EÖTVÖS LORÁND UNIVERSITY
FACULTY OF PEDAGOGY AND PSYCHOLOGY
DOCTORAL SCHOOL OF PSYCHOLOGY, COGNITIVE PSYCHOLOGY PROGRAM
Head of Doctoral School: Dr. Oláh Attila
Head of the Cognitive Psychology Program: Dr. Czigler István

Nagy Márton

Recollection in the light of eye movements:

The relational eye movement effect and its role in recollection

Supervisors:
Dr. Király Ildikó
Dr. Kónya Anikó

BUDAPEST, 2017

Introduction

The introduction of the dissertation will have six main parts. We will start with the review of the concept of episodic memory in chapter 1. In the next part we will introduce two separate approaches to study episodic memory: the phenomenological (chapter 2) and the so-called representational approach (chapter 3). In both of these chapters we will introduce the main concepts and methods the approaches use to study episodic memory. In chapter 3 our special focus will be on the process of recollection and how this process is related to episodic memory retrieval. We will also introduce the two-stage model of recollection, which can be viewed as a theoretical framework that merges the two main approaches of episodic memory. In chapter 4 we will present evidence that eye movement measures can be successively used to signal various memory processes. We will end the introduction with an important finding (Hannula & Ranganath, 2009) that suggested that eye movements indicate automatic, unconscious relational memory retrieval. Based on previous findings (Hannula, Ryan, Tranel & Cohen, 2007; Hannula & Ranganath, 2009) we will present the main assumption that the relational eye movement effect is a universal and necessary indicator of relational memory retrieval that can be regarded as the behavioural marker of the first stage of recollection. In chapter 5 we will present our experiments, which tested specific predictions derived from the main assumption. In the final chapter (chapter 6) we will draw our general conclusions related to the relational eye movement effect and the process of recollection.

The field of memory research is multidisciplinary with special focus on neuroscience to answer questions about the neural implementation of different memory functions. Although, the forthcoming experiments were inspired by neuroimaging results, they use a purely experimental approach. We will use our results to highlight important aspects of the methodology and theory making related to recollection and we will limit our conclusions to the cognitive level of theorizing.

Chapter 1: Episodic memory

Early years

The concept of episodic memory was introduced by Endel Tulving (1972) at a symposium on the organization of memory, hosted by Wayne Donaldson and him (Tulving & Donaldson, 1972).

The introduction of episodic memory to the community was a reaction to authors who were using the term 'semantic memory' (e.g., Collins & Quillian, 1969) to refer to a structured network of concepts, words and images which is capable of making inferences, comprehending languages and solve problems (Tulving, 1972). These authors were interested in creating a unitary framework of long term memory (LTM) that stores our knowledge of the world and uses language. As a reaction to these models Tulving (1972) expressed a possible taxonomic distinction between episodic and semantic memory systems. He defined episodic memory as a memory system, which receives, stores and retrieves information about temporally dated events and temporal-spatial relations among these events. According to his definition the episodic system represents a more or less faithful record of personally experienced events in terms of their perceptible properties (Tulving, 1972). He also provided some examples of memory expressions based on episodic memory: (a) I remember seeing a flash of light a short while ago, followed by a loud sound a few seconds later; (b) Last year, while on my summer vacation, I met a retired sea captain who knew more jokes than any other person I have ever met; (c) I remember that I have an appointment with a student at 9:30 tomorrow morning; (d) One of the words I am sure I saw in the first list I studied was LEGEND; (e) I know the word that was paired with DAX in this list was FRIGID.

On the other hand the semantic memory system in Tulving's interpretation is a mental thesaurus representing a person's knowledge about words, symbols, concepts, rules, formulas and algorithms for the manipulation of these symbols and concepts (Tulving, 1972). Here are some examples of information retrieval from the semantic system: (a) I remember that the chemical formula for common table salt is NaCl; (b) I know that summers are usually quite hot in Kathmandu; (c) I know that the name of the month that follows June is July, if we consider them in the order in which they occur in the calendar; (d) I think that the association between the words TABLE and CHAIR is stronger than that between the words TABLE and NOSE.

In the 1972 paper Tulving also speculates that the two systems may differ in several points: (1) the nature of the stored information (personal events vs. knowledge of the world, concepts; (2) autobiographical vs. cognitive reference (episodic memories refer to ones own past vs. semantic memory refers to an existing cognitive structure); (3) conditions and consequences of retrieval (episodic memory can only retrieve an event which was encoded before vs. semantic memory is capable of inferences and generalization to retrieve something which was not personally learned); (4) their vulnerability to interference (interference causes retrieval deficit in episodic memory vs. semantic memory may be less vulnerable to interference); (5) their dependence upon each other (they are in interaction but the specific link is not well understood). This separation of memory systems was meant to inspire and guide future research in this area and the semantic-episodic dichotomy had an enormous impact on forthcoming memory research. The distinction became a part of the general taxonomy of human memory, which is widely accepted nowadays in the research community (Squire, 1992, 2004; Henke, 2010).

Interestingly, the part of the introductory 1972 paper that describes a typical episodic memory task had a major influence on research: 'the participant must remember that such and such an item occurred at such and such a time, in such and such a temporal relation to other items' (Tulving, 1972, p. 10). This description of an episodic memory task was later used as a behavioural criterion for episodic retrieval: the retrieval of what-where-when information related to a study event may signal the existence of episodic memory (Clayton, 2015). This 'wide' criterion – because of the ease it can be translated to measureable behaviour – was used to test whether animals have the capacity for episodic memory. The measurement problem of episodic memory with purely behavioural tasks in the laboratory is one of the main questions of memory research and we will address the question later on in this introduction (see chapter 3).

During the forthcoming years the concept of episodic memory was elaborated by Tulving (1985a, 1985b, 1995, 2002, 2005) and its defining characteristics shifted to focus on the self, autonoetic awareness and subjectively sensed time (Baddeley, 2001). In the following section we will summarize the important parts of Tulving's concepts, which work encompasses more than three decades. We will also mention some problematic issues of the definition of the concepts ('semantic', 'episodic') and address important questions regarding the relationship of the proposed memory systems. The following summary of the model is not intended to be a full review, instead we will try to give the reader the broad points of the theory with the intention to make the origin of the concept of episodic memory more understandable.

Multiple memory systems approach

The original episodic-semantic dichotomy was broadened to a 'monohierarchical' ternary classification (Tulving, 1985a, 1995, 2001). A memory system is thought of as an organized structure of more elementary operating components, basically mental processes. The components of a system always have a neural basis and behavioural, cognitive correlates. However there is no one-to-one match between memory tasks and memory systems they are systematically related and the contribution of systems to tasks are thought to be measurable. Tulving's ternary classification identifies three different major memory systems: perceptual, semantic and episodic. The three systems constitute a *monohierarchy*, which means that moving up in the hierarchy the upper system is a specialized subsystem, which depends on and supported by the lower level system.

It is interesting how the system at the bottom of the hierarchy changed throughout the years (original idea see Tulving, 1985a; later see Tulving, 1995). At first, the lowest level system was the procedural system. This system was thought to be responsible for the acquisition, retention and utilization of perceptual, cognitive and motor skills. The acquisition in this system required behavioural responses and the stored representation was a blueprint for future action without any information about the past (Tulving, 1985a). The procedural system was later replaced by the perceptual representation system (PRS) (Tulving, 1995) at the bottom of the hierarchy (Figure 1). In this updated version of the hierarchy the procedural system was described as the 'action' system with the capacity of motor and cognitive skill learning and it was separated from the 'representation' systems (PRS, semantic, episodic) organized in a hierarchy.

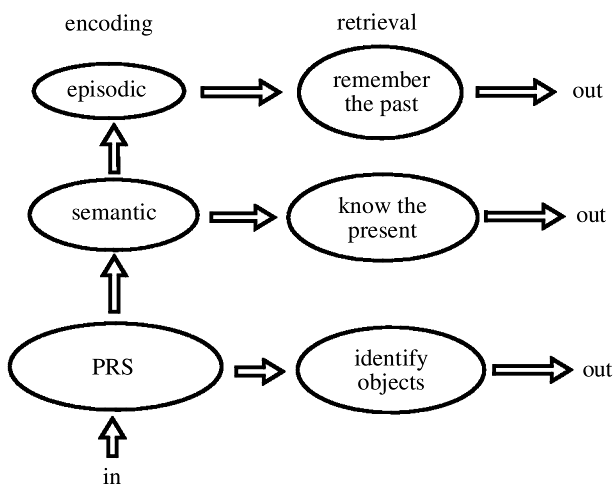

Figure 1. Tulving (2001). The ternary classification scheme (PRS, semantic, episodic) and the relation between systems at memory encoding and retrieval (SPI model).

The PRS (Tulving, 1995, 2001) operates with the information about the perceptual features of the physical environment (e.g. object identification). It also makes possible stimulus-response type of associative learning on the perceptual level and it is also responsible for perceptual priming effects (Tulving, 2001).

The semantic system makes possible the acquisition, retention and retrieval of our knowledge of the world without any contextual detail. It makes us capable of applying cognitive operations on aspects of the world beyond immediate perception (Tulving, 1995). It holds the symbolic representation of our knowledge of the world in general, dealing with 'facts' of the world (Tulving, 2001). Semantic memory goes beyond language or meaning (Tulving, 1995, 2001).

The episodic system supports the acquisition, retention and retrieval of personally experienced events and their relationship to other events in subjective time

(Tulving, 1985a). Later (Tulving, 2002, 2005) it was specified as the conjunction of three characteristics: (1) there has to be someone how is capable of time travel, this is the self; (2) the rememberer has to be capable to subjectively sense time, (3) there is a unique type of conscious awareness during retrieval of an episode, which is called autonoetic (self-knowing) consciousness (Tulving, 2002). Retrieval from episodic memory is the conscious re-experiencing of a previously encountered event, which is called 'recollective experience' (Tulving, 1983). Our capacity to episodic retrieval makes us capable of 'mental time travel', which is an expression used to refer to the mental projection of the self backwards in time (episodic memory) and into the future (episodic future thinking) (Tulving, 2002; Maguire & Mullaly, 2013).

The relation between memory systems: The SPI model

The proposed relationship between the memory systems is process-specific (Tulving, 1995, 2001). Process-specificity states that the type of relationship depends on the memory process: encoding, storage or retrieval. At encoding the relationship between the systems is serial (the 'S' part of SPI stands for serial). Information is passed from the lowest level to the highest: first the PRS receives an input and it can – although not necessarily - pass the information to the semantic system, which might also pass it to the episodic system. This serial relationship holds that there is no 'direct' input from the perceptual system to episodic memory. Information has to go through the PRS first to be able to serve as an input to the semantic system. Then it also has to go through the semantic system to enter the episodic system. This serial encoding has important implications to amnesia. Regarding anterograde amnesia (AA) - which is the condition of memory impairment after brain damage - there can't be intact episodic memory and impaired semantic memory. This double dissociation is not possible, because episodic memory at encoding needs input from semantic memory.

The storage of information in the systems is parallel (the 'P' part in SPI). Parallel storage holds that different aspects of incoming information are stored separately in the systems. Perceptual information about the input is stored in the PRS while its conceptual and semantic aspects are in the semantic system, and the information about the involvement of the self in the episodic system (Tulving, 2001).

At memory retrieval the three systems work independently from each other (the 'I' part in SPI). This assumption holds that when information is retrieved from one system it doesn't need to have any implications for retrieval of information from another system (Tulving, 2001). Most of the time the systems work together in a memory task but retrieval can occur only from one system. This independent retrieval assumption has also implication to amnesia. Retrograde amnesia patients - who have memory impairment for information before the brain damage - can show double dissociation regarding semantic and episodic memory. Episodic or semantic memory can be impaired selectively while the other one functioning at a normal level. This is possible because the retrieval from one system is independent from other systems.

Major questions regarding Tulving's framework

Tulving constantly expressed in his works that his model is a hypothetical, psychological model of memory (he even used the term 'armchair speculation' related to his original proposals (Tulving, 1983)) and extensive research is needed to clarify the details of the concepts and their relations. The existence of episodic memory in the sense that there seems to be a capacity of memory, which can encode, store and retrieve unique events from the personal past, is not questioned in modern memory research. Additionally, the separation of semantic and episodic retrieval have become a widely used heuristic in memory research: a great number of researchers agreed that there is a difference between retrieving decontextualized, general knowledge from memory (semantic) and remembering unique personal events (Greenberg, 2011). Neuropsychological case studies showed that memory for facts can be intact and episodic memory impaired (cases like N.N. (Tulving, 1985b), K.C. (Tulving, 2002)).

However the details of Tulving's proposals related to the existence of multiple memory systems, the clear-cut distinction between semantic and episodic memory and their specific relationships are all under lively debate in the literature. Here we only mention the major questions raised by some researchers.

The general critics stated that the framework is vague and it violates the law of parsimony of the field (Tulving, 2002). Some had problems with introducing memory systems and then make up complex relationships between them. Others didn't see it justified talking about facts about the self and facts about the world in terms of sharply different categories. Below, we will only focus on questions and problems related to the framework if one accepts the theoretical distinctions and tries to think about the relationship between memory types.

Questions:

1. *Where does our general knowledge of the world (semantic memory) come from?*

Tulving hadn't specified the development of such knowledge. He mentions that 'human semantic memory has evolved from the spatial learning and knowledge of the ancestors of humans' (Tulving, 1995, p. 841), however he is not talking about the developmental source of the *content* of semantic memory but its origins as a memory system. There is no mention in his framework about the development of its content. One way of thinking about the origins of our general knowledge of the world is that it is based on the unique events we encounter during our life, episodic memory. This framework suggests that general, conceptual knowledge is based on decontextualized episodic memories (Baddeley, 1988; Conway, 2009). However the SPI model goes against this view: the encoding to episodic memory has to go through the semantic system. There seems to be some developmental evidence in favour of the view that the episodic system is not a prerequisite for building up our general knowledge. The neuropsychological case studies of developmental amnesia (Vargha-Kadem et al., 1997) show that these patients (especially one of them with bilateral hippocampal damage at birth, Jon) have normal levels of factual knowledge and language use along a pronounced impairment of their ability to recollect personal events. However these cases show that semantic knowledge can develop in these patients it is unclear whether they have any basic forms of episodic capacity intact which could help them develop their knowledge (Conway, 2009). Additionally, there is also evidence that episodic memory helps the gradual formation of general knowledge (Conway, Gardiner, Perfect, Anderson & Cohen, 1997). The exact mechanism of how our conceptual knowledge develops still remains an open question.

2. *When we retrieve episodic memories can these memories be independent of semantic memory?*

Tulving's standpoint in the SPI model is that episodic retrieval can be independent of retrieval from any other system. However, Reder, Park and Kieffaber (2009) for example suggest that episodic memories are the bindings of semantic concepts to specific contexts in which they appeared. In other words in this framework episodic memory is the synergy of semantic memory and contextual information (Glenberg & Verfaellie, 2010) and it is not possible that an episode is ever retrieved completely separately from its semantic building blocks.

3. *Can we think about the episodic and semantic distinction as a continuum of memory expressions?*

This question is related to other criticism about Tulving's separation of semantic and episodic memories as part of distinct memory systems. We could also think about the semantic-episodic distinction as the results of different retrieval conditions from a single memory system (Baddeley, 1988). For example the answer to a memory task is different whether we are asked to remember what we did yesterday evening or whether we are asked what is the capital of Hungary. The differences between our answers do not necessarily mean that we are using different memory systems to retrieve information. We can think about semantic and episodic retrieval as a continuum of memory expressions (Cabeza & St Jacques, 2007). One extreme endpoint of this continuum is the general, decontextualized memory (e.g. lifetime periods) and the other endpoint is the specific event memory (e.g. unique episode). There is the possibility to retrieve memories from the middle of the continuum (e.g. repeated events). These frameworks pose questions to the separate memory systems approach and raise the possibility that memories can be more general (simple?) or more specific (complex?) without the need to presume different systems behind their expression.

4. *Can there be a direct input to the episodic system from the perceptual system?*

According to the SPI model the episodic system receives input only from the semantic system and there is no direct link between the perceptual system and the episodic system. Simons, Graham, Galton, Patterson and Hodges (2001) showed that there could be a direct link between perceptual and episodic system. In one of their experiments (2001, experiment 2), semantic dementia patients learned pictures of known and unknown celebrities. Later, they were tested in two different test conditions. In one, the test faces were perceptually the same as those seen initially; in the other the test faces were perceptually different. The logic was that if patients encountered unknown faces, their semantic memory couldn't contribute (or contribute less) to recognize them later independently of the test conditions. The results showed that the patients had no difficulty recognizing unknown or known faces in the test condition with perceptually identical faces, while they had severe deficit in recognizing unknown faces compared to known faces when the test contained perceptually different pictures of the faces. The authors suggested that the results support their view that episodic learning can rely on information coming directly from the perceptual system and the SPI model needs to be changed accordingly. Nevertheless, the results point to that direction, there are concerns about the experimental methodology. Simons et al. (2001) used simple yes-no recognition tasks to test episodic memory, which is a problematic measurement (Tulving, 2001). However the suggestion that a person can encode purely perceptual episodes makes good common sense (Tulving, 2001).

5. *What is the relationship between the framework's memory systems and other taxonomies of memory?*

Traditionally, semantic and episodic memories are both thought to be part of declarative (explicit) memory (Squire, 2004; Henke, 2010). However the semantic system is vague and it is not always clear whether the same system is behind the conscious retrieval of 'general facts' of the world (Squire, 1992, 2004) and conceptual priming. Priming is treated as part of non-declarative memory (Squire, 1992, 2004) and conceptual priming experiments are clearly relying on semantic memory in the sense that they are based on the conceptual information of a presented word to elicit unconscious memory effects. The categorization of semantic memory to declarative or non-declarative memory seems to depend on the task. When one is asked about general facts about the world the answers are conscious (e.g., Budapest is the capital of Hungary) and semantic retrieval is treated as part of declarative memory. However in the case of a free association priming task, when the participant in biased to give highly associated answers in a free association task (e.g. noon) to previously learnt words (e.g. lunch), the effect is unconscious, thus semantic memory can be treated as part of non-declarative memory.

The categorization of episodic memory to declarative/explicit memory is widely accepted. The capacity to rapidly encode and consciously retrieve personal events and the ability to share these unique events with others are all regarded as the hallmarks of declarative/explicit memory (Squire, 2004; Henke, 2010).

Chapter 2: The main contribution of Tulving's framework: The phenomenological approach to episodic memory

The framework introduced by Tulving has major unanswered questions and problems, however its semantic-episodic separation clearly became one of the most influential and widely used dichotomies in modern memory research. We believe that its success relies on the insightful observation that retrieving general facts from memory has a different subjective experience than remembering yesterday's lunch. This observation was the basis of a paper, which described the special type of conscious awareness during recollective experiences (Tulving, 1985b).

In his phenomenological approach, Tulving (1985b) linked information retrieval from the memory systems with different levels of conscious awareness. The retrieval from the lowest level system, the perceptual representation system is 'anoetic' (non-knowing), which means that we are not aware of the information retrieved from the system. The semantic memory retrieval is accompanied by 'noetic' (knowing) consciousness, which allows us to be aware of the retrieved memory content, and operate on objects and events in the absence of them and independently of the self. Episodic retrieval is accompanied by 'autonoetic' (self-knowing) consciousness ('autonoesis'), which gives the special phenomenal flavour of remembering a personal event connected to the self in time. This special kind of consciousness is a necessary feature of episodic memory and it can be measured (Tulving, 1985b). He illustrated the lack of autonoesis with the case study of N.N., a patient who suffered a closed head injury and consequently lost his ability to retrieve personal episodes of his own past (Tulving, 1985b). Tulving also introduced the remember-know task to measure the phenomenal experience of retrieval, in which participants have to introspectively distinguish between remembering (autonoetic) or knowing (noetic) that certain items appeared before in a study test. This task is meant to study the subjective states of memory retrieval during a recognition task in the laboratory and we will discuss it in detail as part of the dual-process models of recognition (see chapter 3).

In sum, episodic memory is our memory for personally experienced events that is explained by the conjunction of three ideas (Tulving, 2001, 2002): the self, autonoetic awareness, and subjectively sensed time. These three characteristics are behind the phenomenological approach (see Figure 2) of episodic memory, which holds that this kind of memory is human-specific and it is the remembering of personal events accompanied by 'autonoetic' awareness. We are capable to sense subjective time and extend it both backwards with 'remembering' past events and forward into the future by 'thinking about the future' (Tulving, 2001). The focus is on the unique experience of remembering (autonoetic awareness) and episodic future thinking, in other words 'mental time travel'. How is it possible to measure episodic memory?

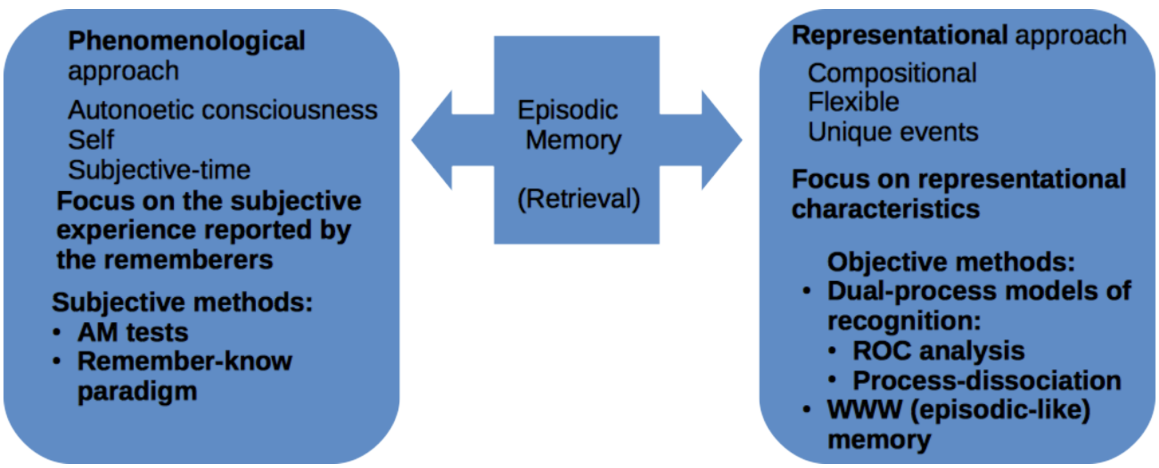

Figure 2. The phenomenological and the Representational approach of episodic memory retrieval.

Subjective methods to measure episodic memory: personal memories from outside the laboratory

Based on the phenomenological approach of episodic memory, one research program focuses on the subjective experience of complex and constructive episodic retrieval when participants recall their personal past (Conway et al., 1999; Levine, Svoboda, Hay, Winocur & Moscovitch, 2002; Kopelman, Wilson & Baddeley, 1989; Piolino, Desgranges & Eustache, 2009). This research is concentrated on autobiographical memory (AM) instead of laboratory memories (LM).

Autobiographical memory is thought to be an abstract hierarchical knowledge structure of our personal lifetime periods and general events, which gives our episodic memories a conceptual context and which is also connected to our conceptual self (Conway, 2009). Episodic memories provide specificity to autobiographical memories and they are part of the autobiographical knowledge structure. In autobiographical research participants give their verbal accounts of specific past events,. There are different autobiographical test methods (e.g., *Autobiographical Memory Inventory* (AMI, Kopelman et al., 1989); *TEMPau task* (Piolino et al., 2009); *Autobiographical Interview* (AI, Levine et al., 2002); *Galton-Crovitz task* (Conway et al, 1999). In all these tasks participants are asked to retrieve events from different life periods of their personal past, which has to be unique and specific regarding its time and place. Usually the retrieval condition is free and only the lifetime period is given to participants as restriction for recall (e.g., AMI, AI) or they have to answer to a specific word cue (e.g., Galton-Crovitz task). Sometimes after free recall the participants are also prompted with more specific cues to retrieve as much episodic detail from the experience as possible (e.g., AI). Based on verbal reports, the transcripts are rated by coders on different categories, which focus on various elements of episodic details (e.g., spatiotemporal, perceptual, emotional, cognitive) and also on non-episodic (repeated, semantic, general) details of the event. Sometimes the rememberers are also asked to complete a self-assessment rating of their memories (TEMPau task) to provide subjective measures of their recollective experiences. This approach is important because it helps us describe the nature of episodic memory as *memories*, which we retrieve and share with others in the form of speech. Using these methods it is also possible to collect personal memories and ask participants to remember them while we measure their brain activity (for a review see: Cabeza & St. Jacques, 2007) to assess different brain-areas contributing to memory retrieval. Cabeza and St. Jacques (2007) highlighted three important advantages of AM research over laboratory methods. Firstly, it is the best method to measure remote memory. Laboratory research is usually not creating remote (10-20-years old) memories. Using AM and neuroimaging can help us measure which brain regions are involved in the retrieval of remote personal memories and what are the differences and similarities between brain networks behind remote and recent memory retrieval. Secondly, AM research is focused on complex and constructive processes related to memory retrieval, which is difficult to capture with usually simple laboratory stimuli. Thirdly, methods used to study AMs are well suited to study recollective qualities, which are typically absent or impoverished in laboratory memories, such as emotional content, self-relevance and vividness.

This method also has a number of limitations. Asking participants to provide personal events makes it difficult to assess the accuracy of these memories and there is also a high risk that they will tend to report events, which were retrieved several times prior to the test. This could result in elaborated, well-rehearsed verbal reports based on multiple retrieval occasions and not 'raw' episodic memories. Multiple retrieval occasions are also a problem for a number of neuroimaging studies, where memories are collected via pre-scan interviews and later they are retrieved again in the scanner. Additionally, using free recall the experimenter has little control over the age and the content of retrieved memories (Cabeza & St. Jacques, 2007). Altogether, the phenomenological approach collecting verbal reports (AM tests) sacrifices experimental control to subjective experiences of complex memories. The subject of its analysis is the verbal report of the rememberers, which is closely bound to the self and the subjective experience of remembering. It is also important to highlight that this approach, despite its focus on subjective reports, uses verbal indicators to infer the level of autonoetic consciousness or re-experiencing of the event. These indicators are usually the perceptual, spatiotemporal, cognitive and affective contextual details of the original event (Levine et al., 2002; Piolino et al., 2009). These characteristics are used to pinpoint episodicity of the experience and to separate the episodic and semantic parts of autobiographical memories.

Another method to measure the phenomenological experience of memory retrieval was introduced by Tulving (1985b) as the remember-know task. This paradigm is an introspective method, which is based on the participants' judgments of their subjective experiences during memory retrieval. Because this task was extensively used in laboratory studies of recognition memory and because its results are relevant to dual-process models of recognition, we will discuss it later in chapter 3.

The measurement problem of autonoetic consciousness: towards objective criteria

If one is interested in the components and the fundamental mental processes involved in how the remembering of a personal event is 'put together', then laboratory induced memories are needed. Creating memories in the laboratory gives us the tools to control most of the aspects of memory encoding and memory retrieval. Using controlled experiments we are able to test causal relationships between variables and ultimately test theories of memory. The question is how can we measure the phenomenal characteristic of episodic retrieval with laboratory methods? What behavioural criterion can we use to measure autonoetic awareness? Defining a behavioural criterion to measure the subjective experience of retrieval is not a new problem. Tulving (2002) expressed that 'there is no necessary correlation between behaviour and conscious experience' (page 4). Tulving (2002) points out that in traditional memory experiments the researchers assumed that the measured behaviour (e.g. a 'yes' response in a recognition task) is a faithful index of the participants' conscious experience of remembering. However later on, for example implicit learning studies showed that there could be behavioural evidence of memory in the absence of any conscious awareness of the study episode or without a conscious experience of any memory usage. Additionally, recognition research showed that participants could introspectively use two different types of conscious awareness to consciously recognize an item as 'old' (Gardiner, 2008). This shows that there can be at least three different sources behind a simple 'old' judgment in a recognition memory task: implicit responses, conscious but 'anoetic' response ('know') and autonoetic responses ('remember'). The challenge for memory research is to come up with a behavioural task, where the participant can succeed only if he/she possesses the capacity of autonoetic consciousness. This task would be the objective measurement of a subjective mental experience.

Tulving (2005) suggested such a task, which he named the Spoon Test (Tulving, 2005, page 43). It is a future-based test of autonoetic consciousness, which doesn't need introspection and language. Tulving's idea for this task was based on an Estonian children's story. A little girl dreams that at a friend's birthday party they serve her favourite pudding, but she is not allowed to eat from it unlike all the other guests, because she doesn't have her own spoon. The next day the little girl, not to have the same disappointing experience again, goes to sleep holding a spoon. Tulving's view is that the little girl shows that she can mentally travel in time and that she has autonoetic consciousness because she solves the problem of the spoon by bringing her own the next day to bed.

Contrarily to Tulving, we think that the girl's behaviour can be explained without any need to project herself into the future and imagine the she will need the spoon to eat from the pudding. She can just know, without any autonoetic consciousness (self-projection to the future), based on her unpleasant experience, that a spoon generally could be a helpful tool for her and choose to hold a copy of the item close to her. Interestingly, there are experimental findings, which show that chimpanzees (Osvath & Osvath, 2008) and scub-jays (Raby, Alexis, Dickinson & Clayton, 2007) can pass the Spoon Test (see later in chapter 2).

There is no sufficient behavioural task yet to test autonoetic consciousness and many think that it is not possible to design such a task, because of its introspective nature. However, there is one good reason to be optimistic about the future of finding such a task. It is reasonable to think about the emergence of our capacity of episodic memory as an evolutionary advantage, compared to those who lacked such ability. This would mean that there has to be a behavioural advantage for those, who possess this special kind of awareness, consequently the phenomenon could be measured.

Chapter 3: The representational approach: episodic memories as flexible and unique relational representations of events

We will introduce the representational approach to study episodic memory, which puts aside the measurement problem of the unique subjective experience ('mental time travel') accompanying episodic retrieval and concentrates on the main cognitive preconditions of episodic memory. This approach focuses on three characteristics of episodic representations, which together provide the basis of this capacity. These aspects contain the structure, the access and the emergence of the representations (Figure 3).

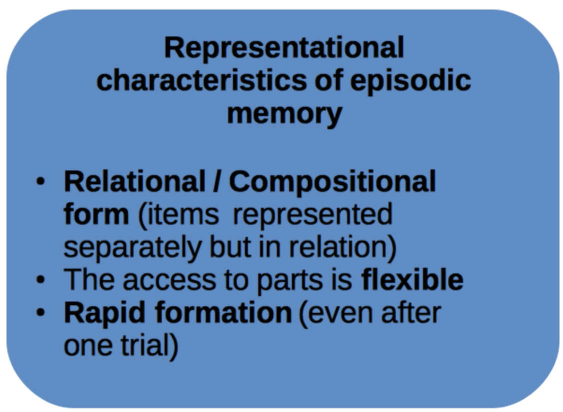

Figure 3. The representational characteristics of episodic memory

The structure of the representation: multi-element, relational representation

Episodic memories include multiple features in combination (Eichenbaum, 2004). These representations contain summary records of sensory-perceptual-conceptual –affective processing (Conway, 2009) and consist of sensory, emotional, temporal and spatial information (Henke, 2010). One of their fundamental characteristics is that they are relational representations (Yonelinas, 2013; Henke, 2010; Richmond & Nelson, 2009; Eichenbaum, 2004). They are constructed by individual elements of experience, which are bind together in a complex, multi-faceted form. Episodic memories are compositional, which entails the simultaneous and independent representation both of the separate elements and their relations (Henke, 2010; Cohen, Poldrack and Eichenbaum, 1997). Because of this structure, the elements of episodic representations are not blended into an inseparable or unitized representation (Henke, 2010). These representations are also linked to each other via relational networks (Eichenbaum, 2004).

There is a specific line of research, which focuses on the structure of episodic-like memory of animals and infants (Clayton & Dickinson, 1998; Clayton, 2015). This framework based its criterion of episodic memory on an earlier definition (page 10) by Tulving (1972). In that early paper Tulving described episodic memory as the capacity to encode and retrieve the information of *what* happened, *where* and *when*. Clayton and Dickinson (1998) used the term *episodic-like memory* to emphasize the difference between this behaviourally measureable representational capacity and the phenomenological definition of episodic memory (the capacity for 'mental time travel'). They presented several experiments to demonstrate that scrub-jays are capable to retrieve and act upon their memory for what-where-when information (WWW memory). We can regard the WWW memory framework as a specific type of the representational approach of episodic memory, which is focused on the representational abilities to form multi-element event memories containing what, where and when information.

The access of the representation: flexible

Episodic memories are flexible representations, which is a consequence of their compositional structure. The elements and the relations of these elements are stored independently, which enables their separate access during retrieval. Episodes can be flexibly reactivated: their components can cue the retrieval of any combination of components. This flexibility permits the inferential use of memories in novel retrieval situations, independent of the circumstances in which the information was initially acquired (Henke, 2010; Cohen et al., 1997). Conway (2009) also proposes the possibility, that episodic representations could be easily accessed by cues without necessarily entering conscious awareness. The access or activation of episodic memories without reaching conscious awareness is an interesting topic of theories about the process of 'explicit/conscious recollection', which we will discuss later (the two-stage model of recollection: Moscovitch, 2008; Moscovitch, Cabeza, Winocur & Nadel, 2016).

The emergence of the representations: rapid formation

Episodic memories can be formed rapidly, after a one-trial learning episode (Henke, 2010). This capacity for one-trial learning is behind episodic memory's special feature to be able to represent unique events (Conway, 2009). Because of their uniqueness, episodic memories make autobiographical remembering specific (Conway, 2009). There are other processes of learning, which can rapidly encode information but these represent single or unitized elements and they can be accessed only as a unit, without the capacity to activate separate elements (e.g., priming, familiarity).

The advantage and limitation of the representational approach

The representational approach to episodic memory has the advantage to provide us with behavioural criteria to study the phenomenon. Contrarily to autonoetic consciousness, the properties of flexible and specific relational memory representations are behaviourally measureable in the laboratory. Although this approach has clear limitations (e.g., avoiding the question of subjective aspects of episodic retrieval), it gave the field important methods to measure behavioural correlates of episodic memory. Focusing on the encoding and retrieval of specific relational representations can set the ground to integrate theories of fields of human memory, visuospatial cognition and also memory across species (Rubin & Umanath, 2015). This research focus has the potential to lay down a common ground between human and animal models of memory. For example we can think about the possible function of episodic memory as a type of memory, which helps the organism to keep a specific summary record of experiences related to recent goal processing (Conway, 2009). Defining episodic memory as a species wide capacity, which might have adaptive value by informing the organism about its progress in recent goal-processing is a clear departure from Tulving's views, which states that episodic memory is human specific with autonoetic consciousness as its key feature. Our standpoint is that focusing on specific relational representations as core elements of episodic memory can bridge the gap between animal, infant and adult memory in identifying the features they share and by doing this, it can be useful to describe the cognitive building blocks of the complex phenomenon of episodic memory shared across species.

Access of information from episodic representations: The process of recollection

Recollection is a widely used term in memory research to refer to a specific memory process by which we access relational representations of events. This process is a broad category of memory retrieval inferred when we retrieve additional information about an item beyond its general oldness (Yonelinas, 2002; Konkel & Cohen, 2009), such as specific details about its original occurrence or the location and time of the encounter. The retrieval of '*additional information*' has different names in the literature: contextual retrieval, relational memory (Konkel & Cohen, 2009; Ryan, Althoff, Whitlow & Cohen, 2000; Hannula & Ranganath, 2009), source memory (Yonelinas, 1997). As we will demonstrate it later, the experimental paradigms used to measure recollection show high similarities. Recollection in general is thought to be the retrieval process behind episodic memory (Aggleton & Brown, 1999; Yonelinas, 2002; Moscovitch et al., 2016) as it provides access to the individual elements and their relations from a specific event.

In the dual-process framework of recognition memory the process of recollection is contrasted with familiarity (see below, Yonelinas, 2013), which helps to recognize items based on their global memory strength without retrieving any specific aspects of the event, where they were originally experienced. In the next section we will review the evidence provided by the dual-process models, which support the distinction of recollection and familiarity. Our aim is to introduce the main measurement methods and results of this framework to highlight the experimental evidence supporting the existence of the recollection process.

Dual-process models of recognition memory

This important field of memory research is focused on explaining how humans and animals can recognize previously seen items. A typical recognition memory experiment consists of two parts: a study phase and a test phase. During the study phase the participant is presented with a list of stimuli (e.g. words, pictures, sounds). Later, during the test phase the participant is again presented with a list, which contains items from the study phase intermixed with new items. The typical recognition task is to indicate whether the item presented in the test phase is old or new.

There are different models, which try to explain what memory process or processes are behind the recognition of a studied item. According to the widely accepted dual-process models recognition memory is based on two distinct memory processes: *recollection* and *familiarity* (Yonelinas, 2002; Yonelinas, Aly, Wang, Koen, 2010). There are three commonly used methods to measure these processes: (1) the analysis of receiver operating characteristics (ROCs), (2) the remember-know paradigm and (3) the process dissociation procedure. As we will see it later, the convergence of results of these methods shows the validity of the distinction of recollection and familiarity (Eichenbaum, Yonelinas & Ranganath, 2007).

Familiarity:

According to dual-process models familiarity provides a global signal of memory strength or stimulus recency (Yonelinas, 2010). It is often described in terms of the signal-detection theory, such that old and new items have different but overlapping Gaussian familiarity distribution levels (Yonelinas, 2001). When presenting items at the study phase of a recognition task the level of familiarity will be enhanced for these items compared to new ones. During the test phase of the task it is possible to rely only on the strength level of item-familiarity and accept items as old when their familiarity level exceeds a criterion (Figure 4). Figure 4 shows that new and old items differ in their familiarity level. We can quantify this difference as d', which is the difference between the old and new item distributions. By setting a response criterion (c) the participant accepts only items exceeding this criterion as having been studied. Familiarity-based recognition provides ‘quantitaive’ information (level of strength) without any ‘qualitative’ information about the item (which is provided by recollection).

Single-process models assume that recognition judgments rely exclusively on a single familiarity measure (Donaldson, 1996; Dunn, 2004; Yonelinas, 2001). However these single-process models have difficulties explaining associative and source recognition task ROC results (Yonelinas, 1997; Yonelinas et al., 2010), where the task is to retrieve information about the studied item beyond its oldness (e.g., its pair or whether it was part of list A or B). We will get back to these differences between single and dual process models when we summarize the evidence for dual-process models.

Figure 4 assumes that the variance of the old and new items will be equal after the study phase (equal-variance signal-detection) but there are other models, which accept that the old and new item distributions will be different (unequal-variance signal-detection).

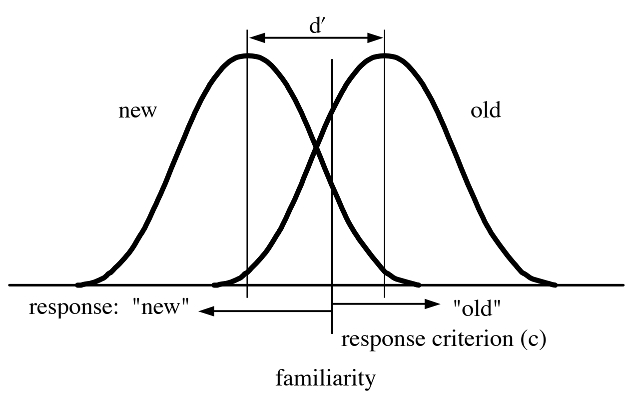

Figure 4. Equal-variance signal-detection theory of the familiarity process (Yonelinas, 2001).

Supporting evidence for the separation of recollection and familiarity

Recognition ROC curves

A quantitative method to study recognition performance is to calculate the participants hit rate and false alarm rate and visualize them as a receiver operating characteristics curve (ROC). The hit rate is the proportion of correct responses to old items (the test item is old and the participant's response is 'old'), while the false alarm rate is the proportion of incorrect responses to new items (the test item is new and the participant's response is 'old'). ROC is the function that relates hit rates and false alarm rates. To visualize performance the ROC is plotted across different levels of response confidence. For this method the participants have to make their recognition judgments on a confidence scale ranging from 'sure studied' to 'sure new', typically containing 5 or 6 different levels. Yonelinas (2001) presents two hypothetical ROCs (Figure 5). The leftmost point of the curves includes the responses with the highest confidence level and subsequent points to the right include less and less confident responses in a cumulative way.

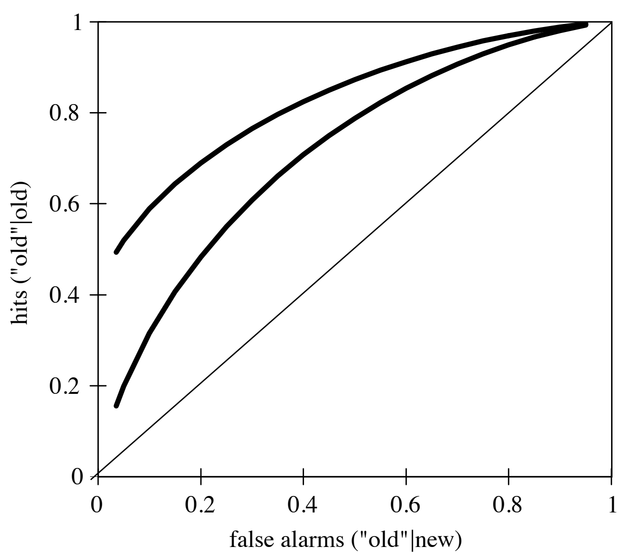

Figure 5. ROC curve examples from Yonelinas (2001).

The signal-detection model illustrated in Figure 4 produces the lower curve. It is symmetrical to the negative diagonal because of the assumption that the variance of old and new items is equal. However the upper curve shows a typical result from a recognition memory experiment. The curve is curvilinear but it is asymmetrical to the negative diagonal. The curve is pushed up along the y-axis. The asymmetry of the ROCs could be still explained with a single-process model that uses a single d' parameter for recognition accuracy and another fixed constant for the variance of the old items, which is always greater then the new item variance. This would mean that increased accuracy always have to result in greater asymmetry of the ROCs. However, ROC studies showed that the accuracy (d') and the asymmetry of the curves are functionally independent (Yonelinas, 2001). In some cases the accuracy increases while the asymmetry of the curve stays relatively constant. This independence applies that we have to use two separate memory components to explain recognition memory performance.

One can use an *unequal-variance signal detection model (UVSD)* to explain ROCs in recognition memory. This approach uses a parameter to account for the accuracy (d') and another parameter for the ROC asymmetry (the variance of the old item distribution relative to the new items). This account can explain the functional independence of accuracy and asymmetry but as we will see later it fails to predict linear ROC curves in source or associative recognition tasks (Yonelinas, 2001; Yonelinas et al., 2010).

Yonelinas (1994) proposed a *dual-process signal-detection model (DPSD)* of recognition. In this model familiarity is an equal-variance signal-detection process where the memory strength of the studied items is greater than the distribution of new items and the variance of old and new items are equal (Figure 4). Recollection on the other hand is assumed to be a threshold process. The participant either retrieves different aspects related to the item from the original event or fails to. It is believed that on some trials when recollection is below the threshold it doesn't provide any evidence that an item has been studied before. On other trials, when recollection is above the threshold it provides some contextual evidence that the item have been studied which will result in high confidence answers. The threshold theory does not make any specific assumptions related to the distribution of the recollective strength (Yonelinas et al., 2010). The model is using the degree of the curve in the ROC to estimate familiarity while the y-intercept of the ROC to estimate recollection (Yonelinas et al., 2010). It predicts that if decisions are based on familiarity, then the ROC will be curved and the y-intercept (recollection) will be 0, meaning that the ROC is symmetrical. On the other hand if recognition is based on recollection, then the ROCs y-intercept will be grater then 0 and it will be linear, because of its threshold feature. See Figure 6 for the illustration of the model's predictions.

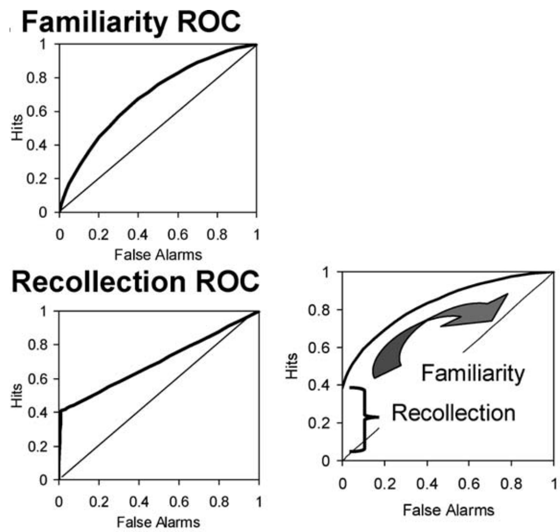

Figure 6. ROC curves of familiarity and recollection from Yonelinas (2010).

