Dissertation

zur Erlangung des Grades

einer Doktorin der Philosophie

der Philosophischen Fakultät lll

der Universität des Saarlandes

Memory and Metamemory of Word-pair Learning

An electrophysiological investigation

vorgelegt von

BAI, Cheng-Hua

aus Yungho (Taiwan)

Saarbrücken
2015

Dekan:

Prof. Dr. Roland Brünken

Berichterstatter:

Prof. Dr. Axel Mecklinger

Prof. Dr. Hubert Zimmer

Tag der Disputation:

31 07 2015

Memory and Metamemory of Word-pair Learning

An electrophysiological investigation

BAI, Cheng-Hua

The research reported in this dissertation was carried out at the Experimental Neuropsychology Unit, Department of Psychology, Saarland University, Saarbrücken, Germany, with financial support from German Research Foundation (DFG) in the first three years & from Doctoral Scholarship at Saarland University (Promotionsstipendien der Universität des Saarlandes) in the fourth year.

List of Content

LIST OF TABLES

LIST OF FIGURES

Abstract

This dissertation investigates two factors related to human learning and memory processes. The first factor is the testing effect: in comparison to repeatedly study learnt materials, repeated testing is known to be beneficial for future remembering. According to the episodic context account of the testing effect, this beneficial effect of testing is related to a process which reinstates the previously learnt episodic information. The second factor is the delayed judgment of learning (JOL) effect: the metacognitive judgment is more accurate in predicting later remembering when there is a temporal delay between the learning and the judgment than without. Following the cue-utilization account, the delayed JOL is based on the accessibility of the learnt materials whilst the immediate JOL is based on item-characteristics. It is hypothesized that there is a similar retrieval process engaged during testing and/ or during a delayed JOL which enhances the memory performance or the prediction of its accuracy. The neural correlates of testing and JOLs were examined using the electrophysiological method. In the current study, word-pair learning was used as an example of learning and memory: Swahili-German word pairs in the testing experiment (Chapter 2) and German-German semantically-related word pairs in the JOL experiment (Chapter 3). The results showed that the event-related potential (ERP) of the immediate retrieval at testing resembles the ones of the subsequent remembering. However, the ERP of the delayed JOL did not show a similar pattern. In summary, the current data did not support the hypothesis that the similar mechanisms underlie the testing and the delayed JOL effects.

Keywords: Testing effect, Judgment of learning, ERP, Reinstatment, Memory retrieval

Chapter 1

Introduction

Objective

Learning is a vital characteristic of human intelligence. This PhD project aims to investigate two factors which enhance or predict memory performance during learning. These factors have generally been approached independently in the memory literature but this PhD project will address, among other issues, whether the two are supported by common neural mechanisms.

The first factor is the testing effect. Testing during learning is a common method in the classroom to assess students' performance. It can serve as an evaluation for the students and to allow teachers to provide feedback, as well as to adapt the pedagogical method for individual students. This conventional view on the role of testing in a learning process is rather passive. Applied experimental psychology studies have revealed that testing does not only play a role in assessing learners' performance, but also serves an active role in making learning more effective and this advantage of testing is labeled "the testing effect" (Butler & Roediger, 2007; Roediger & Karpicke, 2006).

The second effect of interest is the "judgment of learning" effect, which refers to the fact that the metamemory judgments during learning might have an impact on later memory performance (Dunlosky & Nelson, 1992; Efklides, 2008; Veenman, Hout-Wolters, & Afflerbach, 2006). A theoretical question raised here is that whether this "judgment of learning" effect is comparable to the testing effect in their underlying mechanisms. For instance, a hypothesis is that the requirement to judge an item's later memorability might function as an active mediator in enhancing memory (and under some circumstances also require actual retrieval of an item) rather than play a passive role in evaluating students' performance. Or alternatively, the judgment of learning should not only be considered as a memory process, but also as an operation on the metamemory level which monitors the mnenomic proccesses. To date, few studies have investigated the underlying mechanism of these beneficial effects in learning.

In the current project, a series of experiments employing behavioral and neuroimaging methods are conducted to investigate the active roles of testing and judgment of learning in word-pair learning. Behavioral measures across experiments will outline the impact of testing and judgment of learning on later memory performance. The highlight of the current project is to use the event-related potential (ERP) method to examine the neural correlates underpinning these learning-enhancing effects. It is hypothesized that the mechanisms underlying the testing and judgment of learning effects are related to a retrieval process and this process enhances memory strength, which allows information to be more retrievable in the future. The high temporal resolution of the ERP technique can be served as the ideal method to answer a question as such. The current PhD project will provide insights into the mechanisms of these learning-enhancing effects so that these methods can be improved in the applied settings.

Theoretical background

Declarative memory

In Squire (1992)'s model of memory classification, the episodic memory and the semantic memory are both considered as explicit memory, also known as declarative memory, which can be retrieved with intention.

Figure 1.1. A schema illustration showing the taxonomy of declarative memory, episodic memory and the dual processes of the episodic memory based on Squire's mode of memory system.

In Squire (2004)'s model of the multiple memory systems, the declarative memory is defined in contrast to the non-declarative memory: "Declarative memory is representational. It provides a way to model the external world, and as a model of the world it is either true or false. In contrast, non-declarative memory is neither true nor false. It is dispositional and is expressed through performance rather than recollection. Non-declarative forms of memory occur as modifications within specialized performance systems. The memories are revealed through reactivation of the systems within which the learning originally occurred."

Episodic memory

Under the category of declarative memory, memories for specific content of individual events and episodic are termed "episodic memory" which, in Endel Tulving's proposal, is different from "semantic memory." Episodic memory is more related to personal experiences on encoding an episode with time and space information, whilst semantic memory is more related to an encyclopedia type of stored factual knowledge about the world. Taking word learning for example, the situation of which someone learns a word is more related to an episodic memory and the word itself is more like a semantic memory. Whether the two memory systems fall under the category of declarative memory system or not is an ongoing debate. Tulving and Markowitsch (1998) proposed that the episodic memory is a subsystem within the declarative memory system and that it possesses functions which are unique among the other memory systems.

The unique function of the episodic memory system is that human episodic memory allows us to travel through time. This ability of encoding and retrieving an episode enables us to recall the past and to re-experience the episode through retrieval process. The studies in Chapter 2 investigated whether the testing effect is related to this concept “episodic re-experience during retrieval”; this is further discussed under the sub-section “Retrieval and Computing” in this Introduction. Whether or not this ability of time traveling is unique to human beings is out of the scope of the current study, for animal experiments on time traveling, one can refer to the studies conducted by Nicola Clayton and colleagues (such as Clayton and Dickinson, 1998; Griffiths, Dickinson and Clayton, 1999) which presented behavioural data from food-storing jays and claimed that birds also have episodic memory.

Encoding and Storing

Human learners encode information in a selective fashion. An early meta-analysis on memory for schema-relevant memory formation (Rojahn and Pettigrew, 1992) has found a slight overall memory advantage for schema-inconsistent information. The direction of the effect was found to be depending on the type of measures. Among all the measures, the results shown recognition tests corrected for guessing and recall tests reveal consistently better memory for schema-inconsistent information. In the contrary, recognition memory tests uncorrected for guessing consistently uncover better memory for schema-consistent information. In a more recent review, van Kesteren and colleagues (2012) proposed a framework on schema and novelty on memory formation. Their findings supported the idea that in some cases, when the formation is consistent to prior knowledge (a schema), it was better remembered, while in other cases, the inconsistent (i.e. novel) information was better remembered.

After encoding, the information is processed through a consolidation process. This consolidation process enables information to be stored into a more stable form for later retrieval. Depending on different models, the process of consolidation is thought to be supported by different brain structures over time. According to the standard model of systems-level memory consolidation, this memory formation process engages the medial temporal lobe structures and then with time the coded information is transferred to the cortical level. The hippocampus first fast selects information by pattern completion and pattern separation (O'reilly and McClelland, 1994), and then with time, the information becomes more dependent on the cortex and less dependent on the hippocampal structures (Squire and Alvarez, 1995; Frankland and Bontempi, 2005; Morris, 2006). That is, the memory is re-organized over time. The hippocampus engages a fast synaptic synthesis process to reconstruct the synaptic connections, whilst on the cortical level, the process is slower and gradual, which plays a role to stabilize memory.

Alternatively, the multiple trace theory proposes that there is no prolonged consolidation process from the fast hippocampus-complex to the neocortical component of the memory trace. The binding between the MTL structure and the neo-cortical level interacts given the strength of the memory trace. Each one of the various regions within the medical temporal lobe structure provides a specific function to process and to consolidate a different type of memory (Nadel and Moscovitch, 1997, Moscovitch and Nadel, 1998). Therefore, the retention and recollection of the episodic memory under this model also depends on the hippocampal-complex (Moscovitch et al., 2005). My current project is related to the hippocampal-cortical network because we are interested in a retrieval processes and its relevance to later memory performance over time. However, the techniques I employ in this project will mainly reveal the temporal envelope of the processes in interest. The deep source neural activities from the hippocampus are usually inferred by the activities detected at the medial temporal regions or the parietal regions.

