividness rated at encoding, correlations between distinctiveness rated at encoding and confidence were only moderate; correlations between distinctiveness and accuracy were small. Despite these differences, the correlation between source memory accuracy and confidence remained moderate to large and was constant regardless of when the quality judgment was made.

From the qualitative data, participants used the same criteria on which to base their judgments of distinctiveness at encoding as they did for memory of the pictures at test.

Discussion of results.

Results from these experiments involving either vividness or distinctiveness judged at encoding, i.e., rated on the appearance of the picture when first experienced rather than its memory when shown again as a retrieval cue, show considerable disparity compared to the previously reported experiments. This is taken as confirming that participants were not simply judging how vivid or how distinctive the photographs were when seen again in the test block. There were no significant differences in memory performance between image sets (for vividness), or *list-types* (for distinctiveness), when judgments were made at encoding.

How vivid or distinctive a picture was considered when first experienced, had no significant correlation with source memory accuracy and had only a small to moderate correlation with confidence. Ratings of memory quality made at encoding had little or no effect on participants' accuracy of recollection for the cross location. Nevertheless, the correlation between accuracy and confidence remained moderate or moderate to large throughout all the source memory experiments, i.e., neither the timing of the quality judgment, the image sets nor the list arrangement of images, affected how participants rated confidence based on their factual accuracy. Participants were equally able to perform the task regardless of when quality judgments were made.

Nevertheless, the qualitative data showed that participants were basing their judgments of either vividness or distinctiveness on the same criteria whether this was for the photograph or for their memory of the photograph, using the same codes and themes. The finding (when rated at encoding) of significantly higher distinctiveness for images viewed in a homogenous *list-type*, and of higher vividness for images in a manipulated 240-image set to reduce visual vividness requires explanation, given the significant impairment of the same judgments when made for the memory of the photographs at test. It could be suggested that some images stand out from their category when shown in this context at encoding (when first experienced) and will be rated higher in vividness or distinctiveness. Bylinskii et al., (2015) termed such images as “contextually distinct”, i.e., distinct with respect to the set of images from which the experimental list arrangement is taken, referring to a large body of literature suggesting that items that stand out from their context are better remembered, although whether images ‘pop out’ when shown in this way is disputed (Bainbridge, 2020).

Finally, I consider the debate as to whether distinctiveness is processed at encoding or at retrieval. It is concluded that distinctiveness effects in episodic memory do not rely on the processing of distinctiveness at encoding. To be clear, this holds for the present experimental paradigm only; it is undetermined whether in an isolation paradigm recognition memory task, distinctiveness has an impact at encoding or at retrieval. Judgments of distinctiveness are of primary importance at retrieval in respect of their association with confidence in memory accuracy for contextual detail (the cross location). If a stimulus is considered distinctive when remembered, although this does not necessarily mean that contextual detail will be accurately remembered, it does mean that the individual will be confident that their memory is accurate.

Chapter 8: Discussion, Conclusions & Recommendations.

What confirms confidence in memory's accuracy? To test my hypothesis that we base our belief in a memory's accuracy on the quality of the experience of remembering, across two different experimental tasks I provide conclusive evidence that the memory's quality is more important for our belief in its accuracy than is its factual accuracy. Belief in a memory's accuracy does not guarantee accurate recall of associated contextual detail (source memory); remembering predicts accurate recognition only if confidence is very high, when it is the quality of the memory which drives confidence.

Summary of the findings.

In a continuous-response source memory recollection task, participants studied real-world natural scenes colour photographs paired with the location of a cross on a circle, rating either the vividness or the distinctiveness of the photographs immediately after they were shown again at test, to cue memory for the cross location. Participants were then presented with a blank circle and were required to indicate where they thought the cross had been positioned, then rate how confident they were that they had remembered the location correctly. The correlation between vividness or distinctiveness of memory for the photograph and participants' actual accuracy for recollection of the cross location was moderate. While the correlation between the participants' actual accuracy and confidence was moderate to large throughout the experiments, importantly, the correlation between vividness or distinctiveness of memory for the photograph and confidence in their memory for recollecting the correct cross location was very large; this remained the case despite manipulation of the images to be less visually vivid by partially desaturating half of them for colour, or less distinctive by arranging them in homogenous lists comprising pictures from a single natural scenes category. Both judgments of memory quality positively moderated the relationship between accuracy and confidence, increasing or decreasing confidence with no change in the level of accuracy. We assume that our vivid or distinctive memories are accurate; however, the relationship between these aspects of memory quality and accuracy of memory acts largely indirectly, mediated through confidence.

Once the experiment was finished, participants were asked for comments on what their judgments of vividness or distinctiveness were based. Although no instructions were given as to what these judgments meant, participants instinctively used external features of the pictures: how colourful, bright, and clear the images were, or what the subject of the picture was (i.e., pictures of flowers compared to forests). Photographs also triggered the spontaneous recall of personal memories or associations, increasing their vividness or distinctiveness. For both judgments, the participant's concept of what vividness or distinctiveness meant to them was used in their rating of the pictures as being vivid, distinctive, or not.

A very large correlation between vividness and confidence was reproduced in a separate recognition memory task. Participants studied a series of photographs, shown singly, some of which (old pictures or targets) were shown again at test, paired with a new (not previously seen) picture. Participants were required to indicate which photograph had been seen before; then rate their memory of the photograph for vividness; how confident they were that they had recognised the photograph correctly; and answer whether they had remembered the picture. High vividness and high confidence decisions were more likely to be associated with remembering the picture than low confidence and low vividness recognition decisions. Remembering the picture moderated the relationship between vividness and confidence, i.e., if remembered, confidence was higher for the same level of vividness. Recognition accuracy (whether the picture recognised was the target (correct) or the paired new photograph (incorrect) was more loosely linked with confidence and vividness ratings; accurate recognition of the target could be associated with any level of confidence or vividness. In a post-test questionnaire, participants who did not remember the picture based their decisions on its familiarity, a guess, or on not recognising the other paired picture, more than simply just knowing the picture had been seen before.

Importantly, if the paired new photograph was similar (but not the same as) a picture that had previously been seen at study, participants were liable to incorrectly recognise the new rather than the old photograph (which was the same picture they had seen before). If participants also remembered the picture, they were more confident but less accurate than if they had not remembered the picture recognised as the target. In effect, participants were misremembering a similar new picture as one they had seen before at study.

Using the same experiment, it was possible to explore the potential interaction of the processes of recollection and familiarity in the task of recognition. Participants rated familiarity of the images recognised as old. In many such experimental tasks the participant is forced to choose between recollecting the picture or if the picture was familiar (but would be unable to choose both, i.e., to say they had recollected the picture as well as thinking it familiar). Rather than just answer yes or no to a "was it familiar" question, I asked participants to rate the strength of familiarity (as for confidence) on a 0 – 100 scale. There was a very large correlation between familiarity and confidence that they had recognised the old picture correctly. High familiarity and high confidence were strongly associated with recollection of a picture; low familiarity and low confidence recognition decisions were more commonly associated with not recollecting (remembering) the photograph. Importantly, my results demonstrated a trade-off between the two processes, information from recollection was more important for judgments of low familiarity, and less important for judgments of high familiarity. Furthermore, recollection moderated the relationship between familiarity and confidence, increasing levels of confidence for the same value of familiarity, if the picture was also recollected.

Interpretation of findings.

Confidence & accuracy.

Despite theoretical convention that confidence and accuracy are strongly related, there is empirical evidence that this may not always be the case. I wanted to explore why we could be confident in a memory which was not accurate and believe in vivid memories for which the detail was lacking or inaccurate. Using a source memory (recollection) paradigm as well as an old or new 2-AFC recognition memory task, results consistently demonstrated that our judgments of belief in a memory are based on memory quality, i.e., on its vividness or distinctiveness, more than on its accuracy. However, this is not new knowledge, and is not so surprising (Gallo, 2013; Scimeca et al., 2011; Wong et al., 2012). If our memory is vague (i.e., not vivid, or not distinctive) we are less inclined to believe it – "I think that happened, but I'm not sure." If the retrieved memory is clear (i.e., it is vivid or distinctive) we are most likely to believe it – "that must be true, the memory is so clear, I remember it vividly." The quality of the experience of remembering influences the participant's confidence, yet they are unaware of the fact (Chandler, 1994).

Both metacognitive judgments of memory quality demonstrate very large positive correlations with confidence in the accuracy of the memory, greater than their correlation with memory's accuracy. Nevertheless, I would not dispute strong theoretical views held elsewhere, that there is a significant correlation between accuracy and confidence, but I have an open mind as to whether this is incontrovertible. Although correlation between source memory accuracy and confidence remained moderate or moderate to large throughout the experiments, high confidence is no guarantee of high accuracy; low confidence does not always indicate inaccurate recollection of the associated memory detail. Moreover, we are aware from our personal experience that our memory can be fallible.

Analytical methods.

Critically, methods of analysis and data presentation can have a significant impact on their interpretation. The difficulties of adequately representing accuracy in a binary old or new recognition memory task are not glossed over. The use of ROC curve analysis is questionable. Unless a serum biomarker for confidence is developed, this type of analysis is inappropriate and may be flawed, e.g., in the way that confidence is used for calculations as a proxy for memory strength, particularly if other factors also influence confidence ratings. Alternatives such as CAC analysis, involving binned and averaged data may well be apposite for recognition memory tasks (Tekin et al., 2021), but the way bins are arranged can lead to misleading conclusions (particularly if later re-arranged to make interpretations stronger).

My work could be dismissed for reasons of being based on “inappropriate” analyses. Arguably, I have used analyses which present data from all trials carried out rather than averaging data for each individual participant. Here, data from all trials is presented, this is even possible for a recognition memory task. The potential effect on how the findings from a correlation analysis may be interpreted, based on whether or not the results are averaged, was highlighted above. When I initially analysed some of the work on vividness, I followed the general practice of averaging and binning data (admittedly, unnecessary for tasks where accuracy is not a binary outcome), which led me to convincingly report very large correlations between confidence and accuracy, confidence and vividness, and vividness and accuracy, thus not upsetting the status quo. More considered analysis of the data revealed that only the relationship between confidence and vividness was very large, and graphs based on binning variables into decile[18] responses plotted against averaged data for each bin can readily lead to misleading data interpretation. Nevertheless, this practice does represent how the CAC (calibration) plots are calculated, my excuse for this being to compare like with like

[18]In descriptive statistics, a decile is any of the nine values that divide the sorted data into ten equal parts, so that each part represents 1/10 of the sample or population

for results between test picture-pair conditions in the recognition memory task, when there was little alternative given the nature of the results.

Memory quality.

Everyday experience tells us that an individual recounting a vivid or a distinctive memory may have an unshakeable belief that the detail of the memory is correct, and this lay understanding of memory can have significant impact. The unambiguous evidence from the source memory experiments shows that variability in confidence is strongly linked to both vividness and distinctiveness; suggesting it is the quality of memory, or the rich experience of remembering, which is important for our belief that the exact detail of the memory is accurate.

In the natural scheme of things, and when the participant is left to determine the vividness or distinctiveness of a stimulus, these judgments of memory quality are of primary importance at retrieval, in respect of their association with confidence in source memory accuracy (recollection of the cross location). If memory for a previously seen photograph is considered vivid or distinctive, although this does not necessarily mean that contextual detail will be accurately recollected, it does mean that the individual will be confident that their recollection is accurate.

The fact that vividness and distinctiveness were less strongly related to the accuracy of source memory in these experiments, provides a potential reason for dissociation between confidence and accuracy, from an experiment not based on the DRM memory illusion or on picture similarity experiments using deceptive lures. Highly vivid or distinctive memories inflate confidence more than they do accuracy, as confirmed by the moderation analyses, leading us to mistakenly believe that the remembered detail of the memory is more accurate than it actually is, i.e., an illusion of memory accuracy, explaining why highly confident memories can sometimes be unreliable.

Repeating the source memory experiments with a judgment of memory quality at encoding confirmed that it is the vividness of a memory which drives confidence, and not the vividness of the experience itself, i.e., when judged at encoding. This echoes results from the flashbulb and false memory literature where vivid events are not remembered accurately, but vivid memories of the events are remembered confidently (Coan,1997; Doss, Picart & Gallo, 2018;2020; Felstead & French, 2021; Roediger & Gallo, 2004; Talarico & Rubin, 2003; 2007). Results also suggested that distinctiveness is processed at retrieval and relates to the memory of an experience rather than the perceived distinctiveness of an event when experienced.

Moderation and mediation.

To answer what is new about my research findings, I highlight results of the moderation and mediation analyses. The implications of these findings, and the strong correlation between memory vividness and confidence, for theoretical accounts of the confidence-accuracy association, provide evidence for their dissociation and may help to explain why, when our recollected experience is vivid, we are confident that the detail of the memory is correct.

Although it was proposed in 1990 that distinctiveness (of faces) might moderate the relationship between confidence and accuracy, Brigham's study predated current methods of moderation analysis. That both memory quality judgments moderated the relationship between accuracy and confidence is an important finding. Mediation analysis demonstrated the relationship between memory quality judgments and source memory accuracy to operate largely indirectly, mediated through confidence, there being only a trivial direct association between the quality of a memory and its accuracy. Our belief in the accuracy of a vivid memory is not due to the accuracy of its recollected detail; it is due primarily to the quality of remembering increasing our confidence that the memory is factually accurate. One could speculate that accuracy increases confidence, and that confidence is responsible for the perceived quality of memory for the picture; this hypothesis was tested but was not supported by the results.

Implication of the findings.

One practical implication of these findings relates to the manipulation of how to-be-remembered information is presented. Our instincts might be to use vivid colours, or contrast, and distinctive lettering, to ensure accurate recollection of the information detail. These results indicate that such manipulations may prove futile, as vividness or distinctiveness of a stimulus at encoding did not relate to either confidence or source memory accuracy. For example, the vividness of an advert, its distinctiveness, or its familiarity due to repetition at each advertising break in a TV programme, might result in better recognition, however, we may confidently but inaccurately recollect the exact details of the advert.

On a separate issue, recollection of real-life personal events, rich in meaning, is speculated to produce memory errors "because of the reconstructive processes involved in their retrieval," (Koriat et al., 2000). These "memory errors" may be produced because vivid recollection of personal events inflates confidence and causes confidence-accuracy dissociation. False memories may be viewed as a conscious fabrication by individuals or witnesses. Illusions of memory accuracy are genuine and unconscious misinterpretations of the truth of remembered details resulting from inflated confidence due to the rich quality of remembering, i.e., the vividness or distinctiveness of the remembered event.

A further explanation for memory errors may lie in the impaired performance resulting from the homogenous list arrangement used for distinctiveness in Experiment 5, which did not affect the very large correlation between distinctiveness and confidence. This can be hypothesized as representing inferior recall of similar events; we remember the unusual, not because these events are more deeply encoded but because we have inferior recall for the less unusual events (Hunt, 1995; 2006).

"We do not remember days, we remember moments. The richness of life lies in memories we have forgotten." (***Cesare Pavese,*** *1908 – 1950).*

The richness of life may instead lie in those remembered moments, the more distinctive or standout events in our day-to-day lives that we recall better than the more mundane or routine events that occurred around them on the same day. Rather like watching a football highlights programme on TV, we are shown only brief clips of the interesting events from a match, the moments (goals, tackles, and controversies) with the routine run of play erased. When watching such matches in real-life, likewise we forget the routine and remember the distinctive.

Recognition memory tasks.

Speculating as to whether source memory tasks are in fact distinct from recognition tasks, I point to the similarity in results for the very large correlations between vividness and confidence in both the continuous response source memory task and the 2-AFC recognition memory task. The natural scenes colour photographs used as retrieval cues in the source memory tasks can either be remembered (recollected) or recognised, (or both). If the retrieval cue picture is not recognised it will not cue recollection of source memory; if the natural scenes picture is recognised or recollected it will cue recollection of the associated detail (source memory), bearing in mind that the associated detail may be incorrect, even if confidently recollected.

In a recognition memory task, it is generally accepted that when a remember or 'know' question is asked, 'know' judgments can sometimes be based on familiarity, or low confidence guesses. Additionally, a remember or 'know' response has been hypothesized due to participants judging their confidence in memory for recognising the target (Dunn, 2004), leading to the conjecture that remember and 'know' responses are both based on confidence judgments, with remember being associated with high confidence.

When confidence is high.

When participants recognise targets, you can check their accuracy, but you cannot check the veracity of their remembering or recollection and have to take the participant's word for it. As shown in the 2-AFC recognition memory task (Experiment 3), when a target is recognised and vividness is high, confidence is high, and the target is more likely to be remembered. If vividness is low, confidence is low, the participant is less likely to remember the target. So far, this is in line with established theory, that a remember response in an old or new recognition memory task is associated with high confidence, but here it is also associated with high vividness, i.e., memory quality. However, despite the relationship between confidence, remembering and vividness being so obvious in the correlation scatterplots, the accuracy of recognising the target is shown to be associated with any levels of vividness and confidence.

CAC plots shed more light on the relationship between confidence, remembering, and recognition accuracy (proportion correct). In all but the mnemonically similar test pair conditions, remembering is associated with increased accuracy. However, this is only the case for the highest bin of confidence. The association between remembering and confidence only translates to accuracy when confidence is very high. This is well-known to some research groups, who tailor their analyses to focus only on high confidence judgments and dismiss the rest of the data as irrelevant (e.g., Ingram et al., 2012).

My results provide confirmation that remember responses are strongly associated with confidence, but that confidence judgments are based on the quality of remembering, rather than on the memory's strength, commonly represented as accuracy (of recognising the target). Here, it is also demonstrated that remembering a recognised image moderates the relationship between vividness and confidence, differentially increasing confidence for the same level of vividness. Although this might suggest that remembering increases confidence rather than being based on judging confidence, I will again point out that moderation analysis only confirms the association and not causality or the direction of the association.

Results from the post-test questionnaire revealed that ‘know’ was less commonly associated with a recognition decision when the target was not remembered than were guesses, decisions based on familiarity, and more commonly, those based on not recognising the other picture. This would equate with the previously discussed 1 – 20 confidence scale, where 1 indicates “definitely not on the list” and can be given with high confidence and accuracy (even if later represented as a low confidence judgment).

The picture similarity task.

The picture similarity task was used in the early 1980s by Tulving to show dissociation between confidence and accuracy and to provide a theoretical framework for episodic memory. Here, there was also a clear indication of an underlying cause for dissociation between confidence and accuracy for perceptually similar lures as compared to mnemonically similar lures. Whilst accuracy is increased for perceptually similar lures due to prolonged decision times (also reducing confidence), reduced accuracy and increased confidence for mnemonically similar lures is theorised as due to confident misremembering of an image similar to one that has been seen before, and (as for the DRM memory illusion), provides a further example of an illusion of memory accuracy. We may be at risk of misremembering events if they are similar to other events held in memory. Gonsalves and Paller, (2000) demonstrated a similar effect resulting from visual imagery, where imagined objects could be recalled as being seen as a photograph at study. Neither situation is exactly described as “remembering events that never happened” and has a parallel in everyday experience:

Your partner enquires, where did you leave the car keys? Your confident response is on the hall table. But the keys are not there, and you are confident you left them exactly in that place, somebody must have moved them (your partner). An argument ensues because your partner is equally adamant, they did not touch them. The keys turn up but happen to be on the kitchen table; your confident memory is shattered. Then you notice the cat in the kitchen, and it comes back to you, you had to let the cat out when you entered the house and put the keys on the kitchen table on the way to the back door, instead of leaving them where you normally leave them, on the hall table. You apologise, but you are bemused by your confident yet distinctive memory. As you always leave your keys on the hall table, you remembered a previous experience of doing just that and it was that experience you recollected rather than the one where you left them in the kitchen. The literature is ambiguous as to whether this would be considered a false memory, an illusion of memory accuracy, or misremembering.

Such illusions of inaccurate remembering may relate more to how the neural networks, memory traces or engrams within the brain encode the memory for routine events as gist rather than detailed memories of every single moment of our lives. Considering diffusion theory, when searching for a match to the memory trace, perhaps the process stops when a neural network or engram provides a reasonable match for the memory's gist. This reprises the introduction to the thesis, and the fact that the lay conception of memory may be far from the truth (Brewin et al., 2019; Roediger & McDermott, 2000). We may all suffer from inferior recall regarding the humdrum events of our lives.

Familiarity & recollection.

In dual-process theory recollection and familiarity are independent memory processes involved in a recognition decision, ('know' is theorised as based on familiarity and is not considered a memory process). but some maintain they do not act together. Others suggest that participants can opt to rely on whichever is stronger, (Cha & Dobbins, 2021). The question asked may only give you the result you are looking for, if a participant cannot complete the task without choosing one or the other (did you recollect it or was it familiar?), how are we able to demonstrate these processes might act together? Asking a "how familiar" question, rated on a 0 –100 scale, with a separate recollect yes or no question, allowed participants to choose both or neither. My results provided no clear support for an "either familiarity, or recollection" dual process theory (Yonelinas,1999; Yonelinas et al., 2010).