The evidence supporting the assumption of the DPSD, that familiarity is a signal-detection process is coming from patients with amnesia. The model predicts that amnesic patients, who rely exclusively on familiarity in a recognition task, will show curvilinear and symmetric ROCs. The overlapping Gaussian distribution of old and new items predicts a curvilinear ROC and the equal-variance assumption for the distributions predicts a symmetrical ROC. Figure 7 shows the results of Yonelinas, Kroll, Dobbins, Lazzara and Knight (1998), where three amnesic patients with MTL damage and matched control participants completed a recognition task. The results showed that while amnesic patients had curved and symmetrical ROCs, control patients showed asymmetrical ROCs (y-intercept > 0 + curve) indicating that they could rely both on familiarity and recollection during the task.

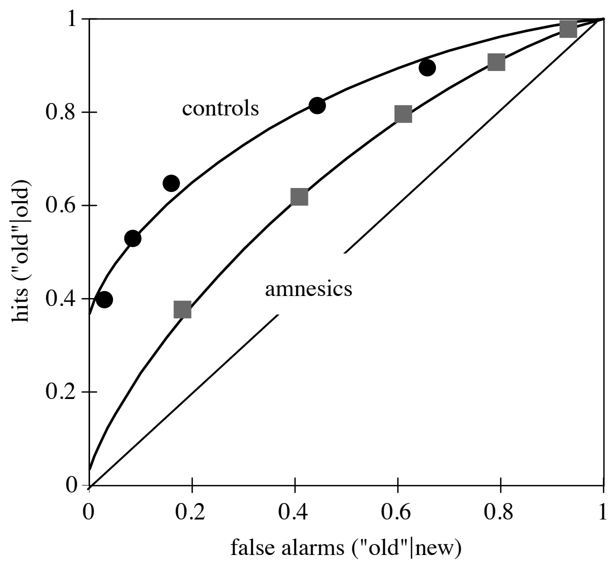

Figure 7. ROC curves of amnesics and control participants from Yonelinas et al. (1998)

The evidence supporting that recollection is a threshold process is coming from ROCs of source and associative tasks. The prediction was that if recognition relies on recollection alone, then the shape of the ROC would be close to linear. Yonelinas (1997) used an associative task to measure recognition ROCs. In the task participants had to recognize which items were paired together in a previous study list. First they studied a list of word pairs, which was followed by a recognition test containing intact and rearranged pairs. The task is to indicate whether the pair is intact or rearranged. This task is meant to rely primarily on recollection because items in intact and rearranged pairs are all studied before, thus item familiarity won't help to discriminate between them. However, the recollection of relational information (item-item information) of the study event will help participants on this associative task. The participants completed associative and item recognition tasks too, where they had to indicate at the test, whether the presented item was studied before or not (Yonelinas, 1997). The results showed that the item recognition ROCs were curvilinear and they had a clear y-intercept indicating that both familiarity and recollection contributed to task performance. However, for the associative recognition tasks the ROCs were linear (Figure 8). The DPSD theory with the threshold recollection process assumption predicts a linear ROC. These results are also important because the unequal variance signal detection theory (UVSD) always assumes curved ROCs, thus it fails to predict the associative task show results, whereas the DPSD model predicts linear ROC (Yonelinas et al., 2010).

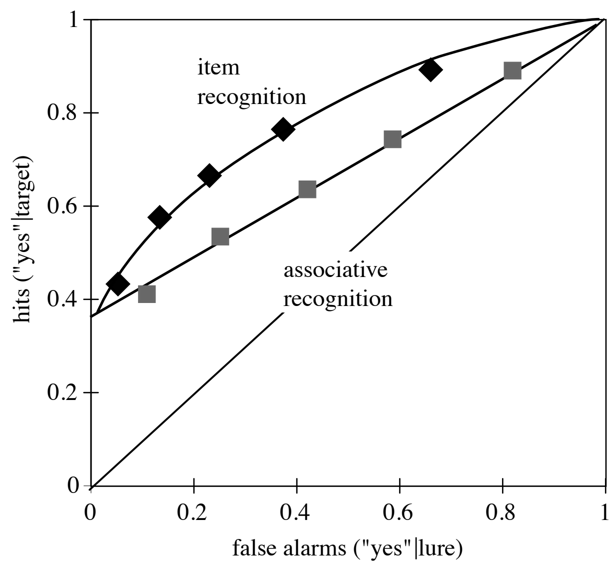

Figure 8. ROC curves of the item and associative recognition tasks from Yonelinas (1997)

In another study Yonelinas (1999) used a source memory tasks to further test the DPSD theory. In the first experiment participants studied words, which appeared either on the left or the right side of the computer screen intermixed randomly during presentation. They were instructed to remember all words and to try and remember the side of the screen on which the words were presented. Half of the participants completed an item recognition task while the other half received a source memory task. For the recognition task they had to make a recognition judgment on a 6-point confidence scale ranging from *sure it was new* to *sure it was old*. For the source memory task participants had to decide whether the presented item on the test list was on the left or the right side of the screen during study. Their task was to indicate their responses on a scale 6-point scale from *sure it was on the left* to *sure it was on the right.*

In the second experiment the source memory task was altered. At study the participants first heard a list of words spoken by a male voice followed by a second list of words spoken by a female. After the study phase they received a recognition and a source memory task. They had to fill in a booklet, which contained a list of randomly mixed new and old words. First they had to indicate on a 6-point scale whether the word was studied or was new. Then they also had to decide on a 6-point scale whether the word was spoken by a male or a female voice.

The predictions were the same as for the associative memory task. The item recognition task is based on familiarity and recollection while the source memory task is based on the recollection of specific aspects of the study trial (Exp. 1: item-location, Exp. 2: item-voice associations). Consequently the ROCs have to be curved for item recognition and not different from linear for the source memory task. The two experiments showed that for the item recognition tasks the ROCs were curved and asymmetrical while for the source memory tasks the ROCs were linear similarly to the associative tasks (similar to Figure 8). The linear ROCs again were supporting the DPSD theory predictions and were not predicted by the UVSD theory.

The remember-know paradigm

The remember-know paradigm was introduced by Tulving (1985b) to obtain subjective reports of autonoetic and noetic awareness of retrieval during memory task. After their memory answers or recognition judgments participants had to indicate whether they " 'remember' the items occurrence in the list of whether they simply 'know' on some other basis that the item was a member of the study list" (Tulving, 1985b, page 8). Tulving used the term 'remembering' for the phenomenal experience of mental time travel (autonoesis) while he used 'knowing' to refer to the experience without any self-recollection, simply the familiarity of a test item. The task was designed to measure the independent retrieval from the semantic and episodic memory systems. The paradigm was subsequently modified and extensively used in the 1990's (for reviews see Gardiner, 2001; Gardiner, 2008). Table 1 shows the usual remember-know paradigm and the typical definitions of the two subjective experience based responses.

Procedure
1. Study tasks
2. Retention interval
3. Recognition tests (*Old/New*) If *Old*, then *Remember* or *Know* (sometimes *Guess* option)
Response definitions
Old/New: Test item occurred/ did not occur in the study list
Remember: Test item brought back to mind some specific recollection of something you thought about when it occurred in the study list
Know: Test item does not bring back to mind something you thought about when it occurred in the study list, but it seemed strongly familiar.

Table 1. The usual remember-know paradigm and Remember, Know definitions from Gardiner (2008).

The typical remember-know task uses a two-step task. First there is a study task and after some retention interval a recognition test is used, where participants have to decide whether the presented item is 'Old' or 'New'. Following every 'Old' answer they are also asked to indicate whether they 'Remember' or 'Know' that the item occurred during the study (sometimes a 'Guess' response is also introduced). There are other variations of the task when a three alternative remember-know-new judgment is made immediately after every test item (one-step procedure, for example see Dewhurst, Holmes, Brandt & Dean, 2006).

The main finding of the remember-know paradigm is that these first-person memory responses are functionally dissociable. Some experimental variables have the same effect on them while other variables selectively induce or reduce the answers. Table 2 is an illustration of four different experiments, which show that the answers are experimentally independent. Based on Gardiner (2008) we will summarize these results.

Manipulation	Condition	Remember	Effect	Know	Effect
Study tasks	Associate	. 72	↑	.15	←→
	Letters	.18		.20	
Study/test modes	Visual/visual	.11	←→	.52	↑
	Visual/auditory	.10		.27	
Study/test items	Words	.28	↑	.16	↓
	NonWords	.19		.30	
Study trials	One	.14	↓	.21	↓
	Three	.37		.32	

Table 2. Effects of some experimental manipulations on remember and know responses from Gardiner (2008). Numbers indicate correct 'old' response proportions for the two answer types. Arrows represent the direction of the effect of the first condition.

The first example is a study, where participants had to either report a meaningful associate for the study word or to report two letters which were not in the presented word (Gardiner, Java & Richardson-Klavehn, 1996). This experimental manipulation is meant to create different levels of processing during encoding of the words at study. This manipulation affected only the proportion of remember responses. After deeper encoding the proportion of remember responses was higher but the study task manipulation had no effect on the proportion of know responses.

In the next study (Gregg & Gardiner, 1994) words were presented rapidly during the study phase. At test half of the test items were presented in the same modality as during the study phase (visual presentation) and half of them were presented in a different modality (auditory presentation). Recognition accuracy was higher when items were presented in the same modality and the modality change appeared to have an effect only on know responses.

In Gardiner and Java (1990) participants studied a list of intermixed items with meaningful words and pronounceable nonwords. The type of study item had no effect on overall recognition performance, however for meaningful words there were more remember responses than know responses. This pattern changed for nonword recognition; where there were more know than remember responses. Overall, nonword items reduced remember responses and induced know responses.

Gardiner and Radomski (1999) presented Polish and English folk songs to participants either once or three times during the study phase. The results showed that both remember and know responses benefited from more presentation of the items, resulting in enhanced recognition.

The functional independence of these responses is also evident in different clinical populations (Table 3). The advantage of the remember-know paradigm is that it can inform us about the relative contribution of the two states of awareness to recognition even when overall recognition levels are the same in two groups or experimental conditions. In general remembering is reduced in amnesics, old adults, patients with schizophrenia and adults with autism. Interestingly this decrease in remembering is compensated by increased levels of knowing in older adults and adults with autism, resulting in comparable levels of overall recognition between these groups and the control participants. In other cases, reduced levels of remembering is accompanied by stable levels of knowing (amnesics and schizophrenics).

Condition	**Group**	**Remember**	**Effect**	**Know**	**Effect**
Amnesia	Patients	. 21	↓	.28	←→
	Controls	.50		.25	
Age	Older adults	.17	↓	.51	↑
	Younger adults	.53		.23	
Schizophrenia	Patients	.23	↓	.34	←→
	Controls	.39		.34	
Autism	Adults with Asperger's	.36	↓	.25	↑
	Controls	.47		.11	

Table 3. Differences between remember and know responses in special populations from Gardiner (2008). Numbers indicate correct 'old' response proportions for the two answer types. Arrows represent the direction of the effect of the first condition.

Problems and advantages of the remember-know paradigm

1. Operational resemblance to recollection and familiarity

In Table 4 we present the typical and most widely used definitions of the remember and know responses (Rajaram, 1993; Yonelinas 2001; Gardiner & Java, 1990;

Gardiner 1988). Remember-know studies rarely report the exact instructions given to the participants (McCabe & Geraci, 2009). Most of the time these studies refer to one of the three instructions in Table 4, but the exact word-by-word report for them is missing. It is clear that in all of the three variants, the criterion for a remember response is the retrieval of *some* specific contextual detail or personal thought (~ mental context, association) from the study event. In our view these definitions are closer to the representational criteria of recollection than to the subjective experience of remembering ('mental time travel').

Instructions:	**Remember**	**Know**
Gardiner & Java (1990)	In this test there are four columns of words; some of these words are from the cards you studied in the first part of the experiment, others are not. Please work carefully down each column, indicating for each successive word whether you recognize it from the study cards or not. If you recognize a word, please encircle it. *Additionally,* as you make your decision about recognizing a word, I would like you to bear in mind the following: Often, when ***remembering*** a previous event or occurrence, we consciously recollect and become aware of aspects of the previous experience.	At other times, we simply ***know*** that something has occurred before, but without being able consciously to recollect anything about its occurrence or what we experienced at the time. Thus in addition to your indicating your recognition of a word from the original study set, I would like you to write either the letter "R" after the encircled item, to show that you recollect the word consciously, or "K" if you feel you simply know that the word was in the previous study set. So, for each word that you recognize, please write "R" next to it if you recollect its occurrence, or " K " if you *simply* know that it was shown on the cards.
Rajaram (1993)	"Please read the following instructions carefully. You will be presented with a booklet containing words. Work carefully down the column and indicate on the first blank next to each word whether you recognize each word from the study list. If you do recognize the word, write "Y" (for "yes"), and if you do not recognize it, then write "N" (for "no"), In addition, at the time you recognize the word, you should also write on the second blank next to the word, whether or not you remember the	**Know judgments**: "Know" responses should be made when you recognize that the word was in the study list but you cannot consciously recollect anything about its actual occurrence or what happened or what was experienced at the time of its occurrence. In other words, write "K" (for "know") when you are certain of recognizing the words but these words fail to evoke any specific conscious recollection from the study list."

	word from the list or you just know on some other basis that the word was on the study list. Please read the following instructions to find out how to make the "remember" (or "R") and "know" (or "K") judgments. **Remember judgments:** If your recognition of the word is ac-companied by a conscious recollection of its prior occurrence in the study list, then write "R." "Remember" is the ability to become consciously aware again of some aspect or aspects of what happened or what was experienced at the time the word was presented (e.g., aspects of the physical appearance of the word, or of something that happened in the room, or of what you were thinking and doing at the time). In other words, the "remembered" word should bring back to mind a particular association, image, or something more personal from the time of study, or something about its appearance or position (i.e., what came before or after that word).	
Yonelinas (2001)	Participants were told that they were to respond ***R*** only if they could remember some qualitative information about the study event. They were told that this could include such things as recollecting what they were thinking about when the word was presented, what the word looked like, or what it sounded like. Moreover, they were instructed that they should respond *R* only if they could, if asked, tell the experimenter what they recollected about that study event.	Participants were told to respond ***K*** if they thought the item was studied but could not recollect any details about the study event. They were told to respond *N* if they thought the word was not in the study list. To ensure that participants understood the test instructions, they were asked to describe the remember- know distinction back to the experimenter, and the instructions were repeated if the participant appeared to have misunderstood the distinction.
Gardiner (1988)	In addition, the test was immediately preceded by instructions explaining that at the time they recognized each word, they were also to write an "R," for "remember," if their recognition of the word was accompanied by a	"**Know**" responses were defined as the recognition that the word was in the booklet but the inability to recollect consciously anything about its actual occurrence or what happened or what was experienced

	conscious recollection of its prior occurrence in the study booklet or a "K," for "know," if they did not consciously recollect the word's occurrence in the study booklet but recognized it on some other basis. "**Remember**" was defined in these instructions as the ability to become consciously aware again of some aspect or aspects of what happened or what was experienced at the time the word was presented (e.g., aspects of the physical appearance of the word, or of something that happened in the room, or of what one was thinking or doing at that time).	at the time of its occurrence. To further illustrate this distinction, the instructions pointed out that if asked one's own name, one would typically respond in the "know" sense, that is, without becoming consciously aware of any- thing about a particular event or experience; however, when asked what movie one saw last, one would typically respond in the "remember" sense, that is, becoming consciously aware again of some aspects of the particular experience. One subject failed to write any "R" and "K" responses in the test, and this subject was replaced.

Table 4. Different widely used definitions of remember and know responses.

There is an important difference between the remember response and the responses based on the retrieval of the source or associative information. The use of a specific source/associative task always restricts the measured recollection to a single dimension. These tasks make the recollection measure narrow: they infer recollection if the participants can retrieve a specific detail of the study episode (the item was spoken by a male voice or the item was presented on the right side). Whenever participants fail to retrieve information from the measured dimension these tasks fail to indicate recollection performance. These tasks can only measure *criterial* recollection. The advantage of the remember response is that one can use it to indicate any type of recollected detail from the study event, thus it can measure *noncriterial* recollection. The inability of source or associative tasks to measure noncriterial recollection results in an underestimation of the level of recollection in these paradigms.

A know response is defined when there is a strong feeling of familiarity about the recent occurrence of an item and a lack of any retrieval of contextual details from the learning episode. For some authors the use of the pure proportion of the remember and know judgment's in the old responses was satisfactory to measure these subjective experiences (Gardiner, 2001; Gardiner, 2008). However the responses are mutually exclusive. One can either choose remember or know. This mutual exclusivity results in an interdependence of the responses: more remember responses will result in a decreased chance for know responses (if we assume fixed recognition levels in the task). This practice goes against the assumption that the two responses are retrieved from independent sources/systems (Tulving, 1985a; Yonelinas, 2001). This is why Yonelinas and Jacoby (1995) introduced the *independence remember/know (IRK)* method to propose an independent familiarity measurement: F=(1-R)/K, where F is the familiarity measurement and R is the proportion of remember responses while K is the proportion of know responses.

2. The role of instructions and other methodological difficulties in the remember-know paradigm

We can think about the remember-know paradigm as a memory task that provides participants with scripts/definitions of possible answers at the beginning of the task. The important role of instructions in this paradigm was illustrated in different studies (Geraci, McCabe & Guillory, 2009; McCabe & Geraci, 2009; Migo, Mayes & Montaldi, 2012).

Geraci et al. (2009) showed that using different instructions in the remember-know paradigm altered the pattern of results. The main finding of Gardiner and Java (1990) was that remember-know and sure-unsure answers were assigned differently to words and non-words. These results were replicated later by Rajaram et al. (2002) with a within subject design. Participants gave more remember responses to words than non-words and they used more know responses to non-words than words, whereas for both words and non-words they used more sure than unsure responses. This key result was later used extensively to argue that remember and know responses are not equivalent to simple sure-unsure memory judgments (Gardiner, 2001, 2008).

Geraci et al. (2009) compared the results of two experiments with different instructions. In the first experiment they used the instructions from Rajaram et al. (2002), which used a highly confident know responses: “certain that you recognize the item” (Geraci et al., 2009, Appendix A, page 707). In the second experiment they used the instructions from Yonelinas (2001), where know responses are simply explained as “you think the item was studied but you cannot recollect any details about the study event” (Geraci et al., 2009, Appendix B, page 708). In both experiments participants had to complete a remember-know task and a recognition task with using sure-unsure responses for items judged as ‘Old”. Experiment 1 using an instruction of highly confident know responses replicated previous results showing that remembering and knowing show a different patter than sure-unsure responses (Gardiner & Java, 1990; Rajaram et al., 2002). However, the results of Experiment 2, which used a different instruction for know responses, showed a similar pattern of remember-know and sure-unsure responses. There were more remember and sure responses than know or unsure responses for both words and non-words.

Another study (McCabe & Geraci, 2009) also highlighted important aspects of remember-know instructions. In Experiment 1 the authors changed the expressions of ‘remember’ and ‘know’ and they used the terms ‘Type A’ and ‘Type B’ for them. This manipulation reduced the remember false alarms, thereby increased the accuracy of remember responses. This result was the first to show that using a neutral terminology to refer to remember and know judgments can improve accuracy, potentially by reducing the effect of pre-existing connotations of the expressions. In Experiment 2 the instruction of the remember response was altered to emphasize that only recollected details from the specific study episode are decisive when choosing this answer. They called this the *source-specific* remember instruction. The results showed that the source-specific instruction reduced both hits and false alarms for remember responses and increased know hits and false alarms relative to traditional instructions. This result suggests that typical remember-know paradigms overestimate the relative contribution of remember responses and underestimate know responses.

Altogether these studies point out that the actual wording of the instructions in a remember-know paradigm can significantly alter the results. Depending on the instructions the responses can behave as sure-unsure judgments (Geraci et al, 2009). Additionally, typical instructions can result in underestimation of know responses and overestimation of remember responses and they can make the interpretation of the responses harder than simplified instructions (McCabe & Geraci, 2009). Studies using this paradigm rarely report the written instructions, which makes the results difficult to compare.

There are other methodological difficulties about the paradigm. For example a study showed that approximately 38 % of know responses are accompanied by some level of detail recollection, while 91 % of remember responses had some recollection (McCabe et al., 2011). Researchers need to address the question whether the participants understand and use the subjective reports accordingly to the instructions. Most of the studies do not report whether and how they have checked for this (Migo et al., 2012).

Overall we can conclude that the remember-know paradigm is a useful and easy way of estimating the level of recollection and familiarity in a recognition task. Previous research using this task was more concerned to provide evidence that the two responses are independent than to develop a theory about recognition and subjective experiences which would predict new or explain results from other research methods (e.g., ROC or process-dissociation experimental results). The researchers using the remember-know paradigm often stay descriptive, concluding that the two responses are independent but they fail to provide an integrative, cross-method theory underlying the subjective responses and other characteristics of recognition memory. As we will present it later, the dual-process signal-detection theory can integrate results from different research methods and give a clear and quantitative model of recognition memory.

Process-dissociation procedure

This dual-process method was introduced by Jacoby (1991) to measure the contribution of familiarity and recollection to recognition memory. The framework distinguishes between an intentionally controlled, slow recollection process and a more automatic and fast familiarity process (Jacoby, 1991; Yonelinas, 2001; Jacoby, Yonelinas & Jennings, 1997). Recollection is the capacity to intentionally retrieve specific aspect(s) of the study event, while familiarity is a more automatic process, which helps recognition without providing any specific detail about previous item occurrence. The two processes are thought to have independent effects on recognition.

The method estimates the processes by contrasting the results of two experimental conditions, where recollection and familiarity are thought to contribute to performance differently. In the *inclusion* condition the experimental task is designed in a way that the two processes are thought to support performance. In the *exclusion* condition the task is different and the two processes are thought to act in opposition. Based on the results of these two conditions the levels of recollection and familiarity can be estimated.

We will illustrate the paradigm with the original experimental setup (Jacoby, 1991), which consisted of three phases (2 study + 1 test). In the first study phase the participants read a list of words under incidental encoding instructions. In the second part, they heard a different list and they had to try and remember the items for a later memory test. At test they were assigned to either an inclusion or exclusion condition. In both conditions they were presented with items from both study phases intermixed with new items. In the inclusion condition their task was to respond 'old' for an item if they thought it was presented either in the first (read list) or the second study phase (heard list) and to respond 'new' to unstudied items. It the exclusion condition they have to respond 'old' for items, which were presented in the second study phase (heard items) and respond 'new' for all studied items, which were presented in the first phase (read words) and unstudied items too. Because both recollection and familiarity can be used in the inclusion condition, the probability of correctly accepting an item from the first list is equal to the probability that the item is recollected plus the probability that it is not recollected, but is accepted on the basis of familiarity (Yonelinas, 2002):

P(inclusion)= Recollection + (1-Recollection)*Familiarity

However, in the exclusion condition, the probability that an item from the first phase is incorrectly accepted as 'old' is equal to the probability that the item is not recollected but familiar:

P(exclusion)= (1-Recollection)*Familiarity

Based on these two probabilities the levels of recollection and familiarity can be easily estimated, for example:

Recollection = P(inclusion)-P(exclusion)

Familiarity = P(exclusion)/(1-Recollection)

Studies using this process-dissociation procedure show that there are several experimental conditions, which have different effect on the processes. In Table 5 we summarized the main findings based on Jacoby, Yonelinas & Jennings (1997) and Yonelinas (2002).

There are encoding manipulations that have different effect on the measures. For example generating words from anagrams compared to reading words during study increases both recollection and familiarity estimates. Longer study times also boost both process measures. In contrast, concentrating only on perceptual stimuli characteristics decreases both processes compared to deep encoding processes. Other manipulations influence only one process, like dividing attention during study greatly decreases recollection, while it leaves familiarity uninfluenced.

Retrieval manipulations also have specific effects on the process estimates. For example speeded recognition tests decrease the level of recollection, while familiarity is not sensitive to short decision times. Moreover, changing the size of the stimuli, thus changing the perceptual characteristics between study and test, decreases both recollection and familiarity (size congruency effect). There is also evidence that old compared to young adults have reduced recollection levels accompanied by unchanged levels of familiarity.

Variables	***Δ Recollection***	***Δ Familiarity***
Generation	↑	↑
Divided attention	↓	←→
Study duration	↑	↑
Shallow processing	↓	↓
Speeded recognition	↓	←→
Perceptual change	↓	↓
Aging	↓	←→

Table 5. Changes in recollection and familiarity based on Yonelinas (2002), Yonelinas (2001) and Jacoby, Yonelinas & Jennings (1997).

These results support the assumption of the process-dissociation procedure that the two processes are independent. They can be independently manipulated by different variables.

Additionally, the estimates of the two processes with the process-dissociation procedure in some cases are more plausible compared to the traditional remember-know paradigm (Yonelinas & Jacoby, 1995). Rajaram and Coslett (1992) used the traditional remember-know paradigm in a size congruency recognition task. They presented line drawings of objects to participants for a later recognition test. At test the drawings were either size congruent (same size at study and test) or they were size incongruent (either larger or smaller at test). At test participants had to complete a remember-know task. The result showed that size incongruent items lead to a decrease in remember responses but interestingly, they increased the know responses. This suggests that size incongruency leads to an *increase* of familiarity. This is against traditional claims that familiarity is decreased by perceptual mismatch at study and test (Yonelinas & Jacoby, 1995).

However in a study that used the process-dissociation procedure to measure the size congruency effect, Yonelinas and Jacoby (1995) found that for size incongruent items both recollection and familiarity estimates decreased, which result was in line with the common assumption of dual-process models that perceptual change leads to decreased familiarity. In another experiment they also replicated previous remember-know results showing that remembering decreases while knowing increases with size incongruency (Rajaram & Coslett, 1992). There was clearly a discrepancy between results of the two measurement methods: PDP and R/K.

Yonelinas and Jacoby (1995) suggested that the problem behind these results is that the simple use of the know responses in the traditional R/K method is not a good measure of familiarity. In the traditional R/K method the proportion of know responses for old items is used to measure familiarity. However the two possible responses remember and know, are mutually exclusive: the participant either answers remember or know, there is no other option. This mutual exclusivity means that the two measures (remembering and knowing) are interdependent. The probability of a know response will decrease with higher levels of remember responses. This feature of the traditional R/K measurement is also problematic to the multiple systems theory (Tulving, 1985a) that inspired the R/K method. According to the multiple systems theory retrieval is independent from the episodic and semantic systems. This discrepancy between the theoretical independence at retrieval and the methodological interdependence is a problem for the traditional R/K method.

To address these problems Yonelinas and Jacoby (1995) introduced the independent remember-know method (IRK method). The IRK procedure estimates recollection with the proportion of 'remember' responses to old items. This method looks at the remember response proportion as a valid measure of recollection. Remember responses according to their descriptions are made when there is any contextual detail retrieved from the study event.

However in the R/K method know responses are made if an item is familiar (F) but not recollected (1-R), they indicate familiarity in the absence of recollection. If the two processes were thought to be independent that would mean that there are items that were both familiar and recollected. In these cases the participant will indicate a remember response, because there is some retrieved aspect of the study event related to the item. Consequently, the proportion of know responses will underestimate that an item is familiar. This is why Yonelinas and Jacoby (1995) suggested that to measure the probability that an item is familiar (F) one must divide the proportion of know responses (K) by the opportunity that the participant is able to make a know responses (1-R):

$$\mathbf{F = K / (1-R)}$$

Given this equation we can use the results from any traditional R/K method to calculate the probability that items are familiar.

Using the IRK method to estimate recollection and familiarity Yonelinas and Jacoby (1995) showed that size incongruency at study and test decreases both of these measures, thus the results are consistent with the PDP measures. They concluded that the IRK is a better method that takes the independence assumption seriously and estimates the recollection and familiarity measures from the remember and know proportions. The convergence of the results also provides evidence of the validity of their assumptions regarding two independent sources of information behind recognition.

Converging evidence of behavioural measures of recognition to support the DPSD theory

In the next section we will argue that the behavioural results using either the ROC, IRK or the PDP methods support a specific dual process model, the signal-detection dual-process model (Yonelinas, 1994; 2001). We report the experimental evidence in a five-point list:

1. *IRK-based ROCs support a threshold recollection and a signal detection familiarity process.*

   The first supporting evidence comes for the study by Yonelinas and Jacoby (1995) that introduced the independent remember-know procedure. Based on the DPSD model familiarity is an equal variance signal-detection process that predicts that the ROCs be symmetrical to the negative diagonal. In Experiment 3 they used a modified remember-know paradigm, where participants had to indicate the confidence level of their familiarity judgment on a 6-point scale. They estimated the familiarity for every confidence level using the IRK method. Based on these estimates they drew up the ROCs for familiarity. The ROC was symmetrical (Figure 9), which supported the DPSD theory's prediction and also showed evidence that using the IRK method familiarity is a signal-detection process.

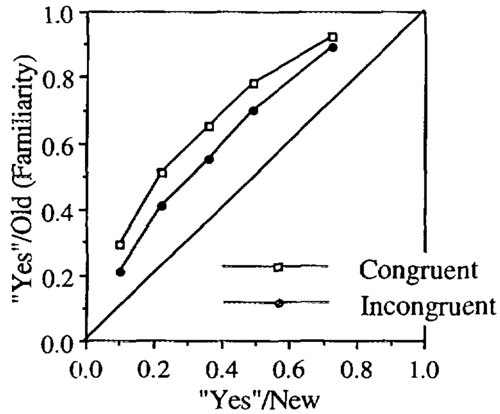

FIG. 4. The average familiarity ROCs for size congruent and size incongruent items for Experiment 3, based on the independence assumption. Estimates of familiarity are plotted against false alarms.

Figure 9. Symmetrical ROC curves from Yonelinas and Jacoby (1995).

2. *The IRK and ROC method gives similar estimates.*

   Yonelinas (2001) in another study in Experiment 1 participants had to indicate their recognition confidence for the test list items and also decide whether they remember or know that the item was old. He used the IRK method and the ROC method to estimate the processes . The results showed that the estimates of the two processes were almost identical and they did not differ statistically. Moreover both estimates were similarly affected by the experimental manipulation, which was full or divided attention during the study phase. Both recollection and familiarity decreased in the divided condition compared with full attention at study (Figure 10).

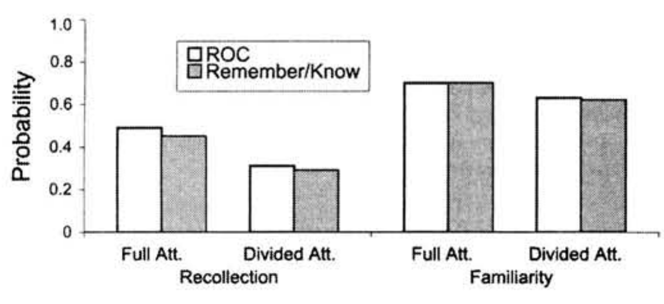

Figure 10. Experimental results of Experiment 1 from Yonelinas (2001). IRK and ROC methods show identical estimates.

3. *PDP-based process estimates show threshold recollection and signal detection familiarity.*

   Yonelinas (1994) used the PDP paradigm to measure recollection and familiarity on different confidence levels. Participants had to answer in the exclusion and inclusion conditions on a 6-point confidence scale ranging from *'sure yes'* to *'sure no'*. ROC results showed that recollection is a threshold process that created a flat ROC for recollection (Experiment 3), while familiarity measures showed a signal detection process with symmetrical ROCs (Figure 11).

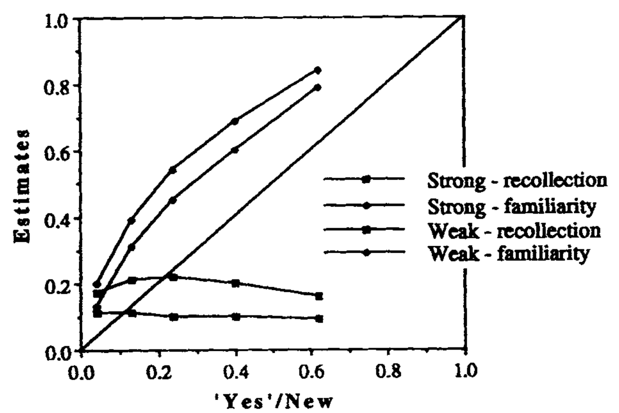

Figure 9. Estimates for recollection and familiarity for weak and strong items as a function of response confidence in Experiment 3.

Figure 11. ROC curves from Yonelinas (1994).

4. *The three process estimation methods (PDP, IKR, ROC) show statistically equivalent results.*

   The same study (Yonelinas, 2001) that showed that IRK and ROC estimates are almost identical (Experiment 1) also measured in Experiment 3 the process estimates from the three different methods. Again, the results showed that the methods produced undistinguishable estimates and that they all showed that recollection and familiarity are both reduced by divided attention at study (Figure 12).

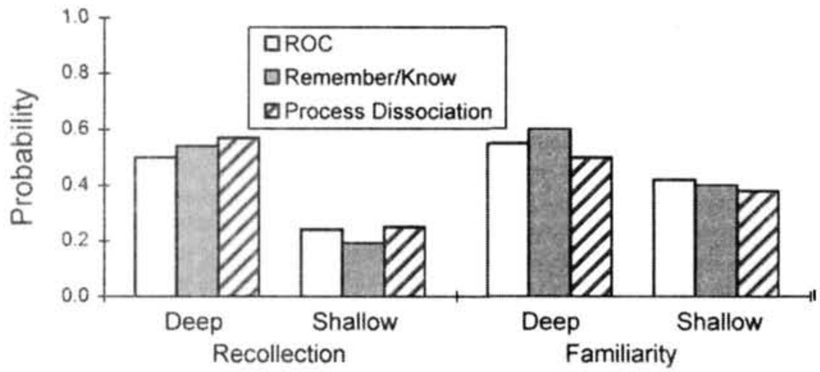

Figure 12. ROCs, IRK estimates and process dissociation show similar patter of results from Yonelinas (2001).

5. *Converging estimates across studies.*

Finally, in his review, Yonelinas (2001) reported the correlation of the estimates across 20 experimental conditions (including the ones mentioned above in our list). The results showed that there is a high correlation between the IRK and ROC method and also between the PDP and ROC method (Figure 13).

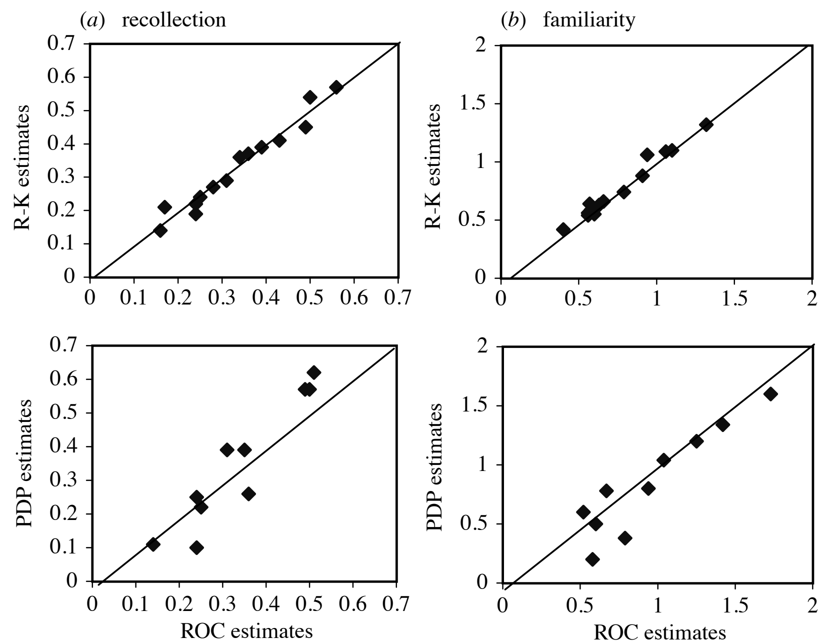

Figure 13. Correlation between the IRK and ROC method and between the PDP and ROC method from Yonelinas (2001).

This high level of convergence indicates that the ROC method, which uses pure parameter estimates to measure the processes, is a psychologically valid method (Yonelinas, 2001). Very different methods as the introspection-based remember-know paradigm and the PDP both show high similarity with ROC results.

Importantly, the results listed above indicate that the dual-process signal-detection (DPSD) model has the potential to integrate previous results from different measurement methods, and it also give a strong quantitative model of the processes behind recognition. The DPSD theory, unlike other theories and methods (e.g., multiple memory systems) can provide memory research with a simple theory and method that can be used across paradigms to test its assumptions: familiarity is a signal detection process and recollection is a threshold process.

Methodologically the ROC analysis has the advantage that based on the parameter estimates it can provide the researcher with estimates of the two processes even in a simple old-new recognition task. None of the previous methods can estimate the two processes in a single recognition task. The RK method needs participants to introspectively respond either with remember or know after every old response and the PDP method needs two different task conditions (the inclusion and exclusion conditions) to estimate the processes.

Summary of the dual-process models and new directions

Dual-process models are supported by a wide range of results from behavioural, animal, neuropsychological, electrophysiological and neuroimaging studies. These studies show that the recollection process is behind the retrieval of contextual information from events and it is independent of item familiarity. By providing access to certain aspects of the original event, recollection is a fundamental process of episodic retrieval. Based on these models, there is a great amount of evidence, which suggests that recollection is a threshold-like, controlled and effortful process.

We also showed that specifically the DPSD theory is a great candidate to integrate previous results and it has the strength to provide a single theoretical framework for understanding a wide variety of test methods in recognition memory research (Yonelinas, 2010).

It is also important to point out that recent research showed that recollection is not the only process, which could contribute to the rapid formation of new associations (Parks & Yonelinas, 2015). Early dual-process models of recognition assumed that item recognition is based on both familiarity and recollection, while associative recognition relies on recollection alone. The logic was that for associative recognition, when the task is to recognize, which items were paired together, the recollection of details about the study episode (such as the original pair of the item) supports discrimination. Meanwhile, the familiarity level of items is not helpful in associative recognition, because all items have the same level of familiarity (they were presented once during the learning phase). However, recently a new concept of unitization was used to argue that familiarity could help associative recognition if the item pair was encoded during the learning episode as a single item (Yonelinas et al., 2010; Yonelinas, 2013; Parks & Yonelinas, 2015). Unitization studies generally use ROC curve analysis (hits and false alarm rates on different confidence levels) to show that conditions with higher level of unitization (high LOU) could selectively increase the familiarity estimate of participants compared to conditions with low level of unitization (low LOU) (Parks & Yonelinas, 2015). Previous research showed that the effect of unitization is not equivalent to effects of levels of processing and, that it can boost familiarity within- (word-word) and across-domain (face-word or fractal-sound) associations. Unitization was used to demonstrate that familiarity could support associative recognition in a wide variety of studies using different behavioural (Parks & Yonelinas, 2009; 2015), ERP (e.g., Diana Van den Boom, Yonelinas & Ranganath, 2011), neuroimaging (e.g. Haskins, Yonelinas, Quamme & Ranganath 2008) techniques. Meanwhile the concept of unitization received a number of critiques (Mayes et al., 2007; Montaldi & Mayes, 2010) highlighting that there is no objective criteria in the experimental literature to measure whether unitization occurred or not. In all the studies, level of unitization is manipulated by contrasting two encoding conditions, which are hypothesised to differ in their relative unitization levels. Unitization has no single operational definition (Parks & Yonelinas, 2015). Moreover, there are important, yet unanswered questions raised by previous results. For example there is no explanation why in some cases unitization also boosts estimates of recollection (Parks & Yonelinas, 2015, Exp 1), or what is the reason for getting strong evidence for unitization effects in word-word and cross-domain (face-word, fractal-sound) stimuli, while there is a lack of unitization effects in using face-face and fractal-fractal within-domain stimuli (Parks & Yonelinas, 2015, Exp 4 and 5). We will discuss the possibility of the unitization effect in our experiments when we will introduce the task we used for our experimental work (see chapter 4).

In sum, dual-process models provide us with important evidence that recollection is a separate memory process, however they lack a detailed description and theory about the main steps or mechanisms of this mental process. It is not surprising, give that the focus of the dual-process research program is the collection of experimental evidence supporting the distinction between recollection and familiarity. These models focus on the mental content provided by recollection, namely the activation of the relational representation but they lack a model to describe the underlying mechanisms. Moscovitch (2008) and Moscovitch et al., (2016) proposed a model of recollection, which is building on behavioural and neuroimaging results and identifies two separate steps of the process and provide us with a testable model. In the following part we will introduce this two-stage model of recollection and also highlight an important characteristic of the model, which we intended to test in our experiments.

The two-stage model of recollection: representational activation and consciousness

The traditional standpoint is that the conscious experience of remembering is a defining feature of recollection (Aggleton & Brown, 1999; Moscovitch, 1995; Squire, 2004, Eichenbaum, Yonelinas & Ranganath, 2007), which is usually referred to as 'recollective experience' (Tulving, 1983), 'explicit recollection' (Moscovitch et al., 2016) or 'conscious recollection' (Squire, 2004). Furthermore, dual-process models (DPSD, PDP, RK) all describe recollection from a perspective focusing on the retrieved representational content provided by the process, namely the consciously available contextual information about the test item. However, there is a difference between the *consciously available result* of a memory process and the *mechanisms/steps* of the process itself. These previous accounts don't separate between two potentially different aspects of recollection: the activation of the relational representation and the conscious access to the retrieved content that could give rise to the experience of remembering. All these accounts assume that there are unconscious steps in the process of recollection (e.g., cue-trace interaction) yet they fail to clearly separate them from the conscious features of memory retrieval and they don't provide any theory how these supposedly unconscious and conscious process elements relate to each other.

The two-stage model of recollection (Moscovitch, 2008; Sheldon & Moscovitch, 2010) addresses the above mentioned issues and separates two parts of the process and suggests that conscious recollection roots in automatic and unconscious relational retrieval. The framework identifies an automatic part of the process that results in the unconscious retrieval of contextual aspects of the original event and a slower, controlled part that makes the relational information explicit so it can be used consciously to guide behaviour. According to the two-step model the first stage of the recollection is a rapid, hippocampus-dependent, automatic interaction between the incoming stimuli (cue) and the previously encoded memory representation. The result of this first stage is not consciously accessible and it can guide behaviour on implicit memory tasks. During the second, slower stage the output becomes accessible to conscious processing so the individual can use it explicitly. The second, conscious stage is dependent on the interaction between prefrontal, parietal cortices and the hippocampus. In the model the first stage is obligatory for the second stage to appear.

Supporting evidence

The model was inspired by behavioural results suggesting that recollection is linked to implicit memory processes. Westmacott and Moscovitch (2003) asked participants to perform fame judgments and speeded reading of famous names. The results showed that for names that participants previously associated with high levels of 'remember' responses (evoked detailed episodic memories related to the famous person, high R names) the fame judgments and reading times were faster than for names with low levels of 'remember' responses (rarely associated with episodic memories, low R names). The study showed that stimuli that are associated with high levels of episodic details also show improved semantic accessibility (enhanced implicit processing). Furthermore, Westmacott, Black, Freedman and Moscovitch (2004) demonstrated that the advantage of high R names in fame judgments and speeded reading disappeared in patients with MTL amnesia but enhanced processing remained intact in patients with semantic dementia. In another study Sheldon and Moscovitch (2010) used the remember-know paradigm (Tulving, 1985b) to measure recollection and demonstrated in several experiments that words, which were recollected ('remember' responses) were associated with better priming effects in two implicit memory tasks (word-stem completion and lexical decision). These results suggest that explicit retrieval processes, presumed to be connected to recollection, contribute to performance on implicit tasks. This is consistent with the notion of the two-stage model of recollection that early processes underlying recollection mediated by the hippocampus may indicate the interaction between conscious and unconscious memory processes (Sheldon & Moscovitch, 2010b).

Another line of research, which inspired the two-stage model of recollection, is based on those neurocognitive findings that highlight the involvement of the hippocampus in associative implicit memory processes. It is well documented in clinical research that the medial temporal lobe (MTL) regions are important in the conscious retrieval of past events (Tulving & Schacter, 1990; Moscovitch, 1995) and others highlight the specific role of the hippocampus in conscious recollection (e.g., Aggleton & Brown, 1999; Yonelinas, 2002). These theories were used to focus on the role of hippocampus in explicit memory and ignored or sometimes excluded the possibility that this region can be involved in implicit processes too. However, other findings noted that specific forms of implicit processes were impaired in patients with hippocampal damage on a variety of tasks, such as perceptual-identification (Yang, 2003) and stem-completion (Schacter, 1987). Based on these studies Schacter et al. (2004) and Schacter, Wig and Stevens (2007) proposed two different types of priming (implicit processes): one that relies on non-associative perceptual representations independent of MTL regions, and another one that relies on MTL-based processes connected to associative representations (Sheldon & Moscovitch, 2010a). This proposal is also compatible with an alternative view on the role of the hippocampus that emphasises its importance in the encoding and retrieval of relational information (Konkel & Cohen, 2009; Henke, 2010; Eichenbaum, 1999; Eichenbaum et al., 2007) even without conscious awareness. This relational memory account expresses the fundamental role of the hippocampus in the binding and subsequent retrieval of arbitrary inter-item and item-context relationships even on tasks that do not require conscious access to memory representations.