Encoding: Subsequent memory effect/ differential memory effect

The subsequent memory paradigm refers to the method which measures the neural responses to events at encoding and classifies these events by the behavioural measures at retrieval according to whether the events were later remembered or forgotten. During encoding operations, there is information which was transformed into memory and will be remembered and there is information which was forgotten. By monitoring the neural activity at encoding and correlating it with the behavioural measurements on whether it will be later remembered or forgotten, the formation of memory can be studied (Davachi, Maril, & Wagner, 2001; Paller & Wagner, 2002; Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980; Wagner, 1998). The research question in the current project focuses more on the memory retrieval process rather than the processes engaged in the encoded stage. Or one can also view the studies in the current project as investigating the interaction between the encoding and retrieval processes at learning phase. The proposed retrieval processes engaged during learning might better later behavioural performance.

Retrieval and computing

"The question of central interest in episodic memory has to do with what the rememberer remembers, the content of his recollective experience." -- Tulving, 1983

Retrieval is a series of processes engaged during a memory test in order to retrieve encoded information for achieving a given task goal. There are different ways to look at the retrieval processes. One is to classify a retrieval process related to a different class of action, such as 'mode', 'effort', 'success' and 'orientation'. These classes are determined according to the task goal or memory test performance. Furthermore, each type of retrieval process classification can be mapped onto activities associated with specific brain regions (Rugg & Wilding, 2000). Alternatively, retrieval process could be viewed under a general memory mechanism which is initiated by different kinds of retrieval cues (Neath and Surprenant, 2005).

Interaction between encoding/retrieval processes

According to Tulving, it is the combination of autonoetic consciousness and episodic memory that allows an individual to engage in mental time travel. Humans are said to have the ability to 're-experience, through autonoetic awareness, previous experiences as such, and to project similar experiences into the future' (Tulving, 1999). Wheeler and colleagues (1997) has proposed a theory of episodic remembering where a specific operational definition for a retrieval process was given as "When a rememberer mentally travels back in subjective time to re-experience his or her personal past, the result is an act of retrieval from episodic memory."

The nature of episodic memory is constructive

Schacter and Addis' (2007) constructive stimulation hypothesis proposed that the operation of memory retrieval is not just to load information which was encoded truthfully; instead, memory retrieval engages a process which reconstructs the encoded information at retrieval. In other words, to retrieve an episodic event is not just to copy and then paste the answer at retrieval. It is more to activate relevant encoded information and use them for a specific retrieval goal. The memory system is rather flexible, which allows information to be recombined and to extract useful information for completing a given task. This constructive and flexible nature of the episodic memory system might be contributing to the word-learning process. As indicated in Duff and Brown-Schmidt (2012), the hippocampus is involved in multimodel language learning due to the following functions associated with the hippocampus: a. different representations are integrated flexibly b. an online-processing of rich representations from multiple domains c. relational binding.

In my view, how to select the information to be encoded might be related to the evaluation of the future relevance of the input information. Studies have shown that recalling past experience (episodic memory retrieval) shares neural networks with imagining or predicting the future (episodic future thinking).

Hassabis and Maguire (2007) define these two cognitive functions as

Episodic memory recall – vivid recollection of personal past event
Episodic future thinking – envisaging a plausible personal future event

A core-network related to memory and imagination has been identified in the following brain regions: lateral and medial prefrontal cortices, precuneus, posterior cingulate, and retrosplenial cortices, lateral and medial temporal areas, including parahippocampal cortex and the hippocampus (D'Argembeau and Van der Linden (2004), Atance and O'Neill, 2001, Hassabis, Kumaran, and Maguire, 2007, Mullally and Maguire, 2014, Schacter, Addis and Buckner, 2007). These findings provide evidence for the constructive stimulation hypothesis for the episodic memory showing that the core network of brain regions that includes the hippocampus supports both mnemonic and simulation-based processes.

How is the understanding between the core-network between memory (past experience) and imagination or prediction (future thinking) related to the research questions in this dissertation work?

To memorize events is related to the future use of information. The adaptive perspective of constructive stimulation hypothesis is related to a benefit of testing over the restudy condition. A fMRI study investigating the testing effect has found greater hemodynamic activity in the striatum region for the practice effect (Van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013). The researchers suggested that the reward circuit activated might be related to an adaptive aspect. In the testing condition, the learners understood that there will be no chance to see the target word later. Therefore, they have to simulate a future event of which they will need to recall the target word given the cue word. The quality of memory formation is also related to reward. Studies have shown that recollection and source memory are better for reward-predicting than non-reward predicting stimuli. The reward system in the brain is modulated by dopaminergic inputs in the mid-brain area, including the substantia nigra. It is hypothesized that the long-term potentiation in the hippocampus can be enhanced and prolonged by the reward system and consequently enhances memory formation and consolidation (Wittmann et al., 2005). Subsequent memory performance is also modulated by the emotional context: free recall performance is better in positive emotional context than in neutral or negative context. In addition, recall in the positive or negative contexts is associated by different neural substrates. The activties in the interior frontal gyrus predicts recall in general. On top of it, the activation in the right anterior parahippocampal ares and right fusiform gyri are associated with the positive encoding context whilst the activation in the amygdala is associated with the negative encoding context (Erk et al., 2003).

The closer the event for encoding is to the later future event, the better the memory performance would be. In addition, in the case of judgment of learning, the judgment has to be made upon a process which is closer to a simulation of future event. The temporal distance between learning the word pairs and judging the word pairs play a role in the accuracy of judging the memory performance at a future time point. A metamemory judgment is then reversely correlated to the temporal distance.

Subsequent memory effect measures the memory formation for later memory performance. Retrieval is thus intertwined with past-and-future thinking. When learners learn an associative word pairs as an incidental memory task, they are aware of the fact that what are encoded at the learning phase will be tested in the future. How do the learners use the retrieval cue to encode the word pair and to build up the memory trace which can be used in a future memory task can be an analogue as the relationship between encoding and future thinking and is therefore related to the present research questions.

‘Reinstatement’ and ‘Reactivation’

In an early MEM framework (Johnson, 1992), the role of ‘reinstatement’ and ‘reactivation’ in a process of episodic memory retrieval was outlined. MEM suggested that these two type of repetition of representation in the memory content both activate and strengthen item and associated relationship between cue-and-target. The differences between ‘reinstatement’ and ‘reactivation’ were the processes involved:

‘reinstatment’ is repetition through perception;
‘reactivation’ is repetition through regenerating a previous perception or reflection (p.273).

Reactivation can be viewed as information encoded during memory formation is then reactivated during retrieval. The transfer-appropriate processing account is similar to the reactivation account which proposed that memory performance is influenced by the overlap between processes at encoding and at retrieval (Graf and Ryan, 1990). How the information can be retrieved is of relevance on how the information is encoded. The concept of context-dependence when learning is also relevant in this context (Gooden and Baddeley, 1975). Context-dependence learning suggests that when the memory is formed in the water, for example, the learner can better recall it in the same context in the water than in another environment.

Retrieval-based learning: an episodic context account

In a recent review, Karpicke, Lehman and Aue (2014) analyzed the conditions under which retrieval practice has shown a memory advantage compared to restudy. Several principles of retrieval-based learning were summarized in the ensuing episodic context account. The core assumption is that retrieval practice places participants into a retrieval mode in which they attempt to reconstruct the past and to re-instantiate the temporal context. This retrieval process causes the item to be updated within its context, such that it may become associated with multiple context cues after extended retrieval practice, making it more retrievable as a consequence. Karpicke and colleagues (2014) provide empirical evidence for this hypothesis by showing that source memory decisions during testing led to better performance in a final recall test than old/new recognition memory decisions, whereas both conditions were better than an elaboration (forming images or generating word associates) condition. According to some dual process models of recognition memory, the reinstatement of an item in its context relies on recollection (Diana, Yonelinas, & Ranganath, 2007). If the effect of retrieval practice on later memory performance improvement is due to re-instatement of the prior episodic context, we should be able to find the neural correlates of recollection at the time point when retrieval practice is actively engaged.

Episodic memory and language processes

In the current work, there are two type of word-pair learning. One is to learn foreign vocabulary by associating it with a corresponding translation, and another is to remember semantic-related pairs in the first language. Both cases are comparable to real learning experiences in daily life. What was shared between both of them was that in both studies, printed words were used as experimental stimuli and the tasks participants had to engage themselves were language learning or word-pair association learning tasks. Here in this context, the language materials were used as an access to understand human memory system, rather than on the level of language processing. Language and memory are two big fields of studies under the umbrella of cognitive science. Throughout decades of researches on the two fields related to cognition, knowledge about how human beings process linguistic information or how the memory system is like has increased. However, the more evidence the scholars gather, the more we understand that the human cognitive system is highly complex. Due to the high complexity in either linguistic or memory research, often researchers avoid complicating the questions even more by taking both into account at the same time. For instance, in the Universal Grammar, Chomsky proposed that children are born with Language Acquisition Device which allows them to acquire human language in the early stage in life. This theory has been controversial because it took a strong view which claimed that some aspects of language are not learnable, but innate (Chomsky, 1957). In contrast, a usage-based model of language acquisition, such as the constructive grammar, argued that the skills required for language learning are not domain-specific, but are generalized across cognitive domains (Tomasello, 2000).