On asking if recollection and familiarity are separate processes or a single process (Slotnick, 2013) concluded that behavioural evidence pointed to a single process. The behavioural evidence collected here suggests that they are separate and independent processes. Familiarity and recollection were both used by participants; recollection moderated the relationship between familiarity and confidence, recollection increased confidence for a given level of familiarity. As illustrated by calibration plots based on familiarity bins rather than confidence bins, if familiarity was high, the additional information provided by recollection was less important for improved performance. When familiarity was low, the additional information provided by recollection was of more importance for improved performance, i.e., participants appeared to be using both processes, but not a single process.

The time to make a recognition decision (response time) was significantly faster if the image was recollected. See Ratcliff et al. (2016), for further discussion of diffusion theory and accumulation of evidence based on response times. Our recognition decisions may be based on a quality heuristic due to either fluency (familiarity) or response time (recollection); the processes of recollection and familiarity thought to be complementary not additive.

There was a very large correlation between familiarity and confidence, which corresponds to findings for both memory strength and recollection strength (e.g., Cha & Dobbins 2021; Wixted & Squire, 2011). It could be suggested that participants who confidently recollected the recognised picture were more likely to rate familiarity as high, one judgment influencing the other. Scatterplots further illustrated the relationships between familiarity, recollection, and recognition accuracy. Results indicating that confidence is strongly associated with responses when recollection was associated with high familiarity. When the image was highly familiar, participants were also highly confident, and were much more likely to have recollected the image; if the image was not familiar, participants were not confident and were less likely to have recollected the image. In contrast, performance in the task (accurate recognition) was associated with any level of familiarity or confidence.

Qualitative data.

Qualitative data was collected in the source memory experiments and demonstrated that the external properties of the image (colour, image quality, and subject), constituted a sizeable proportion of participants' responses. The spontaneous recall of personal memories, (internal vividness) was only used to judge images as vivid; this is unremarkable, and well-known (Ford & Kensinger, 2016). Internal vividness is represented by the spontaneous recall of vivid personal memories and not by the spontaneous recall of mundane *not-vivid* memories. The findings are redolent of an earlier recognition memory study from the end of the 20th century (Gardiner et al.,1998), in which reasons offered by participants for remembering a cue word included self-related involuntary conscious memory (where individual experiences came to mind with no conscious effort on the part of the participant). In the studies presented here, all the images were novel to participants, but nonetheless triggered spontaneous or automatic awareness of personal memories.

Participants' concepts of what was *not-vivid* corresponded with concepts of similarity and lack of distinctiveness. While the same factors were true of the criteria participants used to judge distinctiveness, similar findings to the vividness data are reported here, i.e., the distinctive recall of personal memories being used in addition to external factors and participants' concepts of what distinctiveness meant to them. Evidence from the qualitative analysis confirms the overlap between factors important for both image memorability and image vividness or distinctiveness. When images contain only one or a few objects (e.g., the stand-out features mentioned by participants here as signifying vividness or distinctiveness), it has been shown in image memorability literature that pictures are more likely to be both memorable and distinctive. When vividness or distinctiveness were judged at encoding, the same criteria used for judging memory for the images when seen at test were also used to judge the quality of the photographs when seen at study.

Appendix A takes a more detailed look at the images selected as highest or lowest in quality, and those pictures associated with highest or lowest source memory accuracy, allowing for direct comparison between the image sets used and the judgments of memory quality.

Limitation of the Studies.

Using real-world natural scenes colour pictures in all experiments could mean results might not generalise to other stimuli, such as words or faces, but has the advantage of allowing results to be compared across experiments without the confound of using different stimuli. Changing everything at once is not helpful in understanding which change is significant. Admittedly, some photographs are inherently vivid or distinctive and may be less affected by the context in which they are presented (see Appendix A).

Participants were selected from a population of psychology undergraduates. The limitation is ubiquitous and may affect generalisation of results, while avoiding confounds of age and educational attainment.

The continuous-response source memory task (initially used for testing a some or none account of recollection) and the 2-AFC picture similarity task (involving deceptive lures) can both be criticised as inappropriate. The benefits of the tasks, firstly in providing fine-grained detail for source memory accuracy and secondly for allowing evaluation of the mnemonically similar lures (with little significant difference between the three test picture-pair conditions), I think outweigh these criticisms.

More specifically, I have attempted to address further potential issues for the source memory experiments. Firstly, that vividness or distinctiveness of the retrieval cue picture and not of source memory was assessed, relating to the assumptions made about the cross and picture being encoded as a unified memory trace. Secondly, that participants were merely rating vividness or distinctiveness of the pictures when shown at test rather than their memory for the pictures, relating to repeating the experiments with judgments made at encoding. Finally, that mental imagery rather than vividness of memory for the photographs was used to judge vividness or distinctiveness, relating to judgments being made immediately after the images were shown at test, but not while still on screen, assuming the image was held in short-term visual (working) memory.

For source memory tasks, it is a not uncommon procedure to show either objects or faces on a background natural scenes picture and for the participant to be shown the face to recall the background picture, or the picture to recall the object's location and orientation. An alternative would be to show the natural scenes pictures on different colour backgrounds at study (to act as source memory) or use different locations, i.e., the picture is shown in different quadrants of the screen at study, and when the picture is shown in isolation at test, asking what colour the background had been or where the picture had been located. That the colour of the background may be important in such tasks is confirmed by participants' comments that red, yellow, orange, pink, and purple colours were vivid or distinctive; green, brown, or grey colours were neither vivid nor distinctive. This could lead to confusion between the background and the superimposed picture for participants' ratings of vividness or distinctiveness. Both applications are limited by the lack of detail for source memory accuracy.

2-AFC tasks provide “relative” rather than “absolute” measures, compared to a standard old or new recognition memory task. Whether or not this is a significant limitation in the context of my experiments is debatable, although this point might be raised when results contradict strongly held views. Here, results were not compared between different recognition memory tasks and correlation should not be affected.

Doubt could be cast on the validity of the results by challenging the order in which behavioural questions were asked, whether answers to initial questions would potentially bias later responses. Even if the order of questions were changed (or participants allowed to choose which they answer first) the same criticism could be made.

The 2-AFC tasks were carried out online because of Covid-19 restrictions. The conspicuous presence of anchoring, not apparent in the source memory experiments may be attributable to this, which did not allow for instruction or familiarisation by means of a practice session and personal supervision by the researcher. If there is no opportunity to establish a rapport between participant and researcher, participants may not wholeheartedly engage with the task.

Recommendations.

The recognition memory task could be repeated using a traditional old or new design, to see if results replicate. It would be useful to repeat the recognition task using the memory quality judgment of distinctiveness, particularly to compare blocks of mixed or homogenous picture arrangements, possibly after an interval had elapsed (hours or days), to test the effects of inferior recall on recognition memory.

I would be interested to further investigate the theoretical account of inferior recall for the routine events of our lives highlighted by the pattern of results seen in the distinctiveness manipulation for source memory. Research should ideally involve a population whose lives are regimented, such that every day is the same. Examples of such groups include long stay hospital inpatients, inhabitants of residential homes and nursing homes, and the prison population or those held in detention such as migrants or refugees in the UK. Alternatively you could use a population who were isolating during the Covid-19 pandemic. The ethical issues would depend on whether cooperation could be ensured and what was being requested, i.e. a formal experimental protocol or based on autobiographical recollection (itself problematic if there is a lack of verification for recalled events).

Although it might be an interesting exercise to see if results replicated when participants are asked to judge the memorability of the pictures rather than their quality, this brings its own problems, if you neither recognise nor recollect the picture, how can you rate its memorability? For the source memory task, judgments could be made at encoding, but it would not be possible to show correlation of memorability scores with either accuracy or confidence.

The findings of high familiarity associated with high confidence in Experiment 6 might merit further investigation. As theoretical accounts of déjà vu propose this phenomenon may occur when a recollected memory is associated with strong feelings of familiarity, participants could be asked whether they experienced such feelings during the experimental procedure, to see whether a correlation between familiarity and déjà vu does in fact exist (cf. O’Connor & Moulin, 2013).

A real world paradigm could also be used to see if the familiarity results replicate. For a recognition memory task, photographs of everyday objects or scenes that undergraduates might encounter about the University could be used, taken from the viewpoint of someone walking around the University campus. To make sure that participants had most likely seen all of the scenes or objects, these would have to be from the most common places used by all students (accepting that some may not be recognised). At test pictures could be shown interspersed with similar images from other (unknown and unfamiliar) campuses, with participants asked if they recognised the scene or object, whether they recollected it, how familiar it seemed, and how confident they were. This type of paradigm could potentially be used for different age groups or populations.

Conclusions.

I set out to explain why we can be very confident that we remember an event well (i.e., accurately), even when the details are incorrect or lacking. My research was designed to fill a gap in the literature concerning whether judgments of memory quality can influence the relationship between confidence and memory accuracy. The results show unequivocally that we judge belief in our memories, not necessarily because we are aware of the memory's accuracy, but because of the memory's quality (its vividness or distinctiveness), which both drives confidence and moderates the confidence-accuracy relationship. The quality of remembering inflates confidence for a given level of accuracy. Vivid or distinctive memories may be confidently recalled even if the detail is incorrect; accurate memories may be non-believed because the memory lacks quality. The fact that we regard our vivid or distinctive memories as being accurate is mediated through confidence in the memory's accuracy. Memory quality affects confidence more than it affects accuracy, but the individual is not consciously aware of this.

Performance in the source memory task was significantly impaired when all pictures were from the same subject category. Relevant to our everyday experiences, our past (autobiographical) memories may consist of the distinctive rather than the mundane or routine events. Results from the familiarity experiment are considered consistent with a heuristic for recognition memory based on the fluency of familiarity and the speed of recollection.

Accounts of a very strong and inviolate association between confidence and accuracy in reports of eyewitness testimony can be challenged. Improved performance in the recognition memory task (proportion correct) was only seen at high levels of confidence. Vividness enhances the experience of remembering, thereby increasing confidence in recognition memory accuracy. Confident but incorrect recognition of similar lures as old provides a reason for dissociation of the confidence-accuracy relationship in a picture similarity task, a possible explanation for misremembering or confusing details from similar past experiences and argues against the use of similar fillers for eyewitness identification.

Our personal memories may be more fallible than anticipated. If the distinctive can be confidently but inaccurately recollected, the mundane may equally be misremembered. The findings have implications for accounts of eyewitness testimony and dual process theory and are suggested as a framework for the investigation of the déjà vu phenomenon, false memories, illusions of memory accuracy and misremembering.

Appendix A.

Supplementary data

Comparison of memory quality judgments.

From the results of the source memory experiments, there appeared to be overlap in what participants considered to represent the two memory quality judgments of vividness or distinctiveness used in my studies. This might be due to participants concepts of what the two quality judgments meant, as they were not given pre-instruction on this. Alternatively, it could be suggested that it is the memorability of images that are dictating the memory quality judgments, and results might not replicate with a different set of natural scenes photographs. The images could be recognised due to distinctive features being recollected, or due to the general familiarity of the scene (Reinitz et al. 2011) and these factors could become either more or less important depending on the pictures used. However, the pictures were randomly selected depending more upon the subject category of natural scenes required, rather than having a standout feature or being familiar.

Additionally, results for vividness or distinctiveness were equivalent in terms of correlation, moderation and mediation analyses. As vividness, Experiment 1 uses the same set of images as distinctiveness, Experiments 4 and 5, and as Experiment 2 used the images, but with the addition of partially desaturated pictures, this allows for direct comparisons. I would suggest that context (the list arrangement) would have the same effect on memory performance if participants were judging the photographs for vividness in Experiment 5, rather than distinctiveness, when the images were arranged in a homogenous *list-type*.

The five images that were rated as the most vivid or distinctive from each of the experiments are illustrated in Table 19. Although vividness and distinctiveness could be regarded as different aspects of memory quality perhaps a single metacognitive judgment of memory quality should be used instead (unfortunately, instruction would need to be given, and this would probably refer to both vividness and distinctiveness). It can be readily seen that some (but by no means all) images appear in more than one list; these could be considered to be memorable images, which stand out from their class or category, (e.g., Bylinskii et al., 2015). As mentioned by participants, in the qualitative data, standout features in the foreground, bright colours, flowers, or just attractive pictures are apparent.

Table 19S. Illustrating the five most vivid or most distinctive pictures rated in each of the experiments: vividness, Experiments 1 (120-image) & 2 (240-image); distinctiveness, Experiments 4 (mixed list-type) & 5 (homogenous list-type)

VIVID 120-image					
Vividness rating	1	2	3	4	5
VIVID 240-image					
Vividness rating	1	2	3	4	5
DISTINCTIVE Mixed list 120-image					
Distinctiveness rating	1	2	3	4	5
DISTINCTIVE Homogenous list 120-image					
Distinctiveness rating	1	2	3	4	5

In contrast, Table 20 illustrates the images rated as lowest in either vividness or distinctiveness. The similarity of the lowest rated pictures is evident and reflects the qualitative data, where participants commented that trees, forests, and images showing expanses of sea or sky, or those that were similar, were rated as *not distinctive* or *not vivid*.

Table 20S. Illustrating the five least vivid or least distinctive pictures rated in each of the experiments: vividness, Experiments 1 (120-image) & 2 (240-image); distinctiveness, Experiments 4 (mixed list-type) & 5 (homogenous list-type)

VIVID 120-image					
Vividness rating	116	117	118	119	120
VIVID 240-image					
Vividness rating	236	237	238	239	240
DISTINCTIVE Mixed list 120-image					
Distinctiveness rating	116	117	118	119	120
DISTINCTIVE Homogenous list 120-image					
Distinctiveness rating	116	117	118	119	120

Next, it might be asked whether images highest or lowest in vividness or distinctiveness were also those associated with the highest or the lowest source memory accuracy (they might be expected to differ, given the only moderate association between judgments of memory quality and source memory accuracy). For the sake of completeness, these images are also detailed.

The results in Table 21 illustrate that participants may be relying on remembering a feature within the picture to recollect the cross location, i.e., confirming the assumption made for the experiment that the location and the image were coded together as a unified memory trace. For example, many of the pictures have distinct lines within them or a standout feature which might make them easier to associate with a cross location. The Association of some pictures with both high memory quality and high accuracy (Tables 19 & 21) is consistent with the moderate relationship between vividness or distinctiveness and accuracy.

Table 21S. Illustrating the five pictures associated with the highest accuracy score in each of the experiments: vividness, Experiments 1 (120-image) & 2 (240-image); distinctiveness, Experiments 4 (mixed list-type) & 5 (homogenous list-type)

VIVID 120-image					
Accuracy	1	2	3	4	5
VIVID 240-image					
Accuracy	1	2	3	4	5
DISTINCTIVE Mixed list 120-image					
Accuracy	1	2	3	4	5
DISTINCTIVE Homogenous list 120-image					
Accuracy	1	2	3	4	5

On the contrary, those pictures associated with lowest source memory accuracy, Table 22, have more featureless areas within them. In this case, it may either be that the picture is not recognised because it is similar to others previously shown, or that the picture is recognised, but there is no feature to trigger recollection of the cross location. On the other hand, it could be speculated that the pictures that do have features within them may have caused participants to mistakenly put the cross in the wrong place. I will point out at this juncture that the picture and the cross-locations remained the same for the experiments.

Table 22S. Illustrating the five pictures associated with the lowest accuracy score in each of the experiments: vividness, Experiments 1 (120-image) & 2 (240-image); distinctiveness, Experiments 4 (mixed list-type) & 5 (homogenous list-type)

VIVID 120-image					
Accuracy	116	117	118	119	120
VIVID 240-image					
Accuracy	236	237	238	239	240
DISTINCTIVE Mixed list 120-image					
Accuracy	116	117	118	119	120
DISTINCTIVE Homogenous list 120-image					
Accuracy	116	117	118	119	120

Summary.

Vividness and distinctiveness, while both being considered judgments of memory quality, may not be that separate. From the images detailed above and from the qualitative analyses, the same criteria are used for both ratings. When we judge our memory, confidence arises from both vividness and distinctiveness in the same manner; both judgments moderate the relationship between confidence and accuracy in the same manner; furthermore, the relationship between either vividness or distinctiveness and source memory accuracy is largely mediated through confidence. Results from an isolation list paradigm may however differ, if the isolate were to be predetermined and were not to be vivid, i.e., a photograph low in vividness being shown in a list of images all of high vividness. In this case, the image should be rated as distinctive. But would this be ‘natural' or manipulated distinctiveness?

Correlational analysis (all trials data).

As a further comparison of the similarity between judgments of vividness and distinctiveness in the context of these experiments, Figures 65 and 66 reproduce the scatterplots of confidence consequent on either vividness or distinctiveness for each of the four experiments (1-2; 4-5) carried out.

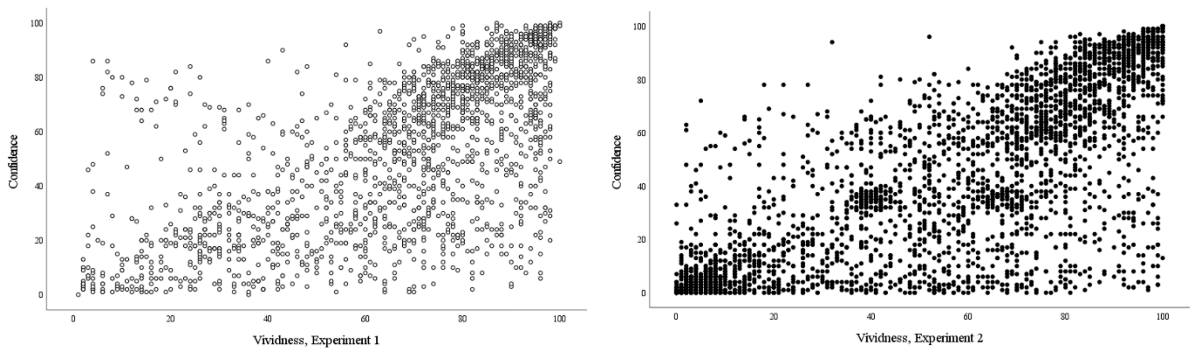

Figure 65S. Vividness, Experiment 1 (120-image set), N = 1920 & Experiment 2 (240-image set), N =3120. Scatterplots of confidence consequent on vividness ratings for the natural scenes photographs shown either in full colour (Experiment 1) or half in full colour, with half partially desaturated for colour (Experiment 2)

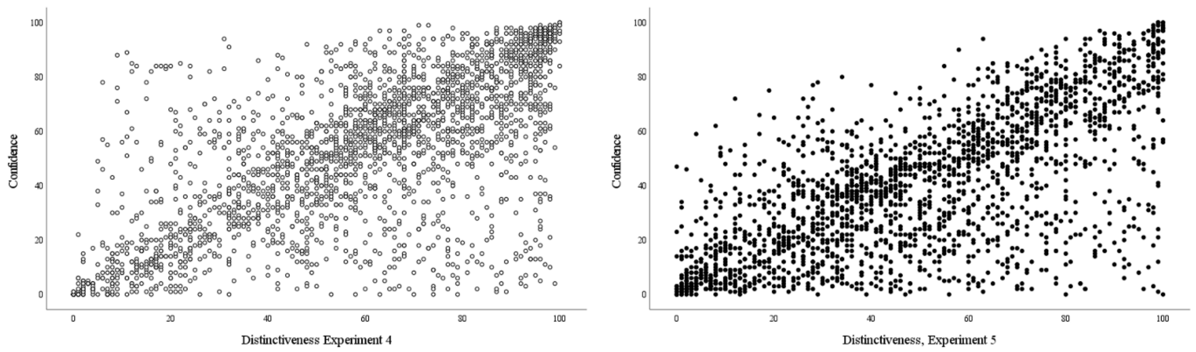

Figure 66S. Distinctiveness, Experiment 4 (mixed list-type), N = 2400 & Experiment 5 (homogenous list-type) N =2160. Scatterplots of confidence consequent on distinctiveness ratings for the 120-image set of full-colour natural scenes photographs

Appendix B.

Supplementary data.

Correlational analysis (averaged within-participant data).

Vividness: Experiments 1 & 2

For clarity, and as discussed under Chapter 3, General methods, data analysis plan (correlational analysis), the results from all trial data correlations have been contrasted with individual within-participant correlations showing relationships between the three variables of interest. Results are detailed, together with significance values for each participant, in Tables 23S and Table 24S.

Results detailed in Table 23S show within-participant correlations from Experiment 1 (120-image set, in full colour). It is evident that the great majority of correlations between vividness and confidence were overwhelmingly very large, positive and significant. Only one participant demonstrated a different pattern, with a large but negative correlation between vividness and confidence, and vividness and accuracy, but a positive correlation between accuracy and confidence. A different participant showed a trivial (nonsignificant) negative correlation between accuracy and confidence. Notably, there was less individual variation in correlations between vividness and confidence, and confidence and accuracy, than seen between vividness and accuracy. Two participants showed negative correlations between vividness and accuracy (one large and significant; one trivial and nonsignificant).