Chun & Phelps (1999) reported supporting evidence for the relational memory theory using a contextual cueing task with controls and mainly hippocampus damaged amnesics. In the experimental task participants searched visual displays for rotated T targets that were presented among rotated L distractors. The displays were randomly generated but some of them were repeated throughout the experiment. In these repeated, old displays the targets always appeared in the same location so the visual context of the displays predicted target locations. Both controls and amnesics mainly with hippocampal damage showed a context-independent learning effect of faster performance through the experiment, which was a sign of their intact skill learning. However, only controls showed a context-dependent learning effect, which was demonstrated by faster search performance for repeated displays than randomly generated displays. The lack of this effect in amnesics suggested that the hippocampus is important in encoding implicit contextual information throughout several learning trials from the environment.

More support for the involvement of hippocampus in implicit relational learning comes from Greene, Gross, Elsinger, & Rao (2006) who used a transitive inference (TI) task. In a TI task participants have to choose between shapes A or B (A?B) and learn by trial and error to choose A (A > B). There are other pairs (B > C, C > D, D > E) and participants eventually learn the structure of the relations and can perform above chance when tested on new pairs (B?D). The results showed that hippocampal activity during both learning and test predicted performance on new pairs even when participants did not have explicit knowledge about the structure of the shape relations.

Henke et al. (2003) used a one-trial masked pair-learning paradigm, where participants studied face-occupation pairs subliminally. Both the hippocampus and the perirhinal cortex were activated during the implicit encoding and retrieval of the pairs. Using a similar paradigm Degonda et al. (2005) also showed evidence of enhanced hippocampal activation for subliminally presented face-occupation pairs compared to face + non-word pairs. They interpreted this difference as a result of implicit formation of semantic face-word associations. The above mentioned behavioural and neurocognitive experimental findings all support the two-stage model of recollection.

Other studies using eye movement measures and neuroimaging techniques also supported the link between relational memory and hippocampal activation in the absence of conscious experience during retrieval (Hannula & Ranganath, 2009). We will discuss in detail these results in the eye movement section (see chapter 4).

The two-stage model as a synthesis of the two approaches of episodic memory

The two-stage model of recollection can be viewed as an attempt to the theoretical synthesis of the representational and phenomenological approaches of episodic retrieval. The first stage of the process is thought to be rapid and unconscious, which results in the retrieval of the content of episodic memories that can guide behaviour in certain conditions. This first stage is the activation of the relational representations of previous events. This activation process is in the core of the representational approach, which infers episodic memory retrieval from the ability to guide behaviour based on relational information retrieval. The function of the first stage is supported by all the previously mentioned studies showing that medial temporal activation accompanies relational memory processing both at encoding and retrieval.

The second stage of the model brings back the focus on the subjective aspects of episodic retrieval. The result of the second stage is that the activation of the episodic memory representations becomes consciously available to the individual. The conscious experience in the model includes subjective qualities such as confidence during retrieval (Moscovitch et al., 2016). There are experimental results showing that high source accuracy can be accompanied by low retrieval confidence in parietal lobe patients (Hower, Wixted, Berryhill & Olson, 2014; Simons et al., 2010). This dissociation suggests that these patients are selectively impaired in the conscious components of recollection while their episodic representations can still guide their retrieval accuracy. Yazar, Bergström and Simons (2014) also demonstrated this confidence-accuracy dissociation in normal adults by disrupting the angular gyrus (part of the parietal lobe) by continuous theta burst stimulation. These results all point to the direction that certain cortical regions (e.g., parietal cortex, prefrontal cortex) are associated with the subjective qualities of relational memory retrieval, which can potentially enable the subjective experience of remembering at the second stage of recollection (Moscovitch et al., 2016).

The obligatory order of the two stages

An important part of the two-stage model of recollection is the suggested relationship between the two steps. The conscious part is based on the first, automatic stage. The initial activation of the relational representation is providing the bases and precondition of the subsequent conscious processes. In the absence of the automatic retrieval of any hippocampal representation there is no memory content that can be apprehended consciously. This suggests that in the model the first stage of recollection is obligatory for the second stage to appear. Therefore, the conscious experience of recollection always has to be based on the successful retrieval of a relational memory representation. This order restriction assumption of the model could be tested, if there was a reliable measure of the first stage and one would use a task, which relies on recollection. The prediction based on the order restriction is that reliable memory performance in a recollection-based task has to be accompanied by the measureable indicator of the first stage of recollection. To introduce a potential measure of the first stage of recollection we will continue by introducing how eye movements can help us in memory research. We will start with the review of previous studies that used eye movements to indicate different memory phenomena. Later we will introduce a specific paradigm, which shows that eye movements can be used to measure relational retrieval in the absence of explicit recollection (Hannula & Ranganath, 2009), suggesting that this eye movement effect can be used as a measure of the first stage of recollection. In our experiments we tested whether the relational eye movement effect could be used as a reliable behavioural measure of the first stage of recollection.

Chapter 4: Eye movements and memory research

Our gaze control has two different sources when viewing scenes (Henderson, 2003). There is stimulus-based gaze control, when the currently available visual input characteristics determine our eye movements (e.g., saliency map models (Itti & Koch, 2001)). On the other hand there is knowledge-driven gaze control, when our prior experience of the world guides our eye movements during perception. The knowledge-driven gaze control effects include short and long-term memory, the characteristics of the task and the plans of the individual (Henderson, 2003).

Memory studies use a wide range of measures to demonstrate memory effects in eye movements. Based on the review by Hannula et al. (2010) we will briefly introduce the most important measures that are used in the field (see Table 6). We will refer to these measures later on in this chapter when we summarize the main findings of memory research with eye movements.

The *overall* viewing measures are based on the level of the entire experimental display. For example we can calculate the overall fixation number directed to an entire display, which contains multiple items. Alternatively, *directed* viewing measures can be used, which categorize eye movements according to their targets (specific items on the display), and summarize them separately. These directed measures divide the display to multiple area of interests (AOIs). These measures can be combined with temporal indices to examine the emergence of particular effects during a test display (time-course analysis). Temporal indices can also be used relative to a particular event (e.g. participant's response) to measure time-locked effects (e.g. response-locked analysis).

Usually, both overall and directed viewing measures are calculated based on fixations. The definition of a fixation may vary between research groups, what makes the results hard to compare. There is no standard in the literature regarding the calculation of fixations from raw data recordings. Event detection is used to identify different phenomena in raw data recordings (e.g., fixations, saccades, blinks, smooth pursuit). There are two common methods to event detection in eye tracking research (Holmqvist et al., 2011). The *identification by dispersion threshold* (I-DT) method is using the raw sample's positional information to identify fixations. In this case, temporally adjacent samples must be located within a spatially limited region (typically 0.5 – 2°) for a minimum duration (typically 50-250 ms) to form a fixation (Holmqvist et al., 2011). The *velocity* methods use the raw samples' velocity values to identify events. These methods define fixations as continuous portions of the raw data where the gaze velocity does not exceed a predefined threshold (10-50°/s). In both calculation methods the actual parameter settings may alter event detections and consequently the overall or directed viewing results too. Many studies using eye movement recordings do not report what parameter values were used to identify eye movement events. In our opinion it is essential to compare the results for different fixation filtering parameters or, if possible, compare the fixation-based results to raw data results. In our experiments we will report both fixation-based and raw-data-based results to check for their potential effect on data analysis. We will get back to these issues later in the methods section.

Apart from the problem of event detection, there are other concerns regarding the comparison of eye movement results. For example there are different methods to calculate the *number of regions sampled* or the *viewing time proportion* to an AOI (Hannula et al., 2010). In some cases the *number of regions sampled* is defined based on the fixation pattern of each individual. Fixations that fall within a predefined distance are treated to be within the same region (Althoff & Cohen, 1999). Others use fixed regions on the display for every stimuli and participant (Smith & Squire, 2008). This method has the advantage to control for the number of possible regions that could be fixated but it may have disadvantages too. For example fixed regions are insensitive measures when participants view two different parts within the same region (e.g., two objects sampled). In this case the fixed region method underestimates the number of regions sampled (Hannula et al., 2010).

In some works the *viewing time proportions* are calculated in reference to a baseline, which exclude the amount of viewing time that is not directed to the stimulus (e.g., excluding any time spend looking outside of the display and blinks or saccadic movements; see: Ryan et al., 2000; Hannula et al., 2007). In other studies the proportion is calculated using the entire trial period of the trial as the baseline (Smith & Squire, 2008). The latter method has a tendency to give us lower estimates of viewing time proportion than the former calculation method.

That methodological differences exist and these differences may alter the results is important to take into consideration for the following part, where we will summarize the main findings of the field.

Measures of overall viewing

- **Number of fixations[1]:** the number of discrete pauses of the eyes for a display.
- **Fixation duration[1]:** the length of time in which the eye pauses on a display, typically between 200–300 ms long. Median or mean fixation duration to a display can be calculated.
- **Saccade amplitude:** the distance traversed between successive fixations, reported in degrees per second.
- **Number of regions fixated[2]:** the number of discrete regions sampled within a display.
- **Number of transitions between regions:** the number of transitions made by the eyes between discrete regions.
- **First return fixation:** the number of fixations made before returning to a previously sampled region.
- **First-order entropy[3]:** the predictability of the transitions between the locations of a given fixation and the preceding fixation.
- **Second-order entropy[3]:** the predictability of the transitions to a given fixation location based on the location of the two immediately preceding fixations.
- **Chi-square, Asymmetric lambda[2]:** other measures used to quantify the randomness of an eye movement transition table.

Measures of directed viewing

- **Proportion of fixations:** the proportion of total fixations that are directed to an experimenter-defined ROI.
- **Proportion of time[2]:** the proportion of total viewing time that is directed to an experimenter-defined ROI.
- **Number of transitions into/out of a critical region:** the number of gaze transitions into/out of an experimenter-defined ROI.
- **Duration of the first gaze:** total viewing time to an experimenter-defined ROI on the first gaze that is directed into that ROI.
- **Number of fixations in the first gaze:** number of fixations to an experimenter-defined ROI during the first gaze that is directed into that ROI.

Table 6. Examples of overall and directed viewing measures from Hannula et al. (2010).

Eye movements and prior exposure

Pre-experimentally known faces and buildings are viewed with fewer fixations and with fewer regions sampled when compared to unknown (novel) faces or buildings (Althoff et al., 1998). Althoff and Cohen (1999) also showed that participants' eye movements had a higher level of constraint in transitions among successive fixation locations for non-famous than for famous faces, meaning that consecutive fixation locations were more predictable for non-famous (novel) faces. This higher predictability of the fixations was paired with fewer regions sampled.

Eye movement-based memory effects are also evident when novel stimuli are used and repeated during one experimental session. Althoff (1998, cited by Hannula et al.,

2010) repeated non-famous (novel) faces 1, 3 or 5 times during an experiment and found that with increasing prior exposure participants sampled fewer regions of the studied faces (Figure 14).

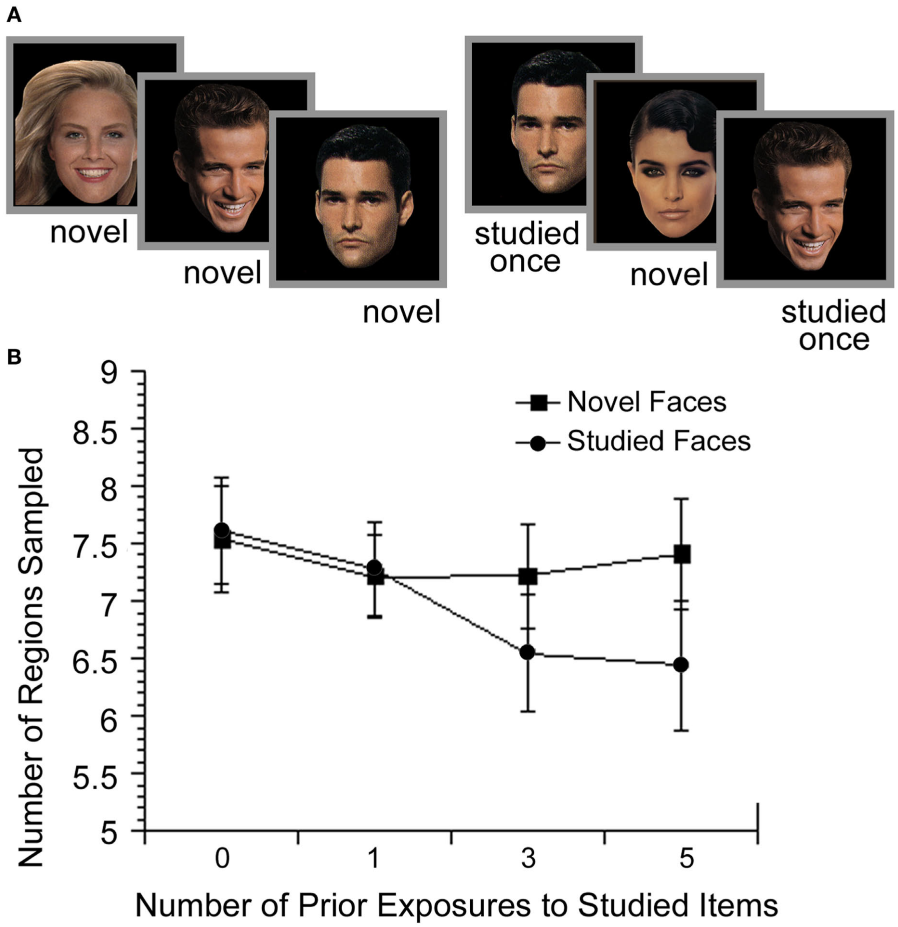

Figure 14. The effect of prior exposure to unknown faces from Althoff (1998, cited by Hannula et al., 2010).

Other results also showed that repeated items have distinct eye movements compared to non-repeated (novel) items. Heisz and Shore (2008) used 2, 4 and 6 prior exposures to unknown faces in one experiment. The results showed that increasing prior exposure decreased the overall fixation count but the proportion of fixations at the eyes region of the faces increased. In another study (Heisz & Ryan, 2010) where they used famous and non-famous faces and varied the prior exposure levels (1, 3 or 5) supported previous findings. The results too showed that for both type of faces (famous and non-famous) additional exposure during the experiment decreased the fixation numbers but increased the fixation proportions to the eyes region.

Natural scenes also elicit eye movement-based memory effects, for example Ryan et al. (2000) used repeated and novel scenes and found a repetition effect in eye movements. For repeated scenes fixation numbers and the number of regions sampled during scene viewing both decreased compared with novel scenes.

Eye movements and relational memory

Apart from item memory, eye movements can also signal relational memory. Ryan et al. (2000) not only showed repetition effects for scenes, but they also found a relational manipulation effect in eye movements. In experiment 1 they presented participants with natural scenes. After two study blocks, in the third, critical block there were three different types of scenes: novel scenes, repeated scenes or manipulated scenes. The manipulated scenes were also seen before but in the critical block they were changed. The change was either an addition of a new object, a deletion of an object or a left-right switch of an object. Each scene had a critical region designated, where the manipulation took place. During the critical block the eye movements of the participants were recorded. The results showed that participants directed higher proportion of their total fixations and higher proportion of their total viewing time to the critical region of the manipulated scenes compared with the repeated scenes. They also made more transitions into and out of the critical regions for manipulated than for novel or repeated scenes. These findings suggest that changes of the original relations among elements of the scenes cause specific eye movement effects, showing relational memory for constituent elements of the scenes.

There is also experimental evidence that eye movement behaviour can indicate memory for temporal relations among items. Ryan and Villate (2009) sequentially presented three objects in unique spatial configurations to participants. At test the objects appeared simultaneously in new absolute but in the same relative positions.

The eye movement results showed that participants tended to inspect the objects in the order that matched their original presentation sequence.

Another line of work also demonstrated that eye movements are sensitive measures of relational memory. Hannula et al. (2007) introduced a scene + face learning paradigm with eye movement recording. Because the findings of this paradigm gave the basis of our experiments, we will introduce its methods and findings in more detail. In the study phase participants were instructed to encode arbitrary scene + face pairs for later recognition. During the next phase participants were tested with 3-face displays superimposed on a scene. The test displays either contained three previously studied faces, one of them studied together with the scene (match displays) or they contained three studied faces, none of them studied with the scene (re-pair displays). Novel displays were also presented during the test phase, which contained three new faces, none of them seen before (novel displays). The participants were instructed to identify the matching face from each test display and even when they thought that none of the faces had been presented together with the scene they had to choose a face (re-pair and novel displays). Participants showed a relational eye movement effect (hereafter we will use the term 'REME'), which manifested in the disproportionate viewing of the chosen matching faces compared to faces chosen from re-pair or novel displays (Figure 15). The effect is caused by the eye movement preference of the matching face compared to other, non-matching faces in other displays. As we will introduce them later, there are other methods that can be used to show the REME, which is always based on the preference of the matching face compared to other non-matching faces. In this study, the REME was evident as early as 500-750 ms after test display onset.

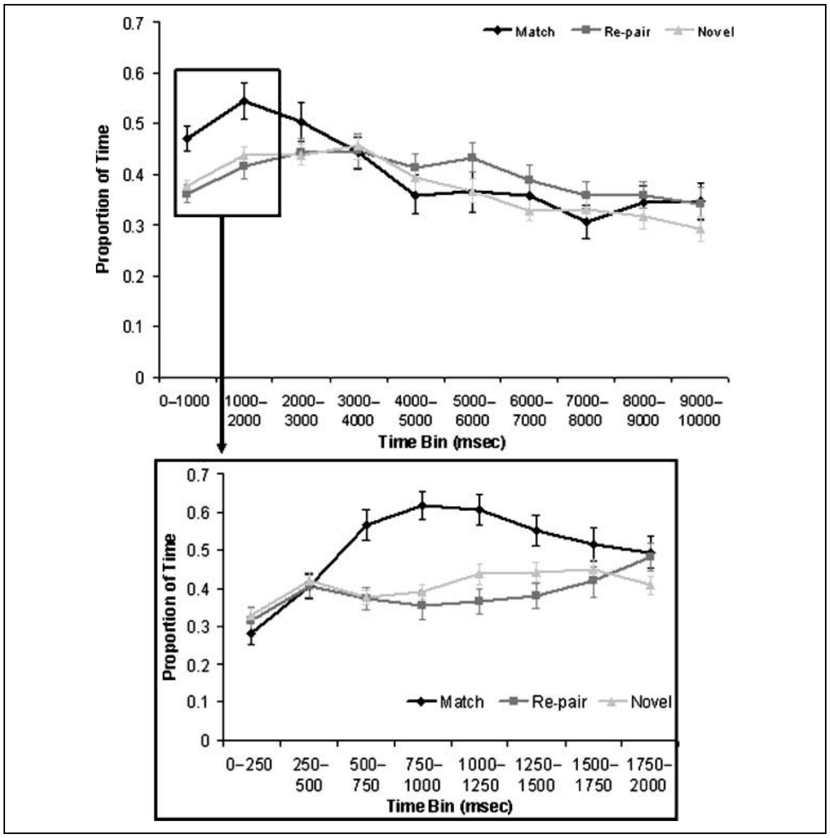

Figure 15. Disproportionate viewing of chosen matching faces (black lines) compared with chosen faces from re-pair (dark grey) and novel (grey) test displays (from Hannula et al., 2007, Experiment 1).

Subsequent response-locked analysis showed that the REME appears 500-1000 ms before explicit responses (Figure 16).

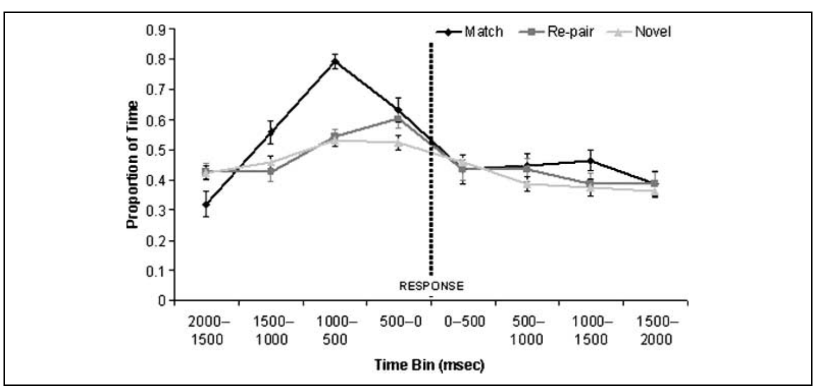

Figure 16. Response-locked analysis of viewing time proportions for match (black line), re-pair (dark grey line) and novel (grey line) test displays (from Hannula et al., 2007, Experiment 1).

In Experiment 2 in the same study the authors changed the task. In this new experiment participants completed the same learning phase but during the test phase they were told to study the three faces along with the presented scene for another recognition test. Only the match displays were analysed and the results showed that the average viewing time proportion of the matching faces was significantly above the 0.33 chance level (0.33 viewing time proportion is expected if there is no preference between the three faces). This effect was present in the first 2000 ms of the trials. Another time-course analysis using 250 ms bins showed that the effect was evident as early as 500-750 ms after test display onset (Figure 17).

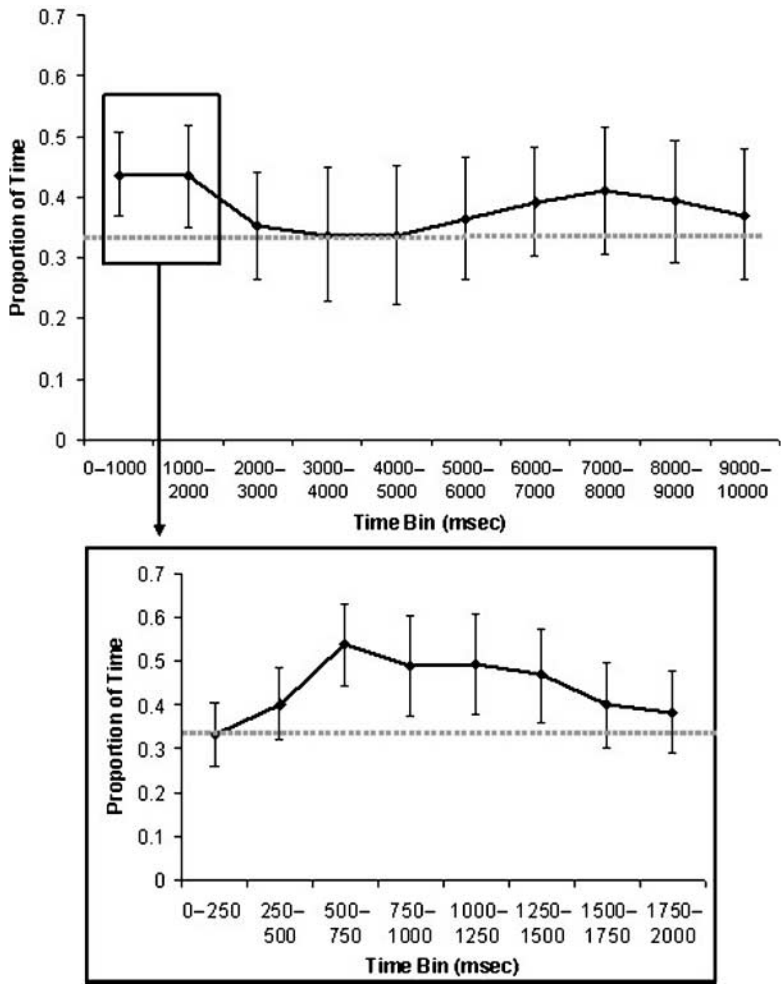

Figure 17. The viewing time proportion of the matching faces compared to the 0.33 chance level in Experiment 2 (from Hannula et al., 2007). Above: viewing time proportions for separate 1000 ms time bins. Below: viewing time proportions during the first 2000 ms for separate 250 ms time bins.

In sum, these results were interpreted by the authors that there is a rapid effect of relational eye movement effect (REME), which appears 500-1000 ms before explicit responses are made and which occurs obligatorily, regardless of response requirements (Hannula et al. 2007). The effect was thought to be dependent on the hippocampal memory system, based on their last experiment (Experiment 4), where this relational memory effect was absent in amnesic patients with hippocampal damage (Figure 18).

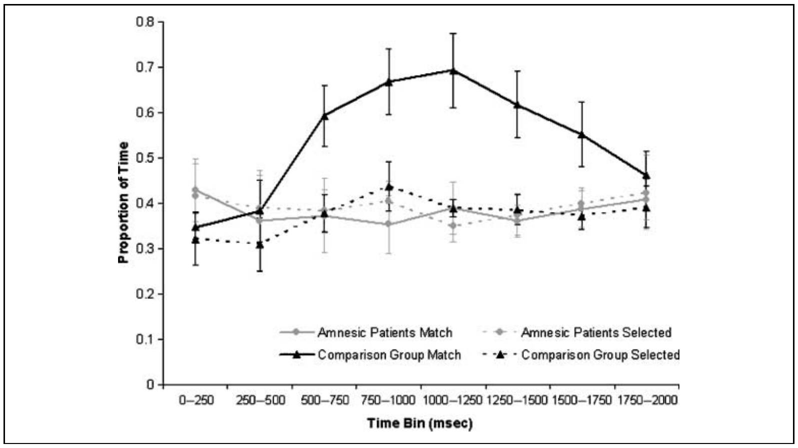

Figure 18. Viewing time proportion of the selected faces: matching (black lines) and non-matching (grey lines) in amnesics (dashed lines) and the comparison group (solid lines) (from Hannula et al., 2007, Experiment 4).

In a second study using a similar face-scene learning paradigm with functional neuroimaging, Hannula and Ranganath (2009) showed that the REME could signal relational memory retrieval even in the absence of explicit recollection.

In their study, during an initial learning phase participants were shown scene + face pairs for later remembering. At test – after a brief presentation of a scene alone (scene cue) – participants were presented with 3-face displays superimposed on a scene. One of the three faces was the original pair of the scene, while the other two faces were initially paired with different scenes (Figure 19).

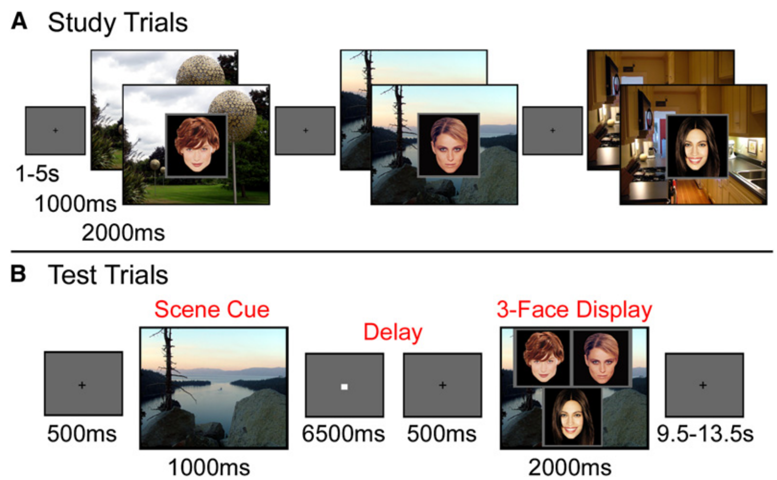

Figure 19. Illustration of the paradigm used by Hannula & Ranaganath (2009). Above: Study trials with face-scene pairs. Below: Test trials with a scene cue followed by the three-face test displays containing one matching face and two non-matching faces (from Hannula & Ranaganth, 2009).

All three faces were equally familiar, with the same amount of prior exposure and there were no spatial cues to guide eye movements to the matching face, because all the faces on the test display were appearing in new spatial locations. The task was to identify the matching face on the test displays. Our opinion is that to identify the target face, participants rely on recollective processes. The results showed a REME, which was demonstrated by the disproportionate viewing of the successfully identified matching faces (correct responses) compared to incorrectly chosen non-matching faces (incorrect responses). This viewing effect was evident for the overall 2000 ms of the test trial and it was also apparent just 500-1000 ms after display onset in a time-course analysis (Figure 20). These results supported previous studies showing that eye movement measures could be successfully applied to measure relational memory retrieval.

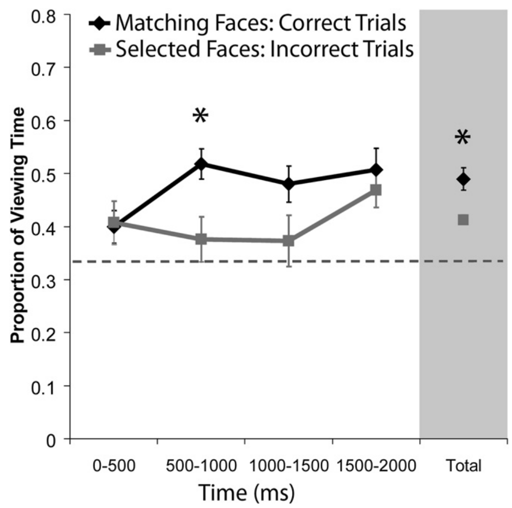

Figure 20. Overall and time bin separated viewing time proportion of correctly identified matching faces (black line) compared to incorrectly identified non-matching faces (grey line) from Hannula & Ranaganath (2009).

The main aim of the study was to test whether hippocampal activity is associated with relational memory even when explicit retrieval is absent. To provide supporting evidence, the authors analysed only the incorrect responses, when there is a lack of conscious relational retrieval. First, they calculated the viewing time proportions of the matching face for every incorrect trial. They categorized the trials to either high proportion viewing or low proportion viewing using a simple median split method based on the viewing proportions of the matching faces. The results showed higher hippocampal activity during the scene cue for high proportion viewing responses compared to low proportion responses (Figure 21).

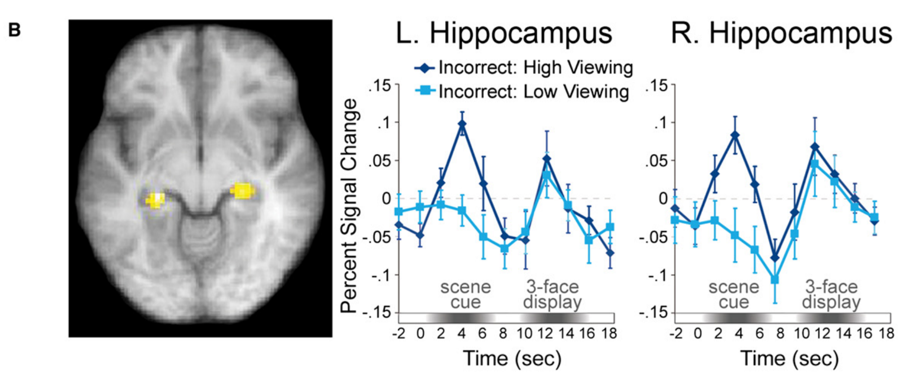

Figure 21. High and Low viewing incorrect responses and signal change (percent) in the hippocampus during the presentation of the scene cue (from Hannula & Ranganath, 2009).

The authors discussed that this result implicates that the hippocampus is involved in the retrieval of the information about the face-scene relationship and this retrieval directs eye movements to the matching face, even when explicit recollection is absent. This interpretation suggests that unconscious relational memory retrieval – by directing the eyes towards the matching face – is the basis of the REME and this effect could be used as the behavioural index of relational retrieval.

Taken together we can summarize the characteristics of the relational eye movement effect (REME) in 5 main points:

1. The REME **can emerge rapidly** after stimulus onset. Hannula et al. (2007) and subsequently Hannula and Ranganath (2009) showed that the eye movement effect can appear within the first 2 seconds of test trials.
2. The REME appears **before explicit responses**. Under response instructions, when participants have to choose from the three-face test displays the REME appears between 500-1000 ms before explicit responses are made (Hannula et al., 2007).
3. The effect emerges **spontaneously and obligatorily**. It appears even when participants are not required to make explicit responses. Hannula et al. (2007) demonstrated the effect in Experiment 2, when the participants' task was to learn the three-face displays for a future recognition test.
4. The REME can appear **in the absence of explicit recollection** (Hannula & Ranganath, 2009).
5. The emergence of the effect is **hippocampus-dependent** (Hannula et al., 2007; Hannula & Ranganath, 2009).

The feature overlap between the REME and the first-stage of recollection:

It is an unanswered and important question for experimental work what are the behavioural markers of unconscious access to relational representations. These markers would allow us to test the theoretically important questions about whether recollection is best characterized as a two-stage process. According to the two-stage model (Moscovitch, 2008; Moscovitch et al., 2016) the first stage of recollection is a rapid, unconscious, hippocampus-dependent obligatory interaction between a cue and the previously encoded memory trace, which results in the access to the episodic representation. This access at the first stage is not consciously apprehended but it enables conscious access. During the second, slower stage the representation becomes accessible to conscious processes, which can give rise to the unique phenomenological experience and make the content explicit. There is a striking feature overlap between the REME and the first step of recollection (see Table 7).

Measurable index of relational retrieval: Relational eye movement effect (REME)	**Theoretical cocept of recollection:** First stage of recollection
Rapid and obligatory	Rapid and obligatory
It can occur in the absence of conscious recollection	It can occur in the absence of conscious recollection
hippocampus dependent	hippocampus dependent
Appears before overt responses are made	Appears before conscious access to the representational content

Table 7. Feature overlap between the measureable REME and the theoretical concept of the first stage of recollection.

This feature overlap of the theoretical concept and the measurable eye movement effect suggests that the REME could be used as the behavioural index of the first step of recollection. Making this link between the REME and the first-stage of recollection is not new in the literature, Hannula and Ranganath (2009) also expressed that their results are consistent with the two-stage model. Others also highlighted that the REME might signal the first-stage of recollection (Sheldon & Moscovitch; 2010, Moscovitch et al., 2016).

Based on the summary of previous results and the feature overlap of the REME and the first-stage of recollection we can make the following assumption:

The REME is a universal and necessary indicator of relational memory retrieval. It accompanies successful relational retrieval obligatorily and unconsciously. It can be regarded as the behavioural marker of the first-stage of recollection.

This assumption, provides us with a powerful tool to explain previous experimental results in the literature:

1. The REME appearing in the correct-incorrect response viewing time proportion difference (see Figure 20, page 73) can be explained as follows. Successful relational retrieval favours the matching face, which will be manifested in eye movements directed towards the matching face. In case of a correct response, the chosen face is the matching face, so the viewing time proportion of the chosen face will be boosted by the relational retrieval. Conversely, in the case of an incorrect response, the chosen face is a non-matching face. The non-matching face's viewing time proportion is not boosted by the relational retrieval, so there will be an overall difference between the correctly and incorrectly chosen faces. The correctly chosen faces will have higher viewing time proportions than the incorrectly chosen faces (Hannula et al., 2009).
2. The REME appearing in the viewing time proportion difference between chosen faces in the match displays and re-pair (or novel) displays in the Hannula et al. (2007) study can be explained similarly (see Figure 16, page 69). On match displays, relational retrieval will guide eye movements to the matching face, which will boost the viewing time proportion of the matching faces. On re-pair or novel displays there is no matching face present, consequently the relational retrieval will not favour any of the faces. This effect will cause the chosen matching faces to have higher viewing time proportions compared to the chosen faces from re-pair or novel displays.
3. In Hannula et al. (2007) Experiment 2 the REME (measured as a significantly higher than chance level viewing time proportion directed to the matching faces) emerged spontaneously, even when participants did not have to explicitly choose the matching face from the three-face test displays (see Figure 17, page 70). Assuming that the REME is obligatorily accompanies relational retrieval can explain this result. This predicts that when there is relational retrieval in a task, the eyes will be drawn automatically to the matching face resulting in a REME, regardless of task characteristics, which will result in above chance viewing time proportion for the matching faces.

4. Hannula & Ranganath (2009) showed enhanced hippocampal activation in incorrect responses, when the eyes were drawn disproportionally to the matching face. This result suggests that relational retrieval can guide the eyes towards the matching face, even when explicit recollection is absent. This pattern can be explained by our assumption, which states that the process behind the REME is essentially unconscious and this relational retrieval guides eye movements even when participants fail to consciously access the retrieved information.

Above we demonstrated how the assumption that the eye movement preference of the matching face is a necessary unconscious memory effect of relational retrieval could explain all previous results from neuroimaging findings to the behavioural findings of different manifestations of the REME. Based on its explanatory power we accept the assumption and we form basic predictions to test in our experiments.

There are four main predictions based on our assumption that the REME necessarily accompanies successful relational retrieval:

1. *Based on the assumption the REME signals a necessary process for relational retrieval, which predicts that when there is evidence of relational memory retrieval in the task, the REME has to appear.*

The other three predictions are derived from our first general prediction:

2. *The REME is independent of conscious retrieval processes. The effect can dissociate from conscious retrieval.*

3. *The REME is task independent.*

4. *The REME is stimulus-type independent.*

In the following section we will focus our attention on the individual predictions in more detail and we will highlight how we intended to test each of these in our experimental work.

1. Is the REME independent of conscious processes?

Our assumption, which states that the REME signals a necessary process for relational retrieval predicts that the emergence of the REME can dissociate from conscious processes. Previous results are not conclusive related to this question. The Hannula and Ranagath (2009) results point to the direction of a possible dissociation as we described earlier. However, in the experiments of Hannula at al. (2007) participants completed an extensive training on the face-scene pairs, completing five study blocks before the test phase. This exhaustive training resulted in high response accuracy on match displays, when they chose the matching face correctly on 95,2 % of the trials. In this study the REME was calculated by comparing these accurate answers on match displays to chosen faces on the re-pair and novel displays. The high accuracy level on this task can be regarded as a sign of conscious retrieval, which could draw the eyes to the matching face. We can conclude that to accept the possibility of the dissociation between the REME and conscious retrieval still lacks strong experimental proof. Finding evidence for this dissociation is also essential to regard the REME as a potential behavioural measure of the first-stage of recollection.

In our experiments we tested for this potential dissociation. In Experiment 1 we used the same experimental design as Hannula & Ranganath (2009). We specifically wanted to show the dissociation between the REME and conscious retrieval by analysing incorrect responses, when there is presumably a lack of conscious retrieval. In Experiment 2 we changed the choice task to a no-choice task and separated participants based on their level of awareness related to the presence of a matching face on the three-face test displays. We tested the possible dissociation of the REME and conscious retrieval by comparing the aware and unaware participant group.

It is a longstanding question in memory research whether eye movements are capable to signal unconscious memory processes (Ryan et al., 2000, Smith, Hopkins & Squire, 2006; Smith & Squire, 2008). Finding evidence that the REME can dissociate from conscious processes would provide pivotal evidence that in certain experimental paradigms eye movements can signal unconscious retrieval processes. We will get back to these questions in the general discussion part of the dissertation.

2. Is the REME independent of task characteristics?

If we accept that the REME signals a necessary process for relational retrieval then changing the task characteristics should still show the correlates of this effect. Previous results are pointing to the direction that preference of the matching face is task independent. In their second experiment, Hannula et al. (2007) changed the original choice task to a re-learning task. The design of the paradigm remained the same, but instead of choosing the matching face out of the three faces, the new task was to learn the three face displays together with their background scene for another recognition memory test. There was no explicit response needed, the participants just had to encode the three face displays for a suggested recognition test. The authors analysed the data for the match displays. They calculated the viewing time proportions for the matching faces in multiple time bins. Then, they tested whether the viewing time proportions of the matching faces are statistically different from the 0.33 proportion level. They used this level because 0.33 viewing time proportion is predicted if there is no preference between the three faces. The results showed that during the first 2000 ms the matching faces had statistically higher viewing time proportions than the 0.33 chance level. The result appeared as early as 500 ms after test trial onset (Figure 17, page 70). The authors concluded that the REME is early-emerging and it occurs obligatory, regardless of response requirements (Hannula et al., 2007). However in our opinion the criterion used to measure matching face preference in this task was moderate, which raises questions about whether one should accept it as evidence of matching face preference. If one wants to show that the matching face has a viewing time preference in this task, then testing the matching face proportion values against the 0.33 chance level is not a sufficient test method. It can be possible that during the test display presentation, two faces are equally preferred and the third one is neglected. This would result in the two equally preferred faces having more than 0.33 values while the third face would have a value below 0.33. This means that there could be cases when the matching face would show above 0.33 values but it would not be preferred more than a non-matching face. To test whether there is a genuine preference of the matching face in this no-choice task one should test the viewing time proportions of the matching faces against the two other, non-matching faces. Based on these concerns, in Experiment 2 we used a similar experimental design with a no-choice task and tested the matching face viewing time proportions against the two non-matching faces. If the REME is independent of task characteristics then we should see evidence that the matching faces are preferred compared to the non-matching faces in the first 2000 ms of the test trials.

3. Is the REME independent of stimuli characteristics?

Our assumption states that the REME signals a universal and necessary relational retrieval process. In two experiments we tested the prediction that when there is measureable relational retrieval on the task, the relational eye movement effect has to appear independently of stimulus characteristics.

In Experiment 3 we altered the perceptual characteristics of the target faces between learning and test, and tested whether the REME still appears in these conditions.

In Experiment 4 we altered the target stimulus category to objects. All previous experimental results are based on the paradigm were using a specific stimulus category, namely human faces (Hannula et al., 2007; Hannula & Ranganath, 2009) or pictures of toys with faces (Chong & Richmond, 2015) or pictures of 3D creatures with faces (Baym et al., 2014) as target stimuli. Our main question in Experiment 4 was to test whether we can elicit the REME with a target category other than faces or face-containing stimuli. We predicted that if participants express relational memory retrieval in their behavioural responses on the task, than the REME has to appear, irrespective of the target stimulus category.

Chapter 5: Experiments

Overview of our experiments

We conducted four experiments, which were intended to test our predictions about the relational eye movement effect. In this overview we will list the experiments and their main questions related to the characteristics of the REME:

In Experiment 1 we used the same face-scene choice task as Hannula & Ranganath (2009) with our own stimuli. We wanted to replicate previous results showing the REME in the task. Moreover, we asked participants to report their subjective confidence levels after each choice they made. We used these confidence ratings to test the emergence of the REME on different criteria levels. Our first prediction states that the REME has to emerge on confidence levels where there is evidence of relational retrieval in the task. By comparing different confidence levels we tested this prediction and we also wanted to get insight whether the magnitude of the REME is connected to subjective experience of conscious retrieval, what would point to the direction that the REME is tightly linked to conscious retrieval processes. Moreover, we also tested whether we can find any evidence that the REME can dissociate from conscious retrieval (our second prediction). We separately analysed whether we can find a matching face preference in incorrect responses, when there is supposedly no conscious retrieval of the matching face.

In Experiment 2 we changed the task to a no-choice task, where participants did not have to identify the matching face in the three-face test displays, they only had to learn the three faces together with the scene for a later recognition test. We wanted to find evidence for the emergence of the REME in this task, which would confirm our third prediction. Additionally, we separated participants based on their level of awareness for the presence of the matching faces in test displays to a high and a low awareness group. By comparing these two groups we intended to show a possible dissociation between the REME and conscious retrieval of the matching faces (second prediction).

In Experiment 3 we used a choice task and changed the perceptual characteristics of the faces between learning and test and we tested for the emergence of the REME.

Based on our predictions we hypothesized that the REME has to appear when participants can reliable show relational memory performance on the task, irrespective of stimulus characteristics (fourth prediction).

In Experiment 4 we changed the stimulus category from faces to objects in a choice task. Based on our fourth prediction we hypothesized that when participants show relational retrieval in the task the REME has to emerge.

Experiment 1.