Another aspect where language and memory systems are intertwined is within the framework of declarative memory system proposed by Tulving. As mentioned above, the declarative memory system consists of episodic and semantic memory. There are ample debates on how to distinguish episodic memory from semantic memory and their organizational relationship. Independently from how they interact with each other, what is clear here is that linguistic components should be included in the discussion for understanding the declarative memory system in a more complete way.

Despite the fact it is highly likely that mental lexicon storage and computation interplay with general memory function, there is less effort paid in the field between psycholinguistic and memory research to consider the two complex systems via an integrated view. Language processing relies arbitrary mapping among different levels of information, such as morpheme-to-phoneme mapping, word form to meaning mapping. The information flow among multiple level of linguistic representation is bi-directional. Information could flow from the word form level to the semantic level and the other way around during language comprehension or production. In addition, language processing involves other non-linguistic informations, including co-speech gesture, eye gaze and many other micro-gestures. It is beneficial for the cognitive research to consider different modalities when considering how the system functions. Gupta (2001) proposed a model which explicitly points out that language learning demands the confluence of memory systems. Neuropsychologically, Melissa Duff and her lab members have been conducting studies with hippocampal-lesion amnesic and control participants to show that language processing and communication relies on many functions associated with the hippocampus, including relational binding, flexible integration of difference representations, and online processing of rich representation from multiple domains (Duff and Brown-Schmidt, 2012; Rubin et al., 2011; Duff, Gupta, Hengst, Tranel and Cohen, 2011; Duff et al., 2006). It is very likely that word pair learning and other types of associative memory rely on some overlapping mechanism.

Neural correlates of declarative memory processes

Hippocampus and binding

The ventromedial prefrontal cortex (vmPFC) couples with the hippocampus during novel experiences. The hippocampus binds an episodic event which contains item information, temporal and spatial information. The events are coded in the sequential order as in the experience itself. The binding process binds distinct episodic representations by abstracting common semantic information linking overlapping representations to construct a relational memory network.

Nyberg and colleagues (2010) have found that the frontal and parietal cortices, but not the hippocampus support mental time travel. Likewise, Andrews-Hanna and colleagues (2010) showed that the hippocampus and medial temporal lobe (MTL) support imagining scenes, whereas other brain regions involved in episodic retrieval were more related to self and with time. The brain areas which are activated during episodic memory retrieval of autobiographical memory include the bilateral hippocampus, parahippocampal gyrus, retrosplenial, posterior cingulated and posterior parietal cortices and medial prefrontal cortex (a meta-analysis by Svoboda et al., 2006) .

While participants retrieved autobiographical memories, frontal region of the brain exhibited marked negative DC shifts, and then as memory was formed and being on hold in mind, marked negative shifts were detected over posterior temporal and occipital regions (Conway, Pleydell-Pearce and Whitecross, 2001). By observing the slow waves, this study provided evidence for the associated regions when one remembers a life event.

Retrieval from the episodic memory

Retrieval success

Retrieval success effects were discussed in a review paper by Rugg, Otten and Henson (2002). Retrieval success effect is defined as “the neural correlates of retrieval and subsequent processing of recently acquired information. In order to access the successful retrieval, “remember-know” paradigm and “source memory” tasks were developed to disentangle a recollection process from other sorts of memory. Different neuroimaging techniques were employed to investigate this topic from different aspects. The ERP technique informs about the neural signature of “recollection” whilst the PET and fMRI techniques informs about the brain regions related to the processes. Briefly speaking, these studies have shown that a recollection process is supported by brain regions, including the left anterior temporal cortex and the lateral and medial parietal cortex.

According to the dual process model of recognition memory, there are two distinct processes at retrieval, which are familiarity and recollection. Recollection requires the retrieval of contextual information from the episodic memory and is a slower process than familiarity (Yonelinas, 2002; Brandt et al., 2001). There is a family of ERP components which are elicited by different retrieval processes. These retrieval-related ERP components can be sorted by their function or their temporal relevance to the stimuli onset. A paradigm often used to explore the retrieval process is the recognition memory task where participants learn information at the learning phase and where in the testing phase, new information is added additionally to the old information. Participants have to discriminate the old information from the new information. Consequently, the ERPs to old items in standard recognition memory tasks are compared to ERPs to new items is called the ERP old/new effects (Vilberg & Rugg, 2009; Rugg & Wilding, 2000).

ERP signature of familiarity-based recongition: Early mid-frontal effect (300-500 ms)

An early ERP signature according to the dual process model is the familiarity signal onsets at around 300 ms after the stimuli onset. Familiarity-based recognition is underlying a feeling of knowing where no contextual information is required to be recoverable.

ERP signature of recollection-based recognition: The parietal old/new effect (500-800 ms)

The ERP signature of successful recollection was often observed at the left parietal site in a recognition memory task. An item which was learned before was categorized as “old” in contrast with a new item in the test phase in a recognition task. The ERP signal for old items indicates a recollection process of which the encoded information was retrieved. This old/new effect was often found at around 500 to 600 ms after stimuli onset (for reviews, see Friedman & Johnson, 2000; Mecklinger, 2000; Michael D Rugg & Curran, 2007). The effect is larger when elicited by items which were recognized as ‘remembered’ rather than ‘known’ (Smith, 1993), by items attracting correct rather than incorrect source memory judgments (Wilding & Rugg, 1996), and by items undergoing deep as opposed to shallow encoding (Rugg et al., 1998). This ERP index of recollection is observed from 500-800 ms post-stimulus and this effect is often observed over left-parietal site (Rugg & Curran, 2007; for reviews, see Friedman & Johnson, 2000; Mecklinger, 2000; Rugg & Curran, 2007). In addition, studies also showed that the amplitude of the left parietal old/new effect is modulated by the amount of information retrieved (Vilberg, Moosavi & Rugg, 2006).

The parietal effect reflects cortical activity which mediates ‘reinstatement’ or ‘reactivation’ of the retrieved information and these effects were proposed to be mediated by the hippocampus in the medial temporal lobe structure (Rugg et al., 1998). Taken together from studies, the ERP parietal old/new effect might be correlated with the effects found in fMRI studies which investigated the recognition memory (Wagner, Shannon, Kahn, & Buckner, 2005).

ERP signature of recollection process 2: the right frontal old/new effect

Another ERP effect associated with recollection process is the right frontal old/new effect which onsets later than the parietal old/new effect. This right frontal effect was observed in studies which investigate source memory in a recognition task. The retrieved information had to be evaluated after the actual retrieval process. This post-retrieval process involves an operation where the retrieved information had to be maintained, represented and monitored with respect to the task goal (Rugg and Allen, 2000; Rugg and Wilding, 2000).

POST-retrieval: memory monitoring and searching for attributes

Another ERP effect that is frequently reported in recognition memory studies is the late posterior negativity (LPN). The LPN is a late and posteriorly distributed ERP component that is observed mainly in source recognition studies (Johansson & Mecklinger, 2003). It onsets around the time recognition decisions are given and is thought to reflect the assessment and evaluation of information retrieved from memory in situations in which memory features cannot easily be recovered. The LPN is most pronounced when extended retrieval processing is required, for example in situations in which discriminations between multi-featured memory traces are required (Leynes & Kakadia, 2013) or when the to-be-discriminated memory traces are weak or overlapping (Rosburg, Mecklinger, & Johansson, 2011).

The functional role of LPN on behaviour is not conclusive yet. Some studies have found a positive correlation between the length of reaction time and the amplitude of the negative-going slow wave. However, this positive correlation between RT and the amplitude of the LPN was not found in all studies reviewed by Johansson and Mecklinger (2003, Table 1 shows an overview of the studies). Additionally, this positive correlation was also observed without overt key-press responses (Donaldson and Rugg, 1999). This finding challenges the view that the LPN is related to action monitoring and it is correlated with the RT conclusively. Cycowicz and colleagues (2001) offered an alternative view on the LPN. Their account suggested that the LPN reflects the processes which are related to sensory-specific search. The maximal activation observed at the occipital area might be associated with the reinstatement of the sensory-specific information. A similar argument was supported by studies which showed that a pronounced negative-going slow wave is elicited by different types of information asked in the task: the spatial condition evoked slow waves at parietal area whilst the color condition evoked slow waves at occipital-temporal areas. This search/retrieval of attributes account states that the LPN reflects processes where the prior study episode is reconstructed when task-relevant attribute conjunctions are not yet recovered by the retrieval cue at test phase or need continued evaluation. Taken together, the functional roles of LPN are associated with two types of processes: 1) memory task where there is a high action monitoring demand; 2) a retrieval process which tapped into the source memory information.