Table 23S. Vividness, Experiment 1 (120-image set): Individual within-participant correlation analyses showing separate correlations between vividness, confidence and accuracy. The averaged correlation coefficients are shown, together with the averaged correlation, used as a comparison in Table 4

Experiment 1	**Vividness & Confidence**		**Vividness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's r	p value	Pearson's r	p value	Pearson's r	p value
1	0.516	**<.001**	0.250	**0.006**	0.665	**<.001**
2	0.667	**<.001**	0.238	**0.009**	0.373	**<.001**
3	0.736	**<.001**	0.068	0.455	-0.052	0.57
4	0.799	**<.001**	0.49	**<.001**	0.608	**<.001**
5	0.522	**<.001**	0.348	**<.001**	0.557	**<.001**
6	0.744	**<.001**	0.647	**<.001**	0.524	**<.001**
7	0.509	**<.001**	0.213	0.019	0.366	**<.001**
8	-0.545	**<.001**	-0.333	**<.001**	0.343	**<.001**
9	0.916	**<.001**	-0.003	0.967	0.051	0.583
10	0.657	**<.001**	0.423	**<.001**	0.561	**<.001**
11	0.735	**<.001**	0.602	**<.001**	0.638	**<.001**
12	0.648	**<.001**	0.159	0.083	0.166	0.07
13	0.895	**<.001**	0.367	**<.001**	0.44	**<.001**
14	0.813	**<.001**	0.373	**<.001**	0.48	**<.001**
15	0.762	**<.001**	0.469	**<.001**	0.506	**<.001**
16	0.627	**<.001**	0.348	**<.001**	0.371	**<.001**
Average *r*	**0.646**		**0.291**		**0.412**	

Results are further detailed in bivariate scatterplots, Figures 67S-68S, illustrating that individual correlations largely corresponded with the pattern seen for the overall results (cf. Figure 65S), with no evidence for specific subgroups having different correlation trends.

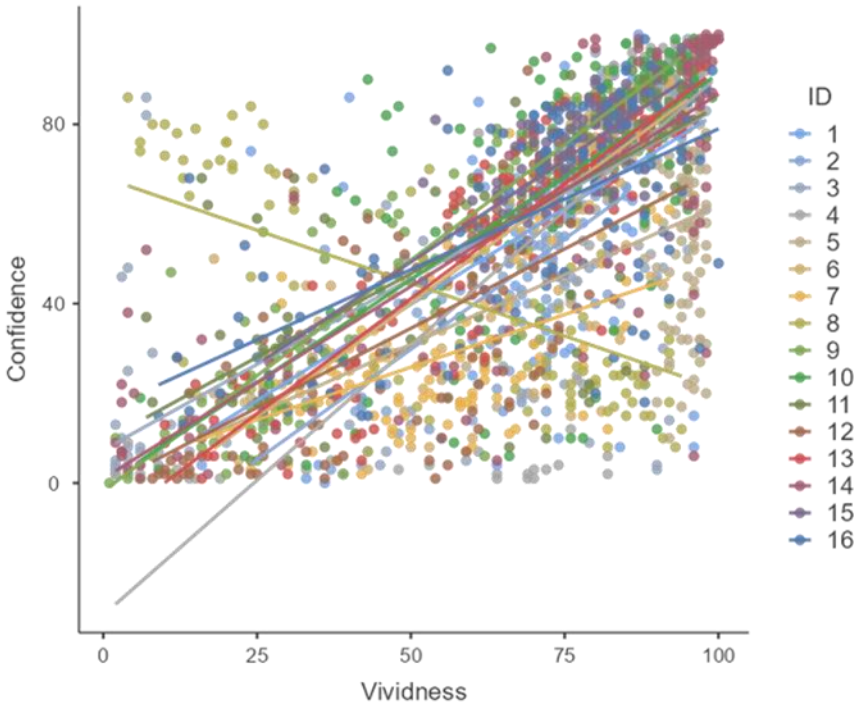

Figure 67S. Vividness: Experiment 1 (120-image set) bivariate scatterplot of confidence dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is overwhelmingly very large and positive

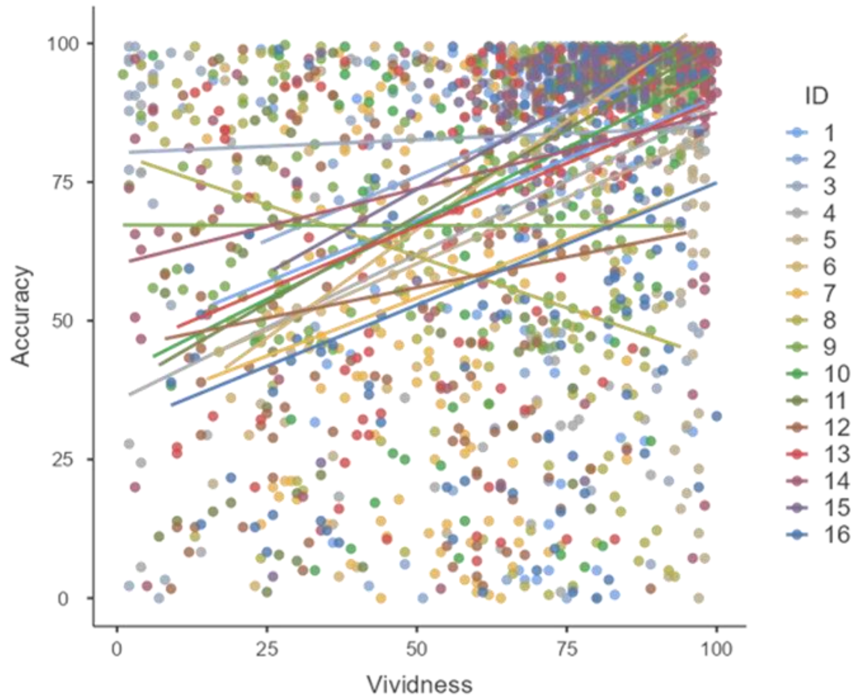

Figure 68S. Vividness: Experiment 1 (120-image set) bivariate scatterplot of accuracy dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, confirming a weaker correlation with more individual variation shown that between vividness and confidence.

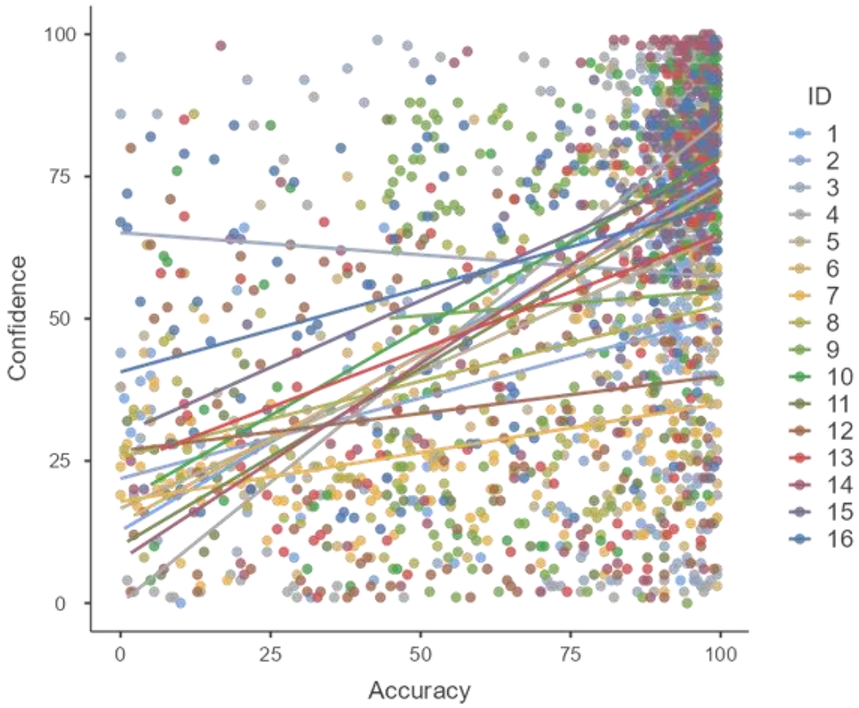

Figure 69S. Vividness: Experiment 1 (120-image set) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results. Although the correlation is generally large, there is more individual variation, i.e., not all participants showed a large and positive relationship between confidence and accuracy

Next, all within-participant correlations are detailed together with significance values for each participant in Experiment 2 (240-image set, half in full colour and half partially desaturated for colour), Table 24S. The same pattern was seen as for Experiment 1. Only one participant demonstrated a small negative (nonsignificant) correlation between vividness and confidence but positive correlations between vividness and accuracy, and accuracy and confidence.

Table 24S. Vividness, Experiment 2 (240-image set): Individual within-participant correlation analyses showing separate correlations between vividness, confidence and accuracy. The averaged correlation coefficients are shown, used as a comparison in Table 4

Experiment 2	**Vividness & Confidence**		**Vividness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's *r*	*p* value	Pearson's *r*	*p* value	Pearson's *r*	*p* value
17	0.684	**<.001**	0.261	**<.001**	0.414	**<.001**
19	0.896	**<.001**	0.548	**<.001**	0.571	**<.001**
20	0.697	**<.001**	0.416	**<.001**	0.506	**<.001**
21	-0.074	0.25	0.015	0.82	0.396	**<.001**
22	0.636	**<.001**	0.63	**<.001**	0.585	**<.001**
23	0.667	**<.001**	0.368	**<.001**	-0.004	0.945
24	0.834	**<.001**	0.503	**<.001**	0.548	**<.001**
25	0.78	**<.001**	0.396	**<.001**	0.48	**<.001**
26	0.845	**<.001**	0.463	**<.001**	0.512	**<.001**
29	0.689	**<.001**	0.23	**<.001**	0.365	**<.001**
30	0.979	**<.001**	0.478	**<.001**	0.467	**<.001**
31	0.613	**<.001**	0.336	**<.001**	0.489	**<.001**
32	0.66	**<.001**	0.097	0.13	0.216	**<.001**
Average *r*	**0.685**		**0.365**		**0.432**	

Results are further detailed in bivariate scatterplots (Figures 70S-72S), illustrating that individual correlations largely corresponded with the pattern seen for the overall results (cf. Table 4; Figure 65S), with no evidence for specific subgroups having different correlations. In both experiments, there was less individual variability in the correlations between vividness and confidence and accuracy and confidence than were seen between vividness and accuracy.

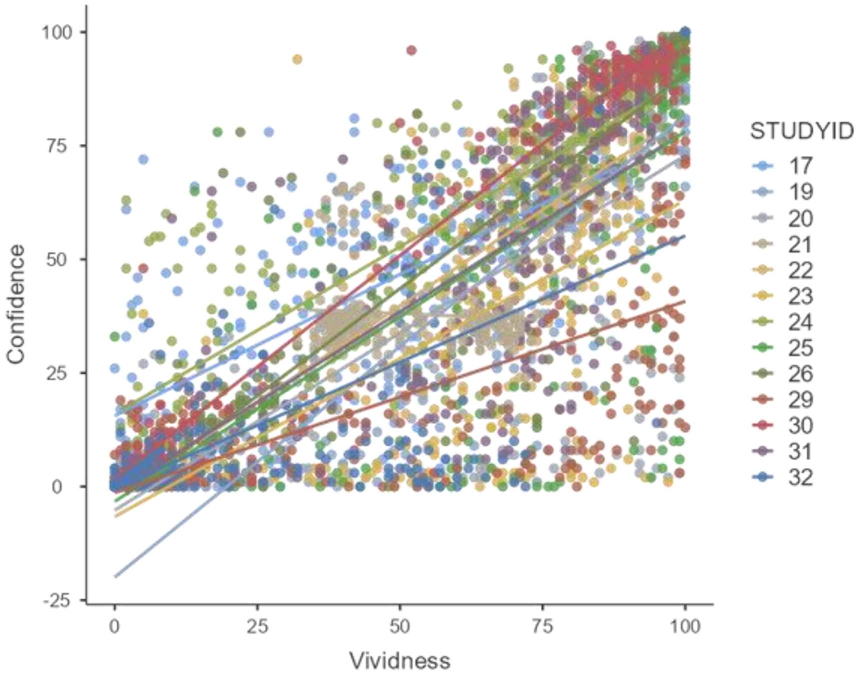

Figure 70S. Vividness: Experiment 2 (240-image set) bivariate scatterplot of confidence dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of correlations, which are very large and positive

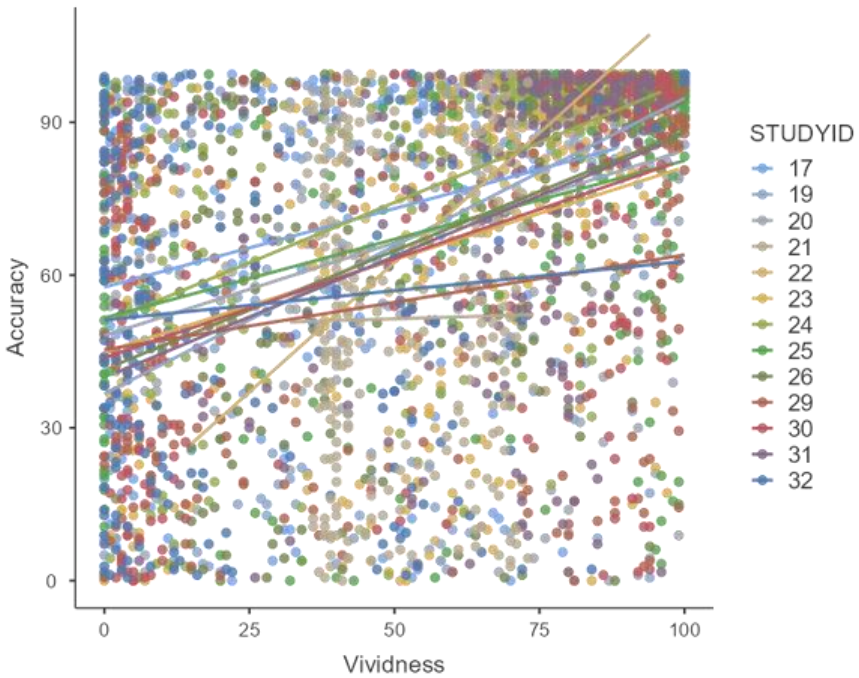

Figure 71S. Vividness: Experiment 2 (240-image set) bivariate scatterplot of accuracy dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which, although generally positive, is weaker than that seen for the relationship between vividness and confidence

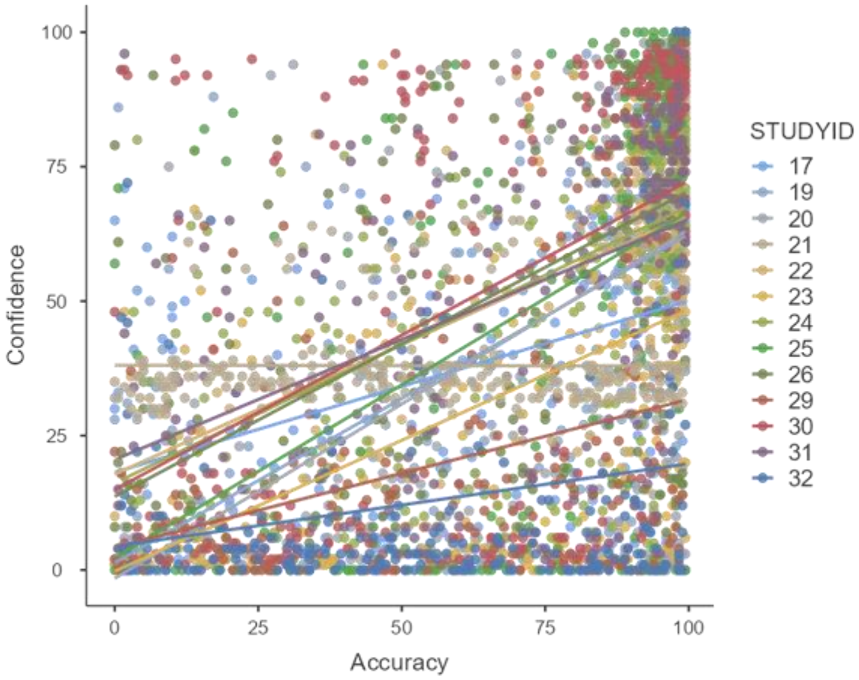

Figure 72S. Vividness: Experiment 2 (240-image set) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, although the majority of correlations are large and positive

Vividness: Experiment 3

Within-participant correlations are detailed in Table 25S. Figure 73S illustrates all regression lines for the 64 participants in Experiment 3 regarding the relationship between vividness and confidence.

Table 25S. Vividness, Experiment 3 (2-AFC task): Individual within-participant correlations between vividness and confidence.

STUDY ID	*r*	*p* value	STUDY ID	*r*	*p* value	STUDY ID	*r*	*p* value	STUDY ID	r	*p* value
1	0.75	<.001	21	0.906	<.001	41	0.748	<.001	61	0.858	<.001
2	0.936	<.001	22	0.698	<.001	42	0.929	<.001	62	0.86	<.001
3	0.604	<.001	23	0.962	<.001	43	0.878	<.001	63	0.938	<.001
4	0.414	0.012	24	0.959	<.001	44	0.942	<.001	64	0.868	<.001
5	0.969	<.001	25	0.846	<.001	45	0.257	0.136	***Average***	**0.785**	
6	0.538	<.001	26	0.812	<.001	46	0.772	<.001			
7	0.885	<.001	27	0.978	<.001	47	0.282	0.101			
8	0.858	<.001	28	0.965	<.001	48	0.901	<.001			
9	0.852	<.001	29	0.849	<.001	49	0.729	<.001			
10	0.792	<.001	30	0.889	<.001	50	0.756	<.001			
11	0.226	0.198	31	0.794	<.001	51	0.644	<.001			
12	0.624	<.001	32	0.901	<.001	52	0.937	<.001			
13	0.972	<.001	33	0.549	<.001	53	0.895	<.001			
14	0.017	<.001	34	0.946	<.001	54	0.879	<.001			
15	0.559	<.001	35	0.81	<.001	55	0.883	<.001			
16	0.903	<.001	36	0.878	<.001	56	0.821	<.001			
17	0.834	<.001	37	0.745	<.001	57	0.925	<.001			
18	0.843	<.001	38	0.985	<.001	58	0.879	<.001			
19	0.757	<.001	39	0.416	0.012	59	0.919	<.001			
20	0.756	<.001	40	0.914	<.001	60	0.876	<.001			

All correlations are positive and overwhelmingly significant. The data do not suggest there are subsets of participants who have different correlations in comparison with the correlations using all trial data (cf. Figure 35). The overall pattern is clearly illustrated in Figure 73S.

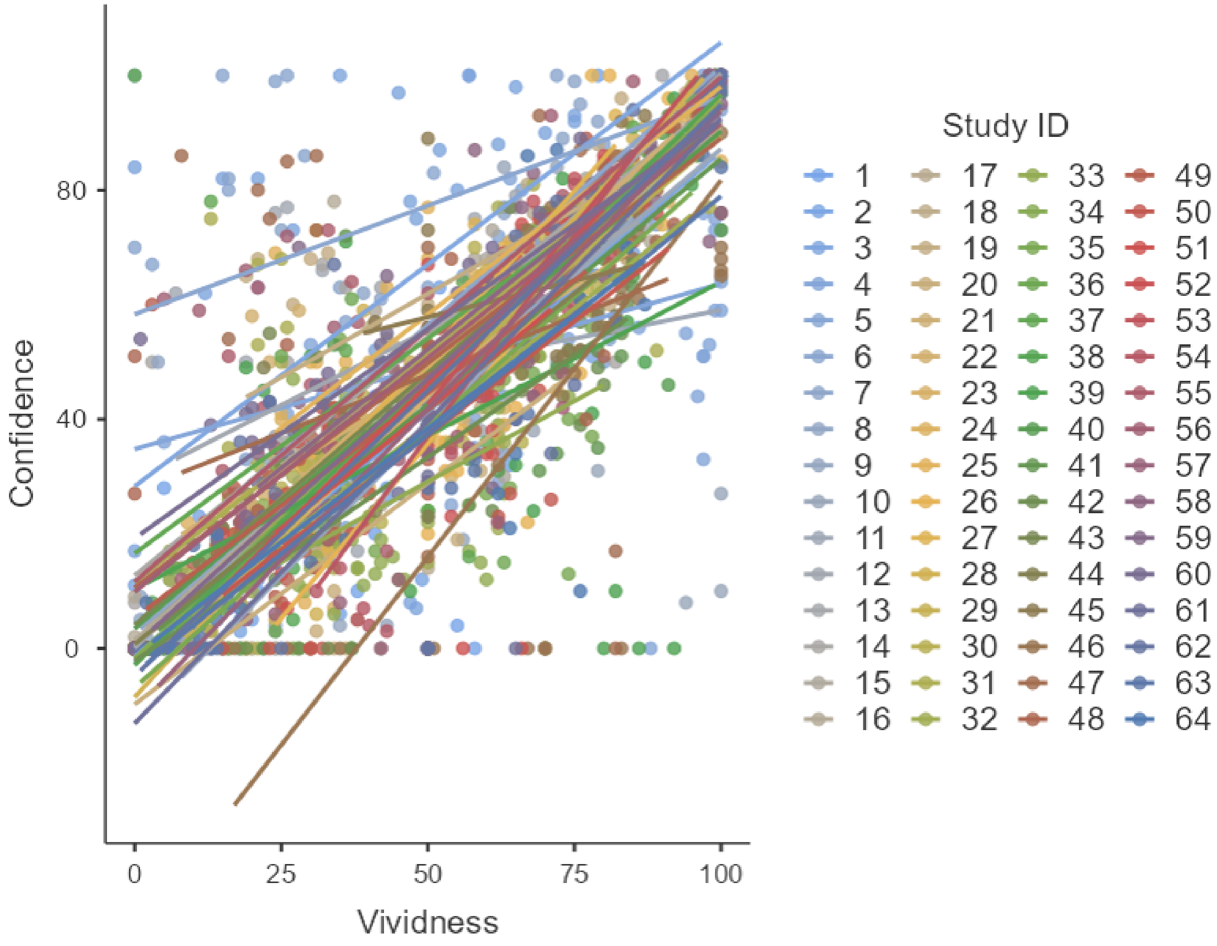

Figure 73S. Vividness: Experiment 3 (2-AFC task)) bivariate scatterplot of confidence dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of correlations, which are very large and positive

Distinctiveness: Experiments 4 & 5

As for the participant's, Experiments 4 and 5 were reanalyzed using averaged within-participant correlations. These are detailed for Experiment 4 (mixed *list-type*), together with significance values for each participant, in Table 25S.