Aims of the experiment

1. Using our own stimuli we wanted to replicate previous results showing a rapid REME in the first 2000 ms of the trials in the face-scene learning task. Hannula & Ranganath (2009) showed a time-course REME by comparing correct and incorrect answers.
2. We also wanted to show a REME before explicit responses using the comparison of correct-incorrect answers. A previous study showed a response-locked REME, appearing 500-1000 ms before responses were made (Hannula et al., 2007). However, that study used a slightly different method to calculate the REME by comparing chosen matching faces and chosen non-matching faces on re-pair displays, when there was no matching face present. We predicted to show the same response-locked REME between incorrect and correct responses in three-face displays containing a matching face.
3. Based on our assumption that the REME is necessary for relational retrieval our first prediction states that when there is evidence of relational memory retrieval in the task the REME has to appear. We tested whether the REME emerges when there is above chance performance on the task (first prediction). To test this prediction we collected subjective confidence judgments for every response in the task and we measured the REME on different response criteria levels: low, medium and high subjective confidence levels.
4. Our method to compare low, medium and high confidence levels was also useful to test whether the REME is a linked to conscious processes. We can regard the subjective confidence as one measure of subjective qualities of memory (Moscovitch et al., 2016), which is possibly related to the consciously available information at retrieval. If we would find a REME difference between confidence levels that could give us an insight whether the REME is linked to conscious processes.
5. Based on our assumption that the REME signals a necessary and unconscious relational retrieval process our second prediction states that the REME is

independent of conscious processes. This independence should manifest in a possible dissociation between the emergence of the REME and conscious retrieval, namely that the REME could appear in the absence of conscious retrieval. We tested for this possible dissociation by analysing whether we can find a matching face preference separately in incorrect responses, when there is presumably no conscious retrieval of the matching face to guide memory responses.

Methods

Participants

Data were analysed for 49 undergraduate students (ages 18-27; 32F/17M) from Eötvös Loránd University, Budapest, who received course credit for participation. All participants gave informed consent for the study approved by the Research Committee of Eötvös Loránd University, Budapest (Ethical No.: 2012/6). Data from 1 additional participant were excluded because of failure of reliable eye movement recording throughout the test phase.

Behavioural paradigm

Matlab (MathWorks Inc.) with Psychophysics Toolbox (PTB-3) extension (Brainard, 1997) was used to present stimuli and collect responses and eye tracking data. The head of the participants was positioned 65 cm away from a 17-inch 1920 x 1200 pixel resolution secondary monitor on which all the instructions and stimuli were presented.

Experimental stimuli consisted of 99 colour face images (50F/49M) and 99 colour scene images. The stimuli were selected from a collection created by Eszter Somos (MA student at the Budapest University of Technology and Economics). The collection used a stimulus bank that was available to undergraduate and graduate students at the Cognitive Science Department of the Budapest University of Technology and Economics. All stimuli were similar to the ones used in previous studies (Hannula et al., 2007; Hannula & Ranganath, 2009). Face images were 190 x 254 pixels with uniform grey backgrounds and scene images were 720 x 720 pixels outdoor photographs (for examples of faces and scenes see Figure 22).

The experiment consisted of three phases. During the learning phase participants viewed 99 face-scene pairs. They were instructed to try and remember the pairs for a later memory test. A learning trial started with a fixation point in the middle of the screen for 500 ms, which was followed by a scene for 2 seconds and immediately after the scene a face appeared superimposed in the middle of the scene for 3 seconds (see Figure 22A). During the learning phase there was a short (max. 2 minutes) break after 50 trials.

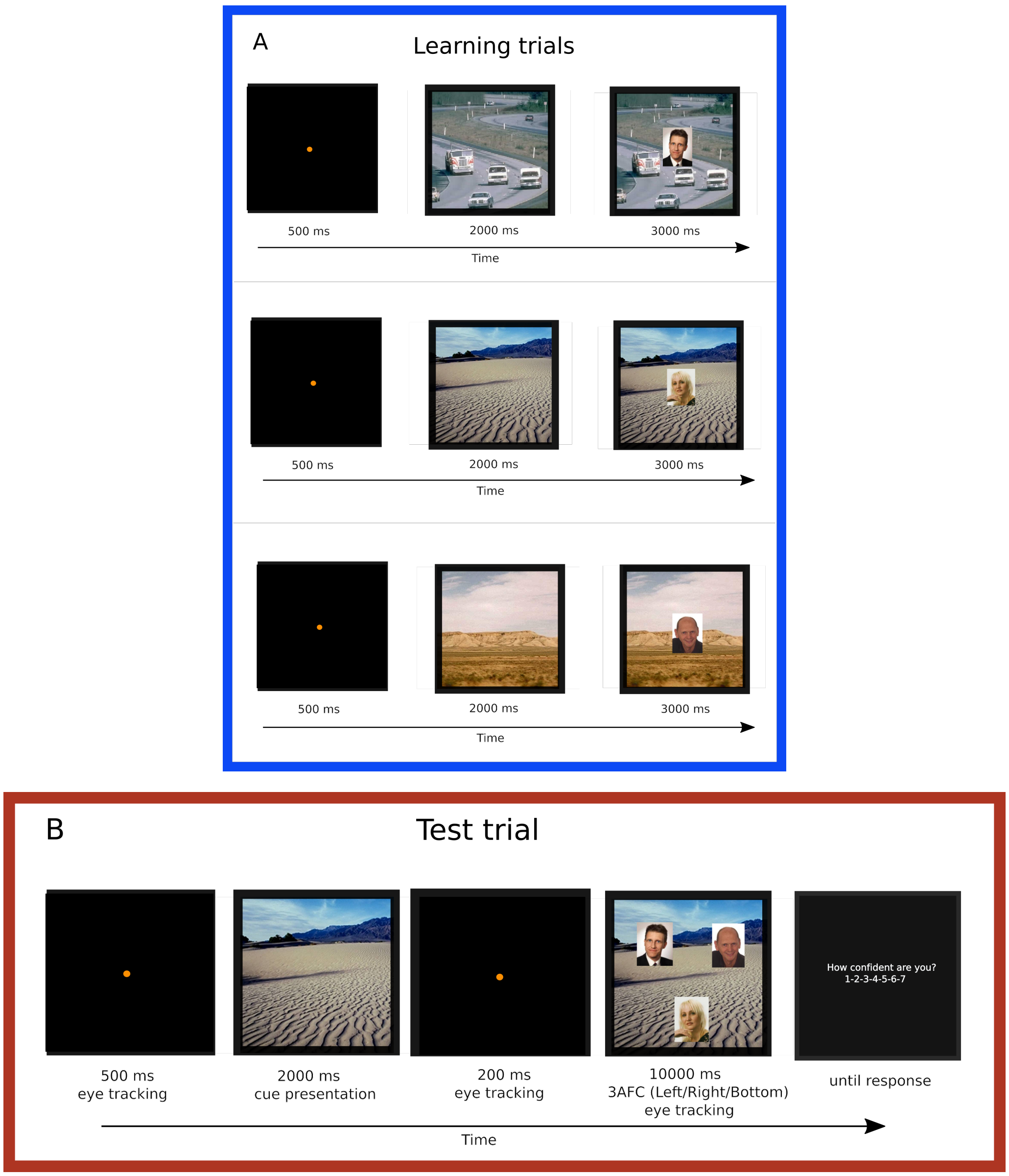

Figure 22. A: Examples of learning trials. B: Example of a test trial.

After the learning phase participants completed a short practice task. The practice phase used eye tracking and a simple detection task with the same response design, that was later used during the test phase. The practice task was 20 trials long. A practice trial started after participants looked in the middle of the screen to a fixation point for 500 ms. After the fixation point disappeared, three 190 x 254 pixels coloured rectangles appeared equally distanced from the fixation point. Two rectangles were placed next to each other above the fixation point and one was placed below it. Two of the rectangles were red and one was blue. The location of the blue rectangle was randomized for every participant. The task was to find the blue rectangle and press the corresponding button on the keyboard (a three alternative forced choice task, 3AFC). The three possible buttons were 'd' (for the left rectangle), 'k' (for the right rectangle) and 'space' (for the bottom rectangle), which followed a similar spatial configuration on the keyboard as the rectangles on the screen. Participants were asked to place their fingers on the response buttons and never look down to the keyboard. After a correct response was made participants were asked to press a number between 1 and 7 on the keyboard to proceed to the next trial. The practice task served as an easy task to familiarize participants with the eye tracking method and with the responses, which they needed to use later in the test phase. After the practice phase participants started the test phase.

The test task consisted of 33 randomly chosen face-scene pairs, which were studied before. During a test trial participants had to look at the fixation point for 500 ms which was located in the middle of the screen. Then a scene image appeared for 3 seconds (*scene cue*). After the scene image, participants had to look again at a fixation point for 200 ms to proceed to the test display. This gaze contingent part of the procedure ensured us that the location of the eyes was always in the middle of the screen at the beginning of the test displays. The test display consisted of three faces superimposed on the scene image (see Figure 22B). The location of the faces was the same as the rectangles in the practice task. Two of the faces were previously paired with different scenes and one of the faces was the pair of the presented scene (matching face). The location of the matching face was counterbalanced across the three possible locations in the 33 test trials. Participants had to choose the face, which they thought was the pair of the scene during the learning phase (3AFC). They had to press the button corresponding to the chosen face on the keyboard while looking at the screen. The active buttons were the same as during the practice task ('d' = left face, 'k' = right face, 'space' = bottom face). The test display lasted 10 seconds and participants had to respond as quickly and as accurately as possible. Responses made after the test display disappeared were not analysed. After the test display participants had to indicate by pressing the corresponding numbers on the keyboard the confidence level of their responses using a 7-unit scale (1 = guessing, 7 = sure).

Eye tracking data acquisition and analysis

A Tobii T60XL eye tracker at 60 Hz was used to record eye positions. Before the practice and test phases participants completed a five-point standard calibration procedure. Our gaze contingent task design - which required eye positions to be measured at the fixation point to proceed to the test displays - ensured us that the recorded eye measurement positions remained reliable during data acquisition. We had to exclude only one participant because of failure to complete the test task due to tracking problems.

We applied two different methods to calculate viewing time proportions of the chosen faces. The first method used raw data and the second method used fixation-based data. The viewing time proportion measure as our dependent variable allowed us to use raw data of gaze positions for our calculations. We decided to use raw data, because there is no standard procedure in the literature to define fixations. There are different methods and there are different parameters, which can be used to segment raw data to fixations (see page 63-64). These potential differences between studies make fixation-based results harder to compare. Our raw data method used three (left face, right face and bottom face) region-of-interests (ROIs). For every raw sample we calculated whether the sample's location is within one of the ROIs. The total number of samples was the sum of all samples within any of the three ROIs. We calculated the viewing time proportions for each

AOI by dividing the number of samples of each AOI by the total number of samples. Based on the participant's responses we could measure the viewing time proportion for correctly or incorrectly chosen faces. We did this calculation of viewing time proportion of the chosen face for overall viewing behaviour before responses and for separate time bins too. For the time-course analysis we divided the 10 second test trial to 500 ms long time bins. Additionally, we used response-locked measures with 8 time bins. We defined 4 time bins before and 4 time bins after the responses during the test display. The length of our response-locked time bins was also 500 ms.

For our fixation-based calculations we segmented the raw samples using the GraFIX (Saez de Urabain, Johnson, & Smith, 2015) software. A fixation was defined by a gaze velocity threshold of 10º/sec and a minimum duration of 100 ms. We used the position and duration of fixations to analyse eye movement data. Fixations made during the 10 second test display were assigned to three (left face, right face and bottom face) regions of interests (ROIs) and the proportions of total viewing time allocated to each ROI were calculated for every trial. Similarly to our raw data analysis, we also divided each trial to 500 ms long time bins. Additionally we used a response-locked analysis with 8 time bins (each 500 ms long), 4 bins before and 4 bins after every response. For every time bin we also calculated the viewing time proportions of the ROIs.

Statistical analyses

We used repeated measures ANOVAs to analyse eye movements in different time bins to check for either the time-course or response-locked REME. The main factors were response accuracy (correct vs. incorrect) and time bins. For our ANOVAs we will always report the partial η^2 values to indicate the effect size of our main factors and interactions.

To compare eye movements on different criteria levels we also used repeated measures ANOVAs. Our main factors were response accuracy (correct vs. incorrect), confidence levels (low, medium and high) and time bins.

For all ANOVAs Mauchly's test of sphericity was used to ensure that the assumption of sphericity was met. We used Greenhouse-Geisser corrections if the assumption of sphericity was violated. For comparing the effects on specific time bins we used post-hoc comparisons with Bonferroni corrected paired *t*-tests.

Results

Task performance

Participants had an average accuracy of 56.65 (*SD* = 0.16) which was significantly above chance level (33%), $t(48) = 10.49$, $p < .001$, $d = 1.48$ (Figure 23). Three participants had chance level accuracy. We excluded chance performers from our data analysis because there was no sign of overall memory performance for these individuals. However, we also ran our analysis on all participants and our main findings stayed the same.

Correct responses had higher confidence levels (*M* = 4.7, *SD* = 0.86) than incorrect responses (*M* = 2.8, *SD* = 0.79), $t(45) = 14.01$, $p < 0.001$, $d = 2.06$). Average reaction time for correct answers was 4444 ms (*SD* = 931) and for incorrect answers, 5412 ms (*SD* = 917). Correct answers were faster than incorrect ones, $t(45) = -8.25$, $p < 0.001$, $d = -1.21$.

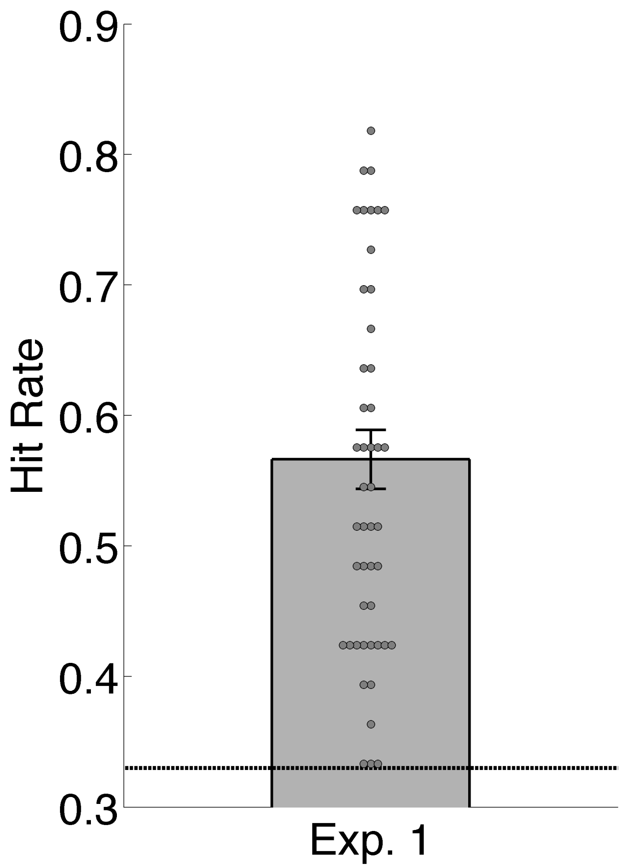

Figure 23. Average accuracy of Experiment 1, points represent individuals. Error bars indicate standard error of the mean (SEM).

We also calculated the average accuracy for each confidence level (Figure 24, Table 8). One-sample t-tests showed that Level 1 and Level 2 did not differ from chance level performance (33 %), however level 3-7 all showed above chance performance. These results indicated that participants used the confidence scale reliably: level 1 and level 2 were treated as guess responses, with no measureable performance on the task. From Level 3, there was a clear tendency of better task performance for subsequent confidence levels. Paired t-tests showed that there was no difference between Level 3 and 4 $t(45) = .22$, $p = .83$, $d = .03$, but Level 5, 6 and 7 were all showing significant increase compared to the preceding confidence levels: Level 4 vs. Level 5: $t(43)= -3.62$, $p = .001$, $d = -.54$; Level 5 vs. Level 6: $t(40)= -3.82$, $p < .001$, $d = -.59$; Level 6 vs. Level 7: $t(41)= -2.90$, $p = .006$, $d = -.44$.

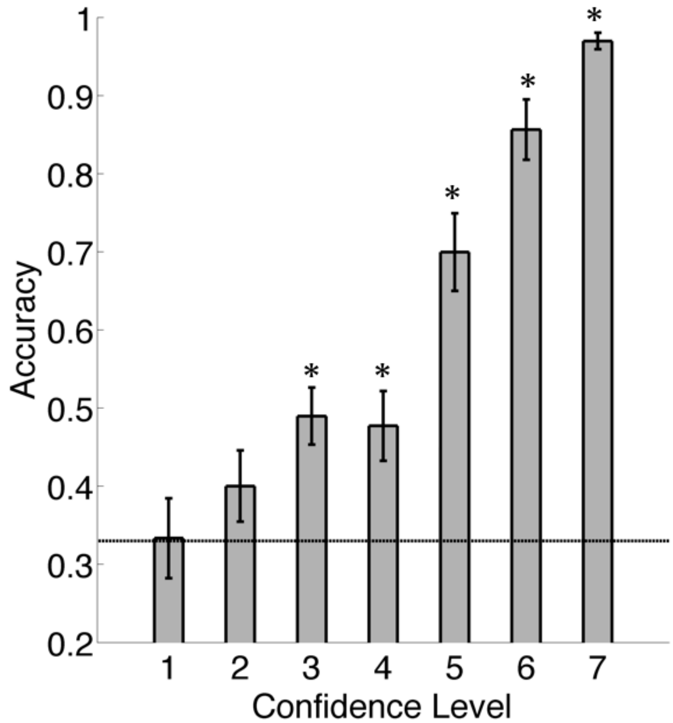

Figure 24. Average accuracy on different confidence levels. Error bars indicate SEM. Stars represent statistically significant differences from chance level performance.

Confidence Level	Mean and Standard Deviation	T-test results (test value = 0.33)
Level 1	$M = .33$, $SD = .32$	$t(37) = .07$, $p = .95$, $d = .01$
Level 2	$M = .39$, $SD = .29$	$t(40) = 1.55$, $p = .13$, $d = .24$
Level 3	$M = .49$, $SD = .25$	$t(45) = 4.39$, $p < .001$, $d = .65$
Level 4	$M = .47$, $SD = .30$	$t(45) = 3.30$, $p = .0019$, $d = .49$
Level 5	$M = .70$, $SD = .33$	$t(43) = 7.44$, $p < .001$, $d = 1.12$
Level 6	$M = .86$, $SD = .25$	$t(41) = 13.61$, $p < .001$, $d = 2.10$
Level 7	$M = .97$, $SD = .07$	$t(43) = 60.82$, $p < .001$, $d = 9.17$

Table 8. Average accuracy on different confidence levels and t-test results.

Next we created three different confidence levels. For Low confidence level we merged Level 1 and Level 2. Our reason for this was that both of these confidence levels showed chance level performance. Our Medium confidence level contained Level 3 and Level 4 answers and the High confidence level consisted of Level 5, 6 and 7 answers. The reason for this separation was that we wanted the Medium and

High confidence levels to have similar number of answers contributing to each category. We calculated the average answer count for Medium (3-4) and High (5-6-7) levels and found no significant difference between them (Medium (3-4) level: M = 11.5 SD = 4.8 and High (5-6-7) level: M = 12.6 SD = 5.7, $t(45) = .95$, $p = .35$, $d = .14$). We compared this result with a different confidence separation method, where Medium confidence level contained Level 3, 4 and 5 and High confidence level contained only Level 5 and 6 answers. For this alternative we found significant difference between the average answers counts (Medium (3-4-5) level: M = 15.4 SD = 5.4 and High (6-7) level: M = 8.7 SD= 5.0 $t(45) = 5.19$, $p < .001$, $d = .77$). Based on these results we used the first separation method, which meant that for Medium confidence level we merged Level 3 and 4 answers and for High confidence level we merged Level 5, 6 and 7. An additional reason to use our confidence separation was that Level 3 and 4 did not differ in task accuracy measures, but Level 5, 6 and 7 were all showing significant increase in accuracy compared to the preceding confidence levels (see page 91-92 for paired t-test results comparing accuracy on different confidence levels).

The accuracy results indicated that at Low confidence level there was no significant task performance (M = .41, SD = .27 $t(45) = 1.97$, $p > .05$, $d = .29$), however Medium confidence (M = .50, SD = .20 $t(45) = 5.63$, $p < 0.001$, $d = .83$) and High confidence level (M = .85, SD = .18 $t(45) = 18.57$, $p < 0.001$, $d = 2.73$) both showed above chance performance (Figure 25).

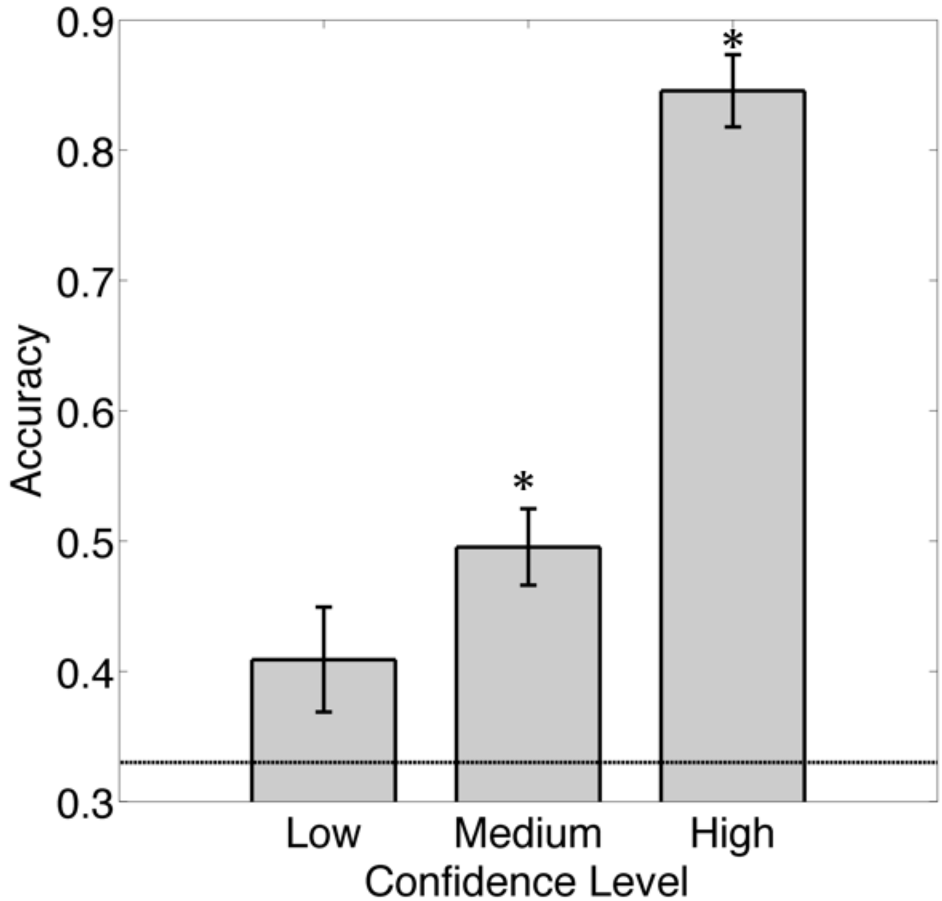

Figure 25. Average accuracy on Low, Medium and High confidence levels. Error bars indicate SEM. Stars represent statistically significant differences from chance level performance.

Eye movement results

For all our analysis we will report in the main text only our raw-data-based results because they showed more robust results compared to the fixation-based method. Both methods indicated the same main effects when we performed our ANOVAs but the raw-data-based values could be used more efficiently to detect significant pairwise differences in individual time bins. This sensitivity difference between raw-data-based and fixation-based results indicates that in certain cases it is better to use raw data to calculate eye movement measures. This is also important because it highlights that fine differences could be left undetected in the data if one would use certain fixation parameters to segment eye movements. In the appendix we will report in full detail our fixation-based results too.

Onset-locked time-course analysis

Firstly, we wanted to show evidence that our participants demonstrated a rapid REME within the first two seconds of the test trials. We were specifically interested in the first 2000 ms of the trials, because this is the time interval where previous results reported a rapid REME (Hannula et al., 2007; Hannula and Ranganath, 2009). We conducted a separate 2 (answers: correct, incorrect) X 4 (time bins) repeated measures ANOVA for the first 4 time bins. There was a main effect of answers, $F(1, 45) = 12.64$, $p = .001$, $\eta^2 = .22$, and time bins, $F(3, 135) = 8.61$, $p < .001$, $\eta^2 = .16$, but no interaction. Bonferroni corrected post-hoc paired *t*-tests showed a significantly greater proportion for correct than incorrect answers 500-1000 ms after trial onset $t(45) = 2.78$, $p = .008$, $d = .40$, while the difference in the fourth time bin (1500 – 2000 ms after trial onset) fell out of the corrected confidence threshold $t(45) = 2.34$, $p = .023$, $d = .34$. The time-course analysis results are shown in Figure 26. These results confirmed the emergence of a rapid REME in our experiment similarly to previous findings (Hannula et al., 2007; Hannula & Ranganath, 2009).

The results of the ANOVA remained the same when we conducted a 2 (answers) X 20 (time bins) repeated measures analysis for the whole trial. There was a main effect of answers, $F(1, 45) = 14.40$, $p < .001$, $\eta^2 = .24$, and time bins, $F(19, 855) = 7.41$, $p < .001$, $\eta^2 = .14$, but no interaction.

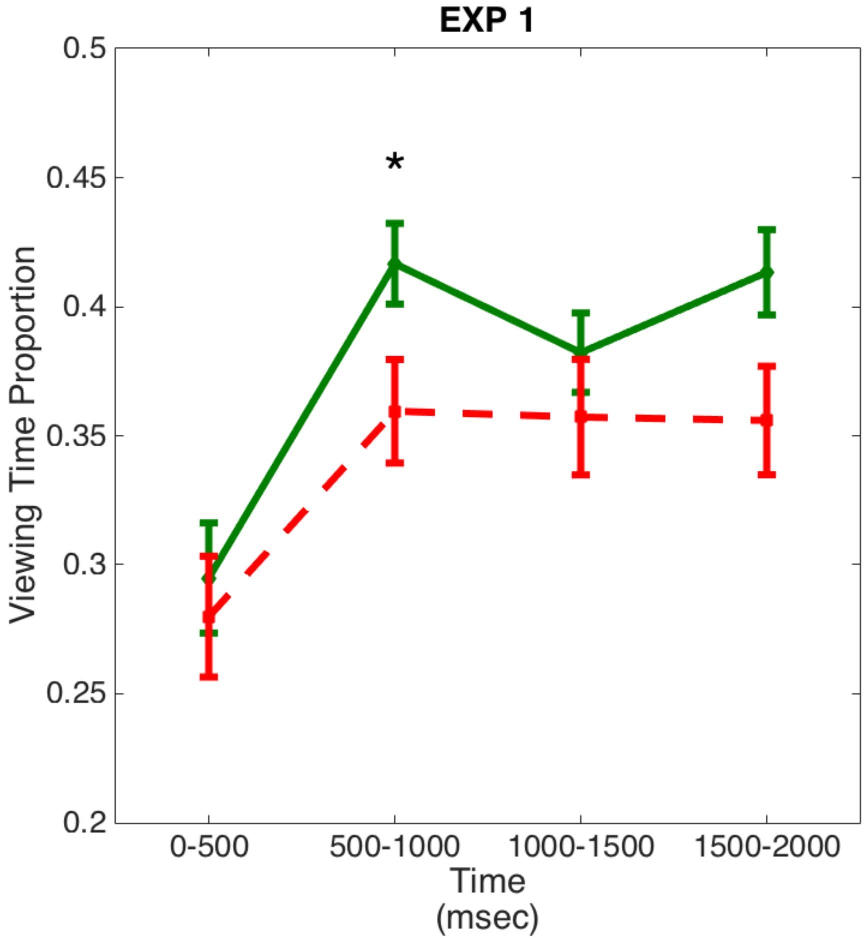

Figure 26. Onset-locked time-course analysis results in 500 ms time bins for the first two seconds of test trials. Green line indicates correct responses; red dashed line indicates incorrect responses. Error bars indicate SEM. Star represents a statistically significant difference.

Response-locked analysis

Overall response-locked eye movement results

We also analysed our eye movement data before overt responses to find evidence that the REME appears before responses are made. First, we analysed our overall eye movement data without segmenting the whole trials to several time bins. We compared the viewing time proportions before every response in the trials between correct and incorrect answers. For every trial we identified the time point of the participant's response and we only calculated the viewing proportions between the onset of the three-face test trial and the response. Using this technique we could control that we measure pre-response eye movements and there will be no contamination of post-response processes in our results, which might have an effect on eye movements. The results demonstrated the REME before responses, which was apparent in greater pre-response viewing time proportions for correct, than incorrect responses $t(45)=4.34$, $p < .001$, $d = .64$ (Figure 27).

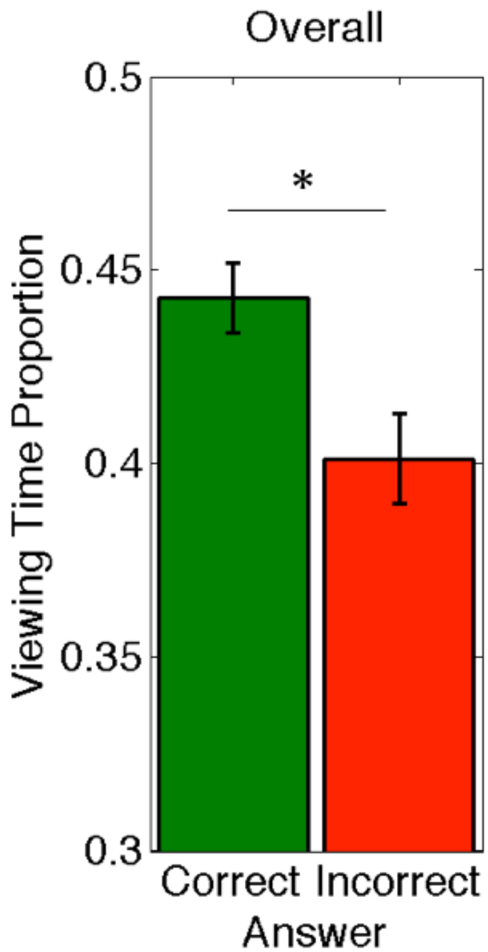

Figure 27. Overall response-locked viewing time proportions for correct and incorrect responses. Green indicates correct responses; red indicates incorrect responses. Error bars indicate SEM. Star represents a statistically significant difference.

Response-locked analysis for several time bins

Based on previous results (Hannula et al., 2007, Hannula et al., 2012) we were interested whether we can show the relational effect in the eye movements in specific time bins before the responses. We conducted a 2 (correct/incorrect answer) X 4 (time bins) repeated measures ANOVA and found a main effect of answer, $F(1, 45) = 9.01$, $p = .004$, $\eta^2 = .17$, and time bins, $F(3, 135) = 8.90$, $p < .001$, $\eta^2 = .17$, but no interaction. Bonferroni corrected post-hoc paired *t*-tests indicated significantly higher proportions for correct responses in two time bins. The first time bin was 500-1000 ms before responses $t(45) = 2.90$, $p = 0.006$, $d = .42$, and the second bin was 1000-1500 ms before responses $t(45) = 2.95$, $p = 0.005$, $d = .43$. The result of our response-locked analysis is presented in Figure 28. Our results were consistent with previous studies, which demonstrated the eye movement effect within the 500-1000 ms time window before overt responses (Hannula et al., 2007, see Figure 16, page 69).

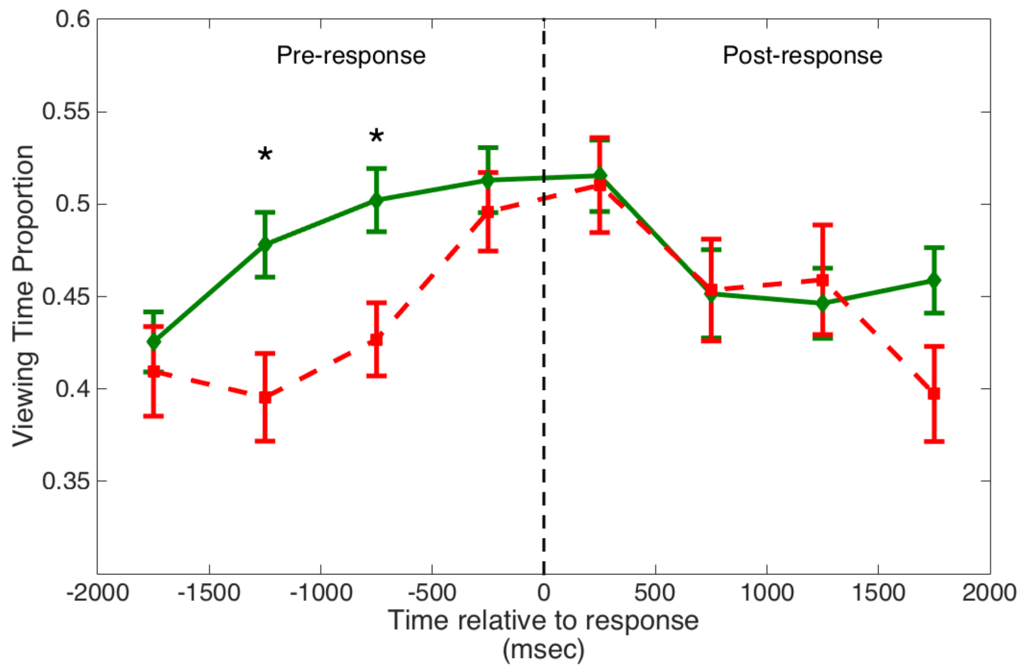

Figure 28. Response-locked results for 500 ms time bins before and after responses. Green line indicates correct responses; red dashed line indicates incorrect responses. Error bars indicate SEM. Stars represent statistically significant differences.

Eye movements on different confidence levels

Overall response-locked results on different confidence levels

We based our overall response-locked analysis on the viewing time proportions before correct and incorrect responses. We compared our three confidence level groups in a 2 (answers: correct vs. incorrect) X 3 (confidence level: low, medium, high) ANOVA, which showed a main effect of answers $F(1, 24) = 8.17$, $p = .009$, $\eta^2 = .25$, a main effect of confidence $F(2, 24) = 5.36$, $p = .008$, $\eta^2 = .18$, and a significant interaction between answer and confidence level $F(2, 48) = 11.36$, $p < 0.001$, $\eta^2 = .32$. Post-hoc Bonferroni corrected pairwise comparisons showed a no significant difference between correct and incorrect responses on Low confidence level $t(24) = -.82$, $p = .42$, $d = .16$, but the difference was significant between correct and incorrect responses on Medium, $t(24) = 2.62$, $p = .015$, $d = .52$, and on High confidence level, $t(24) = 3.91$, $p = .001$, $d = .78$ (Figure 29).

We also run two one-way ANOVAs to see how viewing time proportions are affected by confidence level in correct and incorrect responses. There was a main effect of confidence level on correct responses $F(2, 48) = 14.49$, $p < 0.001$, $\eta^2 = .38$, specifically viewing time proportions increased in correct responses between Low, Medium and High confidence levels (Low vs. Medium: $t(24) = -2.08$, $p = .048$, $d = -.42$; Medium vs. High: $t(24) = -3.50$, $p = .002$, $d = -.7$ and Low vs. High: $t(24) = -4.87$, $p < 0.001$, $d = -.97$). In contrast, confidence was not significantly associated with viewing behaviour in incorrect responses, $F(2, 48) = .67$, $p = .50$, $\eta^2 = .028$.

To check that the REME measured by the correct – incorrect viewing difference is significantly greater in High confidence level compared to Medium confidence level we also run a separate 2 (answers) x 2 (confidence Medium vs. High) ANOVA, which showed a main effect of answers $F(1, 24) = 16.77$, $p < 0.001$, $\eta^2 = .41$, and confidence $F(1, 24) = 5.24$, $p = .03$, $\eta^2 = .18$, and importantly, a significant interaction $F(1, 24) = 5.18$, $p = .03$, $\eta^2 = 0.18$. This interaction was also confirmed by comparing the correct - incorrect viewing time differences in the two confidence groups, which showed a greater difference for High confidence level compared to the Medium level $t(24) =$

2.28, p = 0.03, d = .46. We illustrated this effect size difference separately in Figure 30.

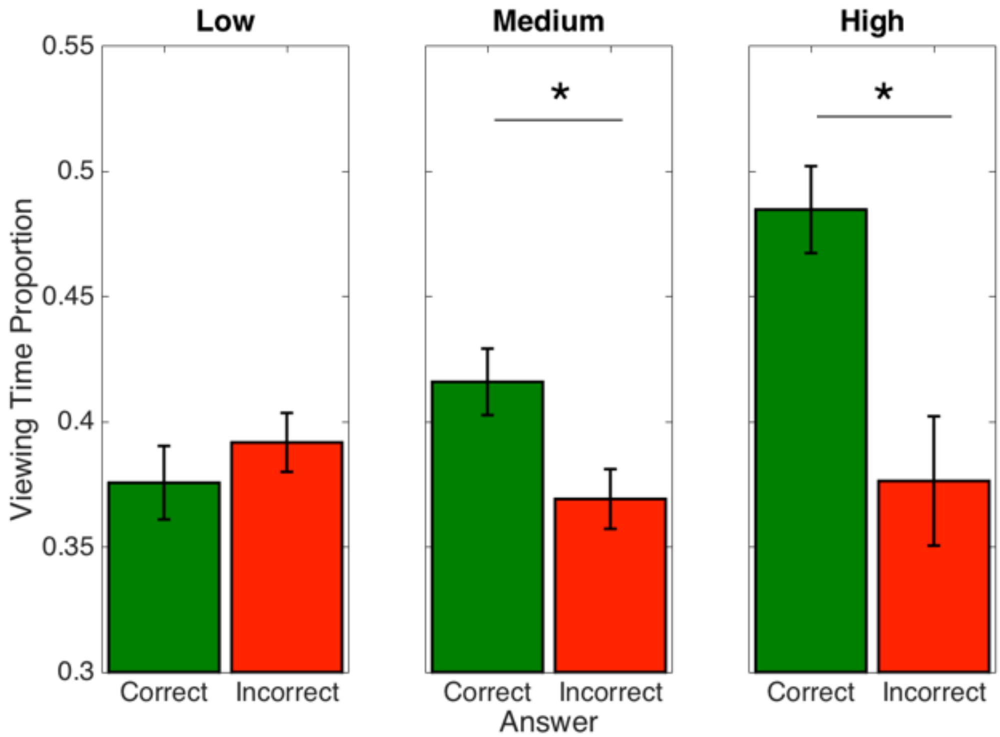

Figure 29. Overall response-locked viewing time proportions in Low, Medium and High confidence responses. Green indicates correct responses; red indicates incorrect responses. Error bars indicate SEM. Stars represent statistically significant difference.

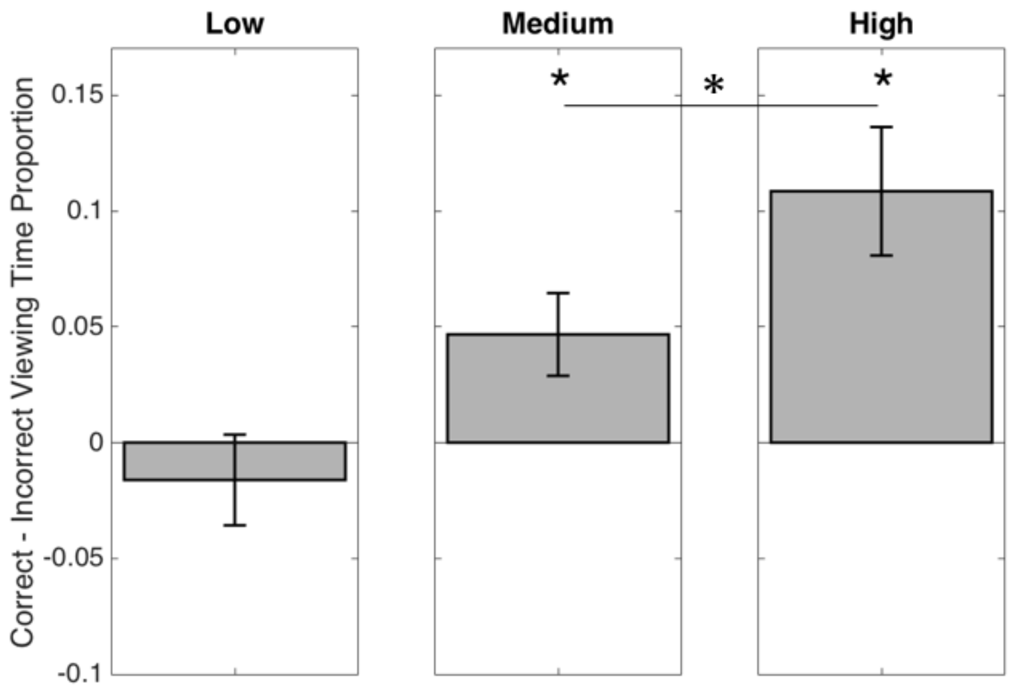

Figure 30. Overall response-locked viewing time proportion differences (Correct – Incorrect) in Low, Medium and High confidence responses. Error bars indicate SEM. Starts represent significant differences.

Matching face preference in incorrect responses

We also tested the possibility that a matching face preference would emerge in incorrect responses, when there is high probability that conscious access to the matching face was absent. We analysed incorrect responses and we separated the three test faces to different categories. The Chosen face was the incorrectly chosen non-matching face, the Matching face was the original pair of the scene and the Third face was the second non-matching face. We calculated the viewing time proportion of these faces separately for 500 ms time bins. We run a 3 (face type: chosen non-matching face, matching face, non-matching face) X 20 (time bins) repeated measures ANOVA and we were interested in finding evidence for a preference of the matching face compared to the other non-matching faces, which would show us that matching face preference can occur even, when overt memory responses lack conscious access to the relational representation. Our results showed a main effect of face type $F(2, 90) = 183.86$, $p < 0.001$, $\eta^2 = .80$, a main effect of time bins $F(19, 855) = 2.23$, p = .002, $\eta^2 = .05$, and a significant interaction $F(38, 1710) = 5.454$, p < 0.001, $\eta^2 = .11$. Our pairwise comparisons showed that there was no time bin where Matching faces had higher viewing time proportions than the other two, non-matching faces (Figure 31, black lines). We could conclude that we did not find any evidence for preference of the matching faces in incorrect responses, when conscious access to the matching face was absent.

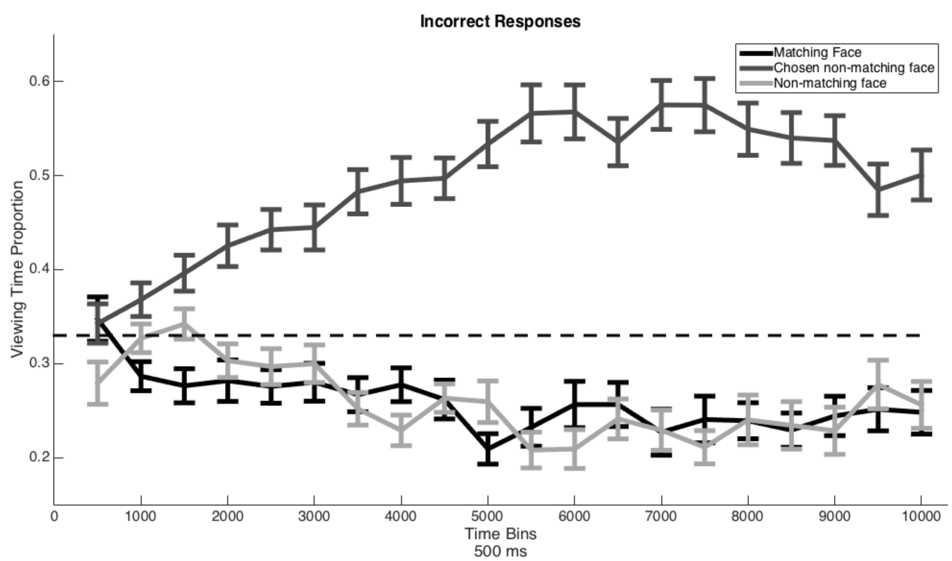

Figure 31. Viewing time proportions in incorrect trials for Matching, Chosen non-matching and non-matching faces. Error bars indicate SEM.

<u>Discussion</u>

Our participants showed reliable relational retrieval effect on the task with an average accuracy of 56 %. This task accuracy level allowed us to test whether we could find evidence for the previously reported relational eye movement effects.

Our first question was whether we could find that the relational eye movement effect emerges rapidly, within the first 2000 ms of the trials (Hannula & Ranganath, 2009). We compared the viewing time proportion of correctly and incorrectly chosen faces in consecutive 500 ms time bins within the first 2000 ms of the test trials and found that viewing time proportion was greater in correct responses compared to incorrect responses as early as 500-1000 ms after the test display onset (Figure 26). This result was consistent with previous studies showing the rapid emergence of disproportionate viewing of the correctly chosen face within the first 2000 ms of test trials (Hannula & Ranganath, 2009; Hannula et al., 2007).