The testing effect

The current dissertation project employed the testing effect because the experiment manipulation in the testing effect studies provides an explicit condition – testing which is open for a direct investigation on its neural correlates. Among a pool of theories about the reasons behind the testing effect, the retrieval-practice hypothesis is most in favor due to increasing evidence.

a. Theory of retrieval practice on longer retention and behavioural data

The testing effect is a robust phenomenon in the field of learning and memory which refers to the fact that in comparison to restudying, the act of testing leads to better later memory performance on the learnt materials. Many studies have shown that testing is not only an assessment of learning, but it plays an active role in learning (for a review, see Roediger & Karpicke, 2006). Many experimental behavioral studies have demonstrated that testing is an important factor for learning (Roediger & Butler, 2011; Karpicke & Roediger, 2008). Specifically, once learned materials are tested, there is a higher likelihood that these materials are remembered better than if the materials would have been merely re-studied. In a testing effect paradigm, the term "testing" means testing during learning in contrast to re-studying. Memory performance for both tested and re-studied items is then accessed via the recall accuracy in a subsequent test phase. Carrier and Pashler (1992) found that more tested items were correctly recalled later as compared to a condition in which the items were only re-studied. The testing effect is sometimes found immediately at the learning phase whilst in some studies, the testing effect is found in a delayed test phase. For instance, Karpicke and Roediger (2008) showed that there is no difference between testing and restudying immediately after the encoding phase. However, after one week, students remembered twice as many items if they had retrieved the items than if they merely restudied the items after learning.

Given that the testing effect has been found repeatedly in different experimental designs and that it has a high value in classroom applied settings, the mechanisms underlying this effect are of obvious interest. Currently, most studies suggest that the benefit of testing in comparison to re-studying results from a retrieval process at the testing condition and thus refer to the effect directly as "the impact of retrieval" (Karpicke & Smith, 2012; Carpenter, 2011; Karpicke & Blunt, 2011). However, to our knowledge, only few studies have endeavoured to empirically show a retrieval process at the time point at testing.

Karpicke and Roediger (2008a) demonstrated that the act of testing is the crucial enhancer for a later retention. A word-pair-associate learning task was employed where the learners had to remember 40 foreign words with the corresponding translations. There are eight study and testing periods. Four groups of learners learn all the word pairs and then were tested in the first two periods. And then, the group condition differed by manipulating the items dropping from the previous performance. In a standard condition, learners in group 1 learned all the word pairs again and then were tested again. This procedure repeated. In the second condition, learners in group 2 learned those non-recalled words only and then were tested with all the words. This procedure repeated for three times. Each time, the amount of item learnt decreased. In the third condition, learners in group 3 learned all the word pairs again and then only those ones which were not recalled in the first test period were tested again. This procedure was repeated three times. The items which were tested were decreased in each round. In the last condition, learners in group 4 learned those non-recalled items and were tested with those non-recalled items. The amount of learn-test items decreased over time. This procedure resulted into a total number of trails of 320, 236.8, 243 and 154.8 in the four conditions.

The result showed that learners' cumulative recall performance were equivalent in the four conditions. However, after 1 week delayed after learning, the proportion recalled items on the final test were distinct across groups. Group 1, where learners learned and were tested all word pairs repeatedly showed similar performance as in group 2, where learners learned dropping-from-test word pairs again, and were tested at all word pairs. In the contrary, group 3, where learners were exposed to a similar amount of item trials as in group 2, showed a much lower performance than learners in group 2. The low performance in group 3 was comparable to the lowest performance in group 4. The results showed that repeated testing is the essential reason which prompts memory performance in a longer period of time after learning.

In the operation of repeated studying the learners were repeatedly exposed to the materials. In the contrary, in the repeated testing condition, part of the information was not shown again to the learners. In order to answer what was in the pairs, learners thus had to actively think about the episode which appeared with the cue word. In this view, one could suggest that repeated testing encourages learners to recall the learned information. Thus, the testing provokes processes which are more effortful, and it encourages a retrieval process which might active the existing encoded information, the memory trace is then strengthened, or it could be the case that when learners attempt to retrieval the associated word, additional memory nodes in the network are activated, inhibited or created.

The beneficial effect of testing on memory is thought to arise because of the consequences of repeatedly retrieving information (Roediger & Butler, 2011). In the view of retrieval-practice as the mechanism underlying the testing effect, it is suggested that in the repeated testing condition in comparison to the restudying condition, a retrieval practice is engaged and that this retrieval process makes the memory more durable over time. One idea is that the retrieval of information enhances the elaboration of the memory trace which makes it more likely to be recalled over time. Testing initiates the process of retrieving information from memory, such as semantically-related information between cue and target in associates learning task, and this process leads to further elaboration or strengthening of the memory trace (Carpenter & DeLosh, 2006; Carpenter, 2009, 2011). The elaborative retrieval account could also be interpreted as the creation of additional memory entries by the retrieval process, which increases the variability of the existing nodes. Consequently, the encoded information becomes easier to access (Bjork, 1975; McDaniel & Masson, 1985). A related idea focuses on the effort attributed to retrieving the information in the testing condition, stating that in comparison to a restudy condition, more effort is allocated during testing, which leads to enhanced reprocessing of encoded information (Pyc & Rawson, 2009). The concept of transfer-appropriate processing is another alternative explanation which indicates the shared processing in testing condition and the final memory task which makes the content of memory more recallable in a later time point. According to the transfer appropriate processing account, processing engaged during testing is similar to the one engaged in the subsequent memory test which makes the materials more memorable in the testing condition (e.g. cued recall task at learning and subsequent test phase) (McDaniel, Kowitz, & Dunay, 1989; McDaniel & Fisher, 1991).

Even if this view of retrieval-practice of the repeated testing holds true, one could argue that learners could also force a retrieval process in the restudying condition which is comparable to the repeated testing condition. Learners could activate the pair-associate connection in the network too as they had to do in the repeated testing condition. The translation word is on the display which serves as a direct feedback. It is arguable whether even a comparable retrieval operation can happen in the restudying condition. Even so, the amount of effort and attempt attributed to the testing should still be higher due to the absence of the translation word. If this argument were true, the underlying process of restudying and testing should be quantitatively distinct.

b. Neuroimaging findings of the testing effect

To date, the testing effect has mostly been observed in behavioral studies, whereas the use of neuroimaging techniques to investigate the mechanisms underlying the testing benefit on learning has only begun in the last few years (Eriksson, Kalpouzos, & Nyberg, 2011; Rosburg, Johansson, Weigl, & Mecklinger, 2015; van den Broek et al., 2013; Wing, Marsh, & Cabeza, 2013).

In line with the accounts proposed and supported by various behavioural studies, functional neuroimaging studies have recently provided additional evidence for the notion that testing benefits memory performance because of memory retrieval processes and elaboration of memory traces. One of the studies conducted by Wing, Marsh and Cabeza (2013) used one single learning trial and one single practice block during scan while another study by Van den Broek, Takashima, Segers, Fernandez and Verhoeven (2013) had 3 practice blocks during scan; in addition, all materials were correctly recalled before scan. The timing of the final memory test differs between the two studies: after 24 hour vs. 7 days. Despite of the differences in design, the two studies have shared the logic of using the subsequent memory paradigm to investigate the testing effect as in the present study. In addition, the two studies reported consistent results. Both studies directly measured the brain activities at the practice (testing, restudy) block and correlated its effect with the subsequent memory effect (successfully recalled or not) in the magnetic-resonant imaging (MRI) scanners. A higher recall rate for testing than for restudy condition at final memory test was found in both studies. The subsequent memory performance was then backsorted according to the two practice conditions.

Wing and colleagues (2013) suggested that the testing effect may be resulting from processes that support memory success at encoding (e.g. relational binding, selection and elaboration of semantic information) in addition to retrieval processes like memory search. They found increased activities in the hippocampus, lateral temporal cortex and medial prefrontal cortex in subsequently remembered tested items than the subsequently forgotten tested items. In addition, the subsequent memory effect (SME) in the testing condition is greater that the SME in the restudy condition, which shows a beneficial function effect of testing on subsequent memory performance. Additional connectivity analyses revealed an increased coupling between the hippocampus and the ventrolateral prefrontal cortex, the medial pre-frontal cortex, and the posterior cingulate cortex. The coactivation with the hippocampus and these regions predicted subsequent memory success to a greater extent for testing than for restudied items.

Van den Broek and colleagues (2013) found that the beneficial effect of testing than restudying may arise because of retrieval processes which are related to the elaboration of semantic information and to strengthening the associates, and that the amount of effort attributed may be relevant for the testing effect on subsequent memory success. Increased activities in the left inferior parietal and left middle temporal areas were found to be greater in the testing SME than in the restudy SME. The two regions were found to be involved in the storage and retrieval of lexical at medial temporal gyrus (MTG) and episodic information at inferior parietal lobe (IPL). It was concluded that there is an increased amount of information retrieved during tested than in restudied trials or that more specific search sets are activated in response to a retrieval cue, which makes tested items more memorable than restudied items. Both fMRI studies argue in favor of the retrieval account of the testing effect, in which retrieval leads to a more elaborated memory trace and may consequently enhance the likelihood of subsequent memory success. Additionally, the beneficial effect of testing may also be related to higher effort allocated in the testing condition than in the restudy condition. The brain regions associated with the reward system, such as striatum and midbrain areas, were also found to be more activated in the testing than restudy condition (Van den Broek et al., 2013; Vestergren & Nyberg, 2013; Wing et al., 2013). Critically, activity in left inferior parietal and left middle temporal areas predicted recall in the final memory test in the testing but not in the restudy condition. The activity in the left inferior parietal and left middle temporal areas was modulated by the amount of information retrieved with higher activity during testing of subsequently remembered words than forgotten words. As both areas have been consistently found to be involved in successful memory retrieval (Vilberg & Rugg, 2008; Diana et al., 2007) or in the allocation of attention to retrieved information (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Hutchinson et al., 2014), this study provides additional support for the view that testing involves the reinstatement of a prior study context by enhancing recollective or relational processing.