Table 26S. Distinctiveness: Experiment 4 (mixed list-type). Individual within-participant correlation analyses showing separate correlations between distinctiveness, confidence and accuracy. The averaged correlation coefficients are shown, used as a comparison in Table 10

Experiment 4	**Distinctiveness & Confidence**		**Distinctiveness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's *r*	*p* value	Pearson's *r*	*p* value	Pearson's *r*	*p* value
1	0.702	**<.001**	0.404	**<.001**	0.445	**<.001**
2	0.936	**<.001**	0.503	**<.001**	0.485	**<.001**
3	0.847	**<.001**	0.554	**<.001**	0.598	**<.001**
4	0.471	**<.001**	0.351	**<.001**	0.654	**<.001**
5	0.768	**<.001**	0.548	**<.001**	0.509	**<.001**
6	0.687	**<.001**	0.354	**<.001**	0.498	**<.001**
7	0.707	**<.001**	0.354	**<.001**	0.52	**<.001**
8	0.645	**<.001**	0.395	**<.001**	0.588	**<.001**
9	0.854	**<.001**	0.531	**<.001**	0.606	**<.001**
10	0.756	**<.001**	0.297	**<.001**	0.538	**<.001**
11	0.691	**<.001**	0.282	0.002	0.441	**<.001**
12	0.717	**<.001**	0.417	**<.001**	0.431	**<.001**
13	0.579	**<.001**	0.253	0.005	0.371	**<.001**
14	0.541	**<.001**	0.372	**<.001**	0.662	**<.001**
15	0.637	**<.001**	0.52	**<.001**	0.602	**<.001**
16	0.67	**<.001**	0.344	**<.001**	0.263	0.004
17	0.763	**<.001**	0.166	0.07	0.568	**<.001**
18	0.675	**<.001**	0.579	**<.001**	0.598	**<.001**
19	0.669	**<.001**	0.609	**<.001**	0.582	**<.001**
20	0.22	0.016	0.184	0.044	0.371	**<.001**
Average *r*	**0.677**		**0.401**		**0.517**	

The pattern of within-participant correlations illustrates positive correlations between all three variables, overwhelmingly very large and highly significant. Of note, the associations are weaker for distinctiveness and accuracy, and strongest for distinctiveness and confidence. This reflects the picture seen for correlations derived from all trial data (cf. Figure 65S; Table 10). The pattern of results are further illustrated in the bivariate scatterplots (Figures 74S-76S).

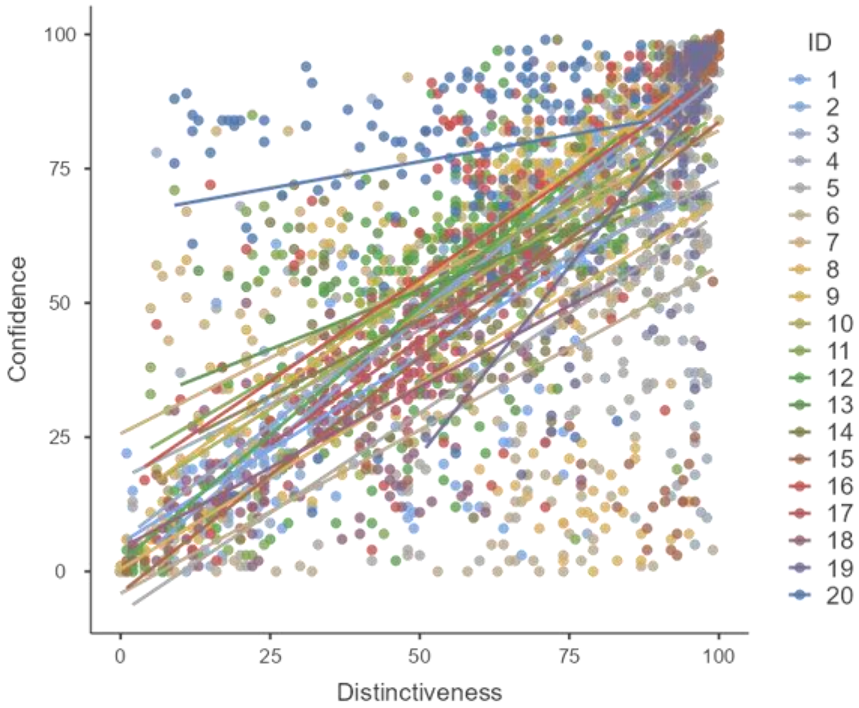

Figure 74S. *Distinctiveness: Experiment 4 (mixed list-type) bivariate scatterplot of confidence dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is overwhelmingly strong and positive*

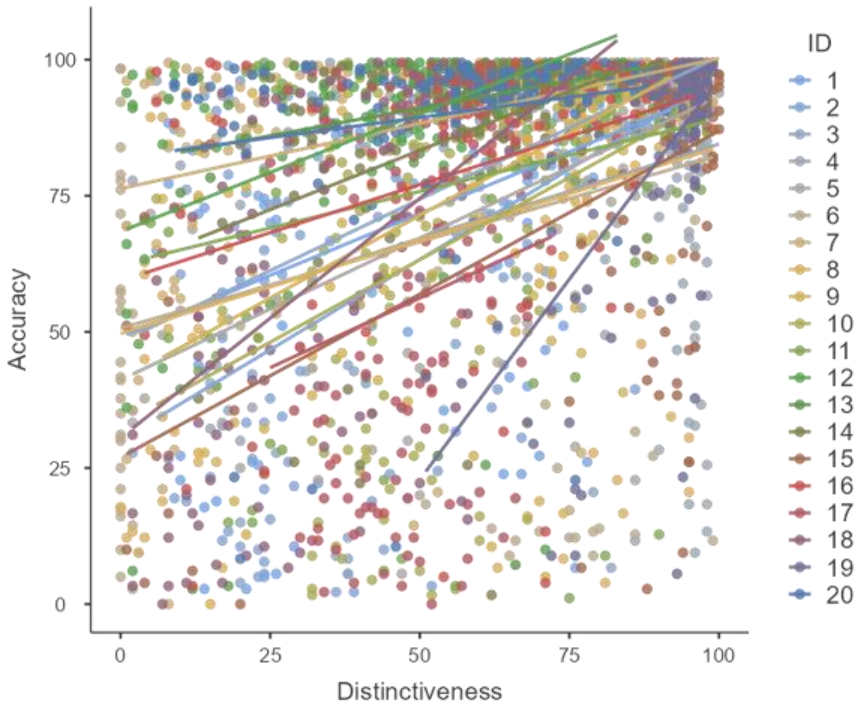

Figure 75S. *Distinctiveness: Experiment 4 (mixed list-type) bivariate scatterplot of accuracy dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which although positive, shows more variability than in Figure 74S*

.

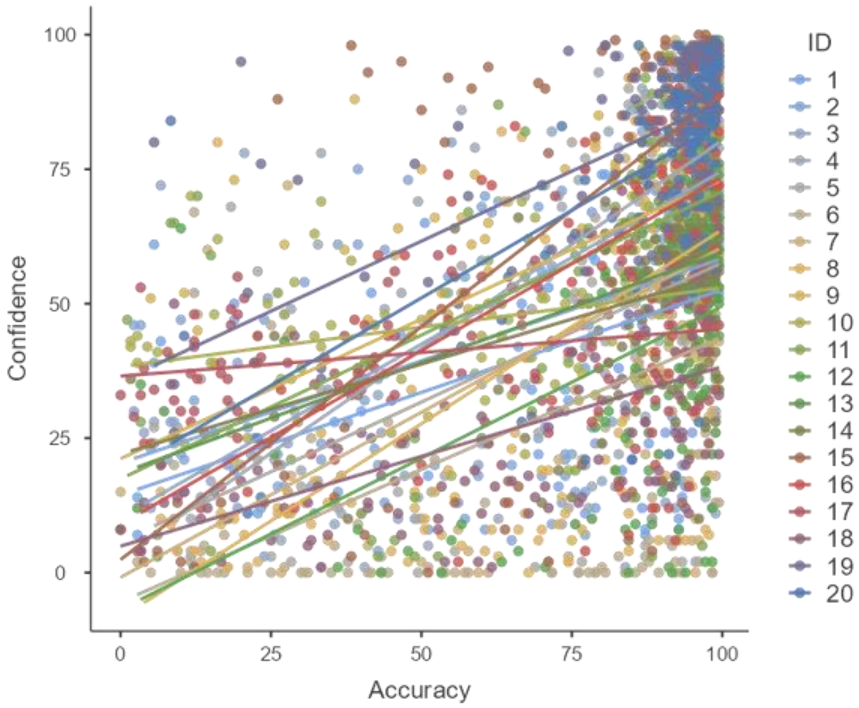

Figure 76S. Distinctiveness: Experiment 4 (mixed list-type) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is moderate and positive

Results of within-participants correlational analyses are next detailed for distinctiveness, Experiment 5 (homogenous list-type), Table 27S.

Table 27S. Distinctiveness: Experiment 5 (homogeneous list-type). Individual within-participant correlation analyses showing separate correlations between distinctiveness, confidence and accuracy. The averaged correlation coefficients are shown, used as a comparison in Table 10

Experiment 5	**Distinctiveness & Confidence**		**Distinctiveness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's *r*	*p* value	Pearson's *r*	*p* value	Pearson's *r*	*p* value
21	0.537	**<.001**	0.252	0.13	0.243	0.005
23	0.644	**<.001**	0.269	0.13	0.383	**<.001**
24	0.849	**<.001**	0.274	0.13	0.264	0.004
25	0.814	**<.001**	0.234	0.13	0.311	**<.001**
27	0.706	**<.001**	0.198	0.13	0.326	**<.001**
28	0.776	**<.001**	0.361	**<.001**	0.425	**<.001**
29	0.191	0.037	0.116	0.13	0.116	0.209
30	0.779	**<.001**	0.211	0.13	0.315	**<.001**
31	0.639	**<.001**	0.109	0.13	0.256	0.005
32	0.579	**<.001**	0.39	**<.001**	0.293	0.001
33	0.764	**<.001**	0.322	**<.001**	0.502	**<.001**
34	0.556	**<.001**	0.115	0.13	0.254	0.005
35	0.807	**<.001**	0.456	**<.001**	0.515	**<.001**
36	0.735	**<.001**	0.193	0.13	0.273	0.003
37	0.763	**<.001**	0.323	**<.001**	0.407	**<.001**
38	0.591	**<.001**	0.183	0.13	0.297	**<.001**
39	0.499	**<.001**	-0.069	**<.001**	0.115	0.212
40	0.748	**<.001**	0.388	**<.001**	0.396	**<.001**
Average *r*	**0.665**		**0.240**		**0.316**	

Although the relationship between distinctiveness and confidence is again very large and positive, the associations are weaker for correlations between distinctiveness and accuracy, and accuracy and confidence. These are illustrated as bivariate scatterplots, Figures 77S-79S. All correlations are positive and once again there is no evidence that subsets exist showing differing trends.

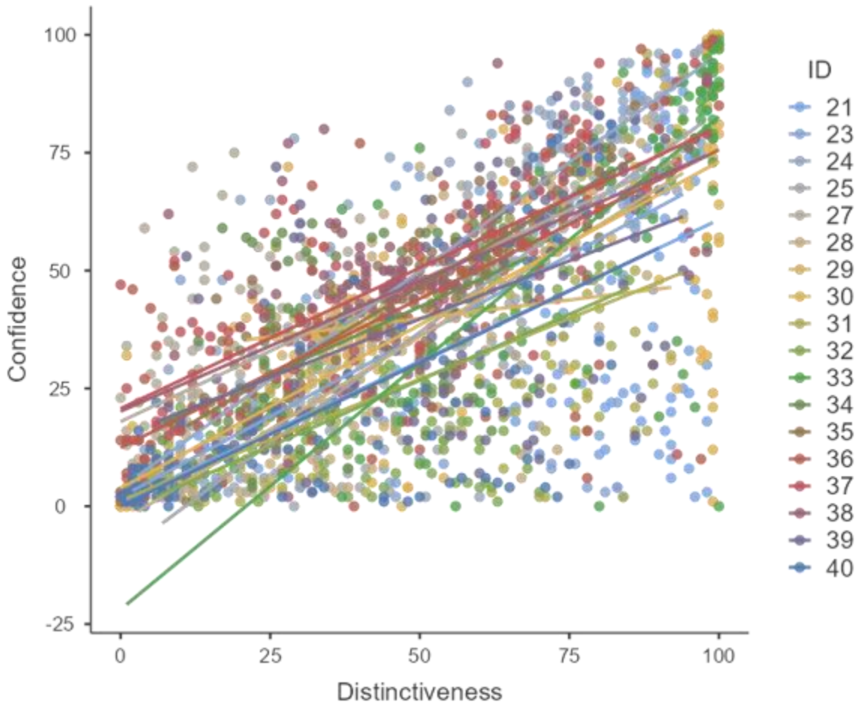

Figure 77S. Distinctiveness: Experiment 5 (homogenous list-type) bivariate scatterplot of confidence dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is overwhelmingly very large, and positive

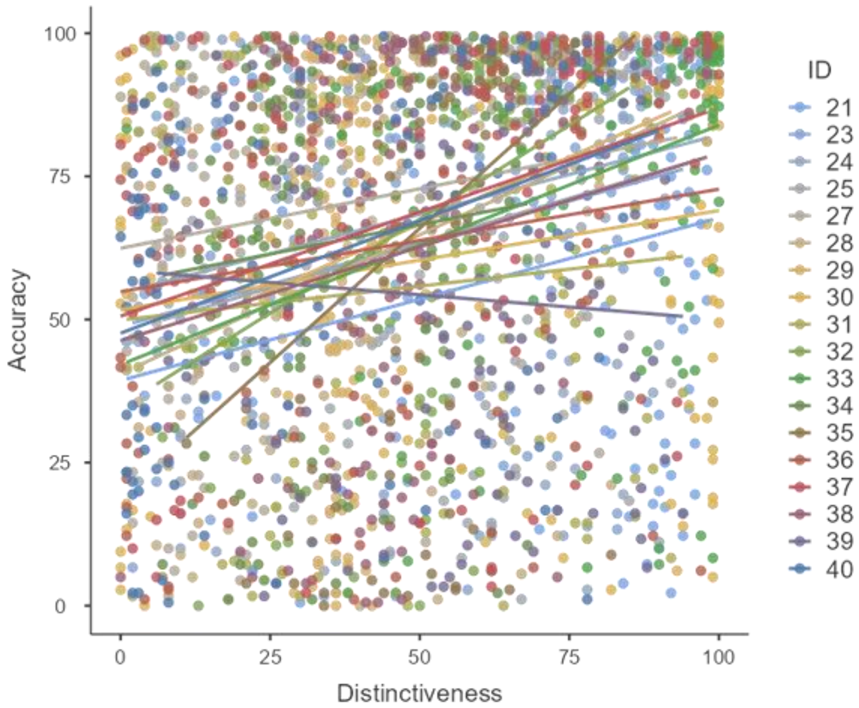

Figure 78S. Distinctiveness: Experiment 5 (homogenous list-type) bivariate scatterplot of accuracy dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is weaker and shows more individual variability than in Figure 77S

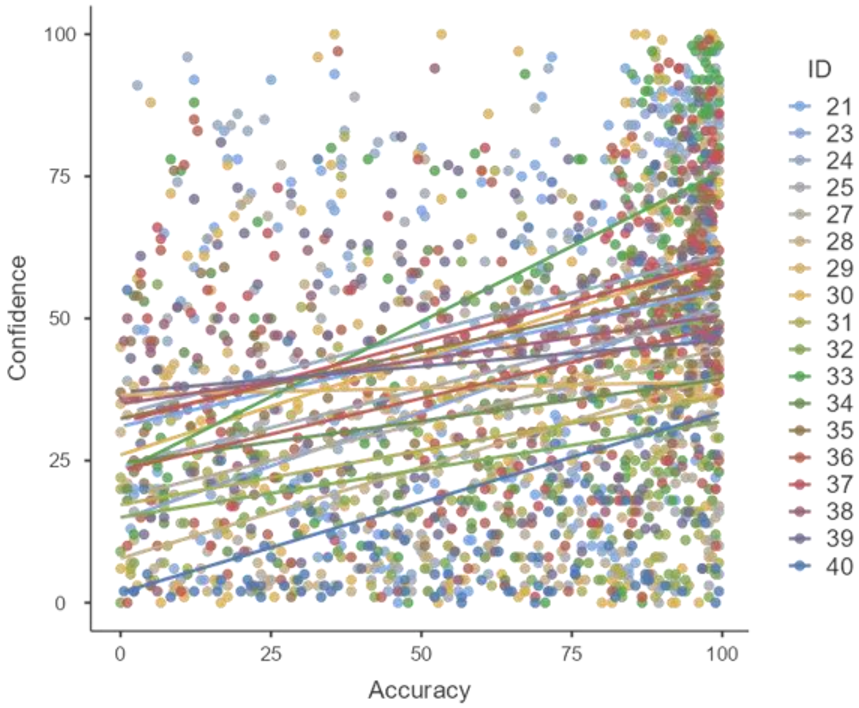

Figure 79S. Distinctiveness: Experiment 5 (homogenous list-type) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is weaker than in Figure 76S

Familiarity: Experiment 6.

Within-participant correlations are detailed together with significance (*p*-values) and the averaged correlation.

Table 28S. Familiarity: Experiment 6 (2-AFC task) Within-participant correlation analyses between familiarity and confidence; the averaged correlation is shown for comparison with the result using data from all trials

STUDY ID	*r*	*p*-value	STUDY ID	*r*	*p*-value	STUDY ID	*r*	*p*-value	STUDY ID	*r*	*p*-value
1	0.788	<.001	21	0.7	<.001	41	0.87	<.001	61	0.502	0.002
2	0.942	<.001	22	0.811	<.001	42	0.92	<.001	62	0.665	<.001
3	0.666	<.001	23	0.699	<.001	43	0.849	<.001	***Average***	**0.775**	
4	0.866	<.001	24	0.849	<.001	44	0.573	<.001			
5	0.869	<.001	25	0.927	<.001	45	0.66	<.001			
6	0.718	<.001	26	0.847	<.001	46	0.751	<.001			
7	0.759	<.001	27	0.685	<.001	47	0.631	<.001			
8	0.897	<.001	28	0.927	<.001	48	0.79	<.001			
9	0.124	0.471	29	0.944	<.001	49	0.663	<.001			
10	0.866	<.001	30	0.462	0.009	50	0.714	<.001			
11	0.053	0.761	31	0.86	<.001	51	0.712	<.001			
12	0.841	<.001	32	0.984	<.001	52	0.802	<.001			
13	0.372	0.028	33	0.872	<.001	53	0.95	<.001			
14	0.832	<.001	34	0.861	<.001	54	0.911	<.001			
15	0.379	0.027	35	0.861	<.001	55	0.884	<.001			
16	0.679	<.001	36	0.902	<.001	56	0.407	0.015			
17	0.777	<.001	37	0.945	<.001	57	0.85	<.001			
18	0.902	<.001	38	0.856	<.001	58	0.919	<.001			
19	0.838	<.001	39	0.993	<.001	59	0.671	<.001			
20	0.838	<.001	40	0.58	<.001	60	0.887	<.001			

All correlations are overwhelmingly very large, positive, and significant. This is further illustrated in Figure 80S.