Our second question was whether the relational eye movement effect is present before explicit responses. In our first analysis we compared the overall viewing data before responses in correct and incorrect responses and our results showed that correct responses had greater viewing time proportions than incorrect responses (Figure 27). This result confirmed that in our task participants showed evidence of the REME before their responses, which was also consistent with previous results (Hannula et al., 2007). Hannula et al. (2007) also demonstrated that the REME could be found 500-1000 ms before explicit responses (Hannula et al., 2007). Based on these results we also wanted to show that in our study we could find the effect emerging in the same time window. We separately analysed viewing data in four 500 ms time bins before responses and found that correct responses had greater viewing time than incorrect responses in two consecutive time bins: 500-1000 ms and even 1000-1500 ms before responses (Figure 28). These results replicated previous response-locked results by showing evidence that disproportionate viewing of the correctly identified matching face was evident 500-1500 ms before explicit responses were made (Hannula et al., 2007).

Taken our results together we could conclude that by using our own stimuli we could replicate previous results in the relational eye movement literature (Hannula & Ranganath, 2009; Hannula et al., 2007). In our study we demonstrated that the REME could appear rapidly, within 500-1000 ms after test display onset. Moreover, we found evidence that the effect appeared before responses and that it could be measured in specific time bins 500-1500 ms before explicit responses.

In this experiment we also wanted to test two of our predictions derived from our assumption. Our assumption states that the retrieval process indexed by the relational eye movement effect is necessary for relational retrieval, which accompanies retrieval unconsciously, outside of our ability to consciously apprehend it (see page 75).

Based on this assumption, that the REME is necessary for relational retrieval, our first prediction states that when there is evidence of relational memory retrieval in the task, the REME has to appear. We tested this prediction in this experiment by using participants' subjective confidence levels to measure the REME on different subjective criteria levels. We created three confidence levels Low, Medium and High. Our accuracy results indicated that participants did not show above chance relational memory performance on the Low confidence level (M = .41, SD = .27) but they showed reliable above chance relational memory performance both on Medium (M = .50, SD = .20) and High confidence levels (M = .85, SD = .18). Based on these performance measures we tested the emergence of the overall relational eye movement effect before responses and found no evidence of the REME on Low confidence level, however the relational eye movement effect appeared on Medium and High confidence levels. These confirmed our prediction that whenever there is evidence of relational memory retrieval on the task, the REME has to appear. We found no choice performance indicating relational memory in Low confidence level and the REME was absent, however on both Medium and High confidence level there was evidence of relational memory in choice performance and participants also demonstrated the eye movement effect.

Additionally, we also found a link between the magnitude of the REME and confidence level. Specifically, our results showed that the magnitude of the disproportionate viewing difference was greater for High confidence level than Medium confidence level, which was driven by the increase of viewing proportions in correct responses (Figure 29). If we regard that the reported subjective confidence level is considerably influenced by the consciously available information at retrieval than our results point to the direction that the process behind the REME might be tightly linked to conscious retrieval. Higher level of conscious access might be associated with greater REME and more confident responses. This possible interpretation is in contrast to our assumption, which states that the REME signals a necessary precondition to relational memory retrieval, which is not accessible to conscious processes. This unconscious property of the REME would predict no difference between Medium and High confidence levels regarding the eye movement effect. Our result also raises questions whether the REME can be regarded as the behavioural measure of the first-stage of recollection, which enables recollection but which is not available to conscious processes. According to the two-stage model of recollection it is the second, slower stage of recollection that is associated with conscious retrieval processes. We further explored the possibility between the tight link of the REME and conscious retrieval in Experiment 2.

Related to these issues, based on our assumption that the REME is an index of a an unconscious process independent of conscious access to the relational information, we predicted that the emergence of the REME could dissociate from conscious retrieval. We tested this prediction by analysing whether the matching face preference in eye movements can emerge in incorrect responses, when there is presumably a lack of conscious retrieval of the matching face. To answer this question, in every incorrect trial we identified the incorrectly chosen non-matching face, the matching face and the second non-matching face and we calculated the viewing time proportions of these faces in consecutive 500 ms time bins throughout the whole test trials. According to our results we found no evidence in any of the time bins of a matching face preference, which would be evident in significantly greater viewing time proportion of the matching face compared to the other two faces. We concluded that contrarily to our prediction, we failed to find a matching face preference in incorrect responses, when there is no sign of conscious retrieval. However, there is a confounding issue related to this null result. One could argue that the lack of dissociation between the REME and conscious retrieval might be driven by the fact that incorrect responses simply represent the complete lack of any relational memory for the face-scene information. In our opinion this is an important issue and we tested the possible dissociation between the REME and conscious retrieval in Experiment 2. In that experiment we changed the task and we separated participants who showed high vs. low levels of awareness regarding the presence of a matching face in test displays. Consistent with our prediction we wanted to find evidence of the REME in low awareness participants, who supposedly lack conscious access to the matching face, but who might show other evidence of relational memory retrieval.

In sum, our study replicated previous results showing that the REME can appear rapidly and before explicit responses. Additionally, based on our assumption that the process indexed by the REME is necessary for relational retrieval, we predicted that when there is behavioural evidence in participants' choices that indicate relational retrieval (above chance performance on the task) than eye movements have to show the REME. Comparing different response criteria levels we found evidence that the REME appears when performance on the task signal relational memory retrieval (Medium and High confidence level). Moreover, our results showed greater REME magnitude for High compared to Medium confidence levels, which raise the question whether the process behind the REME not only enables conscious access but it might be tightly linked to the level of conscious retrieval during retrieval. To complement this possibility we failed to find that the REME could dissociate from conscious retrieval by showing the lack of matching face preference in incorrect responses. This possible dissociation was predicted based on the assumption that the REME is a necessary and unconscious index of relational retrieval, which is independent of conscious processes.

Experiment 2

<u>Aims of the experiment</u>

1. Our assumption states that the REME signals a universal and necessary process for relational retrieval. This predicts that when there is relational retrieval in the task the REME has to appear irrespective of task characteristics. We tested this prediction by giving participants a no-choice task at test in which they did not have to explicitly choose between faces in the test displays. We tested the emergence of the REME by comparing the matching face viewing time proportions against the non-matching face proportions in the first 2000 ms of the test trials.

Previous results point to the direction that the matching face preference is task independent and it might emerge even when overt responses are not necessary (Hannula et al., 2007). In that study the design of the paradigm remained the same, but the task of the participants' was changed. Instead of choosing the matching face out of the three faces, the participants' were told to learn the three face displays together with their background scene for a future recognition memory test (however no subsequent memory task was administered). There was no explicit response needed, the participants just had to encode the three face displays for a suggested recognition test. The authors calculated the viewing time proportions for the matching faces in multiple 500 ms time bins. They tested the matching face viewing proportions against the 0.33 value to indicate relational memory in eye movements. They use this level because 0.33 viewing time proportion is predicted if there is no preference between the three faces. The results showed that during the first 2000 ms the matching faces had statistically higher viewing time proportions than the 0.33 chance level, which preference appeared as early as 500 ms after test trial onset (see Figure 17, page 70).

However, in our opinion if one wants to show that the matching face has a viewing time preference in this task, then testing the matching face proportion values against the 0.33 chance level is not a sufficient test method. There could be two equally preferred faces above 0.33 viewing time proportions when one face is neglected during the test displays. As we also argued earlier (see page 80), to show a genuine matching face preference in this setup, one should test the viewing time proportions of the matching faces against the two other, non-matching faces. Comparing the matching face against the two non-matching faces is the adequate test to show any matching face preference in the no choice task. This was our reason that in this experiment we also tested the preference of the matching face against the two non-matching faces instead of the 0.33 value.

2. Based on our assumption that the REME accompanies relational retrieval obligatorily and unconsciously we predicted that the eye movement effect is independent of conscious processes (second prediction). Our second aim of Experiment 2 was to test this prediction and find evidence that the REME can dissociate from conscious retrieval. In Experiment 1 we failed to find evidence for such dissociation in incorrect responses. Moreover, our previous results showing greater REME for High confidence responses than Medium level responses suggested that there might be a tight link between the emergence of the REME and the level of conscious access to the matching face. In this experiment we wanted to explore the possible link between the REME and conscious retrieval.

Our experimental design was based on Hannula et al. (2007), who found a relational memory effect (above than 0.33 viewing proportion for matching faces) even when participants were asked to learn the three-face test displays instead of choosing the matching face. However, it was not tested in this earlier experiment, whether participants in that study showed conscious access to the matching faces. In the Hannula et al. (2007) study the no choice task had the same design as their first experiment, where participants had 95 % accuracy of choosing the matching face, which gave us reason to think that in the no choice task participants might also showed conscious retrieval. In this experiment we were interested in whether we could separate participants who show high vs. low levels of awareness regarding the presence of a matching face in test displays and we tested whether we can find any difference in the two groups regarding their eye movements. To achieve this goal, at the end of the test phase we asked the question from our participants whether they realized anything about the faces and scenes between the learning and test phase. We predicted that the aware group would show the REME because they can demonstrate relational retrieval by explicitly reporting the presence of the matching face in test displays. In the unaware group we were interested whether we can find any evidence, which suggest relational memory retrieval. If such evidence would be present in the unaware group than we predicted that this group would also have to show the REME, which is assumed to signal a necessary process of relational retrieval independent of conscious processes.

Methods

Participants

There were 38 undergraduates (ages 18-27; 22F/16M) taking part in the experiment from Eötvös Loránd University, Budapest. All of them received course credit for participation and gave informed consent for the study approved by the Research Committee of Eötvös Loránd University, Budapest (Ethical No.: 2012/6).

Behavioural paradigm

We used the same learning and test phase as in Experiment 1. The stimuli were also the same and each participant received randomly assigned scene-face pairs for the task. The only difference compared with Experiment 1 was the nature of the task during the test phase. Based on the instructions by Hannula et al. (2007) participants were told that their task would be to learn the three faces along with the scenes for a forthcoming recognition test (which was not administered). They did not have to use any responses during or after test displays. All they were instructed to do is to concentrate and memorize the scene + three face displays. Because the task did not require participants to use key presses we did not use a practice task between learning and test. Instead of the practice task participants had a 5-minute break between learning and test. The instructions for both the learning and test phases were shown to the participants on the computer screen before each session.

At the end of the test phase the experimenter asked the participants whether they realized anything about the faces and scenes between the learning and test phase. Depending on the participant’s answers the experimenter indicated whether the participant was unaware, partially aware or completely aware of the relationship between the scenes and matching faces during the test phase. Unaware participants could not report any relationship between the scenes and the faces in the experiment. Partially aware participants usually mentioned that there were some test trials where the scene and one of the faces were presented together during the learning phase. Aware participants were the ones who explicitly reported that on all of the test trials there was a matching face present. The experimenter also had the opportunity to take additional notes about the participants answer to clarify them if necessary.

Eye tracking data acquisition and analysis

The same Tobii T60XL eye tracker at 60 Hz was used to record eye positions as in Experiment 1. Before the test phase participants completed a five-point standard calibration procedure. During the test phase we used the same gaze contingent stimulus presentation method as in Experiment 1, which required the average eye positions to be measured at the fixation point for 500 ms to proceed to the next trial. With this method we ensured that participants started viewing the three face test displays in the middle of the screen equal distance from the three faces at the start of every trial.

Similarly to Experiment 1, we used both fixation-based and raw-data-based methods to calculate the viewing time proportion for the three faces. We used 500 ms time bin separations in the trials to compare eye movement results. We will only report our raw-data-based results but we included our fixation-based results in the appendix.

We used a within-display calculation to compare the viewing time proportions between faces. For every trial we identified the matching face and based on the viewing proportions throughout the whole presentation of the test displays, we separated the remaining two faces to Low viewing and High viewing categories. Low viewing faces were the ones, which had lower total viewing time proportions between the two non-matching faces. Contrarily, High viewing faces were the ones with higher total viewing time proportions compared to the other non-matching face. With this method we ended up with three viewing time proportion values for each of the 500 ms time bins for every participant: Matching face proportion, High viewing proportion and Low viewing proportion.

Statistical analyses

We used repeated measures ANOVAs to analyse eye movement results to check for preference for matching faces in different time bins. Our main factors were face type (Matching face, Low viewing face and High viewing face) and time bins (500 ms time bins).

For the ANOVAs Mauchly's test of sphericity was used to ensure that the assumption of sphericity was met. We used Greenhouse-Geisser corrections if the assumption of sphericity was violated. For comparing the effects on specific time bins we used post-hoc comparisons with Bonferroni corrected paired *t*-tests.

Results

Eye movement results for all participants

Firstly we were interested if we could find evidence of the REME by comparing the matching face viewing time proportion to the other non-matching faces (High viewing face, Low viewing face). We argued that to demonstrate a genuine REME one has to find the preference of the matching face compared to the other two faces. We were specifically interested in the first 2000 ms of the test trials, because previous results showed, that this early time window is when the REME appears. We run a 3 (face type: Matching, Low viewing, High viewing) X 4 (time bins) repeated measures ANOVA for the first 4 time bins. This resulted in a main effect of face type $F(2, 74) = 21.50$, $p < .001$, $\eta^2 = .37$ and no significant main effect of time bins $F(3, 111) = 1.63$, $p = .18$, $\eta^2 = .04$ and no interaction $F(6, 222) = .38$, $p = .81$, $\eta^2 = .10$. Most importantly, post-hoc comparisons showed that there was a significantly higher viewing time proportion for the Matching faces compared to the High viewing faces 500-1000 ms after stimulus onset, $t(37)= 2.67$, p = .011, $d = .43$, and there was a tendency for the same difference in the following time bin (1000-1500 ms after stimulus onset), $t(37)= 2.33$, p = .026, $d = .38$, (Figure 32). This result provides us with evidence that the REME can be elicited in the no-choice task, which indicates that the emergence of the relational eye movement effect is independent of task characteristics.

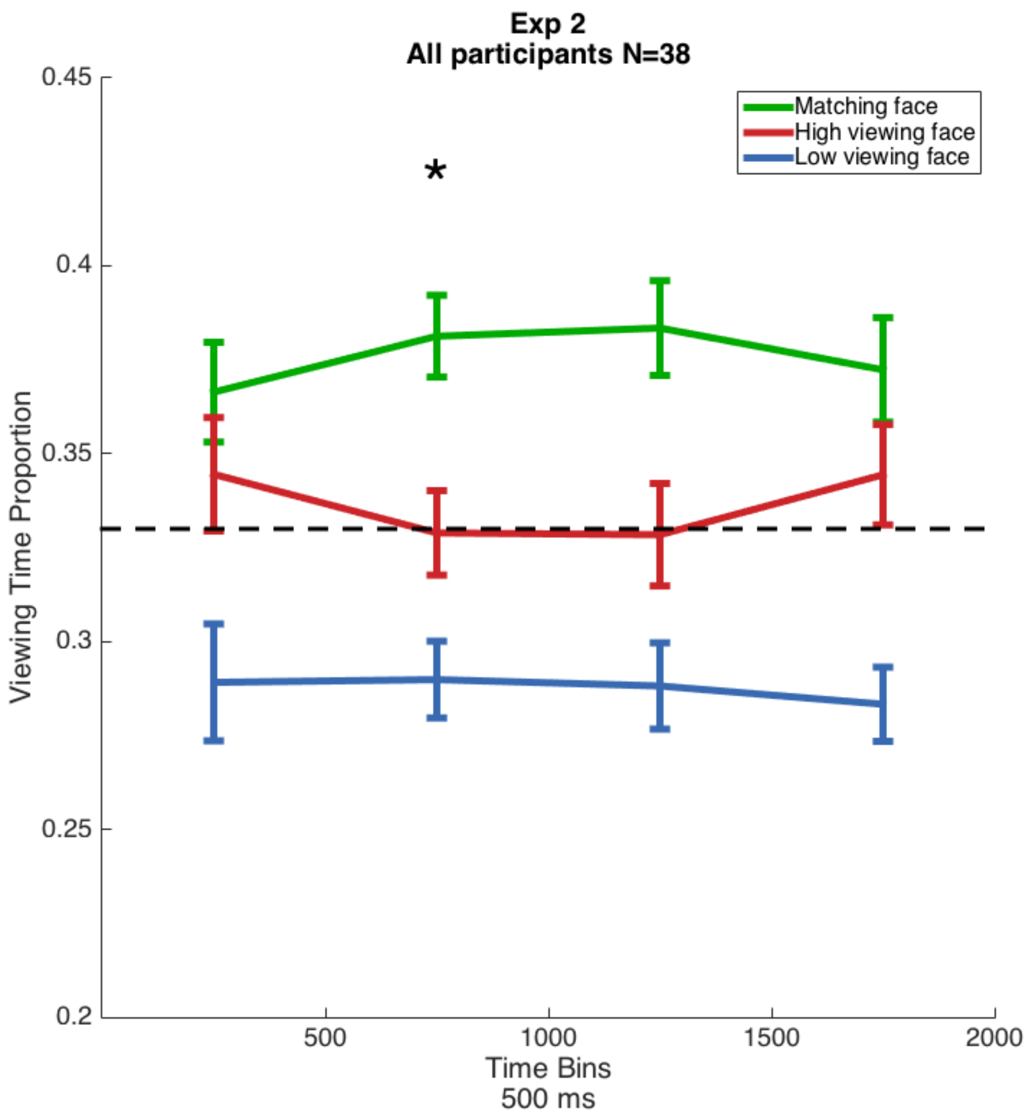

Figure 32. Viewing time proportions of Matching (green), High viewing (red) and Low viewing (blue) faces in 500 ms time bins for the first two seconds of test trials. Error bars indicate SEM. Star indicates significant difference between the Matching and High viewing group.

Eye movements in Aware and Unaware participant groups

To test our prediction that the emergence of the REME can dissociate from conscious retrieval of the matching face we separated participants based on their reported awareness level about the presence of the matching faces in test displays. We created three different awareness groups. The first group of participants explicitly stated that they realized that in all of the test trials one of the faces was the matching face (aware group). Our second group of participants mentioned that in some trials they could recognize that one of the faces was the matching face. These participants constituted the partially aware group. The third group was the unaware group in which participants did not report any kind of relation between the faces and scenes after the test phase (Table 9).

	Aware group (N)	**Partially aware group (N)**	**Unaware group (N)**
Total N = 38	18	5	15

Table 9. Number of participants in separate groups.

We run two separate 3 (face type) X 4 (time bins) repeated measures ANOVAs for the aware and the unaware group to test the emergence of the REME.

In the aware group the results showed a main effect of face type $F(2, 34) = 11.81$, $p < .001$, $\eta^2 = .41$, but no main effect of time bins $F(3, 51) = 1.23$, $p = .31$, $\eta^2 = .07$ or interaction $F(6, 102) = .45$, $p = .85$, $\eta^2 = .03$. Pairwise comparisons indicated that in Bin 2 (500-1000 ms after stimulus onset) $t(17)= 2.85$, $p = .012$, d = .67, and Bin 3 (1000-1500 ms after stimulus onset) $t(17)= 2.89$ $p = .011$, d = .68, the matching faces had higher proportions than High viewing faces (Figure 33). These results showed that for aware participants there was a clear matching face preference between 500 – 1500 ms after stimulus onset indicating the emergence of the REME in these participants.

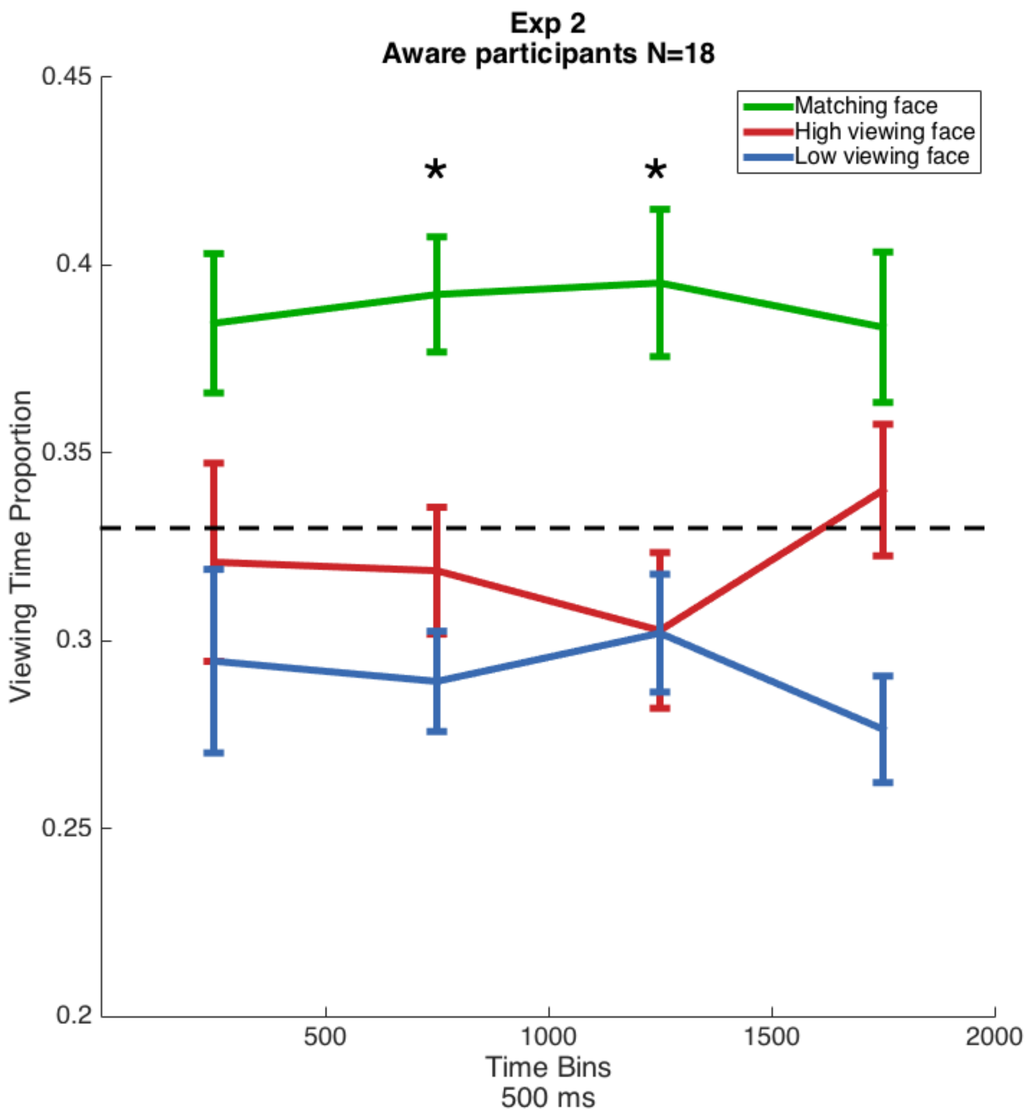

Figure 33. The aware group's viewing time proportions of Matching (green), High viewing (red) and Low viewing (blue) faces in 500 ms time bins for the first two seconds of test trials. Error bars indicate SEM. Stars indicate significant differences between the Matching and High viewing group.

In the unaware group the ANOVA results showed a significant main effect of face type $F(2, 28) = 7.94$, $p = .002$, $\eta^2 = .36$ and no main effect of time bins $F(3, 51) = 1.21$, $p = .35$, $\eta^2 = .04$ or interaction $F(6, 84) = 1.83$, $p = .98$, $\eta^2 = .13$. However, pairwise comparison showed no significant difference between Matching and High viewing faces (Figure 34). Additionally to our first ANOVA, we also run a 2 (face type) X 4 (time bins) analysis where we were specifically interested in whether the main effect of face would disappear when we take out the Low viewing face category and only compare the Matching and High viewing faces. The results showed that the face type main effect was no longer significant $F(1, 14) = .17$, $p = .69$, $\eta^2 = .01$ and there was no main effect of time bins $F(3, 42) = .19$, $p = .90$, $\eta^2 = .01$ or interaction $F(3, 42) = .18$, $p = .91$, $\eta^2 = .01$. These results clearly showed that there was no REME present in the unaware group.

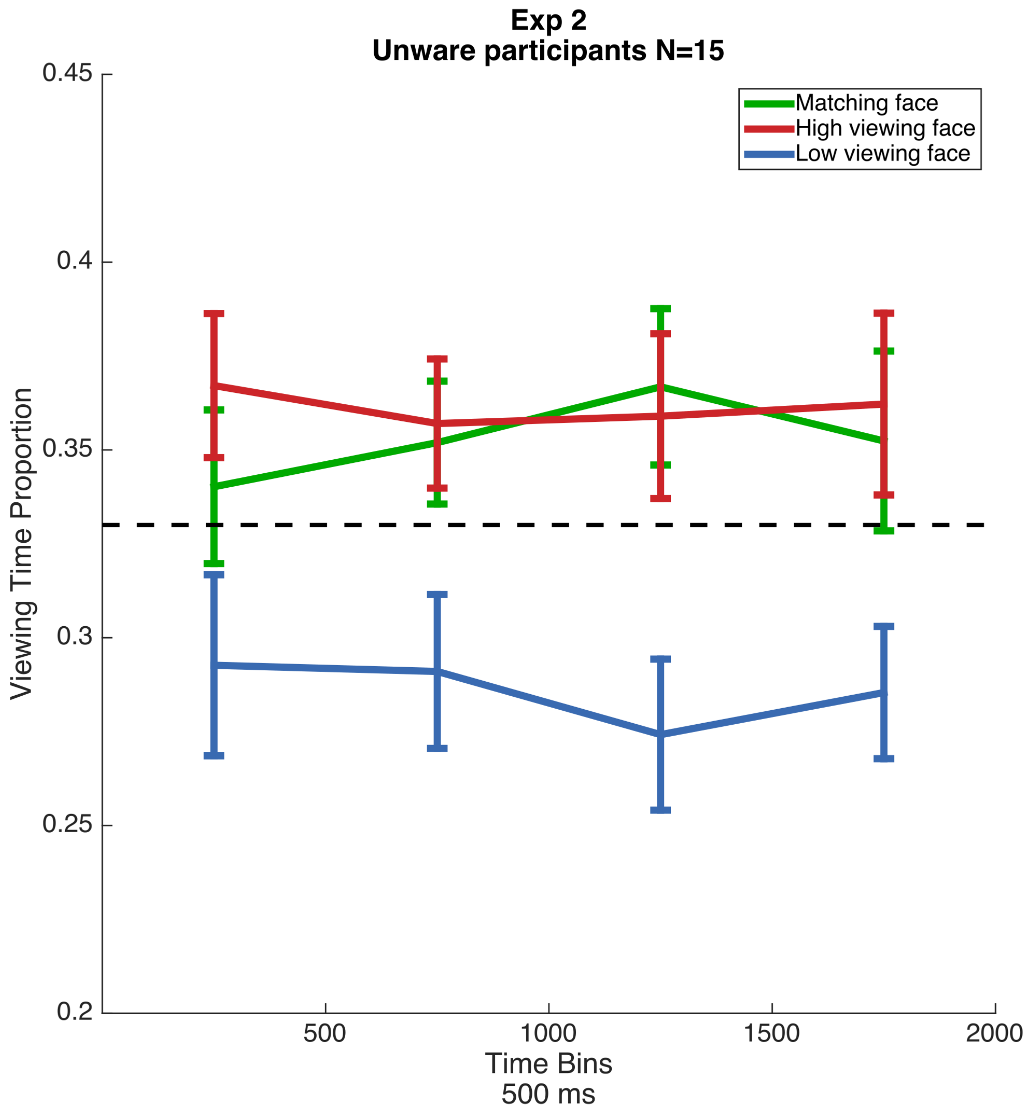

Figure 34. The unaware group's viewing time proportions of Matching (green), High viewing (red) and Low viewing (blue) faces in 500 ms time bins for the first two seconds of test trials. Error bars indicate SEM.

Eye movements indicate relational memory in unaware participants in the absence of the REME

The unaware group in our analysis showed no evidence of the REME. However, we could argue that this absence of the relational eye movement effect is only caused by the complete absence of any relational memory in this group. To check for any evidence that our unaware group shows signs of relational memory we had the following argument. A complete absence of relational memory would result in equal viewing time proportions directed to the three faces on the test displays. However, if participants have any relational memory, an effective strategy would be to allocate more encoding capacity to the non-matching faces compared to the matching face, because the non-matching faces were never seen together with the actual background scene. This relational-memory-based effective encoding would be manifested in a non-matching face preference throughout the test display presentation. To measure any potential difference between encoding capacity allocated to matching and non-matching faces we analysed the viewing time proportions for the three face-types throughout the whole presentation of the test displays. We wanted to find evidence that participants in the unaware group might prefer any or both of the non-matching faces in test trials compared to the matching face. This would suggest that they encoded the faces effectively by allocating more viewing time to the non-matching faces. To answer this question we run a 3 (face type) x 20 (time bins) repeated measures ANOVA on the unaware group. Our results showed a main effect of face type $F(2, 28) = 86.44$, $p < .001$, $\eta^2 = .86$, and a significant interaction $F(38, 532) = 2.83$, $p < .001$, $\eta^2 = .17$. Our corrected pairwise comparisons showed that participants preferred the High viewing non-matching face compared with the Matching face in the following time bins: Bin 15 $t(14)= -3.91$, $p =.002$, $d = -1.0$; Bin 16 $t(14)= -4.05$, $p = .001$, $d = -1.0$; Bin 18 16 $t(14)= -3.55$, $p = .002$, $d = -.92$, Bin 19 $t(14)= -3.56$, $p = .002$, $d = -.92$ (Figure 35). This showed that within several time bins during the second half of the test trials participants preferred the High viewing matching face compared to the Matching face, which can be regarded as evidence of effective coding based on relational memory.

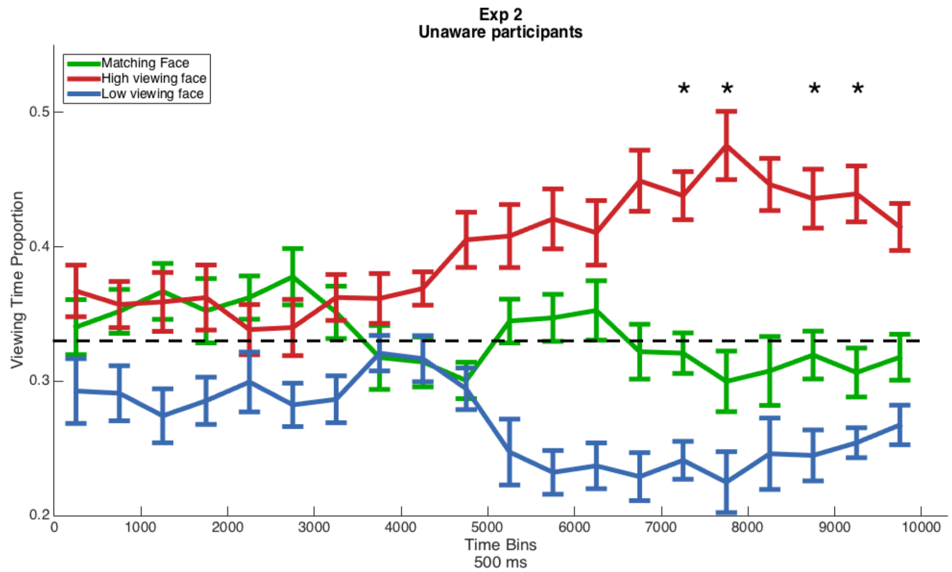

Figure 35. The unaware group's viewing time proportions of Matching (green), High viewing (red) and Low viewing (blue) faces in 500 ms time bins for the whole test trials. Error bars indicate SEM. Stars indicate significant differences between Matching and High viewing faces after correction for multiple comparisons.

These results were also confirmed by a one-way (face type) ANOVA for the overall viewing time proportions without separating viewing behaviour to several time bins. Our analysis showed a significant face type effect $F(2, 28) = 89.48$, $p < .001$, $\eta^2 = .87$, and pairwise comparisons indicated an overall preference of the High viewing non-matching face compared to both the Matching face $t(14)= 7.77$, $p < .001$, $d = 2.0$, and the Low viewing non-matching face $t(14)= 11.65$, $p < .001$, d = 3.0 (Figure 36).

Taken together we found that unaware participants demonstrated a preference of the High viewing non-matching face in their viewing behaviour in test displays, which indicates that they allocated more viewing time to the High viewing non-matching faces compared to the Matching face. We can regard this pattern of viewing behaviour as evidence of relational memory, which guides more viewing time to one of the non-matching faces for effective encoding.

This result is important, because it goes against our first prediction, which predicts that the REME has to appear, when there is evidence of relational memory retrieval. The unaware group showed evidence of relational memory retrieval (by preferring one of the non-matching faces), but this group did not demonstrate the REME. This pattern of viewing behaviour points to the direction that the emergence of the REME is not a necessary for relational retrieval.

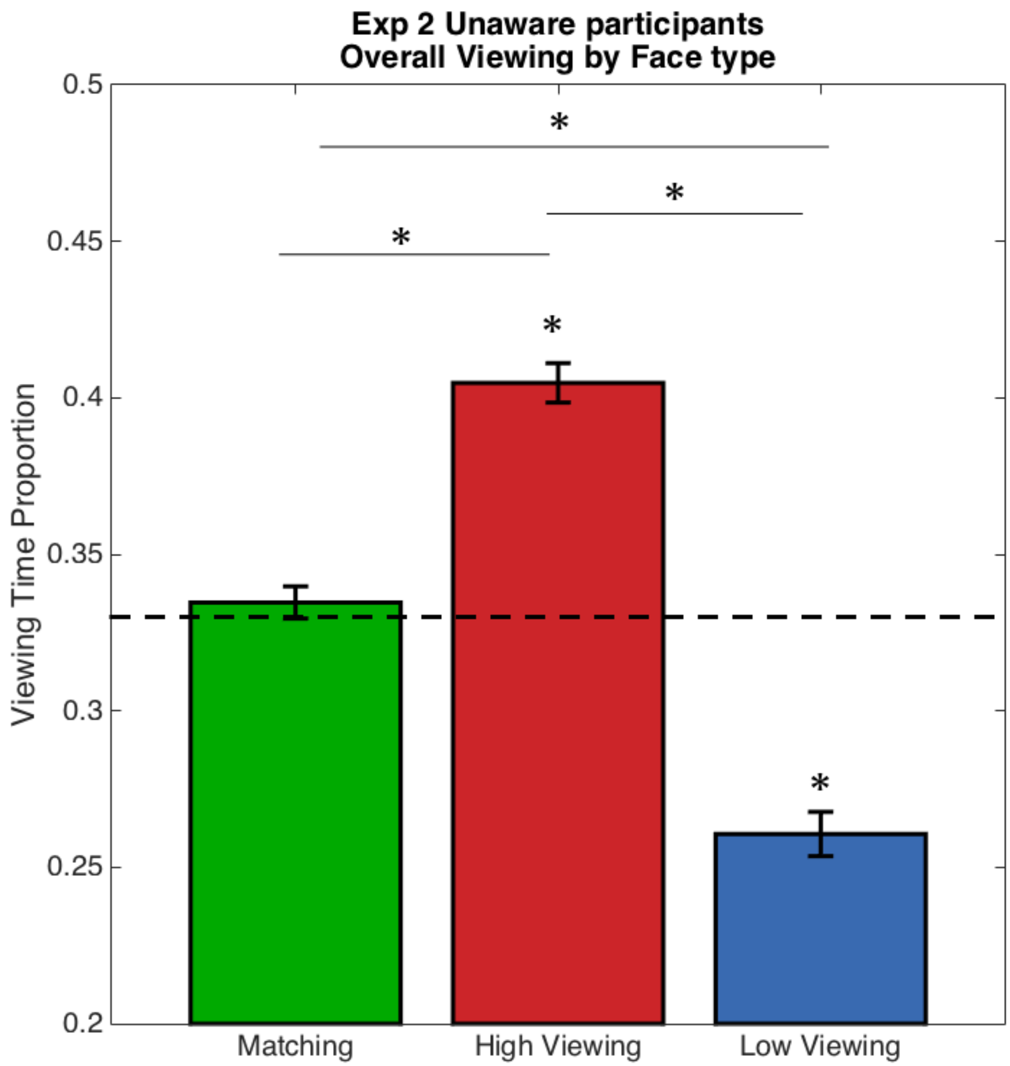

Figure 36. The unaware group's overall viewing time proportions for Matching (green), High viewing (red) and Low viewing (blue) faces. Error bars indicate SEM. Stars indicate significant differences.

Relational-memory-based encoding in aware participants

We also run both of the above mentioned analyses on the aware group and they also demonstarted the same evidence of relational-memory-based encoding in their viewing behaviour as the unaware group. We run a 3 (face type) x 20 (time bins) repeated measures ANOVA on the aware group to test for the same encoding behaviour as we found in the unaware group. Our results showed a main effect of face type $F(2, 34) = 79.74$, $p < .001$, $\eta^2 = .82$, and a significant interaction $F(38, 646) = 3.40$, $p < .001$, $\eta^2 = .17$. Our corrected pairwise comparisons showed that participants preferred the High viewing non-matching face compared with the Matching face in the following time bins: Bin 15 $t(14)= -3.79$, $p =.001$, $d = -.98$; Bin 17 $t(14)= -3.29$, $p = .004$, $d = -.85$, Bin 18 $t(14)= -3.50$, $p = .002$, $d = -.90$ (Figure 37). These results indicate the same High viewing non-matching face preference at the second half of the test display as the unaware group, which can be regarded as evidence of relational-memory-based encoding manifested in eye movement behaviour.

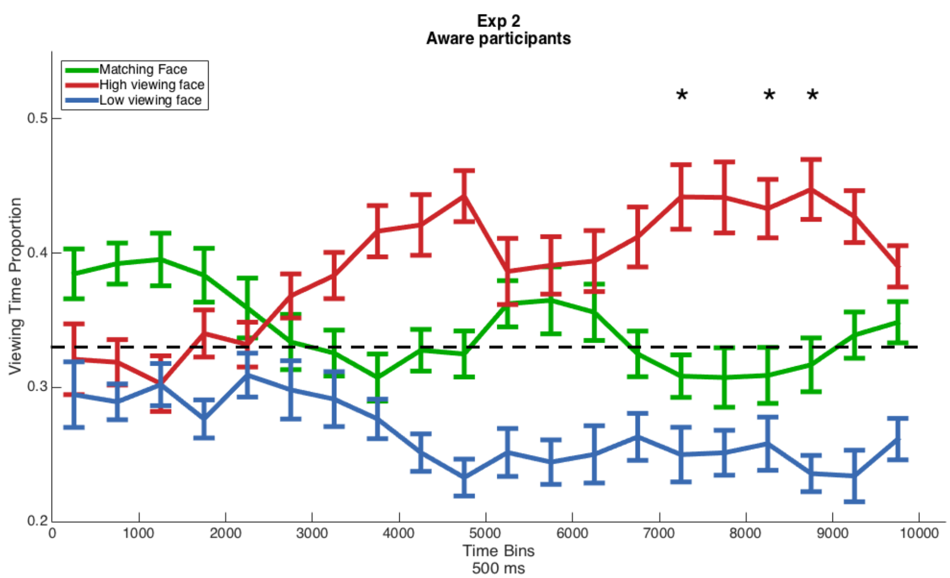

Figure 37. The aware group's viewing time proportions of Matching (green), High viewing (red) and Low viewing (blue) faces in 500 ms time bins for the whole test trials. Error bars indicate SEM. Stars indicate significant differences between Matching and High viewing faces after correction for multiple comparisons.

We also confirmed these results by a one-way (face type) ANOVA for the overall viewing time proportions collapsing eye movement data for the whole test display presentation length. Our analysis showed a significant face type effect $F(2, 34) = 87.75$, $p < .001$, $\eta^2 = .84$, and pairwise comparisons indicated an overall preference of the High viewing non-matching face compared to both the Matching face $t(17)= 4.63$, $p < .001$, $d = 1.1$, and the Low viewing face $t(17)= 23.73$, $p < .001$, $d = 5.6$ (Figure 38). Taken together, we could find the same encoding behaviour in both the unaware and aware group, which was characterised by a preference of one of the non-matching faces compared to the two other faces (most importantly it was greater then the matching face). This suggests that participants in both groups demonstrated a relational-memory-based encoding behaviour in their eye movements, which was evident in the preference of one of the non-matching faces.

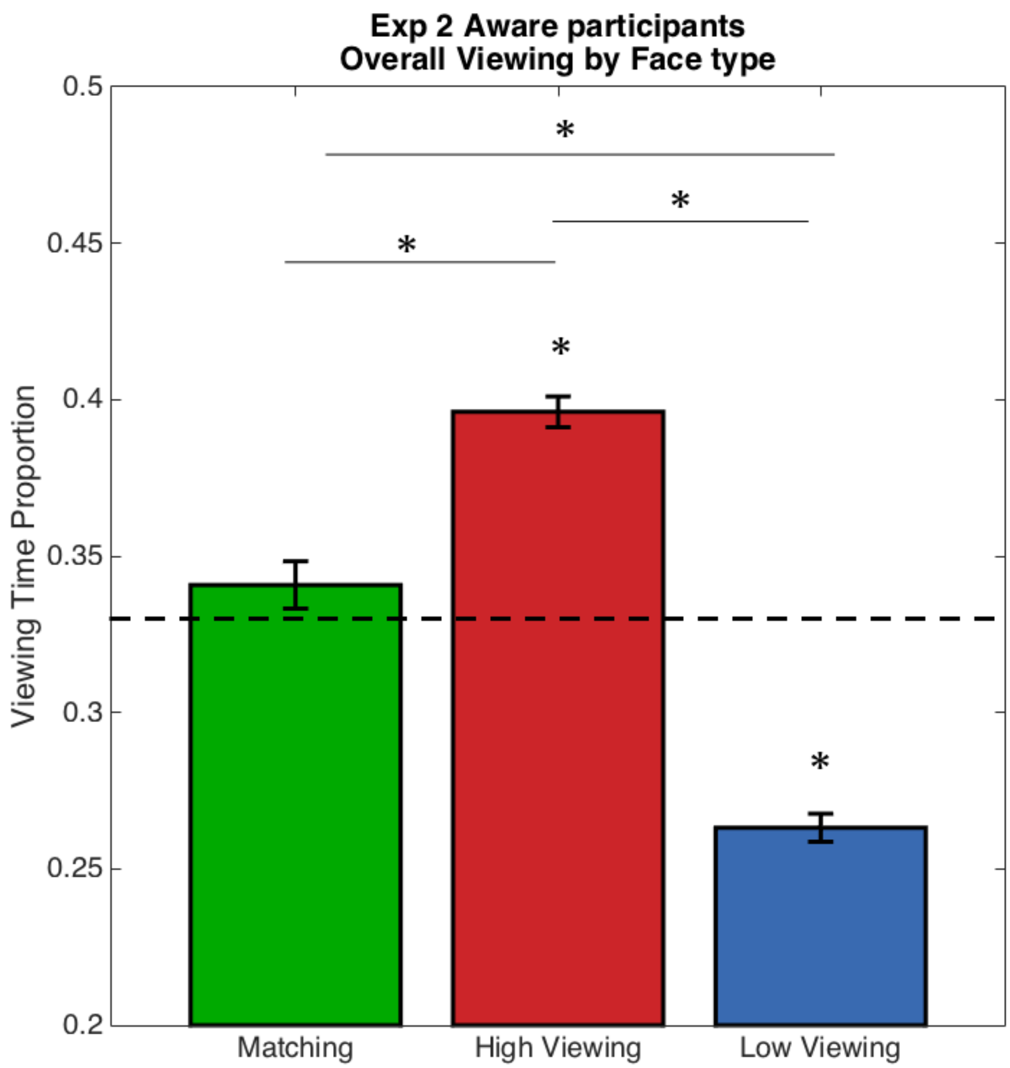

Figure 38. The aware group's overall viewing time proportions for Matching (green), High viewing (red) and Low viewing (blue) faces. Error bars indicate SEM. Stars indicate significant differences.

<u>Discussion</u>

Our assumption states the REME has to appear irrespective of task characteristics. We tested this prediction by changing the task to a no choice task, where participants had to learn the three faces together with the scene for a future recognition task (which was not administered). We checked for a matching face preference by comparing viewing time proportions of the matching faces to the non-matching faces. Our results showed that for all participants (N=38) the REME appeared in the no choice task in the first 2000 ms of test trials. The effect was evident between 500-1500 ms after stimulus onset, which was consistent with previous results suggesting that the eye movement effect could emerge in a no choice task (Hannula et al., 2007). The emergence of the REME was task independent, which supports our prediction that the REME is necessary for relational retrieval. However, when we separated aware and unaware participants, our unaware group demonstrated relational memory in looking behaviour, while they did not show the REME. This pattern of eye movements is against our prediction that the REME is necessary for relational memory retrieval.