These fMRI studies thus provide general support for the retrieval account of the testing effect, in which testing should cause retrieval of prior encoded episodes and a reinstantiation of the item in its context. This update of the memory trace during testing may provide additional cues for the final memory test. Comparable electrophysiological evidence is scarce, however, and the current study was designed to address this gap in the literature. Electrophysiological data is likely to be useful for understanding the mechanisms underlying the testing effect not only because of its greater temporal resolution, but also because decades of work using the event-related potential (ERP) technique in recognition tests have revealed a family of old/new effects thought to map onto distinct retrieval processes (for reviews see Friedman & Johnson, 2000; Mecklinger, 2000; Rugg & Curran, 2007). One such effect is usually referred to as the left-parietal old/new effect. Behavioral conditions that modulate recollection also modulate the left-parietal old/new effect and this effect has been shown to correlate with recollection-based memory judgments in item and associative memory studies (Friedman & Johnson, 2000).

Three studies to date have used the event-related potential technique to investigate neural correlates of retrieval processing. These studies directly contrast a testing-like condition with a restudy or control condition, which is comparable to the core design of the current study;

nonetheless, the previous studies were designed for other research questions which did not directly investigate the testing effect. Johansson, Aslan, Bäuml, Gäbel, & Mecklinger (2007) explored the inhibitory account of retrieval-induced forgetting using a paradigm including two types of reprocessing, retrieval and relearning at the intermediate phase. However, unlike a study which is designed for investigating the testing effect, only part of the materials learnt at the encoding phase were assigned to be re-processed the intermediate phase. At test, the results showed that the retrieval reprocessing at the intermediate phase hinders the retrieval of non-reprocessed items among the same semantic category. Due to an incomparable experimental design, this study did not bring understandings directly to the testing effect. Another study designed for examining the retrieval-induced forgetting in recognition memory (Spitzer, Hanslmayr, Opitz, Mecklinger, & Bäuml, 2009) found that the effect of retrieval-practice is associated with a late parietal positivity (LPP) at around 500-750 ms after onset of the stimulus in the recognition test phase. The LPP is significantly larger for retrieval-practiced compared to control items and this LPP is associated with later parietal old/new effect in the literature which reflects a recollection of episodic information. They suggested that this result is in accordance with the account that repeatedly processing information is strengthened by increasing the associations between the cue and target which is also in line with the elaboration account of the testing effect as stated above. In addition to the previous two studies investigating retrieval-induced forgetting, Rosburg, Johansson, Weigl and Mecklinger (2015) examined the ERP correlates of immediate one-time testing in a recognition task. In the design, there were three phases: study phase, immediate test phase (1st test vs. not-test) and the final test phase (2nd test). The results showed a beneficial effect of repeated testing in the final recall accuracy and reaction time for those items which were tested at the immediate test phase in comparison to items which were not tested. For the ERP result, a left-parietal old/new effect was found at 500-700 ms after stimulus onset for hits (correctly source judgments of old items) at the 1st and the 2nd test. And this effect last in the following 700-900 ms time window for hits at the 2nd test phase. From 500-700 ms, the amplitude of the ERPs for hits to tested items at the 2nd test was larger than hits at 1st test and hits to untested items at the 2nd test. From 700-900 ms, the parietal old/new effect was larger for tested than for untested items. This result suggested that the parietal old/new effect found for hits to the tested items at the 2nd test reflects more recollection. The authors suggested that the elaborative account of the testing effect can be supported by this result where there is a larger amount of memory traces which led to better hit rates and larger amplitude in the ERPs.

Metacognition: knowing about knowing

Flavell (1979) defined metacognition as cognition of cognition that serves two basic functions, namely, the monitoring and control of cognition. Nelson (1996; Nelson & Narens, 1994) defined metacognition as a model of cognition that functions at a meta-level; metacognition represents the object level, that is, cognition. Metacognition, through the monitoring function, is informed by cognition and, through the control function, informs cognition. Both definitions underscore the functioning of metacognition at a "meta" level, which means that metacognition is a representation of cognition, and that metacognition and cognition are connected through the monitoring and control functions (Efklides, 2008).

Information about the state of the object-level is conveyed to the meta-level through monitoring processes, while instructions from the meta-level are transmitted to the object-level through control processes. Thus, if errors occur on the object-level, monitoring processes will give notice of it to the meta-level and control processes will be activated to resolve the problem. This seems an elegant and simple model, including both metacognitive knowledge and skills (Veenman et al., 2006).

Metacognitive monitoring during acquisition or retrieval stage can be observed in JOL, FoK and source memory task (Dunlosky & Bjork, 2008 in Handbook of Metamemory and Memory).

- JOL: Judgments of the likelihood of remembering recently studied items on an upcoming test [Acquisition]
- FoK judgments: Judgments of the likelihood of recognizing unrecallable answers on an upcoming test [Acquisition; Retrieval]
- Source monitoring judgments: Judgments made during a criterion test pertaining to the source of a particular memory [Retrieval]

Metacognitive control during acquisition or retrieval stages classified in different types of selection demanded by different task goals.

- Selection of kind of processing: Selection of strategies to employ when attempting to commit an item to memory [Encoding]
- Item selection: Decision about whether to study an item on an upcoming trial [Encoding]
- Selection of search strategy: Selecting a particular strategy to produce a correct response during a test [Retrieval]

Neuropsychological researches on metacognition

Metacognitive regulation involves attention, conflict resolution, error correction, inhibitory control, and emotional regulation (Fernandez-Duque, Baird, & Posner, 2000) and midfrontal brain regions (Shimamura, 2000). Shimamura identified four aspects of executive control – selecting, maintaining, updating, and rerouting. Each of these aspects is associated with different brain regions. Among the four, updating is an aspect which is critical to memory retrieval task, such as in verbal fluency performance, and is associated with activation in pre-frontal cortex. In Metcalft's CHARM model, the relationship between executive control and the frontal lobe function was proposed. Metacognitive evaluation such as feeling of knowing is based on checking the overlapping between new information and the stored information in the memory system. This "novelty-monitoring" operation modulates to which degree the new information is bound onto the episodic memory (Metcalft, 1993).

Judgments of Learning (JOLs) is a kind of metacognitive judgment: A probabilistic judgment of one's performance before, during, or after performance. Throughout studies JOLs are assessed by a range of scales, such as continuous confidence judgment (no confidence – complete confidence); a dichtomous prediction (successful or unsuccessful); at global or local levels. The timing of assessing JOLs also varies across studies. A JOL could be obtained at study phase in forms like quizzes about trivia, paired-associate learning, text-reading, video-watching or at test phase in forms of quiz, cued recall, inference verification test, verbatim recognition test.

(Delayed) Judgment of Learning Effect

The second factor investigated in this project is the judgment of learning (JOL) on learning. Judgments of learning (JOLs) are taken as reference by learners to monitor their own performance and control their learning schedule (Son & Kornell, 2009; Veenman, Hout-Wolters & Afflerbach, 2006). Understanding the underlying mechanism of the JOLs may be helpful for learners to develop a learning schedule. For instance, a JOL with delayed interval between studying and testing improves the predictability of the JOL on later memory performance. This is referred to as the delayed JOL effect (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). Participants first learn word pairs and are then asked "How confident are you that you will be able to recall the second word when prompted with the first?" (adapted from Nelson & Dunlosky, 1991, p. 268) using a scale from very confident to not at all confident. Furthermore, studies have shown that a JOL made after a delayed interval between learning and JOL provides a more accurate prediction of later memory performance than immediate JOL ratings (JOLs) (Keleman & Weaver, 1997; Nelson and Dunlosky, 1992).

The monitoring-Dual-Memories (MDM) principle explains the low accuracy for the immediate JOL rating by the existence of noise from the short-term memory which makes the monitoring of self later memory performance difficult. In addition, later memory performance is based on memory retrieval from the long-term memory. When participants monitor both their short-term and long-term memory at immediate JOL trials, they could not be as accurate as a JOL rated at delayed trials (Nelson & Dunlosky, 1991).

What is the basis of a judgment on later recall? In an early study, Arbuckle and Cuddy (1969) argued that if memory trace is like other type of signal inputs, one should be able to assess the strength of memory trace. Alternatively, the likelihood of recall accuracy can be made during study when the information was first processed. Participants had to make an appropriate mapping on the likelihood scale. The third possibility is the subjective knowledge on the materials to be remembered itself (King, Zechmeister, & Shaughnessy, 1980).

Assessment of retrieval success proposed by Spellman & Bjork, 1992 is an appealing hypothesis to explain the basis of the delayed JOL effect in my view. They suggested that one strategy to make a JOL is to use the cue to recall the target and by doing so, one can decide whether recall will be successful or not. In line with the findings in the testing effect, spacing effect and generation effect, we know that retrieval practice increases later memory performance and the time between learning and testing are crucial for later remembering. In the scenario where a word pair is learnt and then has to be judge immediately, the likelihood of remembering is then underestimated. This is because after the assessment of JOL or with the time being, the memory strength for this word pair is likely to increase. One can relate this hypothesis also to retrieval attempt. If the assessment of JOL is to use the stimuli to retrieve the response and then to rate the later likelihood of remembering, independently from the recall accuracy at the moment when JOL is assessed, this retrieval-attempt will increase the likelihood for this word pair to be remembered in a later memory task.