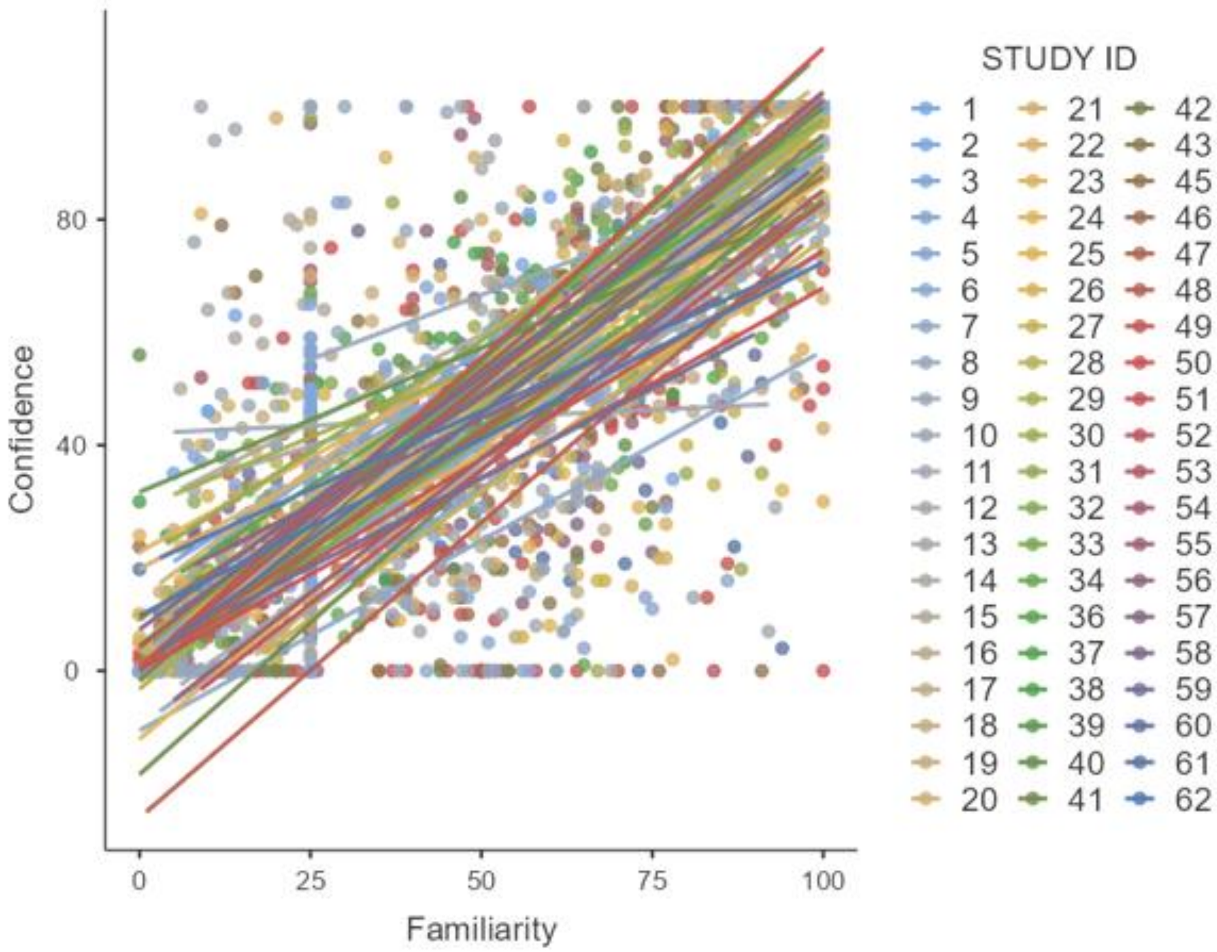

Figure 80S. Familiarity: Experiment 6 (2-AFC task) bivariate scatterplot of confidence dependent on familiarity. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of correlations, which are very large and positive

Vividness: Encoding judgment, 120- & 240-image sets

In the same way as for Experiments 1 and 2, within-participant correlations are detailed for the same experiments carried out with the judgments of vividness at encoding (during the study phase).

Table 29S. Vividness, Encoding (120-image set): Individual within-participant correlation analyses showing separate correlations between vividness, confidence and accuracy. The averaged correlation coefficients are shown, used as a comparison in Table 15

Encoding 120	**Vividness & Confidence**		**Vividness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's r	p value	Pearson's r	p value	Pearson's r	p value
33	0.047	0.61	0.045	0.619	0.472	**<.001**
34	0.202	**0.027**	0.004	0.97	0.491	**<.001**
35	0.041	0.66	0.024	0.79	0.408	**<.001**
36	0.15	0.1	0.119	0.12	0.619	**<.001**
37	0.394	**<.001**	0.153	0.096	0.384	**<.001**
38	-0.116	0.91	-0.011	0.21	0.461	**<.001**
39	-0.058	0.533	-0.12	0.187	0.601	**<.001**
40	0.088	0.38	0.089	0.323	0.417	**<.001**
41	0.402	**<.001**	0.049	0.59	0.385	**<.001**
42	0.174	0.058	0.008	0.93	0.467	**<.001**
43	0.381	**<.001**	0.249	**0.006**	0.564	**<.001**
44	0.555	**<.001**	0.229	**0.01**	0.526	**<.001**
45	0.187	**0.041**	0.224	**0.014**	0.46	**<.001**
46	0.272	**0.003**	0.251	**0.006**	0.284	**0.002**
47	0.243	**0.007**	0.019	0.85	0.482	**<.001**
48	--0.144	0.116	-0.357	**<.001**	0.35	**<.001**
Average *r*	**0.176**		**0.061**		**0.461**	

The pattern of results is clearly very different from those seen in Table 23S. Correlations are generally weaker and more variable for correlations between vividness and confidence, and vividness and accuracy. There is little if any correlation between vividness judged at encoding and source memory accuracy for recollection of the cross location. Results between accuracy and confidence remain positive, large, and significant. The individual correlations are further illustrated in Figures 81S-83.S

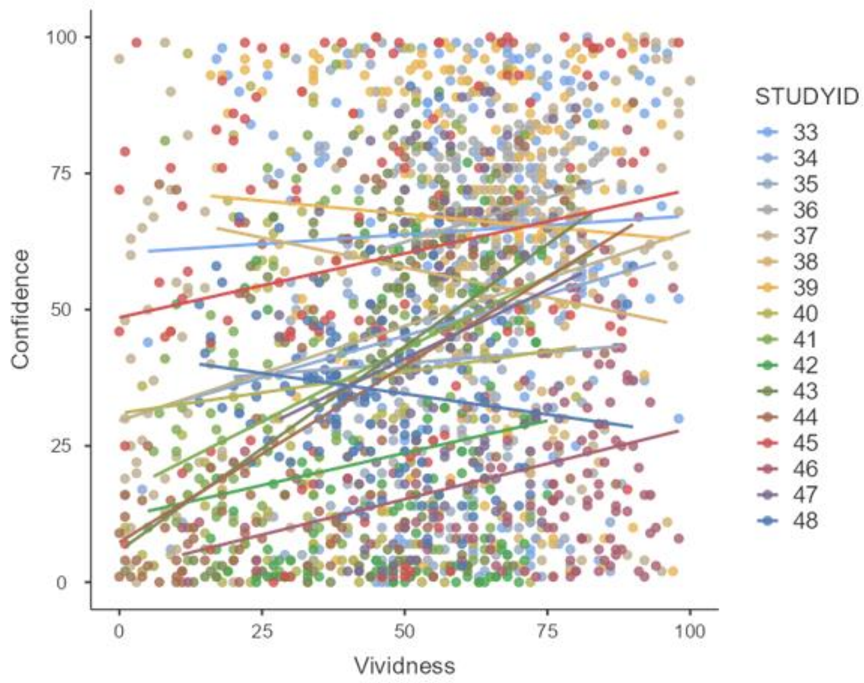

Figure 81S. Vividness: Encoding (120-image set) bivariate scatterplot of confidence dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is more variable in comparison with Figure 67S

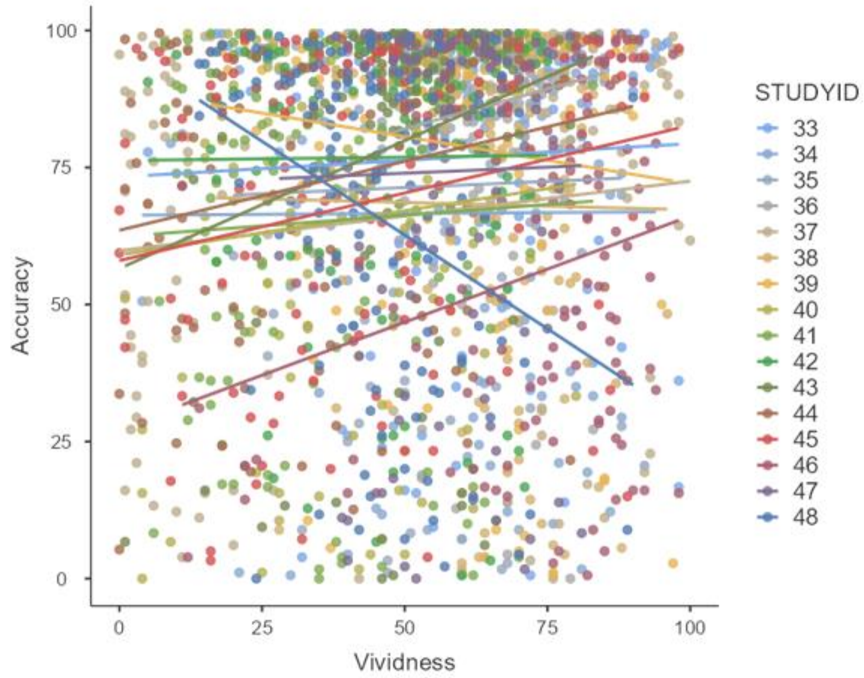

Figure 82S. Vividness: Encoding (120-image set) bivariate scatterplot of accuracy dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which, as shown in Table 27S are not significant

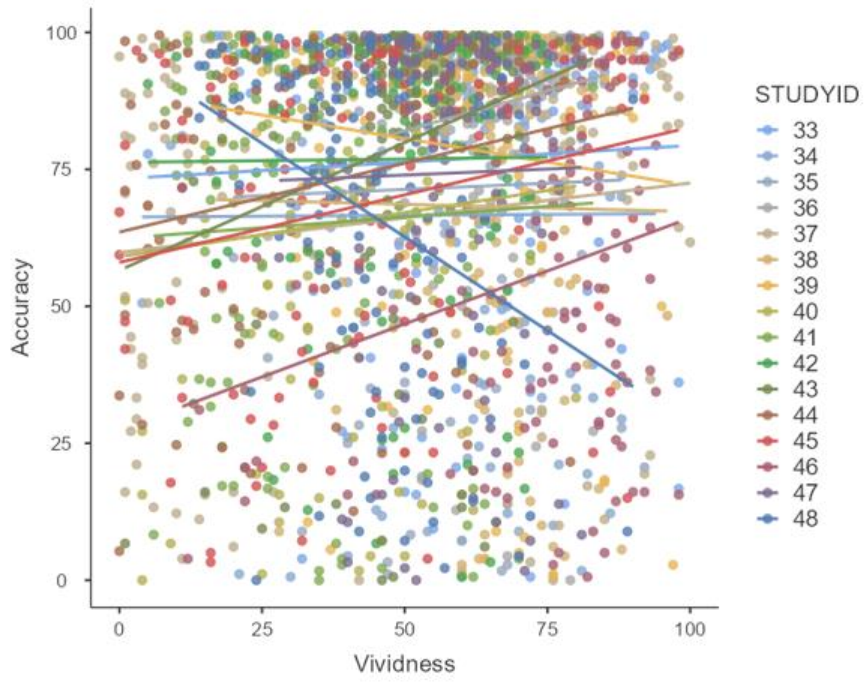

Figure 83S. Vividness: Encoding (120-image set) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which largely shows a positive relationship

Next, results are detailed for the 240-image set (half in full colour and half partially desaturated for colour) used in Experiment 2, but with the judgment of vividness made at encoding, during the study block, Table 30S. In contrast to the results shown for the 120-image set judged at encoding (Table 29S), although the correlations are weaker and similar, there are no negative values. There is little correlation between vividness judged at encoding and accuracy for recollection of the cross location; the correlations between accuracy and confidence are generally large, positive, and significant.

Table 30S. Vividness, Encoding (240-image set): Individual within-participant correlation analyses showing separate correlations between vividness, confidence and accuracy. The averaged correlation coefficients are shown, used as a comparison in Table 15

Encoding 240	**Vividness & Confidence**		**Vividness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's r	p value	Pearson's r	p value	Pearson's r	p value
49	0.348	**<.001**	0.047	0.47	0.312	**<.001**
50	0.238	**0.001**	0.239	**<.001**	0.467	**<.001**
51	0.185	**0.004**	0.163	**0.012**	0.562	**<.001**
52	0.092	0.15	0.12	**0.048**	0.42	**<.001**
53	0.316	**<.001**	0.017	0.79	0.372	**<.001**
54	0.348	**<.001**	0.066	0.31	0.413	**<.001**
55	0.26	**<.001**	0.137	**0.03**	0.175	**0.007**
56	0.181	**0.005**	0.059	0.37	0.526	**<.001**
57	0.283	**<.001**	0.119	0.07	0.214	**<.001**
58	0.347	**<.001**	0.241	**<.001**	0.348	**<.001**
59	0.276	**<.001**	0.14	**0.03**	0.393	0.07
60	0.446	**<.001**	0.244	**<.001**	0.501	**<.001**
61	0.514	**<.001**	0.155	**0.016**	0.454	**<.001**
62	0.2	**0.002**	0.157	**0.015**	0.393	**<.001**
63	0.056	0.392	0.011	0.86	0.278	**<.001**
64	0.208	**0.001**	0.017	0.798	0.215	**<.001**
Average *r*	**0.269**		**0.121**		**0.378**	

Results are further illustrated in Figures 84S-86S. That there are twice as many trials for each participant as in the 120-image set is evident from the concentration of data points. The number of low confidence trials dependent on both vividness and accuracy are clearly shown for all values of the predictor variable.

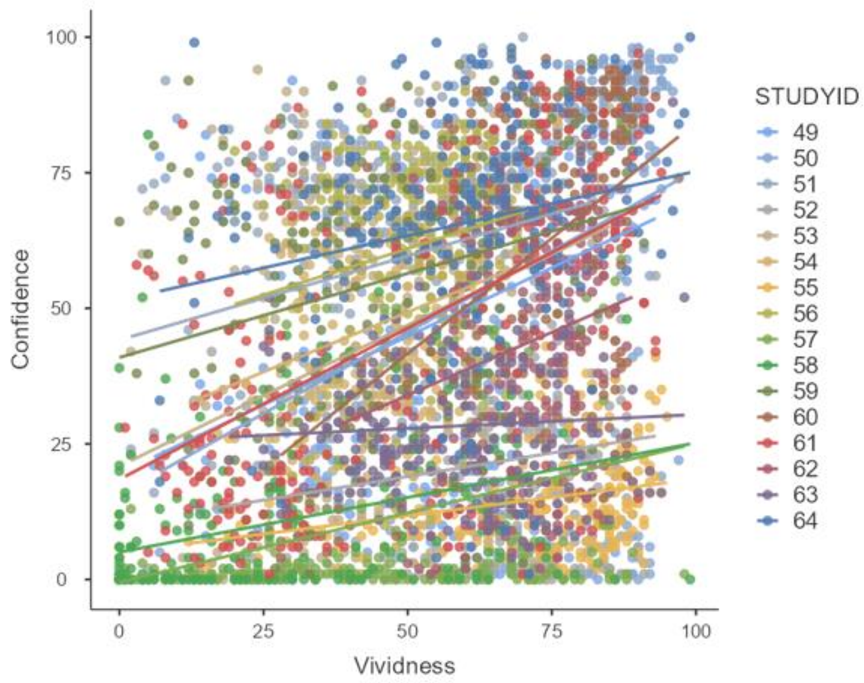

Figure 84S. Vividness: Encoding (240-image set) bivariate scatterplot of confidence dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which are variable but positive

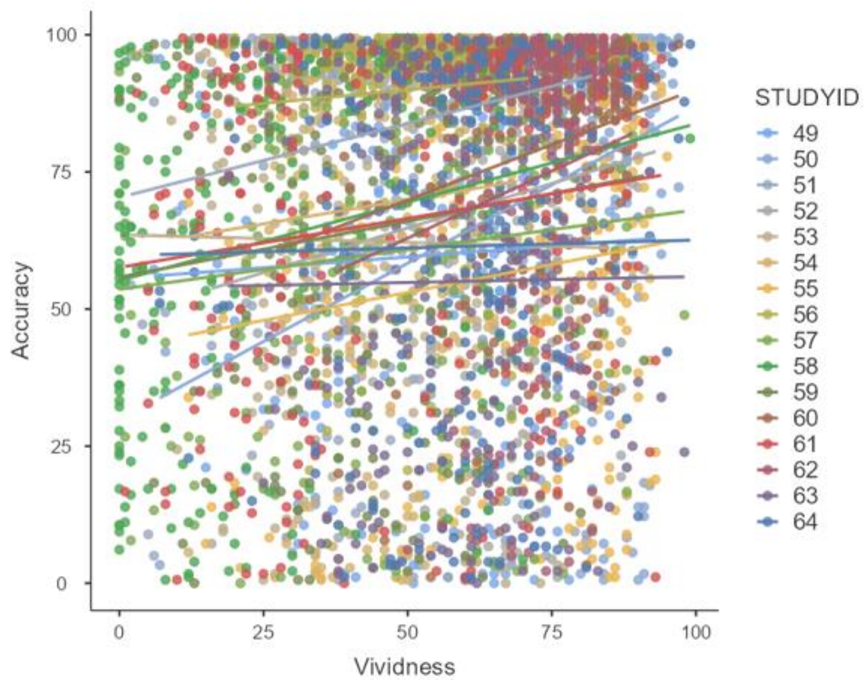

Figure 85S. Vividness: Encoding (240-image set) bivariate scatterplot of accuracy dependent on vividness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which, as shown in Table 28S are not significant

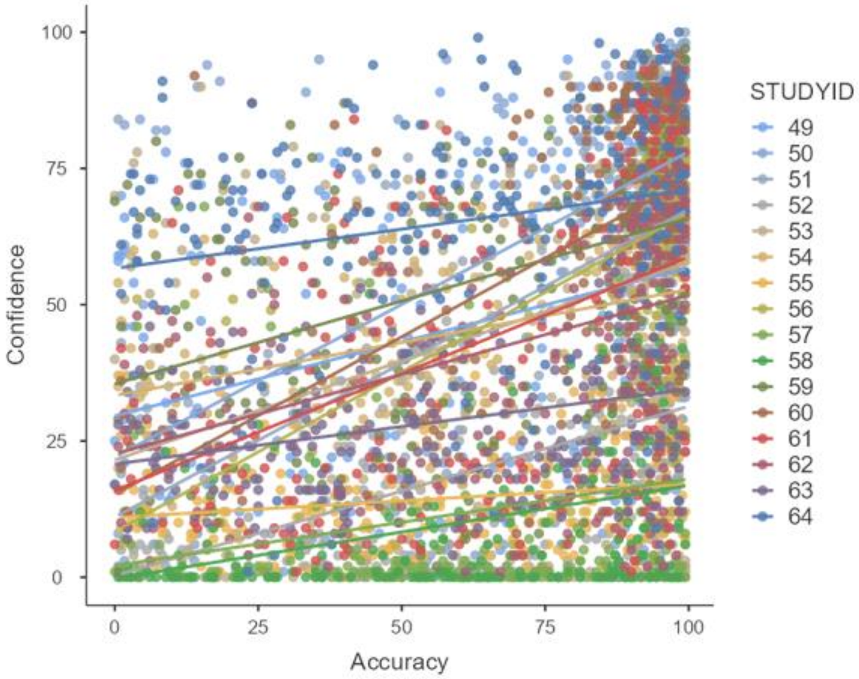

Figure 86S. Vividness: Encoding (240-image set) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, illustrating a positive relationship

Distinctiveness: Encoding judgment, mixed & homogenous list-types

As detailed for Experiments 4 and 5, within-participant correlations are again averaged when the same experiments were repeated with the judgment of distinctiveness made at encoding, during the study block. Table 31S details individual within-participant results and significance (*p*-values) for participants viewing images in a mixed *list-type*.

Table 31S. Distinctiveness: Encoding (mixed list-type). Individual within-participant correlation analyses showing separate correlations between distinctiveness, confidence and accuracy. The averaged correlation coefficients are shown, used as a comparison in Table 18

Encoding Mixed	**Distinctiveness & Confidence**		**Distinctiveness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's *r*	*p* value	Pearson's *r*	*p* value	Pearson's *r*	*p* value
41	0.516	**<.001**	0.208	0.023	0.499	**<.001**
42	0.444	**<.001**	0.187	0.041	0.516	**<.001**
43	0.281	0.002	0.138	0.133	0.468	0.57
44	-0.009	0.92	0.143	0.119	0.514	**<.001**
45	0.362	**<.001**	-0.104	0.259	0.255	0.005
46	0.304	**<.001**	0	1	0.19	0.037
47	0.491	**<.001**	0.237	0.009	0.371	**<.001**
48	0.59	**<.001**	0.282	0.002	0.552	**<.001**
49	0.456	**<.001**	0.147	0.116	0.458	**<.001**
50	0.396	**<.001**	0.249	0.006	0.39	**<.001**
51	0.489	**<.001**	0.099	0.28	0.399	**<.001**
52	0.34	**<.001**	0.266	0.003	0.641	**<.001**
53	0.352	**<.001**	0.189	0.039	0.39	**<.001**
54	0.063	0.496	0.072	0.436	0.49	**<.001**
55	0.433	**<.001**	0.256	**<.001**	0.457	**<.001**
56	0.299	**<.001**	0.032	0.727	0.486	**<.001**
57	0.207	0.024	0.038	0.677	0.107	0.243
58	0.629	**<.001**	0.023	0.805	0.176	0.055
59	0.152	0.097	0.074	0.42	0.341	**<.001**
60	0.397	**<.001**	0.182	0.046	0.36	**<.001**
Average *r*	**0.360**		**0.136**		**0.403**	

Correlations between distinctiveness and confidence are weaker with distinctiveness judged at encoding, reproducing the results seen when all trial data was correlated (Table 18). The correlations between accuracy and confidence remain moderate to large, positive, and generally significant throughout, regardless of when distinctiveness is judged. For further demonstration of the results see Figures 87S-89S. Once again, there is no clear pattern of results to indicate that separate groups exist with differing trends from the overall calculated correlation values.