Our second question was related to the link between the REME and conscious retrieval. Our assumption states that the REME accompanies relational retrieval unconsciously, which predicts a possible dissociation between the REME and conscious relational retrieval. We tested for this dissociation by dividing our participants to different awareness groups based on their verbal reports related to the appearance of a matching face on test displays. We hypothesized that those participants who are aware of the presence of matching faces on test display would show the REME, because they demonstrate evidence of relational retrieval in their verbal answers. Our results confirmed this prediction. Aware participants (N=18) showed evidence of the REME in the first 2000 ms of the test trial by preferring the matching face compared to the High viewing non-matching face. By analysing the unaware group (N=15), who were unable to report the presence of matching faces on the test displays, we wanted to test our prediction that the REME could dissociate from conscious retrieval. Our results showed the lack of the REME in this unaware group. We can conclude that we failed to find supporting evidence for our prediction that the REME is independent of conscious processes. We also tested whether the unaware group demonstrated any sign of relational memory in their eye movements. This was necessary, because we could argue that the lack of the REME in the unaware group is only a consequence of the complete lack of relational memory in this group. We argued that if participants have any relational memory, than an effective encoding strategy in the no choice task is to allocate more encoding capacity to the non-matching faces compared to the matching face, because the task is to learn the three faces together with the scene and the non-matching faces were never seen together with the actual background scene. In our analysis we found that the unaware group demonstrated a preference of the High viewing non-matching face compared to the matching face in the second half of the presentation of the test displays, and also for their overall viewing behaviour. We regard this difference as evidence that based on some relational memory these participants allocated more viewing time to encode one of the non-matching faces than the matching face. This encoding pattern was also present in the aware group.

To summarize our results, we found evidence that the REME is tightly associated with conscious retrieval. Aware participants demonstrated the REME, while unaware participants did not. Moreover, both participant groups showed evidence of relational memory, which affected their allocation of encoding capacity to one of the non-matching faces compared to the matching face. Interestingly, our results showed that in both groups participants favoured only one of the non-matching faces and applied more encoding resources to that face. They did not allocate equal viewing time to the two non-matching faces at the expense of the matching face. This pattern may be a consequence of the fixed presentation time (10 seconds), which might limit participants to favour only one additional non-matching face to be effectively encoded with the background scene.

Our results strongly suggest that the REME is not independent of conscious retrieval. In both Experiment 1 and Experiment 2 we did not find any evidence that the REME can dissociate from conscious retrieval of the matching faces. However, in both of these experiments we found a tight link between the emergence of the REME and conscious processes. In Experiment 1 the magnitude of the REME increased with increasing levels of subjective confidence, while in Experiment 2 the REME only appeared in the aware participant group. These results suggest that the REME signals the conscious experience of relational retrieval. This goes against our assumption that the eye movement effect is an index of a necessary and unconscious relational memory process.

Additionally, the looking behaviour of our unaware group, showing some evidence of relational memory retrieval without the REME, goes against our prediction that the appearance of the REME is needed for relational memory. This result suggests, that in some cases the REME is not needed to demonstrate relational memory in the task. The REME might signal a process, which is not essential for relational retrieval. We further explored this possibility in Experiment 3 and Experiment 4.

Experiment 2.2: Baseline looking behaviour for Experiment 2

Aim of the experiment

In our analysis of Experiment 2 we regarded the significantly higher viewing time proportion of the High viewing non-matching face compared to the matching face in both of the aware and unaware group as a sign of relational memory retrieval. Our argument was that if participants would possess no relational memory in the task, than they should allocate equal amount of viewing time between the three faces, because their task is to encode the three faces together with the scene background. This would result in viewing time proportions not significantly different from the 0.33 chance level for the three faces. Previous studies (Hannula et al., 2007) also used this logic to infer relational memory when the matching face preference exceeded the 0.33 chance level.

However, in our opinion the pattern of viewing time proportions in Experiment 2 could be explained without assuming any memory retrieval effect. According to our results of Experiment 2, the matching face viewing time proportions for the overall trials was not significantly different from the 0.33 chance level. Meanwhile, the High viewing face proportions were significantly higher and the Low viewing face proportions were significantly lower than the 0.33 chance level. This pattern of viewing time proportion could be explained by the spontaneous looking behaviour of participants during the presentation of the three-face test displays without assuming any memory effect. It is possible that participants spontaneously view the test displays when their task is to encode the three faces with the scene background in a way that their looking behaviour will result in significant differences among faces regarding viewing time proportions. These could manifest in a pattern that one face is preferred above the 0.33 chance level, while another face is relatively neglected (below the 0.33 chance level) and the third face's proportion is in-between the other two face (in Experiment 2 not different from chance level). This suggested looking pattern without the assumption of any memory effect could be behind our overall proportion results in Experiment 2 that we used to infer relational memory retrieval. These concerns were behind our motivation to conduct

Experiment 2.2 that measured the spontaneous looking behaviour of participants in a no choice task, when participants were only presented with the three-face displays without a preceding learning phase of face-scene pairs. We argued that a baseline looking behaviour has to be used if one wants to measure the influence of the scene-pair learning phase on subsequent looking behaviour in the three-face test displays. Our objective was to test the baseline looking behaviour of our participants during the three-face test displays, and to use this measure as a reference to be able to infer relational memory retrieval processes that potentially affected looking behaviour in Experiment 2. We predicted that when participants are asked to learn the three-faces together with the background scene for a future recognition test, they would allocate equal viewing time between the three faces. This equal allocation of encoding resources would result in viewing time proportions not significantly different from the 0.33 chance level for of the three faces.

Methods

Participants

There were 15 undergraduates (ages 18-27; 9F/6M) taking part in the experiment from Eötvös Loránd University, Budapest. All of them received course credit for participation and gave informed consent for the study approved by the Research Committee of Eötvös Loránd University, Budapest (Ethical No.: 2012/6).

Behavioural paradigm

We excluded the learning phase and only used a test phase in this task. Background scenes and face stimuli were the same as in Experiment 1 and 2. The task consisted of a single test phase with 33 randomly created three-face test displays with background scenes. Both the faces and the background scene were new to the participants. The test trial design and the presentation times of the stimuli were the same as in Experiment 1 (see Figure 22, page 86). Based on the instructions by Hannula et al. (2007) participants were told that their task would be to learn the three faces along with the scenes for a forthcoming recognition test (which was not administered). They did not have to use any responses during or after test displays.

All they were instructed to do is to concentrate and try to memorize the scene + three face displays.

Eye tracking data acquisition and analysis

The same Tobii T60XL eye tracker at 60 Hz was used to record eye positions as in Experiment 1. Before the test phase participants completed a five-point standard calibration procedure. During the test phase we used the same gaze contingent stimulus presentation method as in Experiment 1, which required the average eye positions to be measured at the fixation point for 500 ms to proceed to the next trial. With this method we ensured that participants started viewing the three face test displays in the middle of the screen equal distance from the three faces at the start of every trial.

We used a raw-data based within-display calculation to compare the viewing time proportions between faces (for fixation-based results see the appendix). For every trial we identified a High, Medium and Low viewing face based on overall looking behaviour during each trial. The High viewing face was the face with the highest overall viewing time proportion among the three faces. Low viewing faces had the lowest viewing time proportions among the three faces, while Medium viewing faces had proportion values in-between High and Low faces.

Statistical analyses

We used one-way ANOVA to analyse eye movement results between the three face categories (High, Medium, Low viewing). For the ANOVA Mauchly's test of sphericity was used to ensure that the assumption of sphericity was met. We used Greenhouse-Geisser corrections if the assumption of sphericity was violated. For comparing the effects we used post-hoc comparisons with Bonferroni corrected paired *t*-tests. Additionally, we used one-sample t-tests to test the viewing time proportion measures of the three face categories against the 0.33 chance level.

Results

Eye movement results for all participants

Our one-way ANOVA showed a main effect of face type (High, Medium, Low) on the viewing time proportions $F(2, 28) = 30.90$, $p < .001$, $\eta^2 = .69$. Post-hoc paired t-tests indicated that High viewing faces had significantly higher viewing time proportions than Medium, $t(14)= 5.74$, $p < .001$, $d = 1.52$, and Low viewing faces, $t(14)= 5.17$, $p < .001$, $d = 1.3$, while Medium faces had higher proportions than Low faces $t(14)= 5.38$, $p < .001$, $d = 1.43$ (Figure 39). Moreover, one sample t-tests showed that viewing proportion of the High viewing faces was significantly above the 0.33 chance level $t(14)= 6.35$, $p < .001$, $d = 1.64$, while the proportion of Low viewing faces were significantly below chance level $t(14)= -4.94$, $p < .001$, $d = -1.28$. The viewing proportion of Medium faces did not differ significantly from chance level $t(14)= 1.05$, $p = .31$, $d = .27$.

These results showed that there was a significant spontaneous viewing behaviour difference among the three faces. Participants on average look disproportionately to faces: there is a preferred face (High viewing face) and a relatively neglected face (Low viewing face), while the third face is in-between the two other faces (Medium viewing face). This significantly different looking behaviour is not predicted if the three-faces would receive equal encoding capacity. Additionally, we demonstrated that the looking behaviour of the participants results in a specific pattern, where the High viewing face is significantly above the 0.33 chance level in the expense of one face (Low viewing face), which is significantly below the 0.33 level. While the third face's viewing proportion (Medium viewing face) is not significantly different from the 0.33 level.

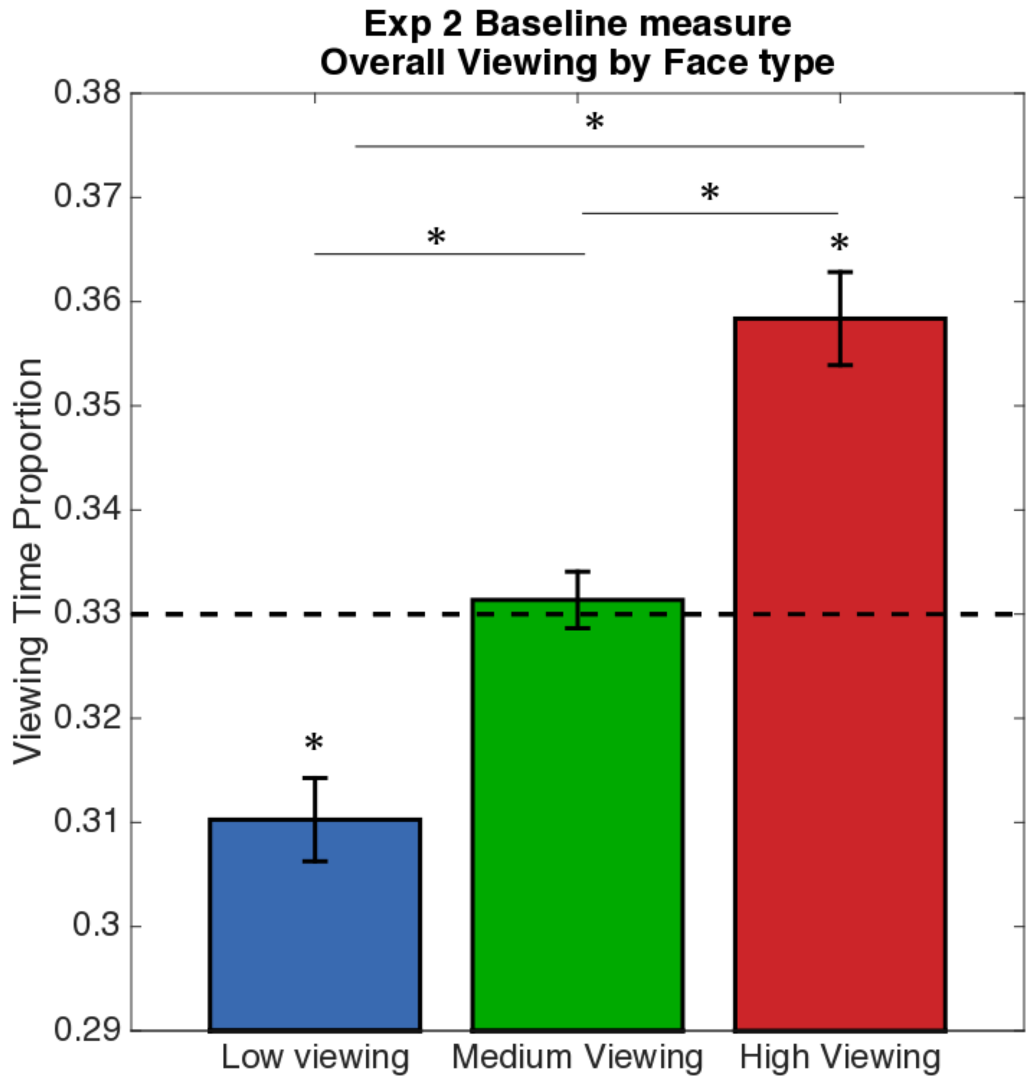

Figure 39. Overall viewing time proportions for Low viewing (blue), Medium viewing (green) and High viewing (red) faces. Error bars indicate SEM. Stars represent significant differences.

Comparing the results of Experiment 2 and the baseline measure of Experiment 2.2:

We were specifically interested whether there are any differences regarding the average viewing time proportions of faces between Experiment 2 and the baseline measures of Experiment 2.2. Our rationale was that any difference of viewing time proportions between the two tasks would be a consequence of the face-scene learning phase in Experiment 2, which would strongly suggest a relational memory effect on eye movements. We argued that the only relative difference between the three faces in the two tasks is that in Experiment 2 one of the faces is the matching face of the background scene, while in Experiment 2.2 there is no matching face present among the three faces. It is important to point out that the *absolute* familiarity level of the faces were different in the two experiments: in Experiment 2 all three faces were presented once in the learning phase, while in Experiment 2.2 all three faces were new. Most importantly, however, the *relative* familiarity level of the three faces were equal in both experiments. In Experiment 2 all three faces were old faces, that were presented for 3 s during the learning phase, while in Experiment 2.2 all three faces were new, never seen by the participants. This relative familiarity equivalence of the three faces assured us that the only difference that could result in viewing time proportion assymetry between the two experimental conditions was the existence (Experiment 2) or absence (Experiemnt 2.2) of a matching face among the three faces in test displays.

In our analyses we compared the overall viewing time proportions for specific pairs of face-types. In both the aware and unaware group of Experiment 2 we used the previously reported overall viewing time proportion results of the Matching face, High viewing face and Low viewing face. Because in both groups the Matching face viewing time proportion values were in-between the High viewing face and Low viewing face values, we choose to compare them with the Medium viewing faces of Experiment 2.2. Accordingly, we compared the viewing time proportions of the following pairs: (1) Matching faces (Exp. 2) vs. Medium viewing face (Exp. 2.2), (2) High viewing faces (Exp. 2) vs. High viewing faces (Exp. 2.2) and (3) Low viewing faces (Exp. 2) vs. Low viewing faces (Exp. 2.2). Our results are depicted in Figure 40 separately for the aware and unaware group.

1. Matching faces (Exp. 2) vs. Medium viewing face (Exp. 2.2):

In the aware group, there was no significant difference between Matching faces and Medium viewing faces of Experiment 2.2, $t(31)= 1.22$, $p = .237$, $d = .22$. Similarly, the unaware group's Matching faces showed no significant difference compared to Medium viewing faces of Experiment 2.2, $t(28)= .614$, $p = .55$, $d = .11$.

2. High viewing faces (Exp. 2) vs. High viewing faces (Exp. 2.2):

In the aware group, the High viewing faces had significantly higher viewing time proportions than the High viewing faces of Experiment 2.2, $t(31)= 5.57$, $p < .001$, $d = 0.98$. Similarly, in the unaware group the High viewing faces also showed significantly higher proportions compared to the High viewing faces of Experiment 2.2, $t(28)= 6.01$, $p < .001$, $d = 1.14$.

3. Low viewing faces (Exp. 2) vs. Low viewing faces (Exp. 2.2):

In the aware group, the Low viewing faces had significantly lower viewing time proportions than the Low viewing faces of Experiment 2.2, $t(31)= -7.70$, $p < .001$, $d = -12.8$. Similarly, in the unaware group the Low viewing faces also showed significantly lower proportions compared to the Low viewing faces of Experiment 2.2, $t(28)= -6.10$, $p < .001$, $d = -1.13$.

These results demonstrated that there was a significant difference of looking behaviour patterns between Experiment 2 and Experiment 2.2. Our baseline measure of Experiment 2.2 suggested that participants spontaneously show significant viewing time proportion differences between the faces, when looking at test displays. However, this baseline looking behaviour was altered in consequence of the presentation of the face-scene pairs during the learning phase. The completion of the learning phase resulted in a significant difference in the High viewing and Low viewing faces between the experiments. In Experiment 2 the High viewing non-matching faces were preferred more, while the Low viewing faces were preferred less than expected by the spontaneous looking behaviour. We could regard this pattern difference of the High viewing and Low viewing faces as evidence of relational memory retrieval in both aware and unaware participants in Experiment 2.

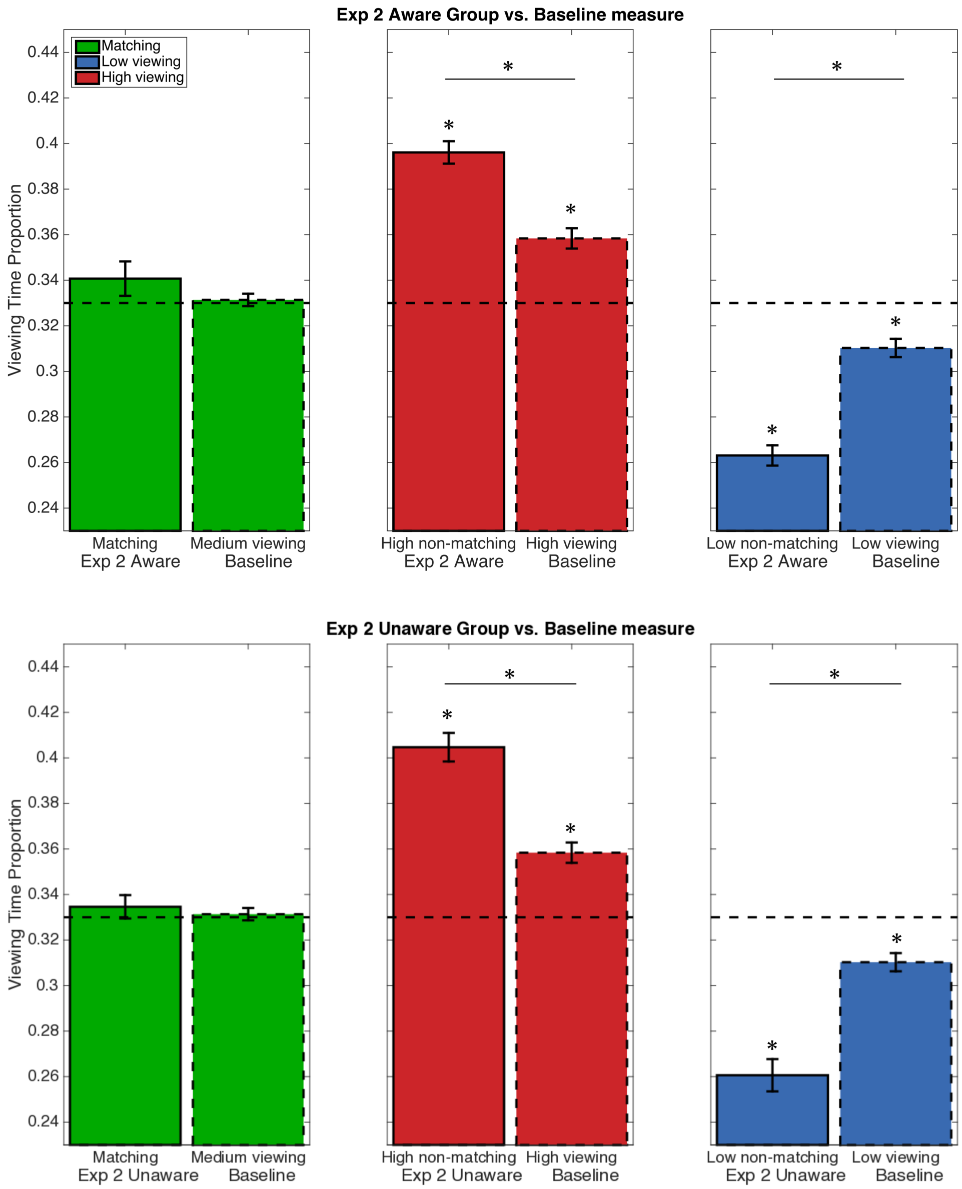

Figure 40. The aware and unaware groups' viewing time proportion results from Experiment 2 (solid lines) compared to the baseline measures of Experiment 2.2 (dashed lines). Green indicates Matching face (Exp 2) and Medium viewing face (Baseline); red indicates High viewing face; blue indicates Low viewing face. Error bars indicate SEM. Stars represent significant differences.

Discussion

The difference in the viewing time proportion pattern between Experiment 2 and the baseline experiment demonstrated a relational memory effect in both the aware and the unaware group. In these groups the High viewing non-matching faces were preferred more, while the Low viewing non-matching faces were preferred less, compared to the baseline level measured in Experiment 2.2. Based on the equivalence of the familiarity level of the three faces in the two experiments, the looking behaviour difference could only be attributed to the presentation of the scene-face pairs during the learning phase in Experiment 2. The pre-exposure to the face-scene pairs in Experiment 2 resulted in the presence of a matching face within the three face test displays. In contrast, there was no matching face in the baseline measure of Experiment 2.2. Consequently, we could conclude that the viewing time proportion difference between the two task conditions was driven by relational memory retrieval. Interestingly, this retrieval resulted in more looking time allocated to the High viewing non-matching face compared to the baseline level of the High viewing face in the expense of the Low viewing face, which received less looking time than the baseline level measure. This could be interpreted, that relational memory retrieval causes the distribution of encoding capacity in a way that one of the non-matching faces receives more encoding capacity than expected, in the expense of the other non-matching face. This pattern might be related to specific task characteristics, such as the presentation time of the test displays. It could be an effective strategy to try to allocate more encoding capacity to both of the non-matching faces in the expense of the matching face, but the time constraint might limit looking behaviour to favour only one non-matching face (High viewing face) and result in a relatively neglected other non-matching face (Low viewing face). In Figure 41 we summarized our results of Experiment 2 and the baseline measure.

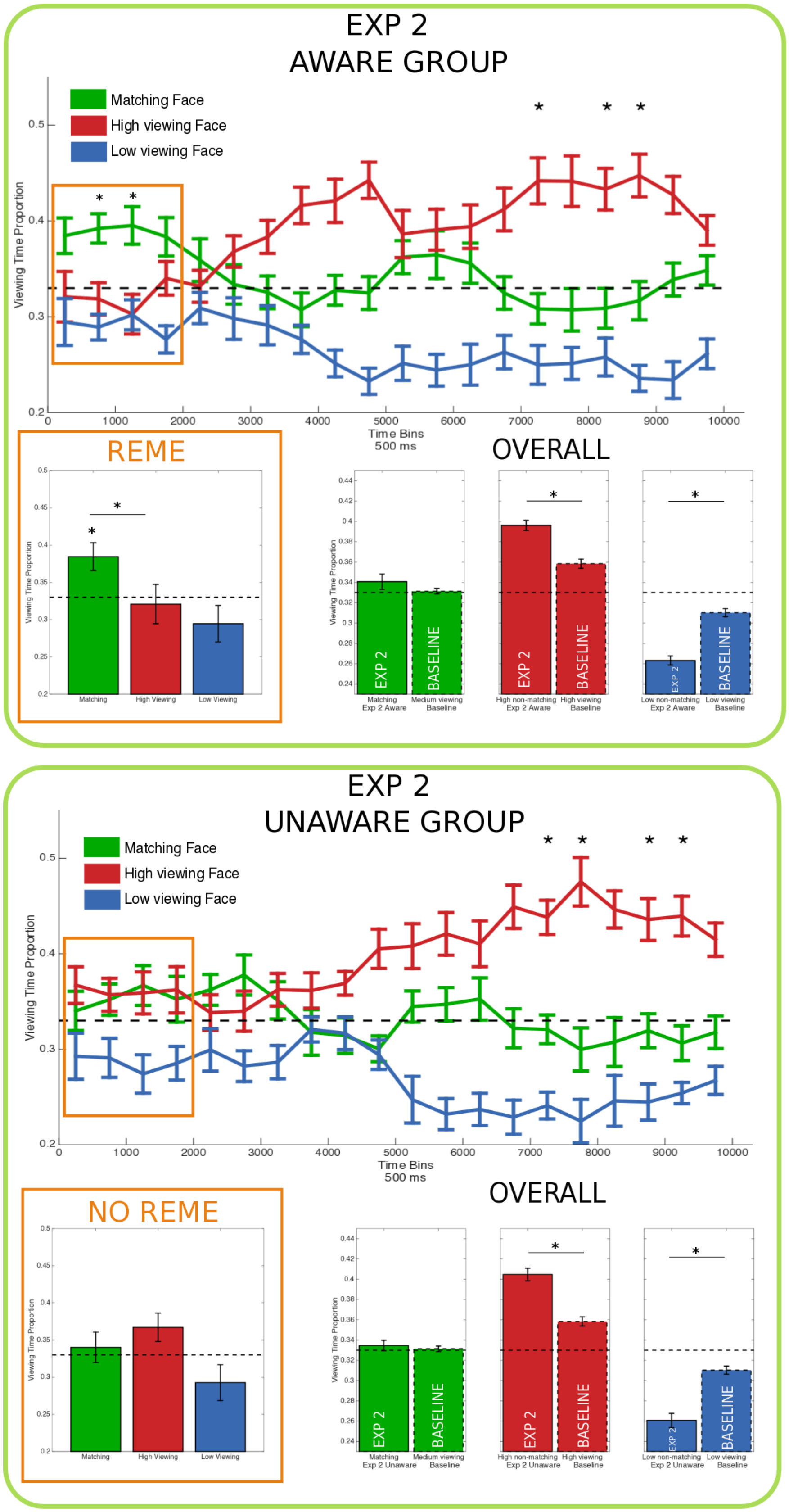

Figure 41. Summary of our results from Experiment 2 compared to the baseline of Experiment 2.2.

Experiment 3.

Based on our previous experiments we could conclude that the REME is tightly linked to conscious retrieval. If the REME signals the conscious retrieval, which is not necessary for relational retrieval than we should be able to show the lack of REME in the task associated with reliable relational memory performance. We tried to elicit this pattern of results by making the identification of the target stimuli harder in the task and see whether this manipulation can diminish the REME by potentially decreasing the level of conscious retrieval of the matching face.

<u>Aim of the experiment</u>

In this experiment we wanted to test whether the REME is necessary to result in relational memory retrieval measured by the above chance on the face-scene task. Our previous results in Experiment 2 suggested that the REME might not signal a necessary process for relational retrieval. In that experiment unaware participants lacked matching face preference but they demonstrated a relational-memory-based encoding effect by directing more encoding capacity to one of the non-matching faces compared to the baseline level. This suggests that there are certain conditions when relational memory performance can be measured in the absence of the relational eye movement effect.

In Experiment 3 we intended to make the choice task more difficult for participants by presenting blurred versions of the faces during the test phase. We hypothesized that, by making the target stimuli harder to identify, task performance would be reduced. Our main question was whether the potential decrease in task performance would obliterate the REME effect. A reliable task performance and the absence of the REME could be regarded as evidence that the REME does not signal a necessary process for relational retrieval. In Experiment 1 we could demonstrate that the REME diminished in Medium confidence level (compared to High confidence level), when task performance was moderate (approx. 50 %). This result also raised the question whether it is possible to reduce task performance to a level, where it still would indicate relational retrieval accompanied by a lack of the REME.

Methods

Participants

37 undergraduate students (ages 18-27; 21F/16M) took part in the experiment from Eötvös Loránd University, Budapest. All of them received course credit for participation and gave informed consent for the study approved by the Research Committee of Eötvös Loránd University, Budapest (Ethical No.: 2012/6).

Behavioural paradigm

The behavioural part of the experiment was exactly the same as in Experiment 1. Participants completed 99 face-scene learning trials, which was followed by a choice task with 33 test trials of three-face displays. The only difference was that we manipulated the perceptual characteristics of the faces between learning and test. At the learning phase faces were not modified. However, during the test phase all the faces appearing on the three-face test display were blurred. The blur was achieved by shrinking the faces to one-tenth of their original size (x 0.1) and in a second step, these modified faces were resized to their original (190 X 254 pixel) resolution. We applied this amount of blur to the faces because in our pilot study 10 participants showed an average of 45 % task accuracy, while eye movement data did not show any tendency of the REME. This pattern of the preliminary results suggested that the amount of the perceptual manipulation could be enough to reduce task performance and show a lack of REME.

Eye movement acquisition and analysis

Both eye movement acquisition and analysis were the same as in Experiment 1. (page 89). Again, we will report in the main text our raw-data-based results but we will present the fixation-based results in the appendix.

Statistical analyses

Similarly to Experiment 1., we used repeated measures ANOVAs to analyse eye movements in different time bins to check for either the time-course or response-locked REME. The main factors were response accuracy (correct vs. incorrect) and time bins. For all ANOVAs Mauchly's test of sphericity was used to ensure that the assumption of sphericity was met. We used Greenhouse-Geisser corrections if the assumption of sphericity was violated. For comparing the effects on specific time bins we used post-hoc comparisons with Bonferroni corrected paired *t*-tests.

Results

Task performance

In contrast to our first 10 participants, average task accuracy after 37 participants was 57.3 (SD = .12), which was significantly above chance level (33%), $t(36) = 12.62$, $p < .001$, $d = 2.1$ (Figure 42. All participants had above chance level performance.

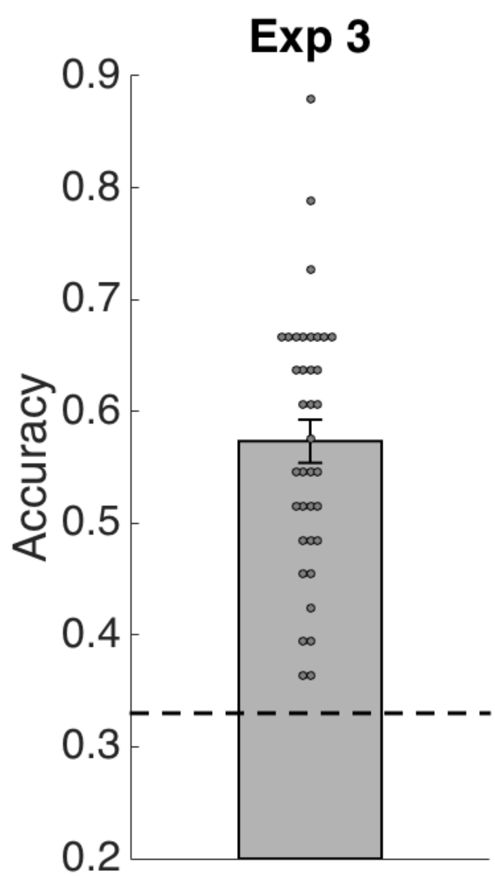

Figure 42. Average accuracy of Experiment 3, points represent individuals. Error bar indicates SEM.

We compared the accuracy in Experiment 3 and Experiment 1 and found no significant difference between performance levels $t(81) = -.28$, $p = .78$, $d = .03$. We also compared the average reaction times and the average reported confidence level between Experiment 3 and Experiment 1, which showed no significant differences (Table 10).

Measure	**Experiment 1**	**Experiment 3**	**Independent samples tests**
Overall RT Mean	4815 ms (SD= 861)	4710 (SD=748)	$t(81)=-5.83$ $p = .56$, $d = -.64$
Correct RT Mean	4444 ms (SD=931)	4409 ms (SD=696)	$t(81)=-1.91$ $p = .85$, $d = -.21$
Incorrect RT Mean	5412 ms (SD=917)	5198 ms (SD=949)	$t(81)=-1.04$ $p = .30$, $d = -.11$
Confidence Mean	3,73	3,86	$t(81)= .83$ $p = .41$, $d = .09$

Table 10. Comparisons of Experiment 1 and Experiment 3.

In Experiment 3 average reaction time for correct answers was 4409 ms ($SD = 696$) and for incorrect answers, 5198 ms ($SD = 949$). Correct answers were faster than incorrect ones, $t(36) = 5.86$, $p < .001$, $d = .96$. The same pattern was observed in Experiment 1.

Eye movement results

Onset-locked time-course analysis:

As in Experiment 1 our first step was to find evidence that our participants demonstrated a rapid REME within the first two seconds of the test trials. We were specifically interested in the first 2000 ms of the trials, because this is the time interval where previous results reported a rapid REME (Hannula et al., 2007; Hannula and Ranganath, 2009). We conducted a separate 2 (answers: correct, incorrect) X 4 (time bins) repeated measures ANOVA for the first 4 time bins. There was a main effect of answers, $F(1, 36) = 11.19$, p = .002, $\eta^2 = .24$, and time bins, $F(3, 108) = 7.09$, $p = .001$, $\eta^2 = .17$, but no interaction. Bonferroni corrected post-hoc paired t-tests showed a significantly greater proportion for correct than incorrect answers 1000-1500 ms after trial onset $t(36) = 2.81$, $p = .008$, $d = .47$. The time-course analysis results are shown in Figure 43. These results confirmed the emergence of a rapid REME in our experiment, similarly to Experiment 1. Interestingly, evidence for the REME appeared in a slightly later time bin (1000-1500 ms) compared to Experiment 1 (500-1000), which result suggests that our perceptual manipulation of the faces delayed the emergence of the effect by 500 ms.

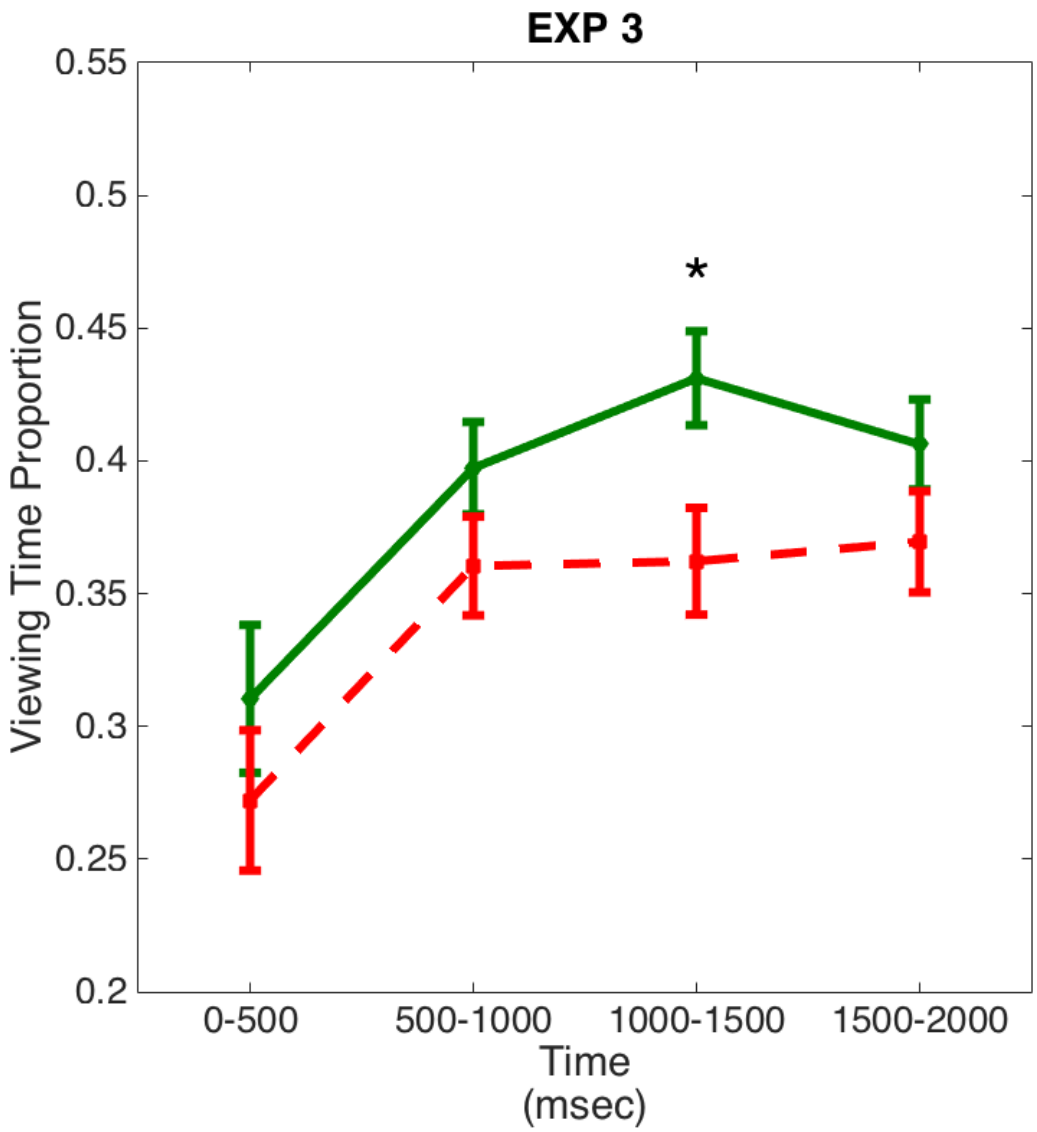

Figure 43. Onset-locked time-course analysis results in 500 ms time bins for the first two seconds of test trials. Green line represents correct responses; red dashed line represents incorrect responses. Error bars indicate SEM. Star indicates significant difference.

Response-locked analysis

Overall response-locked eye movement results

We also analysed our eye movement data before overt responses to find evidence that the REME appears before responses are made. First, we analysed our overall eye movement data without segmenting the whole trials to several time bins. We compared the viewing time proportions before every response in the trials between correct and incorrect answers. For every trial we identified the time point of the participant's response and we only calculated the viewing proportions between the onset of the three-face test trial and the response. Using this technique we could control that we measure pre-response eye movements and there will be no contamination of post-response processes in our results, which might have an effect on eye movements. The results demonstrated the REME before responses, which was apparent in greater pre-response viewing time proportions for correct, than incorrect responses $t(36)=5.12$ $p < .001$, $d = .84$ (Figure 44). We also compared the magnitudes of the REME measured by the average correct vs. incorrect response difference between Experiment 1 and Experiment 2. Statistical comparison showed no difference between them (Experiment 1 M = .053, SD = .068; Experiment 3 M = .047, SD = .056; t(81)= .45, p = .65, d = .05.

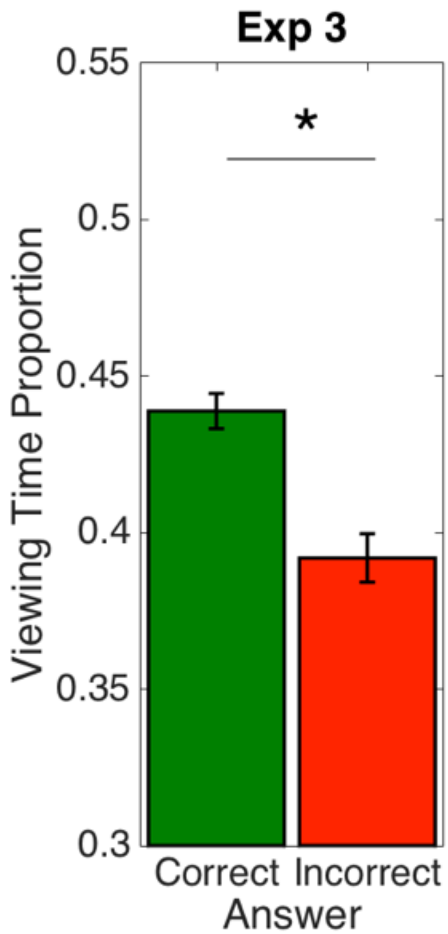

Figure 44. Overall response-locked viewing time proportions for correct (green) and incorrect (red) responses. Error bars indicate SEM. Star indicates significant difference.

Response-locked analysis for several time bins:

We were also interested whether we can show the relational effect in the eye movements in specific time bins before the responses. We conducted a 2 (correct/incorrect answer) X 4 (time bins) repeated measures ANOVA and found a main effect of answer, $F(1, 36) = 9.52$, $p = .004$, $\eta^2 = .21$, and time bins, $F(3, 108) = 21.57$, $p < .001$, $\eta^2 = .55$, but no interaction. Bonferroni corrected post-hoc paired *t*-tests indicated significantly higher proportions for correct responses in the 500-1000 ms time bin before responses $t(36) = 2.63$, $p = 0.012$, $d = .44$, while the 1000-1500 ms time bin before responses just fell out of our corrected acceptance level,

$t(36) = 2.47$, $p = 0.018$, $d = .40$. The results of our response-locked analysis are presented in Figure 45. These were consistent with our previous results of Experiment 1, which demonstrated the eye movement effect within the 500-1500 ms time window before overt responses (see Figure 28, page 98).

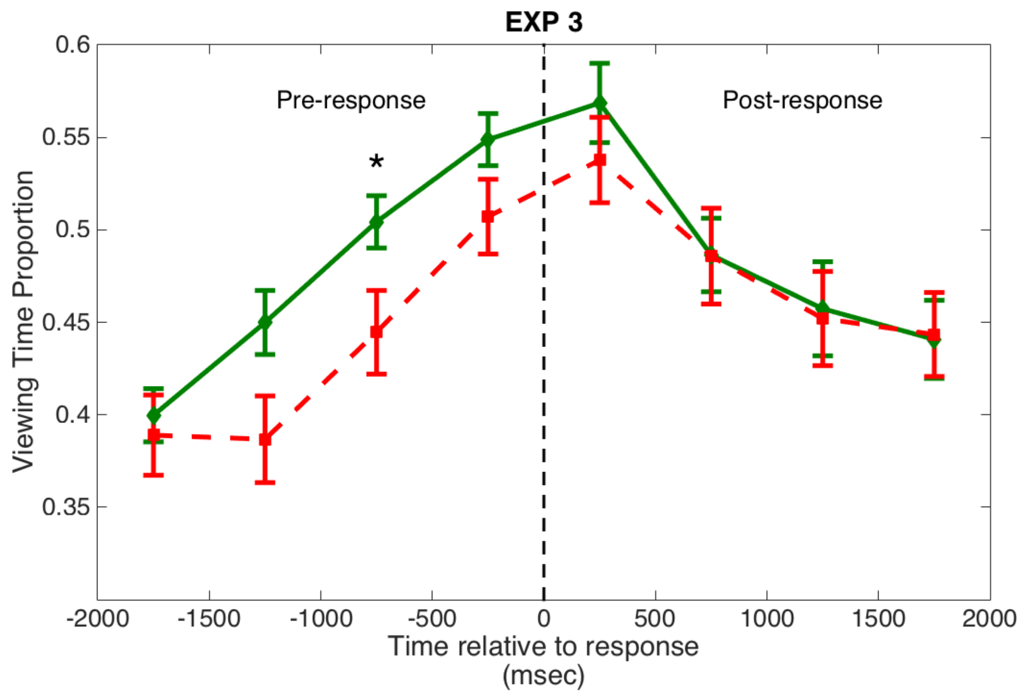

Figure 45. Response-locked results for 500 ms time bins before and after responses. Green line represents correct responses; red dashed line represents incorrect responses. Error bars indicate SEM. Star indicates significant difference.

Discussion

In Experiment 3 we wanted to test whether the REME is necessary to result in relational memory retrieval measured by the above chance on the face-scene task. Our previous results suggested that there are certain conditions when relational memory performance can be measured in the absence of the relational eye movement effect. We introduced a perceptual mismatch of target faces between learning and test to reduce task performance. Our assumption that the REME signals a necessary process for relational retrieval predicts that reliable above chance task performance is coupled with the obligatory emergence of the eye movement effect. However, our question was whether we could find a reduced, but still reliable task performance coupled with a lack of the REME, which would be inconsistent with our assumption, that the REME is necessary index of relational retrieval.

In contrast to our pilot results, after testing 37 participants with the modified task, we could not reduce task performance compared to Experiment 1. Our participants in Experiment 3 had undistinguishable task accuracy compared to Experiment 1 (Experiment 3 average accuracy: M = 57.3, SD = .117 and Experiment 1 average accuracy: M = 56.65, SD = 0.16). To parallel these accuracy results, our participants demonstrated the relational eye movement effect both in the first 2000 ms of the test trials and also before their overt memory responses. The timing of these effects was consistent with our previous results in Experiment 1. The only difference between the eye movement results were that in Experiment 3 the emergence of the rapid eye movement effect appeared to be slightly protracted by 500 ms and the response-locked effect was only significant in one time bin, 500-1000 ms before responses. These marginal differences suggest that the perceptual mismatch could diminish the REME even when it does not alter average accuracy.