A latest account on the basis of JOL is the cue-utilization view (Koriat, 1993, 1995, 1997; Kelemen & Weaver, 1997). Koriat (1997) suggested that the JOLs assessed at different time points rely on the different kinds of information available at the time point when a judgment is made. The different kinds of information here in this context are called "cues" by the author, which is different from the retrieval 'cue' we have been referring to in the context of episodic memory retrieval. When making a JOL immediately after studying, the JOL is based on information which is related to item-characteristics, such as word frequency whereas when making a JOL with some delays after initial studying, the JOL is based on learners' assessment on whether they can remember the items at that point and this checking process induced by the JOLs can better predict the likelihood of which the items will be remembered or not. A successful recollection at this time point might enhance the likelihood that the item is also remembered at a later time point.

Although there are theoretical proposals on the basis of JOL, there is a lack of neuroimaging studies investigating the underlying mechanism of JOL. Spellman and Bjork (1992) proposed that the reason why a JOL made after some time interval (delayed JOL) is more accurate than a JOL made immediately after the learning phase (immediate JOL) is because the retention interval alters what is actually assessed during the JOL. A delayed JOL is presumably more related to the access of retrieval-ability of the learned information and it alters the accessibility of the encoded information so that a delayed JOL has a higher accuracy in predicting later performance (Koriat, 1997; Nelson and Dunlosky, 1992; Spellman and Bjork, 1992). If the delayed JOL effects is supported by a retrieval-related process or not is an empirical question which will be examined in Chapter 3.

ERP correlates of the JOL effect

Two studies using ERP technique to investigate the neural correlates of the JOL effect were reported by Skavhaug and coworkers (Skavhaug, Wilding & Donaldson, 2010, 2013). The earlier one examined the correlation between JOLs and successful encoding whilst the later one studied the correlation between JOLs and retrieval processing at test. The findings in the first study supported the previous findings which indicated that the metacognitve assessment shares overlapping processes with the ones engaged in successful encoding, but the metacognitive assessment cannot be reduced to the processes engaged in successful encoding. The specific temporal information revealed that JOLs and successful encoding are not correlated in later time window. This study is relevant to the current study which will be illustrated in Chapter 3 in the following aspects. First, this 2010 study is the first one which investigated the neural basis of the JOLs. Therefore, the current JOL ERP study adopted its experimental design. Second, their finding confirms that JOLs engage processes additional to the memory operation. In addition, at the crucial time window 550-1000 ms, the processing involved for later memory success is comparable with the ones contributing to JOLs effect. This finding showed that the basis of JOLs is related to the ones engaged for subsequent memory effect.

The later study by Skavhaug and colleagues (2013) is most relevant to the current study. In this study, the researchers investigate the correlation between JOLs and memory accuracy in a recognition memory task. They have found that the JOL effect is more related to a recollection process rather than a familiarity process during memory retrieval. Their hypothesis was that if the basis of JOLs is based on the evaluation of cues present at the time of metacognitive assessment (cue-utilization account: Koriat, 1997), greater predictive performance should be observed in the situations where the encoded information is reinstantiated at test. This study is relevant to the current study in the following aspects. First, this study directly investigated the correlation between the JOLs and the retrieval performance at test. Second, it shares a comparable theoretical account with the present study. The present study proposed that the higher predictive value in the delayed JOL condition is related to a recollection process. More specifically, this recollection process engages the reinstatement of the contextual information.

In this section, I illustrate the design of these two studies and summarize the major research questions, findings and interpretations.

The two ERP studies mainly tackle the correlation between JOLs and memory performance. In both studies, participants learned word pairs and immediately made a JOL rating. Items were categorized into high confidently-rated items “high JOL” and low confidently-rated “low JOL.” The design at the study phase was identical for both studies (Figure 1.2).

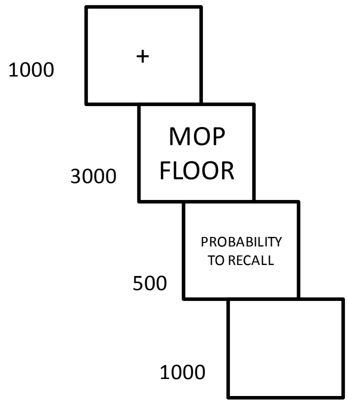


Figure 1.2 The procedure in the STUDY phase (Skavhaug et al., 2010 and 2013). A 1000 ms fixation cross signaled the start of a trial and the participants were asked to focus on the central of the screen. A pair of English related words was then presented for 3000 ms followed by a prompt screen asking for a 1~5 point rating of JOL. A blank screen for 1000 ms was served between the trials.

Unlike the procedure in the study phase, the procedures in the TEST phase were distinct in the two studies. The general presentation time was identical (see Figure 1.3). However, the specific memory task was more complicated in the early study (2010) than in the later study (2013). In the early 2010 study in which the JOLs and successful encoding were investigated, after the presentation of the cue word, the participants had to first make a recognition memory judgment (old or new) and then in the case of identifying an item as old, they were asked whether they could recall the items or not. If yes, they were asked to recall the word. This procedure is partially adopted in the current study. There were some disadvantages due to this complex test structure which I will discuss later in Chapter 3 and 4.

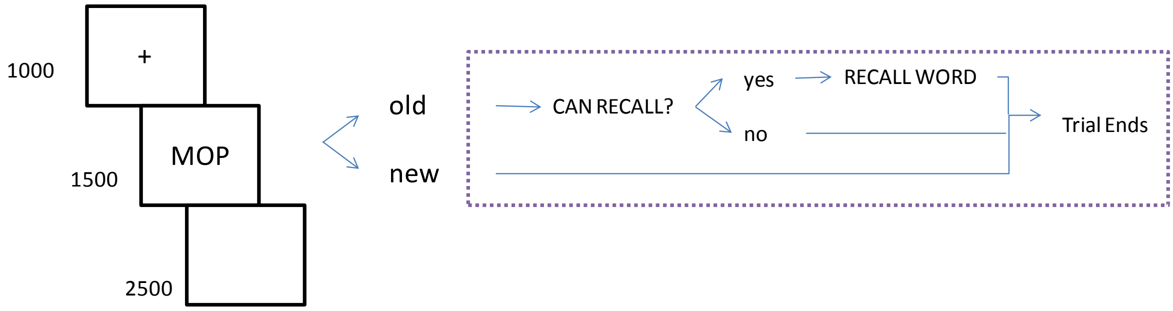


Figure 1.3 The procedure in the TEST phase (Skavhaug et al., 2010, 2013). Participants first focused their attention at the 1000ms fixation cross. A word which was presented on the top position at the study phase is now displayed at the center of the screen. The participants were asked to make an old/new judgment to differentiate items which were presented in the study phase (old items) from those which were not (new items). A blank screen for 2500 ms followed. The response window is in total 4000 ms. In the 2010 study, additional recallability judgment and a cued recall test requested after the old/new judgments as illustrated in the dotted-purple box.

The 2010 study has shown that the neural correlates of JOLs (high versus low confident) and Subsequent Memory Effect (later remembered versus later forgotten) overlap. A similar topographical distribution for these two effects was found at a time window 550-1000 ms from stimuli onset which is associated with the subsequent memory effect. However, the JOLs effect is not restricted to the processes engaged for later memory success. In a later time window, 1300-1900 ms the topographical map of the JOL effect was distinct from the one of the Subsequent Memory Effect. These findings respond to the previous findings that a. JOLs engages additional operations on top of the ones engaged for successful encoding b. JOLs and Memory are dissociable.

Different from the 2010 study, in the 2013 study, the EEG was time-locked at the onset of cued word presentation onset at retrieval. This study was designed to examine the link between JOLs and memory retrieval processes. The ERPs for the hit items (correctly recalled) attracting the high JOLs rating at study phase were compared to the ones attracting the low JOLs ratings. The ERP old/new effect was observed for both high and low JOL hit items. In the early 300-500 ms time window, the mid-frontal old/new effect which reflected the familiarity processing was found. Also in the later 500-800 ms time window, a parietal old/new effect which reflected the recollection processing was also found for both high and low JOLs items. What was crucial in this study was that the magnitude of the neural activities was modulated by the JOL confidence levels only in the later time window, but it was not relevant in the early time window. The amplitude of the parietal old/new effect is more pronounced in the high JOLs than in the low JOLs trials and this difference is absent in the earlier time window. The mid-frontal old/new effect is less relevant to the JOLs ratings than the parietal old/new effect. That is, the JOLs made at the study phase are relevant to a retrieval process and it is specific to the recollection process. The ERPs elicited by "high JOL" were more positive-going than "low JOL" ERPs at 500-800 ms over posterior sites. The magnitude of the recollection ERP signature was larger for high JOLs. Their finding suggested that JOLs made at study correlate with memory retrieval at test phase and the process is specific to recollection. This result provides preliminary evidence for the processes engaged when making JOLs.