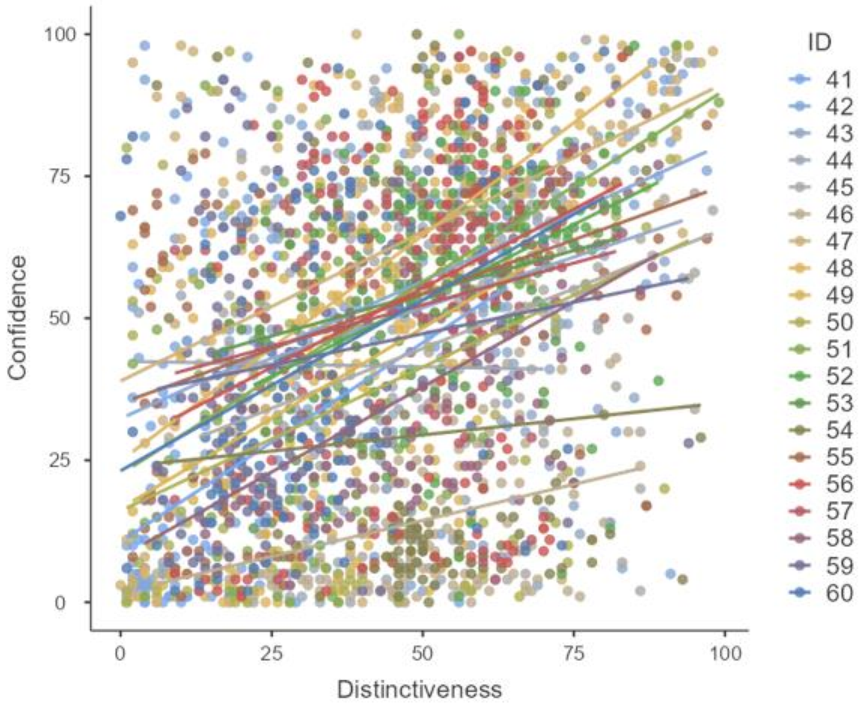

Figure 87S. Distinctiveness: Encoding (mixed list-type) bivariate scatterplot of confidence dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is generally positive

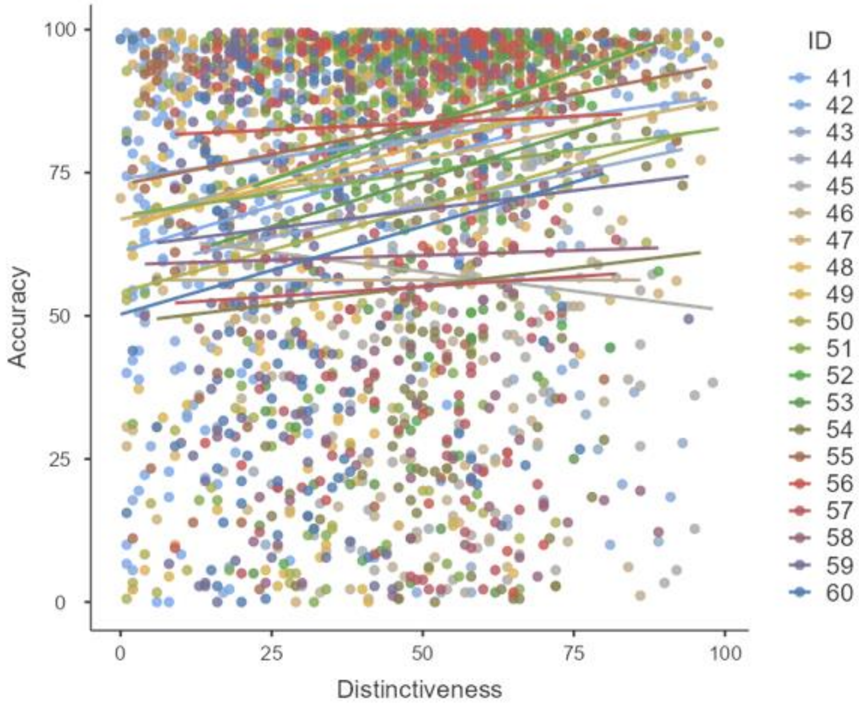

Figure 88S. Distinctiveness: Encoding (mixed list-type) bivariate scatterplot of accuracy dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is not significant

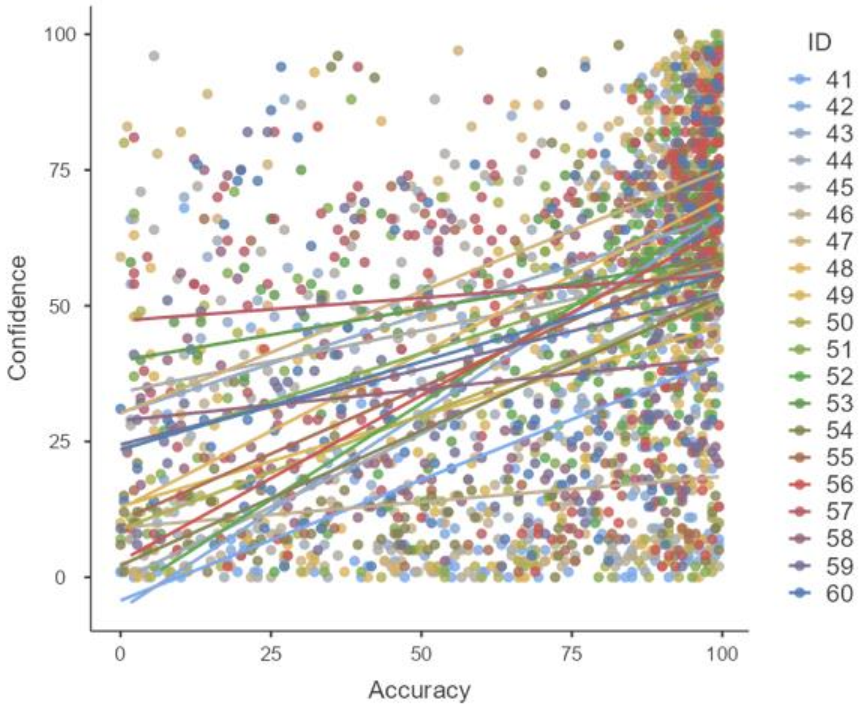

Figure 89S. Distinctiveness: Encoding (mixed list-type) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is generally large and positive

Finally, I present within-participant correlations for distinctiveness judged at encoding with images presented in a homogenous *list-type*, Table 32S. The overall averaged correlations are shown for comparison in Table 18. The association between distinctiveness judged at encoding for images shown in a homogenous list-type and source memory accuracy for recollection of the cross location is not significant. The correlation between accuracy and confidence remained moderate to large and is generally significant.

Table 32S. Distinctiveness: Encoding (homogeneous list-type). Individual within-participant correlation analyses showing separate correlations between distinctiveness, confidence and accuracy. The averaged correlation coefficients are shown, used as a comparison in Table 18

Encoding Homogenous	**Distinctiveness & Confidence**		**Distinctiveness & Accuracy**		**Accuracy & Confidence**	
Participant	Pearson's r	p value	Pearson's r	p value	Pearson's r	p value
61	0.387	**<.001**	0.147	0.109	0.545	**<.001**
62	0.206	**<.001**	0.036	0.694	0.143	0.119
63	0.127	0.168	0.169	0.065	0.111	0.23
64	0.257	0.005	0.183	0.045	0.391	**<.001**
65	0.457	**<.001**	0.179	0.051	0.439	**<.001**
66	0.15	0.01	0.074	0.42	0.31	**<.001**
67	0.119	0.197	-0.072	0.432	0.039	0.674
68	0.181	0.047	0.092	0.318	0.346	**<.001**
69	0.082	0.373	0.033	0.721	0.48	**<.001**
70	0.355	**<.001**	0.108	0.24	0.484	**<.001**
71	0.255	0.005	0.123	0.182	0.525	**<.001**
72	0.539	**<.001**	0.326	**<.001**	0.351	**<.001**
73	0.297	**<.001**	0.097	0.29	0.234	0.01
74	0.041	0.657	-0.107	0.244	0.301	**<.001**
75	0.575	**<.001**	0.126	0.171	0.374	**<.001**
77	0.249	0.006	0.211	0.021	0.59	**<.001**
78	0.254	0.005	0.06	0.514	0.144	0.117
79	0.349	**<.001**	0.104	0.26	0.467	**<.001**
80	0.211	0.04	0.12	0.193	0.55	**<.001**
Average *r*	**0.268**		**0.106**		**0.359**	

While the relationship between distinctiveness and confidence is generally positive, there is more variability between individual participants with trivial or nonsignificant correlations seen in comparison to images presented in a mixed *list-type* (cf. Table 31S). Results are further illustrated in the bivariate scatterplots shown in Figures 90S-92S.

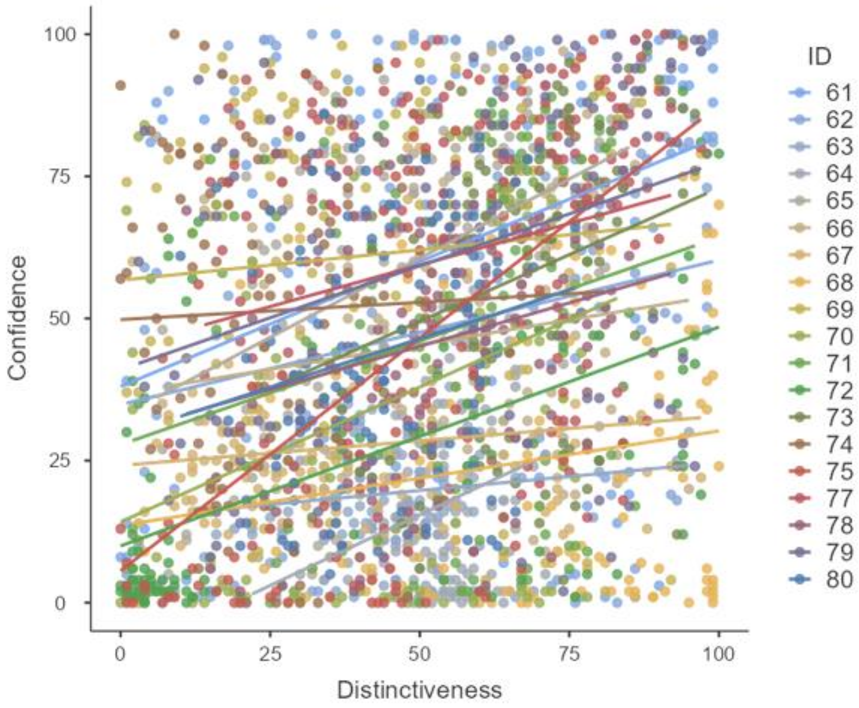

Figure 90S. Distinctiveness: Encoding (homogenous list-type) bivariate scatterplot of confidence dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is weaker and more variable than in Figure 77S

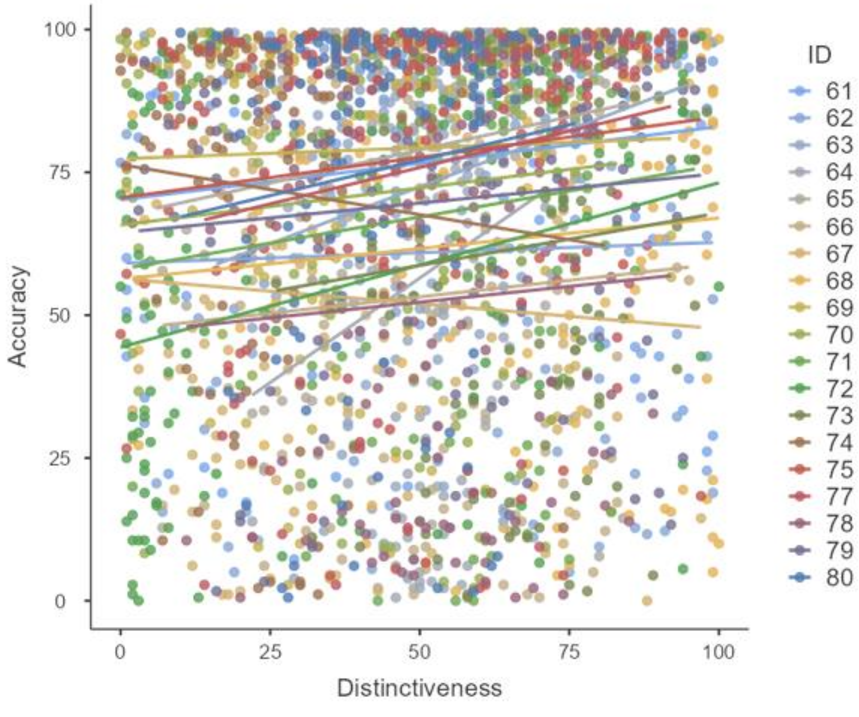

Figure 91S Distinctiveness: Encoding (homogenous list-type) bivariate scatterplot of accuracy dependent on distinctiveness. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is not significant

.

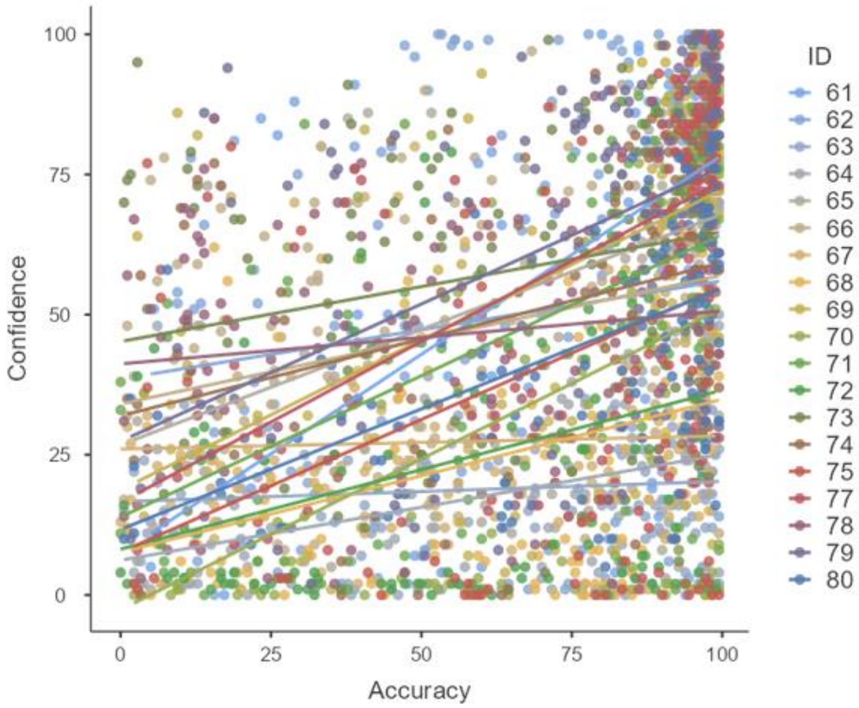

Figure 92S. Distinctiveness: Encoding (homogenous list-type) bivariate scatterplot of confidence dependent on accuracy. Individual linear regression lines are overlaid for each participant, to illustrate the overall pattern of results, which is generally large and positive

Summary.

These detailed tables and correlation plots for within-participant correlations corroborate results obtained from correlation analysis across all trials. Additionally, they give an insight into individual variations in the pattern of data. These variations are not sufficient to suggest that there are separate groups within the participants showing trends associated with Simpson’s paradox. The supplementary data confirm the conclusions and interpretation of the findings discussed in Chapter 8.

Bibliography

Ally, B.A. (2012). Using pictures and words to understand recognition memory deterioration in amnestic mild cognitive impairment and Alzheimer's disease: a review. *Current Neurology and Neuroscience Reports, 12(6),* 687–94.

Anastasi, J.S., Rhodes, M.G. & Burns, M.C. (2000). Distinguishing between memory illusions and actual memories using phenomenological measurements and explicit warnings. *The American Journal of Psychology, 113(1)*,1–26.

Anderson, J.R. & Bower, G.H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79(2),* 97-123.

Appelbaum, M., Cooper, H., Kline, R.B., Mayo-Wilson, E., Nezu, A.M. & Rao, S.M. (2018). (Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73(1),* 3-25.

Arndt, J. & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language, 39(3)*, 371–391.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience 4*, 829–839.

Baddeley, A.D. (2020). On trying to prove Endel Tulving wrong: A revised modal model of amnesia. *Neuropsychologia*, *148*, 107590.

Baddeley, A.D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14(6)*, 575–589.

Baddeley, A.D. & Warrington, E.K. (1970). Amnesia and the distinction between long-and short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9(2),*176–189.

Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia, 141,* 107408.

Bakdash, J.Z. & Marusich, L.R. (2017). Repeated measures correlation. *Frontiers in Psychology, 8,* 456.

Banks, W.P (1970). Signal detection theory and human memory. *Psychological Bulletin, 74(2),* 81-98.

Baron, R.M. & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51(6),*1173-1182.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97(4),* 548

Berry, C.J., Shanks, D.R. & Henson, R.N. (2008). A unitary signal-detection model of implicit and explicit memory. *Trends in Cognitive Sciences, 12(10),* 367–373.

Binder, J.R. & Desai, R.H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15(11)*, 527–536.

Bornstein, B.H. & Zickafoose, D.J. (1999). “I know I know it, I know I saw it”: The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied, 5(1),* 76-88. 117815

Boto, E., Hill, R. M., Rea, M., Holmes, N., Seedat, Z. A., Leggett, J., ... & Brookes, M. J. (2021). Measuring functional connectivity with wearable MEG. *NeuroImage, 230,* 117815

Brainerd, C., Gomes, C. & Nakamura, K. (2015). Dual recollection in episodic memory. *Journal of Experimental Psychology: General, 144(4),* 816-843.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in Psychology, 3(2),* 77–101.

Brewer, N. & Wells, G.L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of line up instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12(1),*11-30.

Brewer, W.F. & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language, 67(1),* 59–77.

Brewin, C.R., Andrews, B. & Mickes, L. (2020). Regaining consensus on the reliability of memory. *Current Directions in Psychological Science, 29(2),*121–125.

Brewin, C.R., Li. H., Ntarantana, V., Unsworth, C.& McNeilis, J. (2019). Is the public understanding of memory prone to widespread "myths"? *Journal of Experimental Psychology: General, 148(12),* 2245-2247.

Brigham, J.C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. *Basic and Applied Social Psychology, 11(1),* 101–115.

Broers, N., & Busch, N. A. (2021). The effect of intrinsic image memorability on recollection and familiarity. *Memory & Cognition, 49(5),* 998-1018.

Brown, R. & Kulik, J. (1977). Flashbulb memories. *Cognition, 5(1),* 73–99.

Brunel, L. Oker, A., Riou, B. & Versace. R. (2010). Memory and consciousness: trace distinctiveness in memory retrievals. *Consciousness and Cognition, 19(4),* 926–37.

Buchner, A. & Wippich, W. (2000). On the Reliability of Implicit and Explicit Memory Measures. *Cognitive Psychology, 40*, 227-259.

Buchsbaum, B.R., Lemire-Rodger, S., Fang, C., & Abdi, H. (2012). The neural basis of vivid memory is patterned on perception. *Journal of Cognitive Neuroscience, 24(9),* 1867–1883.

Burt, D.B., Aylward, E.H. & members of the working group for the establishment of criteria for the diagnosis of dementia in individuals with intellectual disability, (2000). Test battery for the diagnosis of dementia in individuals with intellectual disability. *Journal of Intellectual Disability Research, 44(2),* 175–180.

Busey, T.A., Tunnicliff, J., Loftus, G.R. & Loftus, E.F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7(1),* 26–48.

Bylinskii, Z. Isola, P., Bainbridge. C., Torralba, A. & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research, 116,* 165–178.

Cabeza, R. & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences, 11(5),* 219–27.

Carter, J.V., Pan, J., Rai, S.N. & Galandiuk, S. (2016). ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery, 159(6),*1638–1645.

Cha, J. & Dobbins, I.G. (2021). Critical tests of the continuous dual-process model of recognition. *Cognition, 215*, 104827.

Chandler, C., (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & cognition, 22(3*), 273–280.

Chappell, M. & Humphreys, M.S. (1994). An auto-associative neural network for sparse representations: Analysis and application to models of recognition and cued recall. Psychological Review, 101(1),103-128.

Chua, E.F., Hannula, D.E. & Ranganath, C. (2012). Distinguishing highly confident accurate and inaccurate memory: Insights about relevant and irrelevant influences on memory confidence. *Memory, 20(1),* 48–62.