In sum, we failed to reduce task performance in Experiment 3 by manipulating the faces between learning and test and our results still demonstrated the REME. These could not inform us about the potential dissociation between the REME and relational retrieval. However, our results showed the robustness of the REME effect using the face-scene paradigm. The eye movement effect was still present after the perceptual distortion of the target faces. This suggests that the relational retrieval process behind the REME is able to guide eye movements towards the matching faces even when these target faces are perceptually less detectable. This robustness of the eye movement effect might be linked to the specific nature of face recognition, which is supported by holistic processing (for a review see Tanaka & Simonyi, 2015). This holistic face processing might help to identify retrieved faces from memory representations even when individual elements (eyes, nose, mouth, etc.) are less detectable. In Experiment 3 we could replicate our previous results, which showed that the REME is evident in the first 2000 ms of test trials and it can be measured before overt responses are made. Experiment 3 also demonstrated that the REME is robust and it can resist a significant level of perceptual mismatch between learning and test. Unfortunately, we could not reduce task performance in Experiment 3, thus we failed to test whether the REME is necessary for above chance relational retrieval in the face-scene task. However, we could answer the former question in our next experiment where we manipulated another stimulus characteristic in the task, namely the target stimulus category.

Experiment 4

Aim of the experiment

Previous studies demonstrated the REME using only a specific subset of possible target stimuli in the task. These experiments either used human faces (Hannula et al., 2007; Hannula & Ranganath, 2009) or pictures of toys with faces (Chong & Richmond, 2015). In our opinion it is essential to test the universality of the REME in different stimulus categories, because memory research regards the REME results as a general measure of the retrieval process, which contributes to our understanding of relational memory. Likewise, our assumption based on previous results states that the REME is a universal and necessary indicator of relational memory retrieval. Based on this assumption we predicted that the REME is stimulus-type independent (fourth prediction), thus the REME has to appear when relational retrieval is present in the task, irrespective of the category of the target stimuli. In Experiment 4 we tested this prediction by changing the target stimulus category from faces to objects. We wanted to find evidence that the REME is a universal indicator of relational memory, which will be present in an experiment using objects as target stimuli.

We were specifically interested in whether we can replicate our findings in Experiment 1 by using objects instead of faces. By comparing correctly chosen matching faces to incorrectly chosen non-matching faces we wanted to show that the REME emerges rapidly, in the first 2000 ms of the trials. Our previous results showed that the relational eye movement effect emerges 500-1000 ms after stimulus onset on test trials. Moreover, we also wanted to demonstrate that the effect is apparent before overt responses are made. We analysed overall eye movement data before responses and we also separated our response-locked analysis to several 500 ms time bins before responses. In Experiment 1 the REME was apparent for the overall looking data before responses and also in a separate time bin analysis it appeared 500-1500 ms before overt responses were made.

Methods

Participants

45 undergraduate students (ages 18-27; 28F/17M) took part in the experiment from Eötvös Loránd University, Budapest. All of them received course credit for participation and gave informed consent for the study approved by the Research Committee of Eötvös Loránd University, Budapest (Ethical No.: 2012/6).

Behavioural paradigm

The behavioural part of the experiment was exactly the same as in Experiment 1. The only difference was that instead of faces, in Experiment 4, we used images of objects as target stimuli. The BOSS (Bank of Standardized Stimuli, Brodeur, Dionne-Dostie, Montreuil & Lepage, 2010; Brodeur, Guérard & Bouras, 2014) database was used and we selected 99 objects from five different categories: *electronic devices & accessories, furniture, toys & entertainment, kitchen & utensils, hand labour tools & accessories*. We controlled for the level of familiarity and visual complexity of the selected objects. First we calculated the average and standard deviation of both familiarity and visual complexity for the five categories taken together. Then we selected 99 objects out of those potential ones, which had both their familiarity and visual complexity score -1 and +1 standard deviation within the calculated means. Selected objects are listed in the appendix. The photos of the objects were resized (220 pixel X 220 pixel) to have almost identical surface area as the original faces (190 pixel X 254 pixel). The only difference was that the shape of the images for objects was square, while the faces were rectangular. The scenes remained the same as in our previous experiments. During the leaning phase the objects appeared in the middle of the scenes, while during test, the three objects were again equally distanced from the fixation point superimposed on a scene.

Eye movement acquisition and analysis

Both eye movement acquisition and analysis were the same as in Experiment 1. (page 89). Again, we will report in the main text our raw-data-based results but we will present the fixation-based results in the appendix.

Statistical analyses

Similarly to Experiment 1., we used repeated measures ANOVAs to analyse eye movements in different time bins to check for either the time-course or response-locked REME. The main factors were response accuracy (correct vs. incorrect) and time bins. For all ANOVAs Mauchly's test of sphericity was used to ensure that the assumption of sphericity was met. We used Greenhouse-Geisser corrections if the assumption of sphericity was violated. For comparing the effects on specific time bins we used post-hoc comparisons with Bonferroni corrected paired *t*-tests.

Results

Task performance

Average task accuracy was 61.8 (SD = .160) which was significantly above chance level (33%), $t(44) = 11.78$, $p < .001$, $d = 1.76$ (Figure 46). Four participants had chance level or lower accuracy so we excluded them from our data analysis because these participants showed no sign of relational memory performance. However, we also ran our analysis on all participants and our main findings stayed the same.

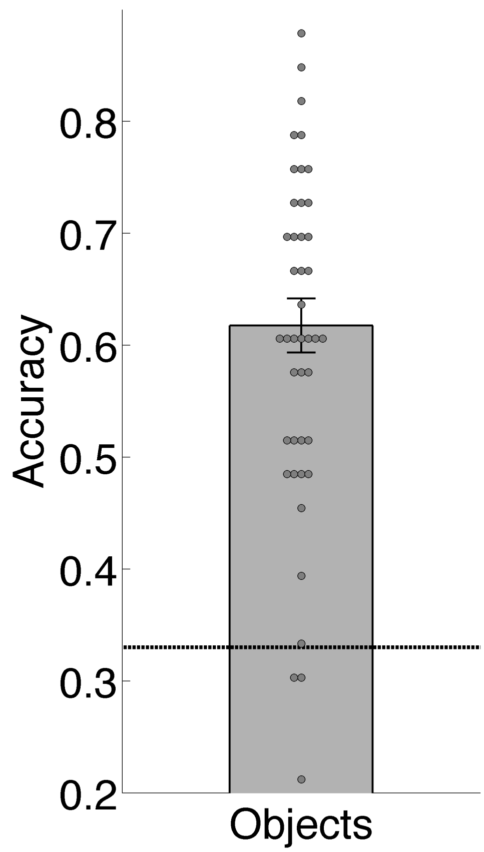

Figure 46. Average accuracy of Experiment 4, points represent individuals. Error bar indicates SEM.

We compared the accuracy of above chance performers in Experiment 1 and Experiment 4 and found that there was a significantly better performance in Experiment 4 $t(85) = 2.24$, $p = .0275$, $d = .24$.

Average reaction time for correct answers was 3902 ms ($SD = 872$) and for incorrect answers, 5222 ms ($SD = 968$). Correct answers were faster than incorrect ones, $t(41) = -8.84$, p < .001, $d = -1.36$. The same pattern was observed in Experiment 1.

We also analysed the behavioural task performance on different confidence levels. We used the same confidence level separation as in Experiment 1.

We also created three different confidence levels using the same separation method as in Experiment 1. For Low confidence level we merged Level 1 and Level 2. Our

Medium confidence level contained Level 3 and Level 4 answers and the High confidence level consisted of Level 5, 6 and 7 answers. The accuracy results indicated that at Low confidence level there was no significant task performance (M = .35, SD = .20, $t(38)$ = .51, p = .61, d = .08), however Medium confidence (M = .52, SD = .23 $t(39)$ = 5.17, $p < 0.001$, d = .82) and High confidence level (M = .85, SD = .11 $t(40)$ = 30.59, $p < 0.001$, d = 4.78) both showed above chance performance (Figure 47). We also used a different method for the separation of the confidence levels, where we put Level 5 answers to the Medium confidence group (Low group= Level 1-2; Medium group= Level 3-5; High group = Level 6-7) and the accuracy results did not differ from our original results.

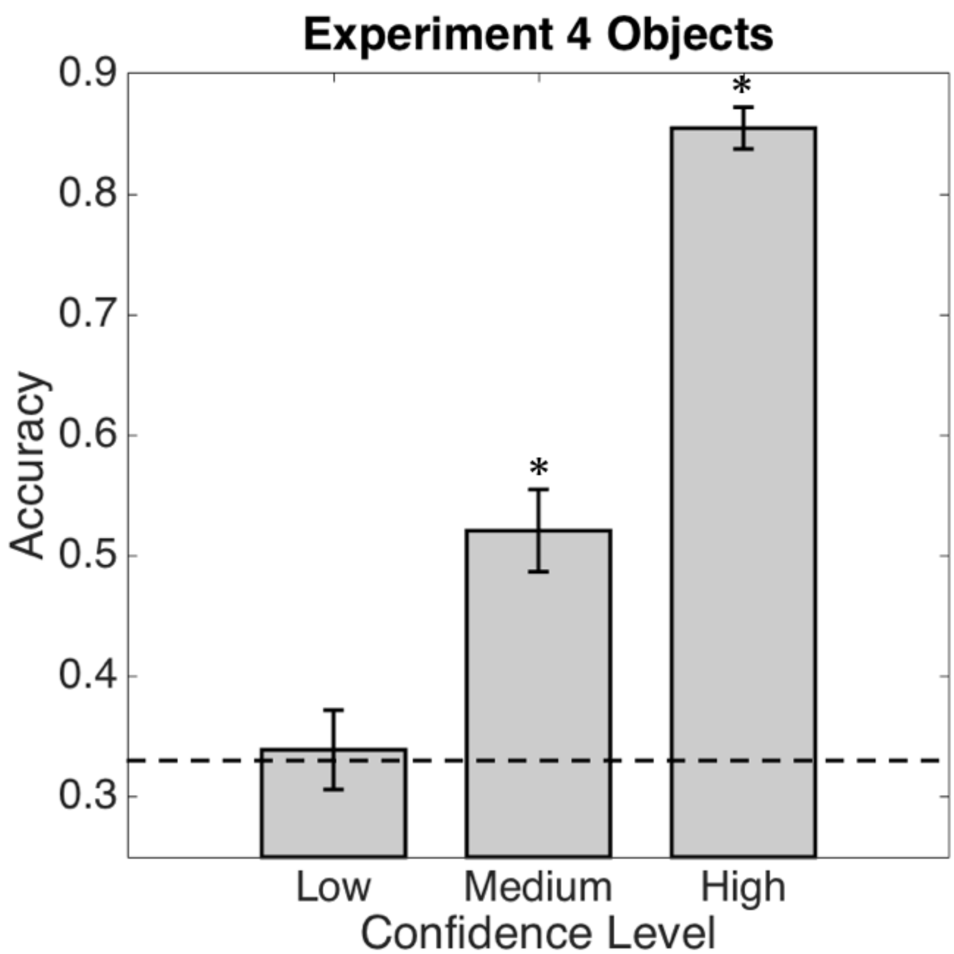

Figure 47. Average accuracy on Low, Medium and High confidence levels. Error bars indicate SEM. Stars represent significant difference from chance level performance.

Eye movement results

Onset-locked time-course analysis:

Firstly, we wanted to show evidence that our participants demonstrated a rapid REME within the first two seconds of the test trials. As in Experiemnt 1 we conducted a 2 (answer) X 4 (time bins) repeated measures ANOVA to test the appearance of the early-emerging REME. The results showed no main effect of answer $F(3, 40) = 3.60$, $p = .065$, $\eta^2 = .08$ or time bins $F(3, 120) = .622$, $p = .60$, $\eta^2 = .02$ and no interaction $F(3, 120) = .630$, $p = .60$, $\eta^2 = .02$. We also conducted Bonferroni corrected post-hoc paired t-tests for every time bin and found no significant difference in either of the time bins (Figure 48). Specifically, compared to Experiment 1, where we saw evidence of the REME within the 500-1000 ms time window, in Experiment 2 there was no evidence of a REME in the same 500-1000 ms time window $t(40) = .56$, $p = .562$, $d = .09$. There was only a tendency of a difference between correct and incorrect viewing proportions between 1500-2000 ms after trial onset, $t(40) = 2.072$, $p = .045$, $d = .32$, but this result was below our Bonferroni corrected p value acceptance level. These showed that in Experiment 4, using objects as target stimuli, we could not find sufficient evidence indicating a rapid REME and failed to replicate our previous results in Experiment 1.

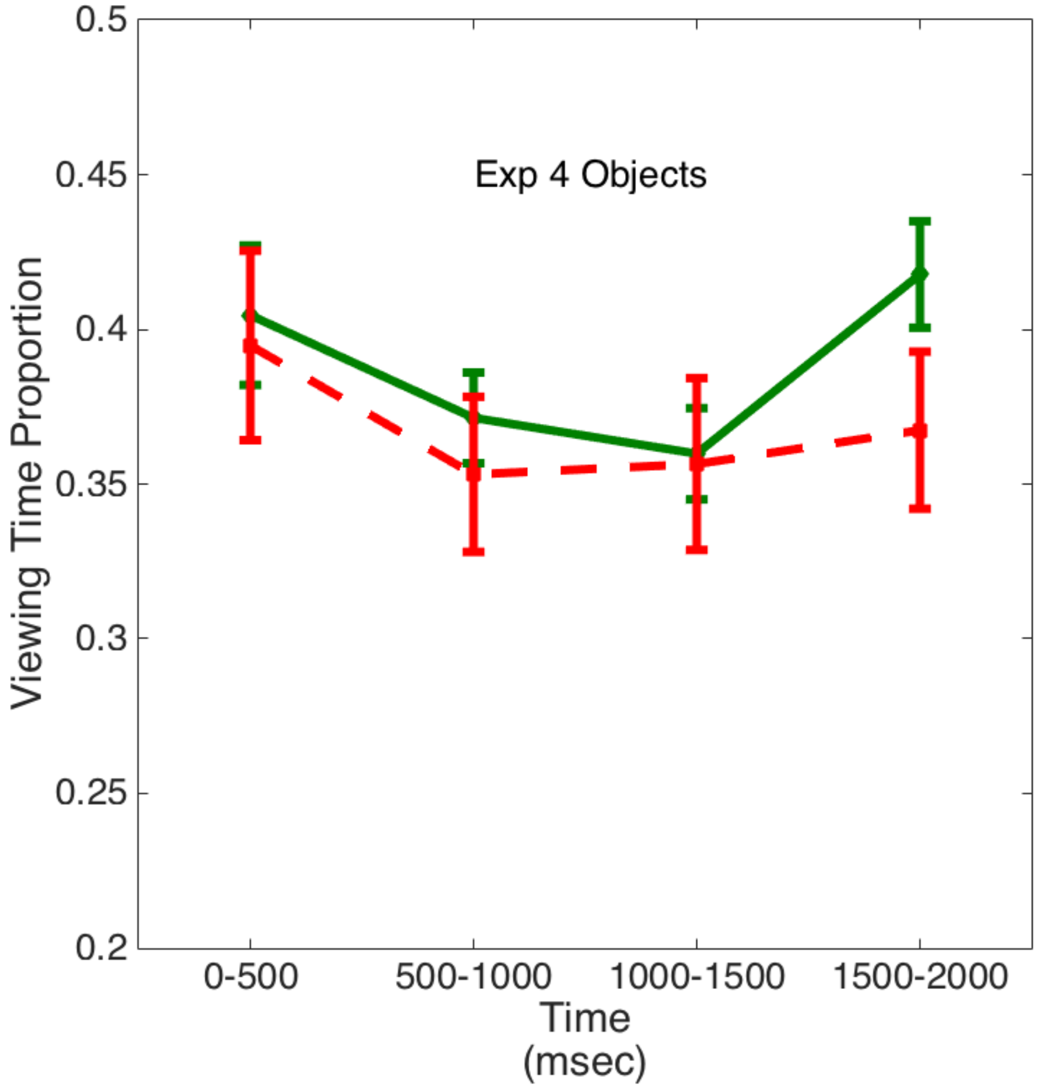

Figure 48. Onset-locked time-course analysis results in 500 ms time bins for the first two seconds of test trials. Green line represents correct responses; red dashed line represents incorrect responses. Error bars indicate SEM.

Response-locked analysis:

Overall response-locked eye movement results:

We also analysed our eye movement data before overt responses to find evidence that the REME appears before responses are made. We compared the viewing time proportions before every response in the trials between correct and incorrect answers. The results showed no difference between correct and incorrect responses, $t(40) = 1.21$, $p = .240$, d = .19 (see Figure 49). As our previous results, these results also showed a lack of the REME in the overall response-locked analysis.

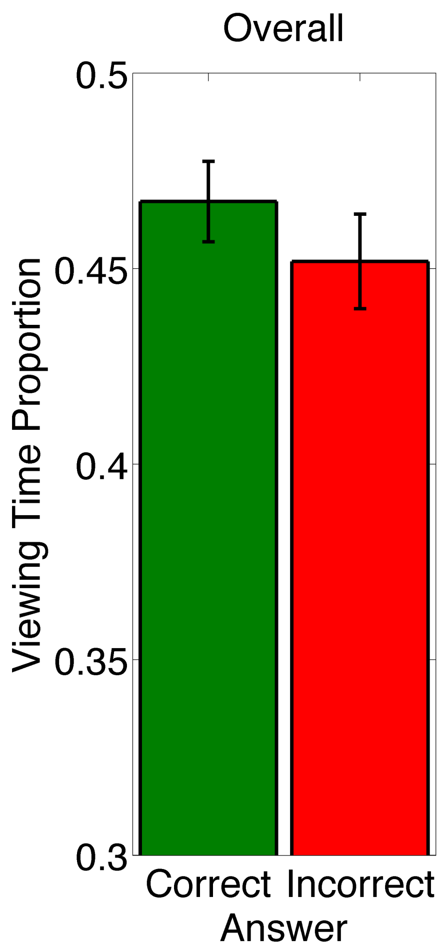

Figure 49. Overall response-locked viewing time proportions for correct (green) and incorrect (red) responses. Error bars indicate SEM.

Response-locked analysis for several time bins:

To check for the response-locked REME in specific time bins before responses we run a 2 (answers: correct, incorrect) X 4 (time bins) repeated measures ANOVA for the 2000 ms period before responses were made. We found a main effect of time bins $F(3, 120) = 13.37$, $p < .001$, $\eta^2 = .25$, but no main effect of answers $F(1, 40) = .59$, $p = .45$, $\eta^2 = .05$, or interaction $F(3, 120) = 1.60$, $p = .19$, $\eta^2 = .04$. Corrected paired t-test showed no significant differences between correct and incorrect viewing time proportions, not even Bin 4 1500-2000 ms before response $t(40) = 1.89$, $p = .065$, $d = 0.29$ (Figure 50). We could conclude that we failed to find any time bin before responses that showed evidence of a REME.

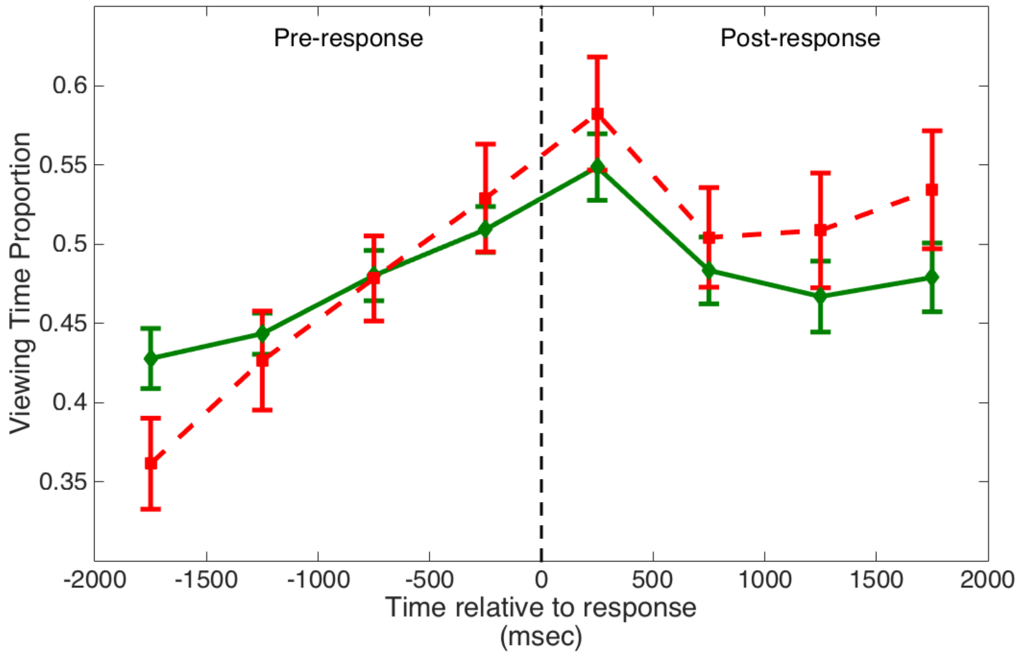

Figure 50. Response-locked results for 500 ms time bins before and after responses. Error bars indicate SEM.

Eye movements on different confidence levels

Before response overall results on different confidence levels:

As in Experiment 1 we also analysed eye movement data on different confidence levels to check for any potential difference between the levels. We based our overall response-locked analysis on the viewing time proportions before correct and incorrect responses. We compared our three confidence level groups in a 2 (answers: correct vs. incorrect) X 3 (confidence level: low, medium, high) ANOVA, which showed no main effect of answers $F(1, 26) = 1.37$, $p = .25$, $\eta^2 = .05$, no main effect of confidence $F(2, 26) = .10$, $p = .09$, $\eta^2 = .004$, and no significant interaction between answer and confidence level $F(2, 52) = .26$, p = 0.77, $\eta^2 = .01$. Post-hoc

Bonferroni corrected pairwise comparisons showed no significant difference between correct and incorrect responses on any of the confidence levels: Low confidence level, $t(26) = 1.34$, $p = .19$, $d = .26$, Medium confidence level, $t(26) = 1.34$, $p = .90$, $d = .26$, and High confidence level, $t(26) = .79$, $p = .44$, $d = .15$ (Figure 51). As we did in Experiment 1, we also illustrated this result by presenting the correct – incorrect viewing difference on the three confidence levels, which were not different from 0 in any of the subjective confidence levels (Figure 52). All these results stayed the same when we used a different separation method (Low group= Level 1-2; Medium group= Level 3-5; High group = Level 6-7). Our analyses paralleled our overall eye movement findings and showed no evidence of the REME on any of the confidence levels.

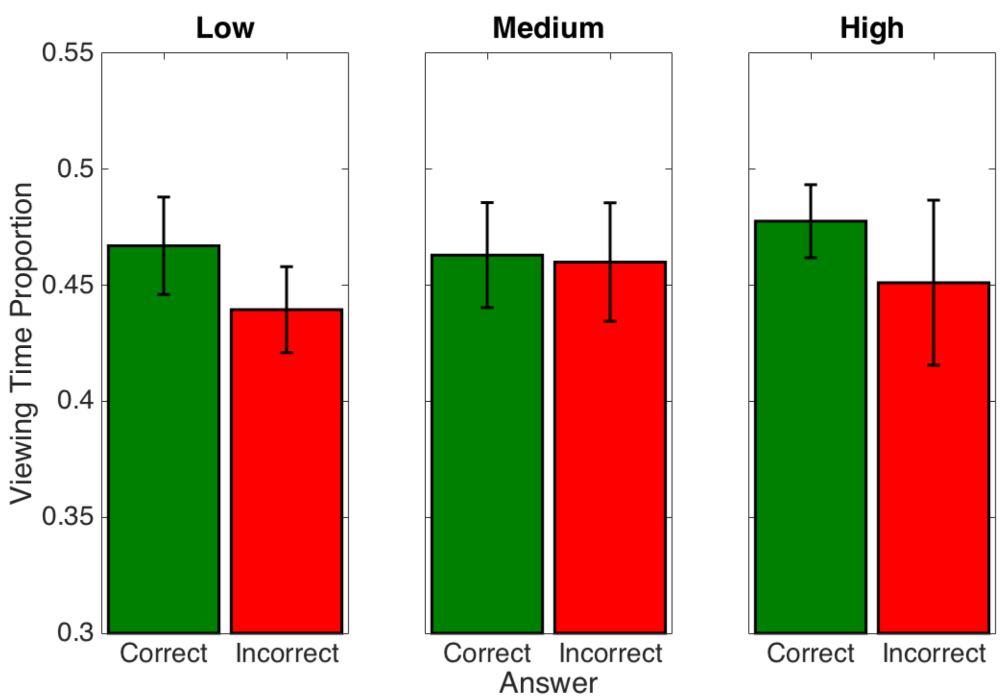

Figure 51. Overall response-locked viewing time proportions in Low, Medium and High confidence responses. Green indicates correct; red indicates incorrect responses. Error bars indicate SEM.

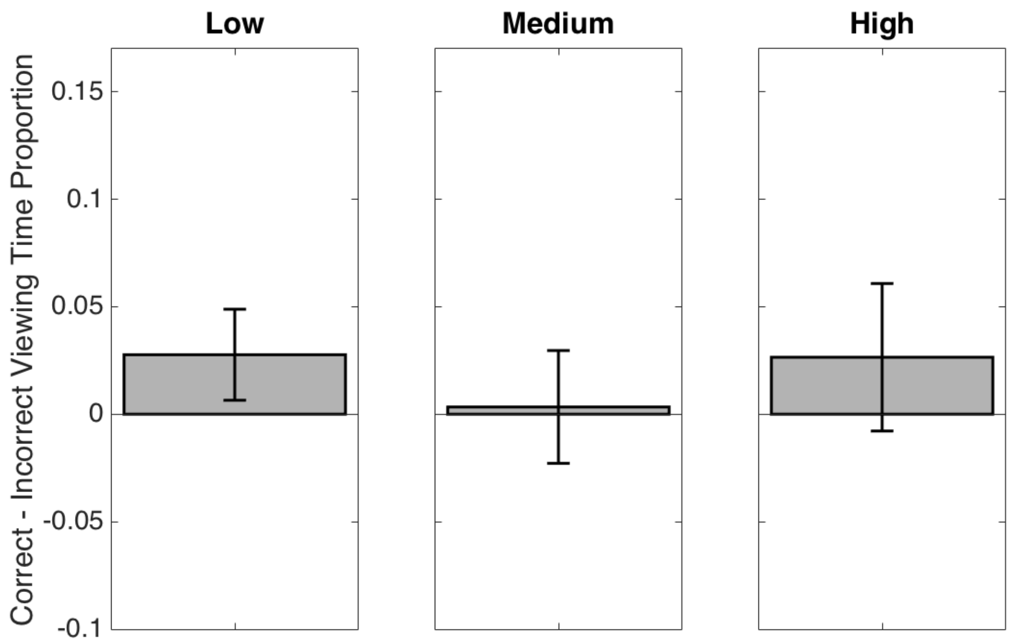

Figure 52. Overall response-locked viewing time proportion differences (Correct – Incorrect) in Low, Medium and High confidence responses. Error bars indicate SEM.

Discussion

Our assumption states that the REME is a universal and necessary indicator of relational memory retrieval. This predicts the emergence of this eye movement effect independently of stimulus-type. In Experiment 4, we changed the target stimuli to objects and predicted that if participants would show relational memory retrieval than the REME has to emerge. Indeed, our participants showed an overall task accuracy around 62 %, which was slightly, but significantly higher than the overall task accuracy of Experiment 1.

Contrarily to our prediction our results showed no evidence of the emergence of the REME. We could not find a rapid REME during the first 2000 ms of test trials between the viewing time proportions of correct and incorrect responses. Additionally, we failed to find the REME before responses. Neither our overall response-locked analysis, nor our time bin separated response-locked analysis yielded in any evidence that the REME was present in this experiment. These results suggest that the eye movement effects depend on the category of the target stimuli.

This lack of any sign of the REME is even more interesting in the light of relational memory performance. Using objects as target stimuli resulted in better memory performance than Experiment 1 using face, but this memory performance was associated with a lack of the REME. However, in Experiment 1 a slightly but significantly lower memory performance was associated with strong REME. This result raise important questions related to the validity of our assumption, which states that the REME is universal and necessary indicator of successful relational retrieval.

The universality of the REME

The results of Experiment 4 suggest that the emergence of the REME is category-dependent: it appears for faces but it does not appear for objects. This category-dependency implies that the REME is not a universal indicator of relational memory; it might be specific to face-like stimuli. These concerns are important, because in the memory research community the REME is referred to as a universal measure of relational retrieval. The results of this paradigm are often used to emphasize that eye movements are able to signal general relational memory processes. However, the results of Experiment 4 highlight the possible limit of such generalization of the eye movement effect. These suggest that the previously reported REME results might be stimulus-type dependent and they might indicate a certain type of relational retrieval specific to the access to face-like stimuli from a relational representation.

If we also consider our previous results in Experiment 1 and Experiment 2 using faces as target stimuli, which showed that the REME is tightly linked to conscious access of the matching faces, we can summarize the REME as a potential indicator of a specific relational retrieval process, which is associated with conscious retrieval of face-like stimuli from a relational memory representation. However, we have to be careful how to interpret our results, because it is possible that other stimulus categories are also associated with the REME apart from faces. This would mean that the REME is not specific to face-like stimuli, but a broader category of stimuli can elicit the effect (e.g., animate kinds). In our opinion these questions are essential for future memory research and they also point to the direction that there might be limitations of the paradigm to contribute to general memory theories.

Is the REME necessary for relational retrieval?

The results of Experiment 4 indicate that the REME can be absent, while participants can still show evidence of relational memory retrieval. Our participants demonstrated reliable relational memory retrieval in their task accuracy level, which was around 60 %. This dissociation between the REME and above chance task performance is not predicted by our assumption, which regards the REME as a necessary indicator of relational retrieval. Based on this assumed necessity, our first prediction states that when there is evidence of relational retrieval in the task, then the REME has to appear. However, our results clearly demonstrated that the REME could be absent, while memory performance can still signal reliable relational retrieval. These results strengthen our previous results of the unaware participant group in Experiment 2, which also suggested that there could be an absence of the REME, while other evidence could still suggest some degree of relational memory retrieval. In Experiment 2 unaware participants lacked the REME, but they still demonstrated relational memory in their looking behaviour favouring one of the non-matching faces compared to the matching face. Taken together we can conclude that our results strongly suggest that the REME is not necessary for relational retrieval, which is problematic if one wants to use the REME as a behavioural measure of obligatory processes for relational retrieval, like the first-stage of recollection. We will further elaborate the consequences of our results in the general discussion.

Chapter 6: General Discussion

Relational memory is described as the memory for arbitrary relations among the elements of an experience (Konkel & Cohen, 2009; Konkel et al., 2008). Relational memory provides us with the ability to retrieve additional information about items beyond their oldness based on their previous co-occurrence with other items or their context (e.g., the names of the faces or the location of objects).

Memory for relations among the constituent elements of our experiences is essential to episodic memory, the remembering of unique personal events (Tulving, 2005). We can describe episodic memory as our memory for complex, multi-modal and multi-element events with specific spatiotemporal contexts accompanied by a unique conscious experience during retrieval (Eichenbaum, 2004; Moscovitch et al. 2016; Henke, 2010; Tulving, 2002). Moscovitch (2008) proposed a two-stage model of episodic retrieval where the access to the representation and the conscious access to the content of the representation are separated. According to the two-stage model the first stage of recollection is a rapid, hippocampus-dependent automatic interaction between a cue and the previously encoded memory trace, which results in the access to the episodic representation. This access at the first stage is not consciously apprehended but it enables conscious access. It can also guide behaviour on a variety of memory tasks. During the second, slower stage the representation becomes accessible to conscious processes, which can give rise to the unique phenomenological experience and make the content explicit. The two-stage model of recollection not only focuses on the outcome of the process, which is explicit recollection, but it identifies potential underlying preconditions, which enable the conscious access to episodic representations. It claims that explicit recollection is preceded by unconscious relational retrieval. An important question is whether behavioural markers of unconscious access to relational representations can be identified. Recently, the relational eye movement effect (REME, Hannula et al., 2007; Hannula & Ranganath, 2009; Baym, Warren & Cohen, 2012) emerged as potential behavioural measure and precursor of relational retrieval. This marker supposed to allow testing the theoretically important questions about whether recollection is best characterized as a two-stage process.

Previous studies showed that eye movements could reflect unconscious memory for items (e.g. faces), scenes and also the relations between items (Ryan, 2000; Hannula et al., 2010; Hannula, Baym, Warren & Cohen, 2012). Using a face-scene paradigm Hannula et al. (2007) and Hannula & Ranganath (2009) demonstrated that eye movement behaviour could index relational retrieval processes. In our opinion this relational memory task measures mainly recollective processes. We discussed earlier that unitizing item pairs (e.g., face-scene, object-scene) to a single element could form rapidly new associations. This unitization at encoding would mean that the identification of the target stimuli at test could be done without the involvement of recollective processes, solely on the basis of familiarity. According to this view, above chance performance on the task could be a consequence of familiarity-based decisions. While we accept the concept of unitization, we have a number of reasons to believe that in our task it is unlikely that familiarity-based associative recognition could play a significant part in helping participants to reach above chance performance.

Firstly, we know from previous research that familiarity is sensitive to perceptual changes in recognition tasks (Yonelinas, 2002). In general perceptual mismatch of stimuli between learning and test leads to a decrease of familiarity but not recollection (Yonelinas, 2002). These previous results were obtained using word stimuli but there is also evidence that using nonword (e.g., drawings) stimuli and manipulating the perspective of the stimuli decreases familiarity-based recognition (Srinivas & Verfaellie, 2000). The experiments, that demonstrated unitization effects always used a perceptual match between learning and test. Usually, an experiment consists of a learning phase of simultaneously presented item pairs followed by a test task, where participants have to discriminate between intact and rearranged pairs. At test, intact pairs are perceptually the same as they were at learning, which brings on the unitization effect that is based on the initial encoding of an item pair as a single item. A perceptual match between learning and test is essential for the effect of familiarity-based associative recognition in these tasks.

In contrast, in the face-scene paradigm there is a mismatch between the learning (scene + face) and test stimuli (scene + 3 faces). Test displays, which consist of three faces on a background scene, are always a completely new constitution of previously seen elements. We think that the presentation of perceptually new displays at test makes it improbable to suggest that familiarity-based associative recognition (via the unitization of the scene-face pair at encoding) can yield to correct answers and ultimately to above chance performance on the task.

Our second point, which supports that the task measures mainly recollection, is that previous research using the face-scene relational learning paradigm showed that the task is hippocampus dependent (Hannula et al., 2007). Previous results show that recollective abilities are based on hippocampal integrity (Aggleton & Brown, 1999; Yonelinas, 2002) and that familiarity is usually dependent on the perirhinal cortex (Aggleton et al., 2005; Yonelinas, Aly, Wang, & Koen, 2010; Haskins et al., 2013).

Based on previous results using the face-scene paradigm and the feature overlap of the relational eye movement effect and the first-stage of recollection (rapid and obligatory, can occur in the absence of conscious recollection, hippocampus dependent) the assumption can be formed that the relational eye movement effect is a universal and necessary index of successful relational retrieval, which can be used as a measure of the first-stage of recollection. In several experiments we tested specific predictions derived from the above assumption. A consequent main prediction should be that the REME is necessary for relational retrieval, which implies that when there is evidence of relational retrieval in the task the REME has to appear obligatorily (1). Based on the main prediction the following suppositions can be formed: the REME has to be independent of conscious processes (2), task independent (3) and stimulus-type independent (4).

In the following section we will introduce our main experimental results related to these predictions.

1. Is the REME necessary for relational retrieval?

Our experimental results provided contradicting evidence that the REME is necessary for relational retrieval. In Experiment 2 participants, who were unaware of the presence of a matching face lacked the REME, however this group showed a different pattern of looking behaviour compared to our baseline measure (Experiment 2.2). Unaware participants (similarly to aware participants) preferred significantly more a non-matching face (High viewing face), which was accompanied by less preference towards the other non-matching face (Low viewing face) compared to the baseline measures of viewing time proportion of these faces (Experiment 2.2). We interpreted these significantly different results between Experiment 2 and our baseline measure (Experiment 2.2) as evidence of a relational-memory-based eye movement effect that guided more encoding capacity towards one of the non-matching faces at the expense of the other non-matching face. This pattern of viewing behaviour demonstrated that the emergence of the matching face preference is not necessary to show relational-memory-based effects in eye movements. Additionally, in Experiment 4 we used objects as target stimuli and participants showed significantly better relational memory performance on the task compared to the faces task (Experiment 1), while we could not find evidence of the REME. Participants in this group lacked the rapid preference of matching faces within the first 2000 ms of test trials and they also lacked any response-locked effects in our overall and time bin separated eye movement analysis. This result also highlighted that there are certain conditions when the REME can be absent but performance can still show evidence of relational retrieval. Both of these results suggest that the emergence of the matching face preference in the task is not necessary for relational retrieval. These findings are in contrast with the assumption suggested by previous results (Hannula et al., 2007; Hannula & Ranganath, 2009) that predicts the obligatory emergence of the REME at relational retrieval.

2. Is the REME independent of conscious processes?

We tested the prediction derived from the assumption suggested by previous results (Hannula & Ranganath, 2009), that the REME is independent of conscious processes. Based on the assumption, we specifically predicted that it is possible to find conditions when the REME dissociates from measures of conscious retrieval. In Experiment 1 we analysed incorrect responses, when participants presumably lack conscious access to the matching face. Our results showed that in incorrect responses there was no evidence of the REME, participants did not show any sign of a matching face preference throughout incorrect trials. Moreover, in Experiment 1 we found a tight link between the level of subjective confidence and the magnitude of the REME. The magnitude of the REME increased between Medium and High confidence levels. This result suggested that there might be a tight link between conscious processes at retrieval and the emergence of the matching face preference. To further test the question of the independence between REME and conscious experience at retrieval in Experiment 2 we compared the eye movements of aware and unaware participants. The aware participant group consisted of those individuals who spontaneously reported that on test displays one of the three faces was the matching face. This group demonstrated the REME, which emerged as the matching face preference within the first 2000 ms of test displays compared to the other two non-matching faces. In contrast, the unaware participant group, who could not report the presence of the matching faces, did not show any preference towards the matching faces. However, this group demonstrated a relational memory effect by preferring one of the non-matching faces more than it was expected based on the baseline measure of viewing time proportions (Experiment 2.2). Taken together, we did not find evidence that the REME could dissociate from conscious retrieval, instead we found evidence that the matching face preference emerges when participants are aware of the presence of matching faces on test displays. We interpreted our results as strong evidence that the emergence of the REME is tightly linked to the conscious retrieval of the matching faces.

3. Is the REME task independent?

In Experiment 2 we replicated previous results (Hannula et al., 2007) showing that the matching face preference emerges rapidly, in the first 2000 ms of test trials when the task is not to explicitly choose the matching face from the three-face test displays. For our data analysis we used both the previously used method to test the matching face preference against the 0.33 chance level, which is predicted if none of the faces were preferred during the test display presentation. Moreover we also tested the emergence of the matching face preference by comparing the average viewing time proportions of the three faces on test displays (Matching face, High viewing non-matching face and Low viewing non-matching face). We used this method because we argued that there is the possibility that on some test trials the matching face is preferred more than the 0.33 chance level, but this does not necessarily have to imply greater viewing proportions than both of the non-matching faces. Both of these analyses indicated that a matching face preference was evident during the first 2000 ms of test trials in the no choice task. These results showed that the matching preference is not associated with response requirements (as it was also showed in Hannula et al., 2007). Moreover, our results demonstrated that this task independence of the REME is paired with conscious retrieval. In Experiment 2 the overall relational memory effect was driven by aware participants, who could explicitly report the presence of matching faces (see point 3. above).

4. Is the REME stimulus-type dependent?

Our assumption based on previous results (Hannula et al., 2007; Hannula & Ranganath, 2009) states that the REME is a universal measure of a process, which obligatorily accompanies relational retrieval. However previous studies only used a subset of possible target stimuli to demonstrate the effect. These target stimuli were usually human faces (Hannula et al., 2007 Hannula & Ranganath, 2009), pictures of toys with faces (Chong & Richmond, 2015) or pictures of 3D creatures with faces (Baym et al., 2014) as target stimuli. In Experiment 4 we changed the target stimuli category from faces to objects and tested the emergence of the REME. Participants demonstrated a slightly but significantly better relational memory performance on the object task compared to Experiment 1, which used faces as target stimuli. Interestingly, participants in the object experiment did not demonstrate the REME in their eye movement behaviour. They lacked both the rapid REME within the first 2000 ms of test trials and they also did not show the relational eye movement effect before their overt responses. Moreover, participants lacked the REME effect on all levels of subjective confidence (Low, Medium and High). These results raised questions about the universality of the eye movement effect. They point to the direction that the relational eye movement effect might be stimulus-type dependent.

At this point it is unclear whether the REME is specifically connected to the retrieval of face-like target stimuli, because there could be other stimulus categories where the REME might be present apart from face-containing targets. Future experimental work is needed to test whether other categories beside faces or face-containing stimuli, like animate kinds can elicit the REME.

In addition, the results of Experiment 4 demonstrated that the REME signals a process, which is not necessary for relational retrieval (see point 1. above). In this experiment relational retrieval performance was paired with a lack of the eye movement effect. Is sum, we can conclude that the eye movement effect might not be universal and necessary for relational memory retrieval.

Taken together our experiments suggested that the REME is (a) not an index of a necessary process for relational retrieval, (b) it is not dissociable from (and tightly linked to) conscious processes and (c) it is stimulus-type dependent.

In the last section we will present three general conclusions suggested by our experimental work.

Conclusion no. 1: **The REME is not a valid behavioural measure of the first stage of recollection.**

Based on previous results and the feature overlap between the REME and the first stage of recollection (e.g., rapid and obligatory emergence, hippocampus-dependence, appearance in the absence of conscious access) it was suggested in our assumption that the REME is a behavioural measure of the first stage of recollection, which enables the conscious retrieval of the representational content, but which is independent of conscious awareness (second stage). If the REME signals a process which is necessary for conscious retrieval of the relational content, than one has to find the emergence of the REME as a prerequisite trigger of conscious retrieval. Additionally, the assumption predicts that the REME can dissociate from measures of conscious retrieval, namely if the REME signals a preconscious process, it can occur without attaining conscious access to retrieval, as pervious results suggest (Hannula & Ranganath, 2009).

Our experimental results suggested that the REME does not necessarily accompany relational retrieval (Experiment 2 and Experiment 4) and it is not dissociable from conscious retrieval (Experiment 1 and Experiment 2). These characteristics of the relational eye movement effect are in contrast with the proposal that the REME is a measure of the first-stage of recollection. We can conclude that the relational eye movement effect is not a behavioural measure of the first stage of recollection. This implies that this eye movement measure is insufficient to test and confirm the predictions related to the first stage of the two-stage model of recollection (Moscovitch, 2008).

Conclusion no. 2: **The REME signals a process, which is obligatorily connected to the conscious access to relational memory representations. This proposes the conscious access framework of the REME.**

In our experiments we demonstrated that the emergence of the REME is strongly associated with conscious access to the relational representational content. As we will show in detail below, our results suggest a conscious access framework of the REME, which asserts that the relational eye movement effect is a precursor for conscious retrieval in a sense that whenever it occurs, it obligatorily brings on conscious retrieval. This predicts that when there is evidence of the REME, than conscious access of the matching face also has to appear. In Experiment 1 our results showed that the REME appeared both in Medium and High confidence judgments and it was absent when no relational performance was present (Low confidence). If we regard subjective confidence level as an indirect measure of conscious access, than this result suggests that the REME is present when there is some evidence of conscious retrieval. Additionally, greater magnitude of the REME was linked to an increase in subjective confidence level (Medium < High), which we interpreted as evidence that the eye movement effect might be sensitive to different levels of conscious retrieval. These results are consistent with the conscious access framework of the REME. More importantly, in Experiment 2 the REME was only present in our aware participants, who could spontaneously report the presence of the matching face in test displays. The eye movement effect was absent in our unaware participants, who did not report the presence of the matching face. These results indicating that the REME is present in case of conscious access to the matching face, and it is absent when no conscious access is reported is consistent with this framework.