Nonetheless, as discussed by the authors, this finding is not sufficient for suggesting a direct link between JOLs made at study and the retrieval cue at test. One possible explanations points to the mediator account and the transfer-appropriate processing from the study to the test phase. At study, the JOL ratings invite participants to develop associates as mediators between the cue and the target. These mediators can lead to higher JOLs and increase the likelihood that the items will be later recollected with the aid of these mediators. In their view, the cue-utilization account (Koriat, 1997) is supported by their data. The cues assessed at JOLs can be served as contextual identifiers at test.

These two ERP studies are, to my understanding, by far the only studies tackling the underlying mechanism of a metacognitive judgment as such. However, there are some aspects which could be modified to further understand the issue. For instance, in the reported studies, the ERP onsets were time-locked either at the onset of word-pair learning or in the onset of a recognition test. The correlation between the JOLs and the memory processes were examined indirectly. Additionally, it is unclear whether JOLs assessed immediately after word pair learning can allow sufficient time for constructing complex contextual information for later recollection.

In the current study, different time points of JOL ratings at the study phase were included for a comparison to examine the cue-utilization account further. Various types of cue might be engaged at different time points after the initial encoding. Koriat (1997) suggests that when assessing a JOL, participants relied on different kinds of cues, including intrinsic, extrinsic and mnemonic cues (note: cues in Koriat's model refers to a more general term than the specific retrieval cues presented at test which are referred to in the current experimental paradigm). Experiments presented in Chapter 3 will examine different processes engaged for making a JOL at two time points. The hypothesis is that some temporal delay between the initial encoding and JOLs will allow richer construction of an episodic-like memory. This contextual information formed should be useful for later retrieval. If this account holds, I should be able to observe that the parietal old/new effect is greater in the delayed condition than in the immediate condition.

Research questions

Previous studies on the testing effect have demonstrated a potential link between the retrieval process and the later memory performance. However, there was to date few using the ERP method. The ERP method employed here will inform further about the temporal characteristics of the testing benefit. The neural correlate of the recollection process, the parietal old/new effect is predicted to be the contribution on the later memory success. The first study in this doctoral project will investigate whether the ERP correlates of the testing benefit is qualitative similar to a process which contributes to the subsequent memory effect or not.

One of the two ERP studies on the JOL has investigating the relationship between the JOL rating and the retrieval processes engaged. The finding suggested that JOLs made at study correlate with memory retrieval at test phase and the process is specific to recollection. In the second study in this doctoral project, the first study using the ERP technique which each event is time-locked on the onset of a JOL rating trial will be reported. This technique can provide further insight on the underlying mechanism which supports a JOL rating online. The study will discuss whether JOL involves non-mnemonic monitoring processes on top of the proposed retrieval-process or not.

Chapter 2

The retrieval process & the benefit of Testing

Introduction

Main messages of this chapter

1. The testing effect is found under conditions
2. Showing the immediate retrieval process is related to later memory success by ERP data

Rationale, Research question & Hypothesis

The ERP study

Method and results

Discussion

This chapter is a modified version of "Bai, C.-H., Bridger, E. K., Zimmer, H., and Mecklinger, A. (2015). The beneficial effects of testing: An event-related potential study (submitted)"

Abstract

The enhanced memory performance for items that are tested as compared to being restudied (the testing effect) is a frequently reported memory phenomenon. According to the episodic context account of the testing effect, this beneficial effect of testing is related to a process which reinstates the previously learnt episodic information. Few studies have explored the neural correlates of this effect at the time point when testing takes place, however. In this study, we utilized the ERP correlates of successful memory encoding to address this issue, hypothesizing that if the benefit of testing is due to retrieval-related processes at test then subsequent memory effects should resemble the ERP correlates of retrieval-based processing in their temporal and spatial characteristics. Participants were asked to learn Swahili-German word pairs before items were presented in either a testing or a restudy condition. Memory performance was assessed immediately and one-day later with a cued recall task. Successfully recalling items at test increased the likelihood that items were remembered over time compared to items which were only restudied. An ERP subsequent memory contrast (later remembered vs. later forgotten tested items), which reflects the engagement of processes that ensure items are recallable the next day were topographically comparable with the ERP correlate of immediate recollection (immediately remembered vs. immediately forgotten tested items). This result shows that the processes which allow items to be more memorable over time share qualitatively similar neural correlates as the processes which relate to successful retrieval at test. This finding supports the notion that testing is more beneficial than restudying on memory performance over time because of its engagement of retrieval processes, i.e. the re-encoding of actively retrieved memory representations.

Rationale

The enhanced memory performance for items that are tested as compared to being re-studied (the testing effect) is a frequently reported memory phenomenon. However, few studies have explored the neural correlates of this effect at the time point when testing takes place. In this study, we utilized the ERP correlates of successful memory encoding to address this issue.

Hypothesis

Participants were asked to learn words in pair before items were presented in either a testing or restudy condition. Memory performance was assessed immediately and one-day later with a cued recall task. Successfully recalling items at test increased the likelihood that items were remembered over time compared to items which were only re-studied. In addition, if the benefit of testing is due to retrieval-related processes at test then subsequent memory effects should resemble the ERP correlates of retrieval-based processing in their temporal and spatial characteristics. That is, an ERP subsequent memory contrast (later remembered vs. later forgotten tested items), which reflects the engagement of processes that ensure items are recallable the next day were topographically comparable with the ERP correlate of immediate recollection (immediately remembered vs. immediately forgotten tested items).

Testing is a process where a retrieval-process is very likely be engaged. An ERP index of episodic recollection is used for understanding the testing operation. Contrasts between ERPs elicited by correctly responded to old and new items in recognition memory tasks have also revealed dissociable old/new effects (Rugg & Curran, 2007). One of these old/new effects, which peaks at around 500~800 ms after stimulus onset with maximal activity at left parietal region, is associated with recollection, the successful recovery of contextual information (Woodruff, Hayama, & Rugg, 2006) and is usually referred to as the left parietal old/ new effect (for a review, see Friedman & Johnson, 2000; Mecklinger, 2000; Rugg & Wilding, 2000). If the processes engaged in testing are also the processes which lead to subsequent memory success, we would predict that the ERPs of testing SME should resemble the ERP correlates of recollection as indicated by the left parietal old/ new effect.

An ERP study

Method and results

Participants

Twenty-six students enrolled at University of Saarland (aged 19 to 29 years old, M = 23.08, SD = 2.23) gave informed consent to participate. Sixteen of the participants were female. Participants were compensated with either course credit or cash (8€ /hour). An additional 10€ were given to the top 25% performers based on their performance at final recall. All participants were right-handed (Oldfield, 1971), reported no history of neurological disorders and had normal or corrected vision. Two participants did not participate in all sessions, three had very poor performance (less than 25% correct at Day 1 recall), one had already participated in a similar experiment, and five had to be excluded due to insufficient artifact-free trials for ERP analysis (< 16 trials). Data from fifteen participants entered the final analysis.

Materials

Stimuli were 220 Swahili-German word-pairs for which the German words had a frequency of between 10 and 100 occurrences per million (Mannheim frequency per Million; Baayen, Piepenbrock & Gulikers, 1995). All words referred to touchable nouns. Swahili words were translations of the German target words. Prenasalized consonants in Swahili (e.g. “mv”) which are difficult for German readers to pronounce were kept to a minimum. Word length was matched so that on average both Swahili and German words were 6-letters long.

Design Overview

The experiment consisted of two sessions separated by one week. Each session comprised five cycles (each consisting of Phase 1 and 2) and a second-day final recall (Phase 3). In each cycle participants studied 22 word-pairs. In the final test all 110 word-pairs studied on the previous day were tested (see Figure 2.1a). During the initial learning phase, word-pairs were presented three times in randomized order. Phase 2 – during which EEG was recorded - followed initial learning. In Phase 2, half of the word-pairs were restudied again whereas the remaining word-pairs were tested. Additionally, at the end of each cycle all 22 word-pairs were tested in a cued recall task (Day 1 recall).

In this test, only Swahili words were presented as cues and participants had to retrieve the associated German words. Participants processed five of these study cycles on Day 1. Approximately 20~28 hours later, participants returned for the final cued recall test (hereafter, Day 2 recall). To obtain sufficiently large trial numbers for the ERP analyses, the same procedure was repeated a week later with a different set of stimuli.

Procedure

Each session began with the application of the electrode cap. All instructions were given both verbally and were shown on the computer screen at the beginning of the actual experiment. Participants began with a practice session comprising six word-pairs to familiarize them with the task procedure. As illustrated in Figure 2.1b, in each learning phase, word-pairs were presented in black against a grey background for 5000 ms on the display followed by a 1000 ms blank screen. Participants were encouraged to memorize word-pairs during this time. Participants were asked to judge how likely it was that they would remember the word-pair after the first presentation of each word-pair. They were instructed to use the right index finger to make a judgment of learning (JOL) on a 5-step scale where 1 means "definitely forget", 2 "probably forget", 3 "unsure", 4 "probably remember", 5 "definitely remember" (Skavhaug, Wilding & Donaldson, 2010). This judgment was given when "Wahrscheinlichkeit Dich zu erinnern" ("likelihood that you will remember") was displayed. The JOL trial terminated when an answer was given or after 2000 ms, and was followed by a 1000 ms blank screen. The JOL data will not be reported here. After initial learning trials were completed for all 22 word-pairs within a cycle, participants studied the same list of word-pairs two more times in randomized order, but no JOL was required for second and third learning presentations. The presentation time of the word-pairs was 3500 ms followed by a 1000 ms blank screen.