Coan, J.A. (1997). Lost in a shopping mall: an experience with controversial research. *Ethics & Behavior, 7(3),* 271–84.

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin., 112(1),* 1155–159.

Colciago, A. & Magnaghi, V. (2016). Neurosteroids involvement in the epigenetic control of memory formation and storage. *Neural Plasticity*, article ID 5985021.

Colciago, A. Casati, L, Negri-Cesi, P. & Celotti, F. (2015). Learning and memory: steroids and epigenetics. *The Journal of Steroid Biochemistry and Molecular Biology, 150*, 64–85.

Colloff, M. F., & Wixted, J. T. (2020). Why are lineups better than showups? A test of the filler siphoning and enhanced discriminability accounts. *Journal of Experimental Psychology: Applied, 26(1),* 124-143.

Cook, G.I., Marsh, R.L. & Hicks, J.L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(4)*, 828-835.

Cooper, E., Greve, A. & Henson, R.N. (2017). Assumptions behind scoring source versus item memory: Effects of age, hippocampal lesions and mild memory problems. *Cortex, 91,* 297–315.

Cooper, R.A., Kensinger, E.A. & Ritchey, M. (2019). Memories Fade: The Relationship Between Memory Vividness and Remembered Visual Salience. *Psychological Science,* 1–12.

Corey, D.M., Dunlap, W.P. & Burke, M.J. (1998). Averaging correlations: Expected values and bias in combined Pearson *r*s and Fisher's *z* transformations. *The Journal of General Psychology, 125(3)*, 245–261.

Craik, F. I. (2006). Distinctiveness and memory: Comments and a point of view. In Hunt, R.R. & Worthen, J.B. (Eds.), *Distinctiveness and memory*, (pp.425-442). Oxford University Press, USA.

Craik, F. I. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Cuperus, A.A. Laken, M., van Schie, K., Engelharde, I.M., van den Houte, M.A. (2019). Dual tasking during recall of negative memories or during visual perception of images: Effects on vividness and emotionality. *Journal of Behavior Therapy and Experimental Psychiatry, 62*, 112–116.

Curran, T. & Cleary, A.M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research, 15(2),* 191-205.

de Zubicaray, G. I., McMahon, K. L., Dennis, S., & Dunn, J. C. (2011). Memory strength effects in fMRI studies: A matter of confidence. *Journal of Cognitive Neuroscience, 23(9),* 2324-2335.

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports, 5(3),* 305–312.

Delay, C.G. & Wixted, J.T. (2020). Discrete-state versus continuous models of the confidence-accuracy relationship in recognition memory. *Psychonomic Bulletin \& Review,* 1–9.

DeSoto, K.A. & Roediger, H.L (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science, 25(3),* 781–788.

Diana, R.A. & Ranganath, C. (2011). Recollection, familiarity and memory strength: confusion about confounds. *Trends in Cognitive Sciences, 15(8),* 337–338.

Dijkstra, N., Bosch, S.E. & van Gerven, M.A. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience, 37(5),* 1367–1373.

Dobbins, I.G., Kroll, N.E. & Liu, Q. (1998). Confidence-accuracy inversions in scene recognition: a remember-know analysis. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 24(5),* 1306–1315.

Dobbins, I. G., Kroll, N. E., Yonelinas, A. P. & Liu, Q. (1998). Distinctiveness in recognition and free recall: The role of recollection in the rejection of the familiar. *Journal of Memory and Language, 38*(4), 381–400.

Dodson, C. S., Bawa, S., & Slotnick, S. D. (2007). Aging, Source memory, and Misrecollections. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(1),* 169-181.

Doerksen, S. & Shimamura, A.P. (2001). Source memory enhancement for emotional words. *Emotion, 1(1),* 5-11.

Doss, M.K., Picart, J.K. & Gallo, D.A. (2018). The dark side of context: Context reinstatement can distort memory. *Psychological Science, 29(6),* 914–925.

Doss, M.K., Picart, J.K. & Gallo, D.A. (2020). Creating emotional false recollections: Perceptual recombination and conceptual fluency mechanisms. *Emotion, 20(5),* p.750-760.

Dunn, J.C. (2004). Remember-know: a matter of confidence. *Psychological Review, 111(2),* 524-542.

Durbin, K.A., Mitchell, K.J. & Johnson, M.K. (2017). Source memory that encoding was self-referential: The influence of stimulus characteristics. *Memory, 25(9),* 1191–1200.

Edery-Halpern, G. & Nachson, I. (2004). Distinctiveness in flashbulb memory: comparative analysis of five terrorist attacks. *Memory*, 12(2), 147–57.

El Haj, M., Nandrino, J-L., Kessels, R.C.P., & Ndobo, A. (2021). High emotional experience during autobiographical retrieval in women with Korsakoff syndrome. *Cognitive Neuropsychiatry, 26(2),* 136–148.

Ensor, T.M., Bancroft, T.D. & Hockley, W.E., (2019). Listening to the Picture-Superiority Effect. *Experimental Psychology, 66(2),*134-153.

Epley, N. & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17(4),* 311–318.

Evans, J.R. & Fisher, R.P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. Applied *Cognitive Psychology, 25(3),* 501–508.

Evans, K.K. & Baddeley, A. (2018). Intention, attention and long-term memory for visual scenes: It all depends on the scenes. *Cognition, 180,* 24–37.

Evans, L.H. & Wilding, E.L. (2012). Recollection and familiarity make independent contributions to memory judgments. *Journal of Neuroscience, 32(21),* 7253–7257.

Fandakova, Y., Johnson, E.G. & Ghetti, S. (2021). Distinct neural mechanisms underlie subjective and objective recollection and guide memory-based decision making. *Elife, 10,* 62520.

Fang, J., Ruther, N., Bellebaum, C., Wiskott, L.& Cheng, S. (2018). The interaction between semantic representation and episodic memory. *Neural Computation, 30(2),* 293–332.

Fitzgerald J, M. (1988). Vivid Memories and the Reminiscence Phenomenon: The Role of a Self Narrative. *Human Development 31*, 261-273.

Fleming, S.M. & Dolan, R.J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1594),*1338–1349,

Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R. J. & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329(5998),* 1541–1543.

Ford, J. H. & Kensinger, E. A. (2016). Effects of internal and external vividness on hippocampal connectivity during memory retrieval. *Neurobiology of Learning and Memory,* 134 Pt A, 78–90.

Ford, J.H. & Kensinger, E.A. (2017). Prefrontally-mediated alterations in the retrieval of negative events: Links to memory vividness across the adult lifespan. *Neuropsychologia,* 102, 82–94.

Forstmann, B.U., Ratcliff, R. & Wagenmakers, E-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual review of psychology, 67,* 641–666.

Fox, J. & Osth, A.F. (2022). Does source memory exist for unrecognized items? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48(2),* 242-271.

Gallo, D.A. (2013). Retrieval expectations affect false recollection: Insights from a criterial recollection task. *Current Directions in Psychological Science, 22(4)*, 316–323.

Gallo, D.A., Cotel, S.C., Moore, C.D., and Schacter, D.L. (2007). Aging can spare recollection-based retrieval monitoring: the importance of event distinctiveness. *Psychology and Aging, 22(1),* 209–13.

Gallo, D.A., Meadow, N.G., Johnson, E.L. & Foster, K.T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58(4)*, 1095–1111.

Gao, A.F. Keith, J.L., Gao, F-q., Black, S.E., Moscovitch, M. & Rosenbaum, R.S. (2020). Neuropathology of a remarkable case of memory impairment informs human memory. Neuropsychologia, 140, p107342.

Gao, A.F. Keith, J.L., Gao, F-Q., Black, S.E., Moscovitch, M. & Rosenbaum, R.S. (2020). Neuropathology of a remarkable case of memory impairment informs human memory. *Neuropsychologia, 140*, 107342.

Gardiner, J.M. (1988). Recognition failures and free-recall failures: Implications for the relation between recall and recognition. *Memory & Cognition,16(5),* 446–451.

Gardiner, J.M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16(4)*, 309–313.

Gardiner, J.M. (2001). Episodic memory and autonoetic consciousness: a first-person approach. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 356(1413)*,1351–1361.

Gardiner, J.M., Ramponi, C. & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7(1)*, 1–26.

Gardiner, J.M., Ramponi, C. & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory, 10(2)*, 83–98.

Geraci, L., McCabe, D.P. & Guillory, J.J. (2009). On interpreting the relationship between remember-know judgments and confidence: The role of instructions. *Consciousness and Cognition, 18(3)*, 701–709.

Ghosh, V.E. & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114.

Gillund, G. & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91(1)*,1-67.

Goetschalckx, L., Moors, J. & Wagemans, J. (2019). Incidental image memorability. *Memory*, 27 (9) 1273-1282.

Gonsalves B. & Paller K.A. (2002). Mistaken memories: remembering events that never happened. *Neuroscientist. 8(5),* 391-395.

Gonsalves, B. & Paller, K.A. (2000). Neural events that underlie remembering something that never happened. *Nature Neuroscience, 3(12),* 1316-1321.

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

Grabman, J. H., Dobolyi, D. G., Berelovich, N. L., & Dodson, C. S. (2019). Predicting high confidence errors in eyewitness memory: The role of face recognition ability, decision-time, and justifications. *Journal of Applied Research in Memory and Cognition, 8(2),* 233-243.

Grande, X., Berron, D., Horner, A. J., Bisby, J. A., Düzel, E., & Burgess, N. (2019). Holistic recollection via pattern completion involves hippocampal subfield CA3. *Journal of Neuroscience, 39(41),* 8100-8111.

Gretz, M.R. & Huff, M.J. (2020). Multiple species of distinctiveness in memory? Comparing encoding versus statistical distinctiveness on recognition. *Memory, 28(8),* 984–997.

Greve, A., van Rossum, M.C.W. & Donaldson, D.I. (2007). Investigating the functional interaction between semantic and episodic memory: convergent behavioral and electrophysiological evidence for the role of familiarity. *NeuroImage, 34(2),* 801–814.

Gronlund, S.D., Edwards, M.B. & Ohrt, D.D. (1997). Comparison of the retrieval of item versus spatial position information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(5),* 1261-1274.

Habermas, T., & Diel, V. (2013). The episodicity of verbal reports of personally significant autobiographical memories: Vividness correlates with narrative text quality more than with detailedness or memory specificity. *Frontiers in Behavioral Neuroscience, 7,* 00110.

Habermas, T., Diel, V. & Welzer, H. (2013). Lifespan trends of autobiographical remembering: Episodicity and search for meaning. *Consciousness and Cognition, 22(3),* 1061–73.

Haist, F., Shimamura, A.P. & Squire, L.R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 18(4),* 691–702.

Harlow, I. M. & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review, 20(2),* 318–325.

Harlow, I. M. & Yonelinas, A. P. (2016). Distinguishing between the success and precision of recollection. *Memory, 24(1),* 114–127.

Hassabis, D. & Maguire, E.A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences, 11(7),* 299–306.

Hayes, A.F. & Rockwood, N.J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour research and therapy, 98,* 39–57

Hayes, A.F. & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24(10*),1918–1927.

Hayes, A.F., Montoya, A.K. & Rockwood, N.J. (2017). The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. *Australasian Marketing Journal, 25(1),* 76–81.

Heathcote, A. Freeman, E., Etherington, J., Tonkin, J. & Bora, B. (2009). A dissociation between similarity effects in episodic face recognition. *Psychonomic Bulletin & Review, 16(5),* 824–831.

Hembacher, E. & Ghetti, S. (2017). Subjective experience guides betting decisions beyond accuracy: evidence from a metamemory illusion. *Memory, 25(5),* 575–585.

Heuer, F. & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18(5),* 496-506.

Hicks, J.L. & Marsh, R.L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5),* 1195-1209.

Hilford, A., Glanzer, M., Kim, K. & Maloney, L.T. (2019). One mirror effect: The regularities of recognition memory. *Memory & Cognition, 47(2),* 266–278.

Hintzman, D.L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review, 87(4),* 398-410.

Hintzman, D.L. (2001). Similarity, global matching, and judgments of frequency. *Memory & Cognition, 29(4),* 547–556.

Hockley, W. E. (2022). Two dichotomies of recognition memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale,* 1196-1961.

Howe, M.L. (2011). The adaptive nature of memory and its illusions. *Current Directions in Psychological Science, 20(5),* 312–315.

Hunt, R.R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review,* 2(1), 105–112.

Hunt, R.R. (2006). The concept of distinctiveness in memory research. In Hunt, R.R. & Worthen, J.B. (Eds.), *Distinctiveness and memory,* (pp.3-25). Oxford University Press, USA.

Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science*, *22*(1), 10–15.

Hunt, R.R. & Lamb, C.A. (2001). What causes the isolation effect? J*ournal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1359-1366.

Hunt, R.R. & Smith, R.E. (1996). Accessing the particular from the general: The power of distinctiveness in the context of organization. *Memory & Cognition, 24(2),* 217–225.

Iida, R., Itsukusima, Y. & Mah, E.Y. (2020). How do we judge our confidence? Differential effects of meta-memory feedback on eyewitness accuracy and confidence. *Applied Cognitive Psychology, 34(2),* 397–408.

Ingram, K.M., Mickes, L. & Wixted, J.T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 38(2),* 325–39.

Ison, M.J., Quiroga, R.Q. & Fried, I. (2015). Rapid encoding of new memories by individual neurons in the human brain. *Neuron, 87(1),* 220–230.

Jacobs, C., Schwarzkopf, D.S. & Silvanto, J. (2018). Visual working memory performance in aphantasia. *Cortex, 105*, 61–73.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30(5),* 513–541.

Jacoby, L.L., Baker, J.G. & Brooks, L.R. (1989). Episodic effects on picture identification: Implications for theories of concept learning and theories of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(2),* 275-281.

Jacoby, L.L., Shimizy, Y., Daniels. K.A. & Rhodes, M.G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12(5),* 852–857.

Jacoby, L.L., Woloshyn, V. & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118(2),* 115-125.

Jacowitz, K.E. & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21(11),* 1161–1166.

Jaeger, A., Queriroz, M.C., Selmeczy, D. & Dobbins, I.G. (2020). Source retrieval under cueing: Dissociated effects on accuracy versus confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46 (8),* 1477-1493.

Jersakova, R., Allen, R.J., Booth, J., Souchay, C. & O'Conner, A.R. (2017). Understanding metacognitive confidence: Insights from judgment-of-learning justifications. *Journal of Memory and Language, 97*, 187–207.

Johns, B.T., Jones, M.N. & Mewhort, D.J. (2021). A continuous source reinstatement model of true and false recollection. *Canadian Journal of Experimental Psychology/Revue Caanadienne de Psychologie Expérimentale, 75(1),* 1 – 18.

Johnson, M.K. (2006). Memory and reality. *American Psychologist, 61(8),* p.764-771.

Johnson, M. K., Kuhl, B. A., Mitchell, K. J., Ankudowich, E. & Durbin, K. A. (2015). Age-related differences in the neural basis of the subjective vividness of memories: Evidence from multivoxel pattern classification. *Cognitive, Affective & Behavioral Neuroscience*, 15(3), 644–661.

Johnson, M.K. (2005). The relation between source memory and episodic memory: Comment on Siedlecki et al. (2005). *Psychology and Aging, 20 (3),* 529 – 531.

Johnson, M.K, Foley, M.A., Suengas, A.G. & Raye, C.L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology. General, 117(4),* 371–6.

Johnson, M.K., Hashtroudi, S. & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin, 114(1),* 3-28.

Johnson, M.K., Kuhl, B. A., Mitchell, K. J., Ankudowich, E. & Durbin, K. A. (2015). Age-related differences in the neural basis of the subjective vividness of memories:

Evidence from multivoxel pattern classification. *Cognitive, Affective & Behavioral Neuroscience, 15(3)*, 644–661.

Jones, T.C., Jacoby, L.L. & Gellis, L.A. (2001). Cross-modal feature and conjunction errors in recognition memory. *Journal of Memory and Language, 44(1),* 131–152.

Josselyn, S.A. & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science, 367(39),* eaaw4325.

Kandasamy, I., Kandasamy, W. B. V., Obbineni, J. M. & Smarandache, F. (2020). Indeterminate Likert scale: feedback based on neutrosophy, its distance measures and clustering algorithm. *Soft Computing, 24(10),* 7459–7468.

Kelley, M.R. & Nairne, J.S. (2001). Von Restorff revisited: Isolation, generation, and memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(1),* 54-66.

Kensinger, E.A., Addis, D.R. & Atapattu, R.K. (2011). Amygdala activity at encoding corresponds with memory vividness and with memory for select episodic details. Neuropsychologia, 49(4), 663–673.

Keogh, R., Pearson, J. & Zeman, A. (202)1. Aphantasia: *The science of visual imagery extremes*. In Handbook of Clinical Neurology. Eds. Barton, J.J.S. & Leff, A. Elsevier, 277–296.

Kim, H. & Cabeza, R. (2007). Trusting our memories: dissociating the neural correlates of confidence in veridical versus illusory memories. *Journal of Neuroscience, 27(45),* 12190–12197.

Kinugawa, K., Schumm, S., Pollina, M., Depre, M., Jungbluth, C., … & Dere, E., (2013). Aging-related episodic memory decline: are emotions the key? *Frontiers in Behavioral Neuroscience, 7,* 2-14.

Koch, G. E., Akpan, E., & Coutanche, M. N. (2020). Image memorability is predicted by discriminability and similarity in different stages of a convolutional neural network. *Learning & Memory, 27(12),* 503-509.

Koriat, A., Goldsmith, M. & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51(1),* 481–537.

Krantz, D.H. (1969). Threshold theories of signal detection. *Psychological Review, 76(3),* 308-324.

Kurdi, B., Diaz, A.J., Wilmuth, C.A., Friedman, M.C. & Banaji, M.R. (2018). Variations in the relationship between memory confidence and memory accuracy: The effects of spontaneous accessibility, list length, modality, and complexity. *Psychology of Consciousness: Theory, Research, and Practice, 5(1),* 3-28.

Leshikar, E.D. & Duarte, A. (2012). Medial prefrontal cortex supports source memory accuracy for self-referenced items. *Social Neuroscience, 7(2),*126–145.

Levine, L.J, Lench, H.C., Stark, C.E.L., Carlson, S.J., Carpenter, Z. K., … and Frithsen, A. (2020). Predicted and remembered emotion: tomorrow's vividness trumps yesterday's accuracy. *Memory, 28(1),* 128–140.

Lockhart, R.S. & Murdock, B.B. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74(2),* 100-109.

Loftus, E.F. (2003). Make-believe memories. *American Psychologist, 58(11),*867-873.

Loftus, E.F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12(4),* 361–366.

Luck, S.J. & Vogel, E.K., (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279-281.

Luck, S.J. & Vogel, E.K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17(8),* 391–400.

Luminet, O. & Spijkerman, R. (2017). “11 November 1918, an exceptional day!”: Flashbulb memories of the World War I Armistice in Belgium from a psychological and a historical perspective. *Memory Studies, 10(3),* 347–362.

MacKenzie, G. & Donaldson, D.I. (2007). Dissociating recollection from familiarity: electrophysiological evidence that familiarity for faces is associated with a posterior old/new effect. *NeuroImage, 36(2),*454–463.

MacKenzie, G. & Donaldson, D.I. (2009). Examining the neural basis of episodic memory: ERP evidence that faces are recollected differently from names. *Neuropsychologia, 47(13),* 2756–2765.

MacLeod, C.A. & Donaldson, D.I. (2014). PRKCA polymorphism changes the neural basis of episodic remembering in healthy individuals. *PloS one, 9(5):* e98018.

MacLeod, C.M. (2020). The butcher on the bus: A note on familiarity without recollection. *History of Psychology, 23(4),* 383-387.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. P*sychological Review, 87(3)*, 252-271.

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science, 3(5),* 390–399.

Mandler, G., Pearlstone, Z. & Koopmans, H.S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8(3)*, 410–423.

Mayford, M., Siegelbaum, S.A. & Kandel, E.R. (2012). Synapses and memory storage. *Cold Spring Harbor Perspectives in Biology, 4(6)*, a005751.

Mazzoni, G., Scoboria, A. & Harvey, L. (2010). Nonbelieved memories. *Psychological Science, 21(9)*, 1334–1340.

McDaniel, M. A. & Geraci, L. (2006). Encoding and retrieval processes in distinctiveness effects: Toward an integrative framework. In Hunt, R.R. & Worthen, J.B. (Eds.), *Distinctiveness and memory*, (pp.65-68). Oxford University Press, USA.

McDaniel, M. A., Dornburg, C. C. & Guynn, M. J. (2005). Disentangling encoding versus retrieval explanations of the bizarreness effect: implications for distinctiveness. *Memory & Cognition*, *33*(2), 270–279.

McDonough, I. M., Enam, T., Kraemer, K. R., Eakin, D. K., & Kim, M. (2021). Is there more to metamemory? An argument for two specialized monitoring abilities. *Psychonomic Bulletin & Review, 28(5)*, 1657-1667.

McElree, B., Dolan, P.O. & Jacoby, L.L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25(3),* 563-582.