Moreover, the claim that the REME is an index of conscious access to the content of relational representations predicts a *lack* of dissociation between the emergence of the relational eye movement effect and conscious access. Our experimental results are also consistent with this prediction. We did not find evidence of a matching face preference in incorrect responses, where it can be presumed that there is an overall lack of conscious access to the matching face, which would guide the participants' overt response to choose correctly. We could not demonstrate the REME in our unaware participants, who did not show preference of the matching face but who showed a relational-memory-based effect, which manifested in more viewing time allocated to one of the non-matching faces compared to the baseline level (at the expense of the other non-matching face). Taken these results together we could not demonstrate that the REME can dissociate from conscious access to the matching face, which is in line with the prediction of the conscious access framework of the REME. In contrast, there are previous results that are not compatible with this framework. Hannula and Ranganath (2009) demonstrated that the matching face preference could emerge in the absence of conscious recollection of the matching face (incorrect trials). The study used the BOLD signal change as an additional measure of relational retrieval and found that incorrect trials, which had higher than the median viewing time proportion of the matching faces (High viewing trials), showed greater hippocampal activation during the presentation of the scene cue compared to incorrect trials with lower than the median viewing time proportions (Low viewing trials). This correlation in incorrect trials between hippocampal activation and the matching face preference in eye movements suggested that the relational eye movement effect is linked to hippocampal activity even in the absence of conscious retrieval. This result gave the basis of the assumption that the REME must be independent of conscious retrieval. In our experiments, we only had the opportunity to measure eye movements to test for behavioural signals of relational memory retrieval. However, the results that a subset of incorrect responses can show neural evidence of relational retrieval connected to the REME does not necessarily support that those incorrect responses lacked conscious awareness of the matching face. As Kumaran & Wagner (2009) suggested it could be possible that in some trials during the test phase of the face-scene paradigm decision-related processes can intervene between the conscious retrieval and overt responses. These decision-related processes in some cases can guide the overt response of the participant to choose a non-matching face contrarily to the correct face supported by conscious retrieval processes. This interpretation of the Hannula and Ranganath (2009) result states that the main finding of the study, that those responses within the incorrect category, which have higher than the median viewing time proportion of the matching faces show hippocampal activation, essentially index the conscious access to the relational representation. If we accept this alternative interpretation of the incorrect response results suggested by Kumaran and Wagner (2009), then it is possible that the findings of the Hannula and Ranganath (2009) study are also compatible with the conscious access framework.

Additionally, the conscious access framework of the REME predicts specific experimentally testable patterns of the REME. According to this framework one prediction is that the REME would be absent when there is no evidence of conscious access to the relational representational content (the matching face). In our opinion this prediction could be tested by analysing the emergence of the REME between correct and incorrect responses, when the average task performance is not different from chance level. When there is conscious access of the matching face in the task than average task performance has to reach above chance level, because participants on average would choose the correct face more often than chance based on available conscious information. However, chance level performance on the task is a sign of the absence of conscious retrieval and the conscious access framework predicts no REME. Interestingly, in Experiment 1 on Low subjective confidence level we did not find above chance performance on the memory task, which was associated with the lack of the REME. This result is also compatible with the conscious access framework, which predicts the lack of REME in case when there is no evidence of conscious retrieval in task performance. In our opinion future research can be directed to test, whether there could be cases when the REME appears at chance level performance. Finding the REME at chance level performance, when there is no evidence of conscious retrieval could be used to argue against the conscious access framework.

Another way to test the conscious access framework is to examine the emergence of the REME in test trials, which are subjectively reported to be 'guess' responses by the participants. We can regard guess responses as memory responses, which lack the subjective experience of conscious access to the matching face. If there would be some conscious access to the matching face, than subjective confidence level would be on average higher than this level. The conscious access framework predicts the lack of the REME in guess responses. In Experiment 1 we only analysed the Low confidence level answers (level 1 and 2 of the 7-point scale), which included guess responses (level 1 of the 7-point scale) and found no evidence of the REME. This result can be regarded as compatible with the conscious access framework. Future experiments can target this question and test, whether there are certain conditions when the subjective confidence level is on guess level and it is accompanied by an emergence of the REME. This pattern would be inconsistent with the predictions of the conscious access framework and suggest that the REME could appear without any evidence of conscious retrieval.

Our REME results are in parallel with other memory studies, where relational eye movement effects reflected conscious awareness. Squire and his colleagues used a scene recognition paradigm in a series of experiments (Smith, Hopkins & Squire, 2006; Smith & Squire, 2008) and demonstrated that multiple eye movement measures differentiate between manipulated and repeated scenes only when participants are *aware* of the manipulation. For example in the study by Smith & Squire (2008) participants saw colour photographs of indoor and outdoor scenes. They were not informed at the beginning of the learning phase that their memory performance will be tested later. At test, participants received a surprise memory test with a novel, a repeated and a manipulated scene. The manipulated scene was always the last one presented and participants had to decide whether that scene was a new, repeated or manipulated scene. If they thought that the scene was manipulated then they also had to indicate what was changed or where did the change take place in the scene. If they answered correctly to either of the questions, their answer was treated to indicate awareness of the change. The results showed that only participants who were aware of the manipulations directed (a) higher proportion of fixations to the manipulated regions, (b) spent more time viewing the manipulated regions and (c) made more transitions in and out of the manipulated regions of the scenes. These results repeated earlier findings (Smith & Squire, 2006) and demonstrated that eye movement measures signal the emergence of conscious awareness. Taken together, these studies demonstrated the same eye movement effect in different task conditions. Smith, Hopkins & Squire (2006) showed that the effect occurred when participants expected their memory to be tested later in the experiment (Experiment 1 and 2). In addition, the effect emerged both when memory of the participants was tested after all the scenes had been viewed (Experiment 1), and also when memory was tested immediately after each scene (Experiment 2). These studies are consistent with our results of the REME and indicate that certain eye movement measures behave as an index of conscious access to relational memory representations. Nevertheless, we also want to mention that in another scene recognition study that we introduced earlier in this dissertation (see page 67), Ryan and her colleagues (Ryan et al., 2000) demonstrated unaware participants spent as much time looking at the changed regions of manipulated scenes as aware participants (Experiment 3). Both of the aware and unaware group spent more time looking at the manipulated regions compared to repeated scenes, but the difference between manipulated and repeated scenes was not significant for either group alone, which prevents us from drawing strong conclusions. However, using the number of transitions in and out of the changed region as a measure of relational memory, Ryan et al. (2000) found that only the unaware group made more transitions compared to repeated scenes. This result raises the question whether specific eye movement measures could be used to indicate unaware relational retrieval in certain task settings.

In sum, we can conclude that previous results using the viewing time proportion measure in scene recognition studies strengthen the conscious access framework of the REME that affirms that the effect reflects the conscious access of memory retrieval.

Conclusion no. 3: **The process indexed by the REME does not necessarily accompany conscious relational retrieval. Conscious retrieval can emerge without the REME.**

It is important to point out that the proposed obligatory link between the emergence of the REME and conscious retrieval in our experiments does not entail that conscious retrieval is always accompanied by the process signalled by the eye movement effect. In Experiment 4 we changed the target stimuli from faces to objects and our results indicated a lack of the relational eye movement effect. The absence of the REME in Experiment 4 was paired with a significantly better relational retrieval performance on the task compared to Experiment 1 using faces as targets. Moreover, the absence of the REME was apparent in all levels of subjective confidence (Low, Medium and High). We can only speculate about this difference between the results of the two Experiments using faces (Exp 1) vs. objects (Exp 4).

Our first point is that, previous research on face processing strongly suggests that faces are represented not only by an analytic representation of their discrete features (eyes, nose, mouth, etc.) but also by a holistic representation, as an amalgamation of their features (for a review see Tanaka & Simonyi, 2016). Results on different experimental paradigms such as the part/whole task (Farah & Tanaka, 1993), the face inversion task (Yin, 1969) and the face composite task (Young, Hellawell & Hay, 1987) all demonstrated that faces are processed holistically. The analytic-holistic representational formats are often seen as a continuum, where faces lie more on the holistic end compared to objects (Tanaka & Simonyi, 2016). The holistic representation of faces could be related to our results that suggest that the REME is an early-emerging index of conscious retrieval specifically related to faces. The hypothetical connection between the REME and holistic processing could be used to explain the lack of such an eye movement effect in other target stimuli categories that are less influenced by holistic processes, such as objects. However, as we stated earlier, it is not yet clear in the field of REME research, whether other categories are able to elicit the REME apart from faces.

Another point could be raised in regard to the difference between our REME results of faces and objects. In our opinion it is possible that in general, objects carry important conceptual information about the type of the individual item, which can help participants in the task to choose the targets based on category membership information alone, without the need to identify the exact token of the object in the three-object test displays. More conceptually driven identification of the targets might result in a lack of eye-movement-based memory effects in our task, which might depend more on perceptual processes. Faces also carry important conceptual information (e.g., gender, age, race), but these could be less diagnostic than the unique perceptual characteristics of the token of the face. The identification of the target faces based mainly on perceptual characteristics could drive eye movements to favour the retrieved target, resulting in the REME effect.

Additionally, the difference between faces and objects in regard to the eye movement effect contribute to our understanding of relational retrieval processes. These results suggest that the conscious access to the content of relational representations does not necessarily involve the retrieval process indexed by the REME. There could be other pathways leading to conscious access apart from the involvement of the process behind the REME. Based on our results, we can hypothesize a unique relationship between the retrieval process signalled by the REME and conscious retrieval. If there is REME, than the process behind the eye movement effect obligatorily brings on conscious retrieval, however conscious retrieval can be obtained without the involvement of the process indicated by the REME. The fact that conscious retrieval can be obtained without the REME suggests that rapid and automatic retrieval processes do not necessarily determine conscious access to the elements of a relational representation. In other words conscious access to the elements of an episode might be attained when there is no priming induced (quasi-)automatic activation of the representational content. This raises the question whether more controlled processes during episodic retrieval can guide retrieval processes to reach conscious awareness in the absence of an early-emerging, (quasi-)automatic activation of relational memory representation.

References

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral and brain sciences*, *22*(03), 425-444.

Althoff, R. R. (1998). *Eye-Movement-Based Memory Assessment: The Use of Eye Movement Monitoring as an Indirect Measure of Memory*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 997.

Althoff, R., Cohen, N. J., McConkie, G., Wasserman, S., Maciukenas, M., Azen, R., & Romine, L. (1998). Eye movement-based memory assessment. *Current oculomotor research: Physiological and psychological aspects*, 293-302.

Baddeley, A. (1988). Cognitive psychology and human memory. *Trends in neurosciences, 11*(4), 176-181.

Baddeley, A. (2001). The concept of episodic memory. Philosophical Transactions of the *Royal Society of London. Series B, 356*(1413), 1345.

Baym, C. L., Khan, N. A., Monti, J. M., Raine, L. B., Drollette, E. S., Moore, R. D., ... & Cohen, N. J. (2014). Dietary lipids are differentially associated with hippocampal-dependent relational memory in prepubescent children. *The American journal of clinical nutrition, 99*(5), 1026-1032.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one, 5*(5), e10773.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. *PloS one, 9*(9), e106953.

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in cognitive sciences, 11*(5), 219-227.

Chong, H. J., Richmond, J. L., Wong, J., Qiu, A., & Rifkin-Graboi, A. (2015). Looking Behavior at Test and Relational Memory in 6-Month-Old Infants. *Infancy, 20*(1), 18-41.

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature neuroscience, 2*(9), 844-847.

Clayton, N. S. (2015). Ways of thinking: from crows to children and back again. *The Quarterly Journal of Experimental Psychology, 68*(2), 209-241.

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature, 395*(6699), 272-274.

Cohen, N. J., Poldrack, R. A., & Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory, 5*(1-2), 131-178.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior, 8*(2), 240-247.

Conway, M. A. (2009). Episodic memories. *Neuropsychologia, 47*(11), 2305-2313.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126*(4), 393.

Conway, M. A., Turk, D. J., Miller, S. L., Logan, J., Nebes, R. D., Meltzer, C. C., & Becker, J. T. (1999). A positron emission tomography (PET) study of autobiographical memory retrieval. *Memory, 7*(5-6), 679-703.

de Urabain, I. R. S., Johnson, M. H., & Smith, T. J. (2015). GraFIX: A semiautomatic approach for parsing low-and high-quality eye-tracking data. *Behavior research methods, 47*(1), 53-72.

Degonda, N., Mondadori, C. R., Bosshardt, S., Schmidt, C. F., Boesiger, P., Nitsch, R. M., ... & Henke, K. (2005). Implicit associative learning engages the hippocampus and interacts with explicit associative learning. *Neuron, 46*(3), 505-520.

Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and cognition*, *15*(1), 147-162.

Diana, R. A., Van den Boom, W., Yonelinas, A. P., & Ranganath, C. (2011). ERP correlates of source memory: Unitized source information increases familiarity-based retrieval. *Brain Research, 1367*, 278-286.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*(4), 523-533.

Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological review, 111*(2), 524.

Eichenbaum, H. (1999). Conscious awareness, memory and the hippocampus. *Nature neuroscience, 2*(9), 775-776.

Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron, 44*(1), 109-120.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci., 30*, 123-152.

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. (2008). Remembering and Knowing. In Menzel, R. & Byrne, J. H. (Eds) (2008). Learning *and memory: a comprehensive reference*, Vol 2.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*(1), 23-30.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*(1), 23-30.

Gardiner, J. M., & Radomski, E. (1999). Awareness of recognition memory for Polish and English folk songs in Polish and English folk. *Memory, 7*(4), 461-470.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 50*(1), 114.

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.

Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: evidence from neuropsychology. *Journal of the International Neuropsychological society, 16*(05), 748-753.

Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2006). An FMRI analysis of the human hippocampus: inference, context, and task awareness. *Journal of cognitive neuroscience, 18*(7), 1156-1173.

Gregg, V. H., & Gardiner, J. M. (1994). Recognition memory and awareness: A large effect of study-test modalities on "know" responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology, 6*(2), 131-147.

Hannula, D. E., & Ranganath, C. (2009). The eyes have it: hippocampal activity predicts expression of memory in eye movements. *Neuron, 63*(5), 592-599.

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in human neuroscience, 4*(166).

Hannula, D. E., Baym, C. L., Warren, D. E., & Cohen, N. J. (2012). The eyes know eye movements as a veridical index of memory. *Psychological science, 23*(3), 278-287.

Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *Journal of cognitive neuroscience, 19*(10), 1690-1705.

Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron, 59*(4), 554-560

Heisz, J. J., & Ryan, J. D. (2010). The effects of prior exposure on face processing in younger and older adults. *Frontiers in aging neuroscience, 3*, 15.

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision, 8*(1), 9.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences, 7*(11), 498-504.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience, 11*(7), 523-532.

Henke, K., Mondadori, C. R., Treyer, V., Nitsch, R. M., Buck, A., & Hock, C. (2003). Nonconscious formation and reactivation of semantic associations by way of the medial temporal lobe. *Neuropsychologia, 41*(8), 863-876.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hower, K. H., Wixted, J., Berryhill, M. E., & Olson, I. R. (2014). Impaired perception of mnemonic oldness, but not mnemonic newness, after parietal lobe damage. *Neuropsychologia, 56*, 409-417.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience, 2*(3), 194-203.

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., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. *Scientific approaches to consciousness*, 13-47.

Konkel, A., & Cohen, N. J. (2009). Relational memory and the hippocampus: representations and methods. *Frontiers in neuroscience, 3*, 23.

Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1989). The autobiographical memory interview: a new assessment of autobiographical and personal semantic memory in amnesic patients. *Journal of clinical and experimental neuropsychology, 11*(5), 724-744.

Kumaran, D., & Wagner, A. D. (2009). It's in my eyes, but it doesn't look that way to me. *Neuron, 63*(5), 561-563.

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychology and aging, 17*(4), 677.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: a manifesto for change. *Journal of Experimental Psychology: General, 142*(4), 1180.

McCabe, D. P., & Geraci, L. D. (2009). The influence of instructions and terminology on the accuracy of remember–know judgments. *Consciousness and cognition, 18*(2), 401-413.

McCabe, D. P., Geraci, L., Boman, J. K., Sensenig, A. E., & Rhodes, M. G. (2011). On the validity of remember–know judgments: Evidence from think aloud protocols. *Consciousness and Cognition, 20*(4), 1625-1633.

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and cognition, 21*(3), 1435-1455.

Montaldi, D., & Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus, 20*(11), 1291-1314.

Moscovitch, M. (1995). Recovered consciousness: a hypothesis concerning modularity and episodic memory. *Journal of clinical and experimental neuropsychology, 17*(2), 276-290.

Moscovitch, M. (2008). The hippocampus as a" stupid," domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 62*(1), 62.

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annual review of psychology, 67*, 105-134.

Osvath, M., & Osvath, H. (2008). Chimpanzee (Pan troglodytes) and orangutan (Pongo abelii) forethought: self-control and pre-experience in the face of future tool use. *Animal cognition, 11*(4), 661-674.

Parks, C. M., & Yonelinas, A. P. (2015). The importance of unitization for familiarity-based learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(3), 881.

Piolino, P., Desgranges, B., & Eustache, F. (2009). Episodic autobiographical memories over the course of time: cognitive, neuropsychological and neuroimaging findings. *Neuropsychologia, 47*(11), 2314-2329.

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub-jays. Nature, 445(7130), 919-921.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*(1), 89-102.

Rajaram, S., & Coslett, H. B. (1992, November). *Further dissociations between "remember" and "know" responses in recognition memory*. In Annual Meeting of the Psychonomic Society, St. Louis, November.

Rajaram, S., Hamilton, M., & Bolton, A. (2002). Distinguishing states of awareness from confidence during retrieval: Evidence from amnesia. *Cognitive, Affective, & Behavioral Neuroscience, 2*(3), 227-235.

Reder, L. M., Park, H., & Kieffaber, P. D. (2009). Memory systems do not divide on consciousness: Reinterpreting memory in terms of activation and binding. *Psychological bulletin, 135*(1), 23.

Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: evidence from eye tracking. *Developmental science, 12*(4), 549-556.

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological review, 122*(1), 1.

Ryan, J. D., & Villate, C. (2009). Building visual representations: the binding of relative spatial relations across time. *Visual Cognition, 17*(1-2), 254-272.

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science, 11*(6), 454-461.

Schacter, D. L. (1995). Memory distortion: History and current status. *Memory distortion: How minds, brains, and societies reconstruct the past*, 1-43.

Schacter, D. L., Dobbins, I. G., & Schnyer, D. M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience, 5*(11), 853-862.

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current opinion in neurobiology, 17*(2), 171-176.

Sheldon, S. A., & Moscovitch, M. (2010). Recollective performance advantages for implicit memory tasks. *Memory, 18*(7), 681-697.

Simons, J. S., Graham, K. S., Galton, C. J., Patterson, K., & Hodges, J. R. (2001). Semantic knowledge and episodic memory for faces in semantic dementia. *Neuropsychology, 15*(1), 101.

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.

Smith, C. N., & Squire, L. R. (2008). Experience-dependent eye movements reflect hippocampus-dependent (aware) memory. *Journal of Neuroscience, 28*(48), 12825-12833.

Smith, C. N., Hopkins, R. O., & Squire, L. R. (2006). Experience-dependent eye movements, awareness, and hippocampus-dependent memory. *Journal of Neuroscience, 26*(44), 11304-11312.

Srinivas, K., & Verfaellie, M. (2000). Orientation effects in amnesics' recognition memory: Familiarity-based access to object attributes. *Journal of Memory and Language, 43*(2), 274-290.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of cognitive neuroscience, 4*(3), 232-243.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory, 82*(3), 171-177.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory, 82*(3), 171-177.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology, 46*(2), 225-245.

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: A review of the literature. *The Quarterly Journal of Experimental Psychology, 69*(10), 1876-1889.

Tulving E. (1983). *Elements of Episodic Memory*. Oxford: Clarendon.

Tulving, E. (1972). Episodic and semantic memory. From E. Tulving & W. Donaldson,(Eds.) *Organization of memory* (pp. 381-403).

Tulving, E. (1985a). How many memory systems are there? *American psychologist, 40*(4), 385.

Tulving, E. (1985b). Memory and consciousness. *Canadian Psychology/Psychologie canadienne, 26*(1), 1.

Tulving, E. (1995). Organization of memory: Quo vadis. *The cognitive neurosciences*, 839847.

Tulving, E. (2001). Episodic memory and common sense: how far apart? *Philosophical Transactions of the Royal Society of London 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. (2005). Episodic memory and autonoesis: Uniquely human. *The missing link in cognition: Origins of self-reflective consciousness*, 3-56.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*(4940), 301.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science, 277*(5324), 376-380.

Westmacott, R., & Moscovitch, M. (2003). The contribution of autobiographical significance to semantic memory. *Memory & Cognition, 31*(5), 761-774.

Westmacott, R., Black, S. E., Freedman, M., & Moscovitch, M. (2004). The contribution of autobiographical significance to semantic memory: Evidence from Alzheimer's disease, semantic dementia, and amnesia. *Neuropsychologia, 42*(1), 25-48.

Yang, J., Weng, X., Guan, L., Kuang, P., Zhang, M., Sun, W., ... & Patterson, K. (2003). Involvement of the medial temporal lobe in priming for new associations. *Neuropsychologia, 41*(7), 818-829.

Yazar, Y., Bergström, Z. M., & Simons, J. S. (2014). Continuous theta burst stimulation of angular gyrus reduces subjective recollection. *PLoS One, 9*(10), e110414.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology, 81*(1), 141.

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.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*(6), 747-763.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characterstics. J*ournal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1415.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: the 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*(3), 361.

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. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behavioural brain research, 254*, 34-44.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of memory and language, 34*(5), 622.

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus, 20*(11), 1178-1194.

Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12*(3), 323.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.

Appendix 1. Fixation-based analysis

We will present the results of our analyses for each experiment using the fixation-based calculation of viewing time proportions. We used the GraFIX (Saez de Urabain, Johnson, & Smith, 2015) software to calculate fixations from our raw data samples. A fixation was defined by a gaze velocity threshold of 10º/sec and a minimum duration of 100 ms. We used the position and duration of fixations to analyse eye movement data. Fixations made during the test displays were assigned to three (left face, right face and bottom face) regions of interests (ROIs) and the proportions of total viewing time allocated to each ROI were calculated for every trial. We used the same analyses as in our raw-data-based calculations, which we reported in the main text of the dissertation. As it will be evident from the detailed report presented below, generally we found the same effects as in our raw-data-based analyses. For helping the reader we will present the results that were **different** compared our raw-data-based results in **bold** letters.

The main difference between our raw data analysis results and the fixation-based results was that raw data results showed more robust post-hoc paired comparison results (significant paired comparisons in multiple time bins) compared to a general null result in post-hoc comparisons in the fixation-based analysis. This was our main reason to report the raw data results in the main text of the dissertation. These differences in the two analysis methods also highlighted important issues for future eye movement studies, namely that for viewing time proportion measures it is beneficial to use raw data calculations, which give more robust results than certain fixation-based calculations.

Exp 1

1. Onset-locked time-course analysis

We conducted a separate 2 (answers: correct, incorrect) X 4 (time bins) repeated measures ANOVA for the first 4 time bins. There was a main effect of answers, $F(1, 45) = 8.97$, $p = .004$ $\eta^2 = 0.17$, and time bins, $F(3, 135) = 6.11$, $p = .001$ $\eta^2 = 0.12$, but no interaction. **Bonferroni corrected post-hoc paired *t*-tests showed no significant differences in any of the time bins.** The closest time bin to a significant effect was 500-1000 ms after trial onset $t(45) = 1.79$, $p = .08$, $d = .26$ (the time bin, where the raw data results showed a significant effect).

The results of the ANOVA remained the same when we conducted a 2 (answers) X 20 (time bins) repeated measures analysis for the whole trial. There was a main effect of answers, $F(1, 45) = 13.84$, $p = .001$ $\eta^2 = 0.26$, and time bins, $F(19, 855) = 5.77$, $p < .001$ $\eta^2 = 0.11$, but no interaction.

2. Response-locked analysis

Overall response-locked eye movement results:

The results demonstrated the REME before responses, which was apparent in greater pre-response viewing time proportions for correct (M=0,49 SD=0,05), than incorrect responses (M=0,43 SD=0,05), $t(45)=7.07$, $p < .001$, $d = 1.04$.

3. Response-locked analysis for several time bins

We conducted a 2 (correct/incorrect answer) X 4 (time bins) repeated measures ANOVA and found a main effect of answer, $F(1, 44) = 7.140$, $p = .011$ $\eta^2 = 0.14$, and time bins, $F(3, 132) = 10.763$, $p < .001$ $\eta^2 = 0.197$, but no interaction. **Bonferroni corrected post-hoc paired *t*-tests indicated significantly higher proportions for correct responses in one time bin only**. The time bin 1000-1500 ms before responses $t(44) = 3.299$, $p = 0.002$, $d = .49$.

4. Eye movements on different confidence levels

We compared our three confidence level groups in a 2 (answers: correct vs. incorrect) X 3 (confidence level: low, medium, high) ANOVA, which showed no main effect of answers $F(1, 19) = 1.40$, $p = .251$ $\eta^2 = 0.069$, a main effect of confidence $F(2, 38) = 5.36$, < 0.001 $\eta^2 = 0.568$, and a significant interaction between answer and confidence level $F(2, 38) = 7.426$, $p = 0.002$ $\eta^2 = 0.281$. Post-hoc Bonferroni corrected pairwise comparisons showed **no significant difference between correct and incorrect responses** on Low confidence level $t(19) = -1.405$, $p = 0.176$, $d = .31$, or **on Medium level, $t(19) = -.179$, $p = .859$, $d = .04$**. However, the difference was significant between correct and incorrect responses on High confidence level, $t(19) = 3.249$, $p = .004$, $d = .73$.

Exp 2

1. All participants

We run a 3 (face type: Matching, Low viewing, High viewing) X 4 (time bins) repeated measures ANOVA for the first 4 time bins. This resulted in a main effect of face type $F(2, 74) = 27.731$, $p < .001$ $\eta^2 = 0.428$ and no significant main effect of time bins or interaction. **Most importantly, post-hoc comparisons showed that there was no significantly higher viewing time proportion for the Matching faces compared to the High viewing faces in any of the 4 time bins** (*p* threshold was .0125). There was a tendency for the difference in time bin 500-1000 ms after stimulus onset, $t(37)= 1.923$, $p = .062$, $d = .31$, and the same difference just fell out of our corrected p value threshold in the following time bin (1000-1500 ms after stimulus onset), $t(37)= 2.33$, $p = .025$, $d = .38$.

2. Eye movements in Aware and Unaware participant groups

First 2 seconds:

In the aware group the results showed a main effect of face type $F(2, 34) = 16.772$, $p < .001$ $\eta^2 = 0.497$ but there was no main effect of time bins or interaction of the factors. Pairwise comparisons both indicated a tendency that in Bin 2 (500-1000 ms after stimulus onset) $t(17)= 1.878$, $p = .078$, $d = .44$, and Bin 3 (1000-1500 ms after stimulus onset) $t(17) = 2.63$ $p = .017$, $d = .62$, the matching faces had higher proportions than High viewing faces (Figure 33). **However, these results both fell out of the corrected p value (.0125) for multiple comparisons.**

In the unaware group the ANOVA results showed a significant main effect of face type $F(2, 28) = 9.14$, $p < .001$ $\eta^2 = 0.396$ and no main effect of time bins or interaction of the two factors. Pairwise comparison showed no significant difference between Matching and High viewing faces.

Additionally to our first ANOVA, we also run a 2 (face type) X 4 (time bins) analysis where we were specifically interested in whether the main effect of face would disappear when we take out the Low viewing face category and only compare the Matching and High viewing faces. The results showed that the face type main effect was no longer significant $F(1, 14) = .715$, $p = .412$ $\eta^2 = 0.049$, and there was again no main effect of time bins or interaction of the factors. These results clearly showed that there was no REME present in the unaware group.

Whole trials analyses

Unaware group

We wanted to find evidence that participants in the unaware group might prefer any or both of the non-matching faces in test trials compared to the matching face. This would suggest that they encoded the faces effectively by allocating more viewing time to the non-matching faces. To answer this question we run a 3 (face type) x 20 (time bins) repeated measures ANOVA on the unaware group. Our results showed a main effect of face type $F(2, 28) = 62.377$, $p < .001$ $\eta^2 = 0.817$ and a significant interaction $F(38, 532) = 2.430$, $p < .001$ $\eta^2 = 0.148$. Our corrected pairwise comparisons showed that participants preferred the High viewing non-matching face compared with the Matching face in the following time bins: Bin 14 $t(14) = -3.85$, $p = .002$, $d = -.99$; Bin 15 $t(14) = -5.89$, $p < .001$, $d = -1.5$; Bin 16 16 $t(14) = -3.91$, $p = .002$, $d = -1.04$, Bin 17 $t(14) = -4.03$, $p = .001$, $d = -1.04$ Bin 18 $t(14) = -4.53$, $p < .001$, $d = -1.17$.

These results were also confirmed by a one-way (face type) ANOVA for the overall viewing time proportions without separating viewing behaviour to several time bins. Our analysis showed a significant face type effect $F(2, 28) = 62.592$, $p < .001$ $\eta^2 = 0.817$ and pairwise comparisons indicated an overall preference of the High viewing non-matching face compared to both the Matching face $t(14)= 7.32$, $p < .001$, $d = 1.89$, and the Low viewing non-matching face $t(14)= 9.63$, $p < .001$, $d = 2.48$.

Aware group

We wanted to find evidence that participants in the aware group might prefer any or both of the non-matching faces in test trials compared to the matching face. This would suggest that they encoded the faces effectively by allocating more viewing time to the non-matching faces. To answer this question we run a 3 (face type) x 20 (time bins) repeated measures ANOVA on the unaware group. Our results showed a main effect of face type $F(2, 34) = 71.213$, $p < .001$ $\eta^2 = 0.807$ and a significant interaction $F(38, 646) = 2.902$, $p = .004$ $\eta^2 = 0.146$. Our corrected pairwise comparisons showed that participants preferred the High viewing non-matching face compared with the Matching face in the following time bins: Bin 15 $t(17)= -3.77$, $p = .002$, $d = .89$.

These results were also confirmed by a one-way (face type) ANOVA for the overall viewing time proportions without separating viewing behaviour to several time bins. Our analysis showed a significant face type effect $F(2, 34) = 72.481$, $p < .001$ $\eta^2 = 0.810$ and pairwise comparisons indicated an overall preference of the High viewing non-matching face compared to both the Matching face $t(17)= 4.032$, $p = .001$, $d = .95$, and the Low viewing non-matching face $t(17)= 22.606$, $p < .001$, $d = 5.33$.

Exp 2.2 Baseline measure of face preference

Eye movement results for all participants

Our one-way ANOVA showed a main effect of face type (High, Medium, Low) on the viewing time proportions $F(2, 28) = 31.54$, $p < .001$, $\eta^2 = .72$. Post-hoc paired t-tests indicated that High viewing faces had significantly higher viewing time proportions than Medium, $t(14)= 5.54$, $p < .001$, $d = 1.43$, and Low viewing faces, $t(14)= 5.20$, $p <$

.001, d = 1.3, while Medium faces had higher proportions than Low faces $t(14)$= 5.14, $p < .001$, d = 1.33. One sample t-tests showed that viewing proportion of the High viewing faces was significantly above the 0.33 chance level $t(14)$= 6.30, $p <$ .001, d = 1.63, while the proportion of Low viewing faces were significantly below chance level $t(14)$= -4.85, $p < .001$, d = -1.25. The viewing proportion of Medium faces did not differ significantly from chance level $t(14)$= 1.14, p = .27, d = .29.

Comparing the results with Exp 2

Matching faces (Exp. 2) vs. Medium viewing face (Exp. 2.2):

In the aware group, there was no significant difference between Matching faces and Medium viewing faces of Experiment 2.2, $t(31)$= 1.27, p = .227, d = .22. Similarly, the unaware group's Matching faces showed no significant difference compared to Medium viewing faces of Experiment 2.2, $t(28)$= .655, p = .59, d = .12.

High viewing faces (Exp. 2) vs. High viewing faces (Exp. 2.2):

In the aware group, the High viewing faces had significantly higher viewing time proportions than the High viewing faces of Experiment 2.2, $t(31)$= 5.50, $p < .001$, d = 0.97. Similarly, in the unaware group the High viewing faces also showed significantly higher proportions compared to the High viewing faces of Experiment 2.2, $t(28)$= 5.96, $p < .001$, d = 1.11.

Low viewing faces (Exp. 2) vs. Low viewing faces (Exp. 2.2):

In the aware group, the Low viewing faces had significantly lower viewing time proportions than the Low viewing faces of Experiment 2.2, $t(31)$= -7.34, $p < .001$, d = -1.29. Similarly, in the unaware group the Low viewing faces also showed significantly lower proportions compared to the Low viewing faces of Experiment 2.2, $t(28)$= -5.7, $p < .001$, d = -1.08.

<u>Exp 3 Perceptual-mismatch</u>

<u>1. Onset-locked time-course analysis</u>

We conducted a separate 2 (answers: correct, incorrect) X 4 (time bins) repeated measures ANOVA for the first 4 time bins. There was a main effect of answers, $F(1, 36)$ = 10.138, p = .003 η^2 = 0.220, and time bins, $F(3, 108)$ = 5.80, p = .005 η^2 = 0.139, but no interaction. **Bonferroni corrected post-hoc paired *t*-tests showed no significant differences in any of the time bins.** The closest time bin to a significant effect was 1000-1500 ms after trial onset $t(36) = 2.519$, $p = .016$, $d = .41$ (the time bin, where the raw data results showed significant effect).

The results of the ANOVA remained the same when we conducted a 2 (answers) X 20 (time bins) repeated measures analysis for the whole trial. There was a main effect of answers, $F(1, 36) = 11.439$, $p = .002$ $\eta^2 = 0.241$, and time bins, $F(19, 684) = 10.169$, $p < .001$ $\eta^2 = 0.220$, but no interaction.

2. Response-locked analysis

Overall response-locked eye movement results:

The results demonstrated the REME before responses, which was apparent in greater pre-response viewing time proportions for correct (M=0,476 SD=0,05), than incorrect responses (M=0,423 SD=0,04), $t(36)=6.59$ $p < .001$, $d = 1.08$.

3. Response-locked analysis for several time bins

We conducted a 2 (correct/incorrect answer) X 4 (time bins) repeated measures ANOVA and found a main effect of answer, $F(1, 36) = 6.00$, $p = .019$ $\eta^2 = 0.143$, and time bins, $F(3, 108) = 20.56$, $p < .001$ $\eta^2 = 0.364$, but no interaction. Bonferroni corrected post-hoc paired *t*-tests indicated significantly higher proportions for correct responses in one time bin. The time bin 500-1000 ms before responses $t(36) = 2.71$, $p = .01$, $d = .45$.

Exp 4 Objects

1. Onset-locked time-course analysis

We conducted a separate 2 (answers: correct, incorrect) X 4 (time bins) repeated measures ANOVA for the first 4 time bins. **There was a main effect of answers, *F*(1, 40) = 5.315, p = .026 $\eta^2 = 0.117$**, but no main effect of time bins, *F*(3, 120) = .661, *p* = .576 $\eta^2 = 0.16$ and no interaction of the factors. **Bonferroni corrected post-hoc paired *t*-tests showed no significant differences in any of the time bins.** The closest time bin to a significant effect was 1500-1500 ms after trial onset *t*(40) = 1.84, *p* = .073, *d* = .28.

The results of the ANOVA changed when we conducted a 2 (answers) X 20 (time bins) repeated measures analysis for the whole trial. There was no main effect of answers, *F*(1, 39) = .81, p = .777 $\eta^2 = 0.002$, but time bins showed a significant main effect, *F*(19, 722) = 12.923, p < .001 $\eta^2 = 0.254$, but no interaction.

2. Response-locked analysis

Overall response-locked eye movement results:

The results demonstrated no REME before responses, which was apparent in not significantly different pre-response viewing time proportions for correct (M=0,41 SD=0,053), than incorrect responses (M=0,40 SD=0,05), *t*(40)=1.228 *p* = .227, *d* = .19.

3. Response-locked analysis for several time bins

We conducted a 2 (correct/incorrect answer) X 4 (time bins) repeated measures ANOVA and found no main effect of answer, *F*(1, 40) = .884, *p* = .353 $\eta^2 = 0.022$, but significant main effect of time bins, *F*(3, 120) = 13.002, *p* < .001 $\eta^2 = 0.245$, but no interaction. Bonferroni corrected post-hoc paired *t*-tests indicated no significantly different proportions for correct responses in any of the time bins. The time bin 500-1000 ms before responses was the closest to any difference, *t*(40) = 1.736, *p* = 0.09, *d* = .27.

4. Eye movements on different confidence levels

We compared our three confidence level groups in a 2 (answers: correct vs. incorrect) X 3 (confidence level: low, medium, high) ANOVA, which showed no main effect of answers $F(1, 26) = .528$, $p = .474$, $\eta^2 = 0.069$, a main effect of confidence $F(2, 52) = 8.092$, $p = 0.001$, $\eta^2 = 0.237$, and no significant interaction between answer and confidence level. Post-hoc Bonferroni corrected pairwise comparisons showed no significant difference between correct and incorrect responses on Low confidence level $t(26) = 1.050$, $p = 0.303$, $d = .20$, on Medium level, $t(26) = .219$, $p = .829$, $d = .04$ or on High level, $t(26) = .143$, $p = .887$, $d = .03$.

Appendix 2. List of object pictures

List of the object stimuli used in Exp. 4 from the BOSS database (Bank of Standardized Stimuli, Brodeur et al., 2010, 2014).

Number	Object picture name from BOSS database	Familiarity average	Familiarity standard deviation	Visual complexity average	Visual complexity standard deviation
1	cashregister01	4.48	0.80	2.67	1.16
2	chandelier	4.38	0.85	2.48	1.09
3	ceilingspeaker	3.75	1.24	2.22	1.29
4	nintendods	4.26	0.96	2.36	1.10
5	dvdcase	4.43	0.91	2.00	1.04
6	fanblades	4.17	0.96	2.02	0.87
7	gps	4.17	1.03	2.26	1.11
8	halogenlightbulb	4.02	1.16	2.33	1.07
9	ipad02	4.60	0.63	1.76	1.14
10	lectern01	4.17	1.06	2.02	0.81
11	megaphone	4.23	0.92	2.25	1.01
12	microphone01	4.60	0.70	2.26	1.13
13	papershredder	3.88	1.19	2.10	0.88
14	fanheater	4.26	0.99	2.33	0.93
15	securitycamera	4.19	0.97	2.40	1.27
16	smokedetector02	4.48	0.92	2.02	1.05
17	spotlight	3.83	1.15	2.38	1.06
18	videotape01b	3.73	1.20	2.29	1.04
19	webcam	4.17	1.12	2.43	1.06
20	wiicontroller	4.07	1.09	2.29	0.92
21	armchair02	4.62	0.58	2.31	1.18
22	barrel01	4.14	1.05	2.07	1.13
23	chest01	3.90	1.03	2.15	0.85
24	patiochair	4.60	0.66	1.88	0.83
25	desk02	4.52	0.77	2.60	1.04
26	officechair03	4.67	0.69	2.26	1.06
27	dresser02	4.55	0.59	2.19	0.92
28	filingcabinet	4.64	0.58	1.95	0.88
29	footrest02	4.44	0.74	2.05	0.88
30	keyhook	4.25	1.08	2.22	1.17
31	mirror02	4.52	0.67	2.05	0.76
32	nightstand	4.56	0.71	1.95	0.84
33	parkbench02	4.71	0.55	2.40	1.06
34	safe	4.26	0.89	2.36	1.14
35	shelves	4.43	0.99	1.95	0.94
36	floorlamp	4.57	0.70	2.07	0.95
37	stool01	4.67	0.65	1.81	0.92
38	gardenswing	4.12	0.92	2.45	1.02

39	airhockeytable	4.40	0.86	2.48	1.04
40	chessknight03a	4.52	0.89	2.14	1.03
41	gamecontroller01	4.55	0.77	2.00	0.77
42	horseshoe	3.98	1.15	1.88	0.98
43	jokercard	4.71	0.77	2.45	1.42
44	kidpicnictable	4.12	0.99	2.24	1.03
45	marble	4.32	0.91	2.27	1.25
46	mrpotatohead	4.48	0.83	2.62	1.32
47	paddleball	3.90	1.12	1.90	0.79
48	paperairplane	4.71	0.60	1.76	0.96
49	pingpongtable	4.57	0.63	2.12	0.89
50	smokingpipe	4.10	1.23	1.79	0.92
51	pokerchips	4.21	1.12	2.05	0.96
52	pooltable	4.66	0.69	1.98	0.85
53	toyfiretruck	4.36	0.93	1.95	0.94
54	toyanimal05	3.95	1.03	2.05	0.99
55	toysoldier01b	4.31	1.09	2.12	1.11
56	watercolorpaintset	4.49	0.78	2.07	1.03
57	jumpercables	4.24	1.10	2.33	0.93
58	bowrake	4.48	0.82	1.75	0.95
59	chainsaw	4.38	0.79	2.71	1.33
60	bucket02	4.34	0.94	1.95	1.00
61	ducttape	4.55	0.86	1.79	1.05
62	gastank	4.21	0.92	1.86	0.84
63	dolly01	4.24	0.99	2.05	0.89
64	hedgeshears	4.00	0.95	2.22	0.82
65	hoe	3.88	1.11	2.00	1.01
66	icescraper	4.17	1.02	1.93	0.97
67	ladder	4.60	0.77	1.93	0.92
68	lawnmower	4.55	0.67	2.57	1.21
69	leafblower01	3.88	1.11	2.54	0.95
70	boltcutter	3.85	0.96	2.24	0.80
71	loppingshears	4.07	0.84	2.29	0.92
72	pickaxe02	3.98	1.16	1.79	0.81
73	pipewrench	3.83	1.06	2.10	0.94
74	pitchfork	4.10	0.93	1.88	0.77
75	shopvac	4.38	0.91	2.45	1.04
76	sledgehammer	4.26	0.83	1.83	0.76
77	staplegun	4.24	1.08	2.26	0.89
78	stepladder	4.62	0.82	1.76	0.82
79	swissarmyknife01a	4.36	0.88	2.05	0.85
80	toolbox02	4.36	0.79	2.07	0.95
81	weedwacker	3.68	1.37	2.45	0.90
82	blender	4.52	0.67	2.29	1.02
83	butterknife	3.73	1.30	1.86	1.00
84	coffeemaker	4.63	0.70	2.61	1.16
85	cookingpot	4.57	0.74	1.98	0.98
86	cowbell	3.79	1.34	2.30	0.99
87	electricmixer	4.33	0.93	2.55	1.02
88	freezer02	4.38	0.96	1.79	1.12

89	gasburner	4.17	1.05	2.54	1.16
90	gravyboat	4.10	1.03	2.36	0.93
91	paninigrill	3.74	1.06	2.48	0.97
92	handblender	4.07	1.16	2.40	1.04
93	huntingknife	4.19	0.94	2.12	1.11
94	martiniglass01b	4.60	0.59	1.98	1.16
95	muffintray01	4.69	0.60	1.90	0.88
96	stovetop	4.59	0.67	2.05	1.02
97	saladspinner	3.79	1.22	2.45	1.06
98	minifridge	4.45	0.80	1.93	0.97
99	cutleryset	4.33	0.87	2.57	1.15

ADATLAP[1]
a doktori értekezés nyilvánosságra hozatalához

A doktori értekezés adatai
A szerző neve: NAGY MÁRTON GÁSPÁR
MTMT-azonosító: 10031578
A doktori értekezés címe és alcíme: Recollection in the light of eye movements – The relational eye movement effect and its role in recollection
DOI-azonosító[2]: 10.15476/ELTE.2017.114
A doktori iskola neve: Pszichológiai Doktori Iskola
A doktori iskolán belüli doktori program neve: Kognitív Pszichológiai Program
A témavezető neve és tudományos fokozata: Dr. Király Ildikó, habilitált egyetemi docens
A témavezető munkahelye: ELTE PPK
II. Nyilatkozatok

1. A doktori értekezés szerzőjeként[3]
a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom a Pszichológiai Doktori Iskola hivatalának ügyintézőjét Barna Ildikót, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.
b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[4]
c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (dátum)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[5]
d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követőn egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.[6]

2. A doktori értekezés szerzőjeként kijelentem, hogy
a) az ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;
b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.
3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.
Kelt: 2017.07.11

a doktori értekezés szerzőjének aláírása

[1] Beiktatta az Egyetemi Doktori Szabályzat módosításáról szóló CXXXIX/2014. (VI. 30.) Szen. sz. határozat. Hatályos: 2014. VII.1. napjától.
[2] A kari hivatal ügyintézője tölti ki.
[3] A megfelelő szöveg aláhúzandó.
[4] A doktori értekezés benyújtásával egyidejűleg be kell adni a tudományági doktori tanácshoz a szabadalmi, illetőleg oltalmi bejelentést tanúsító okiratot és a nyilvánosságra hozatal elhalasztása iránti kérelmet.
[5] A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a minősített adatra vonatkozó közokiratot.
[6] A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a mű kiadásáról szóló kiadói szerződést.