Metzger, M.M. (2006). Face distinctiveness and delayed testing: Differential effects on performance and confidence. T*he Journal of General Psychology*, 133(2), 209–216.

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4(2)*, 93–102.

Mickes, L., Hwe, V., Wais, P.E. & Wixted, J.T. (2011). Strong memories are hard to scale. *Journal of Experimental Psychology. General, 140(2),* 239–257.

Mickes, L., Wixted, J.T. & Wais, P.E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review, 14(5),* 858–865.

Morita, A. & Kambara, T. (2021). Color bizarreness effects in object memory: Evidence from a recall test and eye tracking. *Color Research & Application,* col.22697, 1-10

Morris, R.G. Abrahams, S., Baddeley, A.D. & Polkey, C.E. (1995). Doors and People: Visual and verbal memory after unilateral temporal lobectomy. *Neuropsychology, 9(4),* 464-469.

Murray, J. G., Howie, C. A. & Donaldson, D. I. (2015). The neural mechanism underlying recollection is sensitive to the quality of episodic memory: Event related potentials reveal a some-or-none threshold. *NeuroImage, 120,* 298–308.

Murray, J.G., Ouyang, G. & Donaldson, D.I. (2019). Compensation of trial-to-trial latency jitter reveals the parietal retrieval success effect to be both variable and thresholded in older adults. *Frontiers in Aging Neuroscience, 11:* 179.

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. Toward an integrative framework. In Hunt, R.R. & Worthen, J.B. (Eds.), *Distinctiveness and memory*, (pp.27-46). Oxford University Press, USA.

Nelson, T.O. & Narens. L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of learning and motivation*, *26*, 125–173.

Nguyen, K. & McDaniel, M. A. (2015). The picture complexity effect: Another list composition paradox. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1026-1037.

Newman, E.J. & Lindsay, D.S. (2009). False memories: What the hell are they for? *Applied Cognitive Psychology, 23(8),* 1105–1121.

Nieznanski, M. (2020). Levels-of-processing effects on context and target recollection for words and pictures. *Acta Psychologica, 209,* 103127.

Norman, K.A., Polyn, S.M., Detre, G.J. & Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10(9),* 424–430.

O'Connor, A. R., & Moulin, C. J. (2010). Recognition without identification, erroneous familiarity, and déjà vu. *Current Psychiatry Reports, 12(3),* 165-173.

O'Connor, A. R., & Moulin, C. J. (2013). Déjà vu experiences in healthy subjects are unrelated to laboratory tests of recollection and familiarity for word stimuli. *Frontiers in Psychology, 4*, 881.

Olmos, A. & Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception, 33(12),* 1463–1473.

Öner, S. & Gülgöz, S. (2018). Remembering successes and failures: rehearsal characteristics influence recollection and distancing. *Journal of Cognitive Psychology, 30(2),* 207–221.

Onyper, S.V., Zhang, Y.X. & Howard, M.W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology. General, 139(2),* 341–64.

Osth, A.F., Fox, J., McKague, M., Heathcote, A. & Dennis, S. (2018). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language, 103*, 91–113.

Oudman, E., Rensen, Y. & Kessels, R.P. (2021). Confabulations in post-acute and chronic alcoholic Korsakoff's syndrome: a cross-sectional study conducted in two centres. *International Journal of Psychiatry in Clinical Practice*, 1-5.

Paap, K.R., Chun, E. & Vonnahme, P. (1999). Discrete threshold versus continuous strength models of perceptual recognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 53(4)*, 277-293.

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: when and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(5),* 1196-1204.

Paller, K.A., Voss, J.L. & Boehm, S.G. (2007). Validating neural correlates of familiarity. *Trends in Cognitive Sciences, 11(6),* 243–250.

Palmer, M.A., Brewer, N., Weber, N. & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19(1)*, 55-71.

Papesh, M.H., Hicks, J.L. & Guevara Pinto, J.D. (2019). Retrieval dynamics of recognition and rejection. *Quarterly Journal of Experimental Psychology, 72(9),* 2328–2341.

Parker, A., Kember, T. & Dagnall, N. (2017). Reduced effects of pictorial distinctiveness on false memory following dynamic visual noise. *Memory, 25(6),* 845–855.

Parks, C.M. & Yonelinas, A.P. (2009). Evidence for a memory threshold in second-choice recognition memory responses. *Proceedings of the National Academy of Sciences of the United States of America, 106(28),* 11515–9.

Patai, E.Z., Gadian, D, G., Cooper, J.M., Dzieciol, A. M. & Mishkin, M. (2015). Extent of hippocampal atrophy predicts degree of deficit in recall. *Proceedings of the National Academy of Sciences of the United States of America*, *112(41)*,12830–3.

Pickering, R. (2017). Describing the participants in a study. *Age and Ageing, 46(4),* 576–581.

Pillemer, D.B. Goldsmith, L.R., Panter, A.T. & White, S.H. (1988). Very long-term memories of the first year in college. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14(4)*, 709-715.

Poldrack, R.A. & Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual Review of Psychology, 67*, 587–612

Preacher, K.J. & Hayes, A.F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40(3),* 879–891.

Prebble, S.C., Addis, D.R. & Tippett, L.J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin, 139(4),* 815-840.

Rajaram, S. (1993). Remembering and knowing: two means of access to the personal past. *Memory & Cognition, 21(1),* 89–102.

Rajaram, S. (1996). Perceptual effects on remembering: recollective processes in picture recognition memory. *Journal of Experimental psychology. Learning, Memory, and Cognition, 22(2),* 365–77.

Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review, 5*(1), 71–78.

Rasch, B. & Born, J. (2013). About sleep's role in memory. *Physiological reviews, 93,* 681-786.

Ratcliff, R. & Hendrickson, A.T. (2021). Do data from mechanical Turk subjects replicate accuracy, response time, and diffusion modeling results? *Behavior Research Methods,* 10.3758.

Ratcliff, R. Smith, P.L., Brown, S.D. & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20(4),* 260–281.

Ratcliff, R., Sheu, C.-F. & Gronlund, S.D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99(3),* 518-535.

Reinitz, M.T., Peria, W.J., Seguin, J.A. & Loftus, G.R. (2011). Different confidence-accuracy relationships for feature-based and familiarity-based memories. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37(2),* 507–15.

Renoult, L. Irish, M., Moscovitch, M. & Rugg, M. D. (2019). From knowing to remembering: the semantic-episodic distinction. Trends in Cognitive Sciences, 23(12), 1041–1057.

Rey, A.E., Riou, B. & Versace, R. (2014). Demonstration of an Ebbinghaus illusion at a memory level. *Experimental Psychology 61(5)*, 378-384.

Reyna, V.F. (2000). Fuzzy-trace theory and source monitoring: An evaluation of theory and false-memory data. *Learning and Individual Differences, 12(2),* 163–175.

Reyna, V.F., Corbin, J. C., Weldon, R.B. & Brainerd, C.J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition, 5(1),* 1–9.

Rhodes, M.G. & Castel, A.D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology. General, 137(4),* 615–625.

Rhodes, M.G. (2019). Metacognition. *Teaching of Psychology, 46(2),* 168–175.

Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *elife, 5*, e18260.

Rissman, J. Chow, T.E., Reggente, N. & Wagner, A.D. (2016). Decoding fMRI signatures of real-world autobiographical memory retrieval. *Journal of Cognitive Neuroscience, 28(4),* 604–620.

Robinson, M.D., Johnson, J.T. & Robertson, D.A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology. Applied, 6(3),* 207–21.

Roediger, H.L. & DeSoto, K.A. (2014). Confidence and memory: assessing positive and negative correlations. *Memory 22(1),* 76–91.

Roediger, H. L., & Gallo, D. A. (2004). Associative memory illusions. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory.* (pp.309-326). New York: Psychology Press.

Roediger, H.L. & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(4),* 803-814.

Roediger III, H.L. & McDermott, K.B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9(4),* 123–127.

Roediger, H.L. & Tekin, E. (2020). Recognition memory: Tulving's contributions and some new findings. *Neuropsychologia, 139*, 107350.

Rotello, C.M. & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28(6)*, 907–922.

Rubin, D.C., Deffler, S.A. & Umanath, S. (2019). Scenes enable a sense of reliving: Implications for autobiographical memory. *Cognition, 183,* 44–56.

Rust, N.C. & Mehrpour, V. (2020). Understanding image memorability. *Trends in Cognitive Sciences, 24(7),* 557–568.

Santangelo, V., Di Francesco, S. A., Mastroberardino, S. & Macaluso, E. (2015). Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory. *Human Brain Mapping, 36(12),* 5003–5017.

Schacter, D.L., Cendon, D.L., Dodson, C.S. & Clifford, E. R. (2001). Retrieval conditions and false recognition: testing the distinctiveness heuristic. *Psychonomic Bulletin & Review, 8(4),* 827–33.

Schacter, D.L., Israel, L. & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40(1),* 1–24.

Schmidt, C.R., Schmidt, S.R. & Wilson, K.A. (2021). Distinctiveness and priority in free recall of words. *Memory*,29(1), 21–38.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory, and Cognition,11*(3), 565-578.

Schurgin, M.W., Wixted, J.T. & Brady, T.F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, *4(11),* 1156–1172. Journal of Memory and Language, 65(4), 363–377.

Scimeca, J. M., McDonough, I. M. & Gallo, D. A. (2011). Quality trumps quantity at reducing memory errors: Implications for retrieval monitoring and mirror effects. *Journal of Memory and Language, 65(4),* 363–377.

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied*, *24(3)*, 400-415.

Shaw, P. A., Pepe, M. S., Alonzo, T. A. & Etzioni, R. (2009). Methods for Assessing Improvement in Specificity When a Biomarker is Combined with a Standard Screening Test, Statistics in *Biopharmaceutical Research, 1, (1),* 18-25.

Shepard, R.N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6(1),* 156–163.

Shing, Y.L. Werkle-Bergner, M., Brehmera, Y., Muller, V., Shu-Chen Lia S-C, & Lindenberger, U. (2010). Episodic memory across the lifespan: the contributions of associative and strategic components. *Neuroscience and Biobehavioral Reviews, 34(7),* 1080–91.

Shipstead, Z., Lindsey D.R.B., Marshall. R.L., & Engle, R.W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*, 116–141.

Siedlecka, M., Paulewicz, B. & Wierzchon, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology, 7,* 218.

Siefke, B. M., Smith, T. A., & Sederberg, P. B. (2019). A context-change account of temporal distinctiveness. Memory & cognition, 47(6), 1158-1172.

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: should Fisher's z transformation be used? *Journal of Applied Psychology, 72(1),* 146-148.

Simons, J. S., Dodson, C. S., Bell, D., & Schacter, D. L. (2004). Specific-and partial-source memory: effects of aging. *Psychology and Aging, 19(4),* 689-694.

Simons, J.S., Peers, P.V., Mazuz, Y. S., Berryhill, M.E. & Olson, I.R. (2009). Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. *Cerebral Cortex, 20(2),* 479–485.

Sitzman, D.M., Rhodes, M.G. & Kornell, N. (2016). The influence of feedback on predictions of future memory performance. *Memory & Cognition, 44(7),* 1102–1113.

Slotnick, S.D. (2013). The nature of recollection in behavior and the brain. *Neuroreport, 24(12)*, 663–70.

Slotnick, S.D., Klein, S.A., Dodson, C.S. & Shimamura A.P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(6),* 1499-1517.

Smith, S.A. & Mulligan, N.W. (2018). Distinctiveness and the attentional boost effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(9),* 1464-1474.

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

Spence, I., Wong, P., Rusan, M., & Rasteger, N. (2006). How color enhances visual memory for natural scenes. *Psychological Science*, *17(1),* 1–6.

Squire, L.R. (1986). Mechanisms of Memory. *Science, New Series, 232, (4758),* 1612-161.

Squire, L.R. & Wixted, J.T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience, 34*, 259–88.

St. Jacques, P.L. & Schacter, D.L. (2013). Modifying memory: Selectively enhancing and updating personal memories for a museum tour by reactivating them. *Psychological Science, 24(4),* 537–543.

Starns, J.J. & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language*, *70*, 36–52.

Starns, J.J. (2021) High-and low-threshold models of the relationship between response time and confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(4),* 671-684.

Starns, J.J., Rotello, C.M. & Hautus, M.J. (2014). Recognition memory zROC slopes for items with correct versus incorrect source decisions discriminate the dual process and unequal variance signal detection models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(5),*1205-1225.

St-Laurent, M., Abdi, H. & Buchsbaum, B.R. (2015). Distributed Patterns of Reactivation Predict Vividness of Recollection. *Journal of cognitive Neuroscience, 27(10),* 2000–2020.

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions, 8(1)*, 1–10.

Stretch, V. & Wixted, J.T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology. Learning, memory, and cognition, 24(6),* 1397–410.

Strube, M. J. (1988). Averaging correlation coefficients: Influence of heterogeneity and set size. *Journal of Applied Psychology, 73(3),* 559-568.

Sumrall, W., Sumrall, R., & Doss, D.A. (2016) A Review of Memory Theory. *International Journal of Humanities and Social Science, 6(5),* 23-30.

Suzuki, K. & Takahashi, R. (1997). Effectiveness of I in picture recognition memory. *Japanese Psychological Research, 39*(1), 25–32.

Talarico, J.M. & Rubin, D.C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14(5),* 455–461.

Talarico, J.M. & Rubin, D.C. (2007). Flashbulb memories are special after all; in phenomenology, not accuracy. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 21(5),* 557–578.

Tanner J.W.P. & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological Review, 61(6),* 401-409.

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography,* 6(1), 35–39.

Tekin, E. & Roediger, H.L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cognitive Research: Principles and Implications, 2(1),* 1–13.

Tekin, E., DeSoto, K., A., Wixted., J., H., & Roediger, H.L. (2021). Applying confidence accuracy characteristic plots to old/new recognition memory experiments. *Memory, 29(4),* 427-443.

Thakral, P. P., Wang, T. H., & Rugg, M. D. (2015). Cortical reinstatement and the confidence and accuracy of source memory. *Neuroimage, 109,* 118-129.

Thomson, D.M. & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86(2),* 255-162.

Tibon, R., Fuhrmann, D., Levy, D.A., Simons., J. S. & Henson, R.N. (2019). Multimodal integration and vividness in the angular gyrus during episodic encoding and retrieval. *Journal of Neuroscience, 39(22),* 4365–4374.

Tkačik, G., Garrigan, P., Ratliff, C., Milčinski, G., Klein, J. M., Seyfarth, L. H., … Balasubramanian, V. (2011). Natural images from the birthplace of the human eye. *PloS One, 6(6),* e20409.

Troyer, A.K., Winocur, G., Craik, F.I.M. & Moscovitch, M. (1999). Source memory and divided attention: Reciprocal costs to primary and secondary tasks. *Neuropsychology, 13(4)*, 467-474.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior, 20(5),* 479–496.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26(1)*, 1–12.

Tulving, E., (2001). Episodic memory and common sense: how far apart? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 356 (1413),*1505–1515.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53(1),* 1–25.

Tulving, E. (2008). On the law of primacy. *Memory and mind: A festschrift for Gordon H. Bower*, 31–48.

Tulving, E., & Rosenbaum, R. S. (2006). What do explanations of the distinctiveness effect need to explain. In Hunt, R.R. & Worthen, J.B. (Eds.), *Distinctiveness and memory*, (pp. 407-423). Oxford University Press, USA.

Tulving, E. & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80(5),* 352-373.

Tulving, E. & Watkins, M.J. (1973). Continuity between recall and recognition. *The American Journal of Psychology, 86(4),* 739–748.

Umanath, S. & Coane, J.H. (2020). Face Validity of Remembering and Knowing: Empirical Consensus and Disagreement Between Participants and Researchers. *Perspectives on Psychological Science,* 1-23.

Van der Plas, M., Wang, D., Brittain, J-S., & Hanslmayr, S. (2020). Investigating the role of phase-synchrony during encoding of episodic memories using electrical stimulation. *Cortex, 133,* 37–47.

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. & Cogdill, M. (2017). A categorical recall strategy does not explain animacy effects in episodic memory. *The Quarterly Journal of Experimental Psychology, 70(4),* 761–771.

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, 18, 299–342.

Voss, J.L. & Paller, K.A. (2010). What makes recognition without awareness appear to be elusive? Strategic factors that influence the accuracy of guesses. *Learning & Memory, 17(9),* 460–468.

Waddill, P.J. & McDaniel, M.A. (1998). Distinctiveness effects in recall. *Memory & Cognition, 26(1)*, 108–120.

Wais, P.E., Squire, L.R. & Wixted, J.T. (2010). In search of recollection and familiarity signals in the hippocampus. *Journal of Cognitive Neuroscience, 22(1),* 109–123.

Watkins, M.J. & Tulving, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General, 104(1),* 5-29.

Weigl, M., Pham, H. H., Mecklinger, A. & Rosburg, T. (2020). The effect of shared distinctiveness on source memory: An event-related potential study. *Cognitive, Affective, & Behavioral Neuroscience, 20(5),* 1027–1040.

Weissgerber, T. L., Milic, N. M., Winham, S.J. & Garovic, V.D. (2015). Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS biology, 13(4),* e1002128.

Whitehouse, A.J., Maybery, M.T. & Durkin, K. (2006). The development of the picture-superiority effect. *British Journal of Developmental Psychology, 24(4)*, 767–773.

Whittlesea, B.W.A. (1993). Illusions of familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(6), 1235-1253.

Whittlesea, B.W., Jacoby, L.L. & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. Journal of Memory and Language, 29(6), 716–732.

Wichmann, F. A., Sharpe, L. T. & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 28(3),* 509–520.

Williams, H.L. & Lindsay, D.S., (2019). Different definitions of the nonrecollection-based response option(s) change how people use the “remember” response in the remember/know paradigm. *Memory & Cognition, 47(7),* 1359–1374.

Wilson, B.M., Harris, C.R. & Wixted, J.T. (2020). Reply to Pek et al.: Science is not the signal detection problem it is ordinarily thought to be. *Proceedings of the National Academy of Sciences, 117(24),* 13201–13202.

Wilson, J.C. & Westerman, D.L. (2018). Picture (im) perfect: Illusions of recognition memory produced by photographs at test. *Memory & Cognition, 46(7),*1210–1221.

Wilson, T.D. & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin, 116(1),* 117-142.

Wixted, J.T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114(1),* 152–76.

Wixted, J.T. (2018). Time to exonerate eyewitness memory. *Forensic Science International, 292*, e13–e15.

Wixted, J.T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46(2),* 201-233.

Wixted, J.T. & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117(4),* 1025–1054.

Wixted, J.T. & Mickes, L. (2015). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition, 4(4),* 318–323.

Wixted, J.T. & Squire, L.R. (2011). The medial temporal lobe and the attributes of memory. *Trends in Cognitive Sciences, 15(5)*, 210–217.

Wixted, J.T. & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107 (2),* 368-376.

Wixted, J.T. & Wells, G.L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18(1),* 10–65.

Wolfe, J.M. (2020). Visual search: How do we find what we are looking for? *Annual Review of Vision Science, 6,* 539–562.

Wong, J. T., Cramer, S. J. & Gallo, D. A. (2012). Age-related reduction of the confidence-accuracy relationship in episodic memory: Effects of recollection quality and retrieval monitoring. *Psychology and Aging, 27(4),* 1053–1065.

Wu, H., & Leung, S. O. (2017). Can Likert scales be treated as interval scales?—A Simulation study. *Journal of Social Service Research, 43(4),* 527-532.

Yonelinas, A.P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 20(6), 1341–1354.

Yonelinas, A.P. (1999). Recognition memory ROCs and the dual-process signal-detection model: Comment on Glanzer, Kim, Hilford, and Adams. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(2)*, 514-521.

Yonelinas, A.P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46(3),* 441–517.

Yonelinas, A.P., Aly, M., Wang, W-C. & Koen, J. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus, 20(11),* 1178–1194.

Yonelinas, A.P., Dobbins, I., Szymanski, M.D., Dhaliwal, H.S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5(4),* 418–441.

Zawadzka, K., Higham, P.A. & Hanczakowski, M. (2017). Confidence in forced-choice recognition: What underlies the ratings? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(4),* 552-564.

Zhang, W. & Luck, S.J. (2011). The number and quality of representations in working memory. *Psychological Science, 22(11)*, 1434–1441.

Zhao, X., Wang, C., Liu, Q., Xiao, X., Jiang, T., Chen, C., & Xue, G. (2015). Neural mechanisms of the spacing effect in episodic memory: a parallel EEG and fMRI study. *Cortex, 69*, 76–92.

Zhou, J. Osth, A.F., Lilburn, S.D., & Smith, P.L. (2021). A circular diffusion model of continuous-outcome source memory retrieval: Contrasting continuous and threshold accounts. *Psychonomic Bulletin & Review*, 1–19.

Zlotnik, G. & Vansintjan, A. (2019). Memory: An extended definition. *Frontiers in Psychology, 10*, 2523.

Zou, F. & Kwok, S.C. (2022). Distinct generation of subjective vividness and confidence during naturalistic memory retrieval in the angular gyrus. *Journal of Cognitive Neuroscience*, 1-13.

Zylberberg, A., Roelfsema, P.R. & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition, 27*, 246–253.