In Phase 2, 50% of the word-pairs were presented in the testing condition whilst the remainder of the pairs was restudied. The assignment of items to testing/restudy condition was counterbalanced across participants. In the testing condition, participants saw Swahili words above six question marks for 2000 ms and were required to recall the German words. At the offset of the stimuli, they were required to say the German translation for the Swahili word aloud within the 3000 ms deadline. In the restudy condition, participants saw the Swahili-German pairs for a fourth time for 2000 ms. Participants were required to say the German words aloud once the stimuli were removed from the screen within the 3000 ms deadline. Testing and restudy trials were blocked to minimize task-switching demands and the order of testing vs. restudy trials was counterbalanced across participants. At the end of each cycle, a Day 1 cued recall task was completed for the 22 word-pairs. Times of presentation and response requirements were identical to testing condition trials. Participants took a self-paced break and proceeded to the next cycle. Each session took approximately 1 hour.

Approximately 20~28 hours later, participants returned to complete Day 2 cued recall test where all the 110 word pairs from the preceding day were tested. Each trial began with a 500 ms fixation cross and a 2000 ms presentation time with Swahili cue word and six question marks. Afterwards, participants had 6000 ms to provide a response for each Swahili word cue. The task lasted approximately 20 minutes. All responses were recorded via a microphone throughout. Correct and incorrect responses were coded online by an experimenter. No EEG was recorded during the final test.

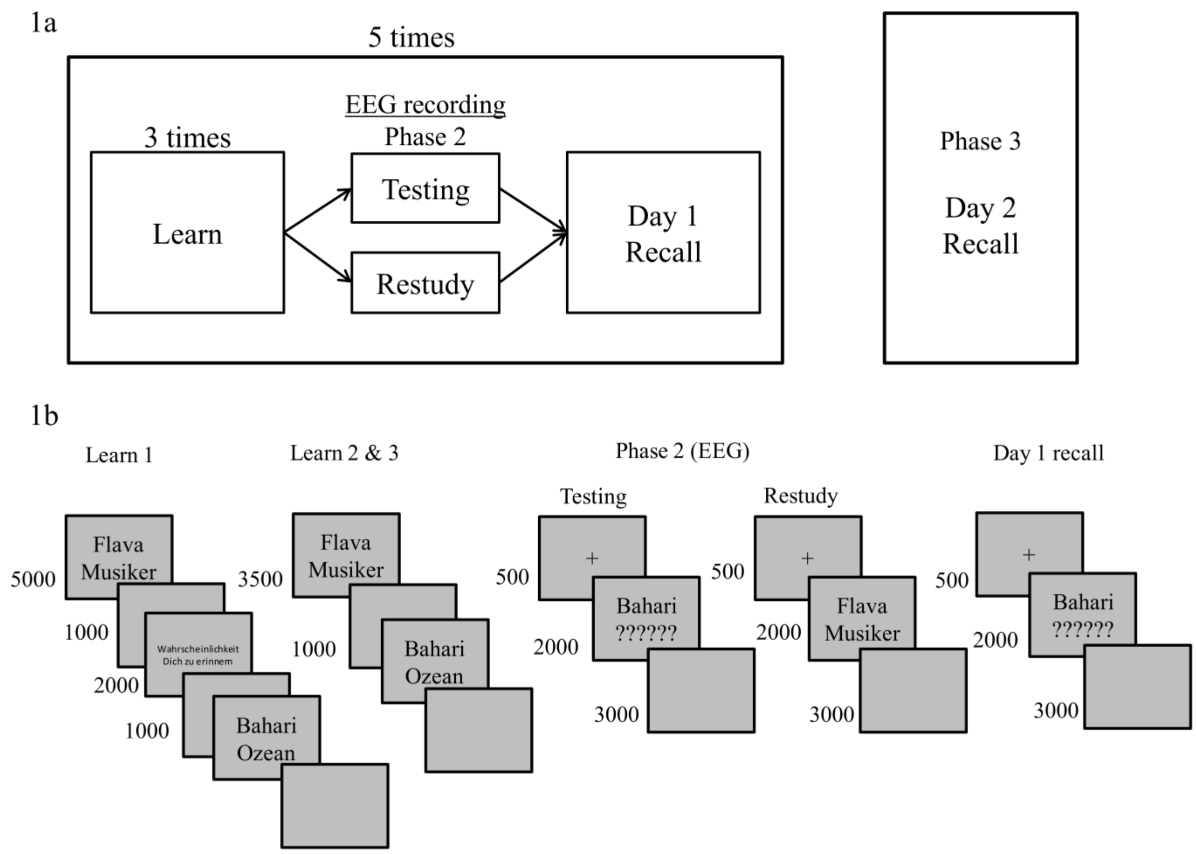


Figure 2.1. a) Illustration of the procedure realized in each session. b) Procedure for one cycle on Day 1. Five such cycles each consisting of 22 different items run on Day 1. The procedure of the final cued recall test on

Day 2 is identical as Day 1 recall with a longer response deadline to 6000 ms and all 220 items were tested, which is not illustrated in this figure.

EEG Acquisition and Analysis

58 Ag/AgCI electrodes were embedded in an elastic cap (Easycap, Herrsching, Germany) based on the extended international 10-20 system. The electroencephalogram (EEG) was recorded continuously with a sampling rate of 500 Hz. Two additional pairs of electrodes were used: one pair was placed on the outer canthi for horizontal EOG. Another two electrodes were placed above and below the right eye for vertical EOG respectively. An electrode placed anterior to Fz served as the ground. EEG was referenced online to the left mastoid. The impedances of the recording electrodes were kept below 5 kΩ. Data was recorded online and processed offline by commercial software Brain Vision Recorder and Analyzer (Brain Products). EEG signals were recorded with a digital bandpass filter (DC-70 Hz) at a rate of 500 Hz with an extra filter applied offline (0.03-30 Hz, 12 dB/ oct). Final epochs extended from 100 ms prestimulus until 1000 ms after stimulus presentation during Phase 2. Data were downsampled to 250 Hz and offline re-referenced to the average of the mastoid signals. Baseline correction started from 100 ms before stimulus onset to stimulus onset. A correction algorithm based on independent component analysis (ICA) was employed for EOG artifact rejection (Makeig et al. 2004).

ERP waveforms were created for five conditions (see 2.1). ERPs to restudied items are labeled as “studied later-remembered (SR)” or “studied later-forgotten (SF)” pairs, depending upon whether they were recalled correctly on Day 2. Tested items were separated into three categories. Tested items recalled correctly at Phase 2 and on Day 2 were labeled as “remember-remember (RR)”; tested items recalled correctly at Phase 2, but forgotten on Day 2 are labeled as “remember-forgotten (RF)”; and tested items which were not correctly retrieved at either Phase 2 or on Day 2 are labeled as “forgotten-forgotten (FF)”. The mean number and range (in parenthesis) of trials entering into each individual’s average were as follows: 35 (16~53) LR; 39 (26~51) LF; 37 (21~57) RR; 30 (16~43) RF; 34 (16~54) FF.

Table 2.1 Condition labels categorized by experimental conditions and subsequent memory performance at three time points

Practice Condition	Phase 2 Practice	Day1 Recall	Phase 3 Day2 Recall	Condition Label
Restudy		R	R	SR
		R	F	SF
Testing	R	R	R	RR
	R	R	F	RF
	F	F	F	FF

Note. R = Remember; F = Forgotten; S = restudy.

ERP analyses are based on the following contrasts: (i) the SME for restudied items was revealed by contrasting SR and SF; (ii) the SME for tested items was revealed by contrasting RR and RF; (iii) the ERP correlate of immediate-retrieval was assessed by contrasting RF and FF which should isolate immediate retrieval success during Phase 2 for tested items. The comparison between this contrast and the SME for tested items (contrast ii) was used to test whether correct recall on Day 2 is associated with the ERP correlate of recollection. The fourth contrast (iv) was between ERPs to studied items (collapsed across SR and SF) and ERPs to RR, RF, and FF pairs, to test whether the LPN is elicited solely by tested items and if so, whether it is modulated by the ease with which memory representations are retrievable. This fourth contrast was specifically tested in the last time window 700-1000 ms due to that fact that the onset of LPN found in previous studies is usually later than the time window of recollection.

ANOVAs were used to test mean amplitude differences for each condition (i.e. SR, SF; RR, RF, FF) from four selected time windows: 150-200 ms, 300-500 ms, 500-700 ms and 700-1000 ms. The first time window 150-200 was chosen because visual inspection of the ERP averages suggested that the ERP waveforms started to diverge at around 150 ms after stimulus. The remaining time windows were chosen to correspond with those used for the conventional analysis of ERP memory old/new effects in line with the assumption that retrieval processes should be evident in the testing but less so in the restudy condition. The 300-500 ms window covers that in which an early mid-frontal old/new effect often associated with familiarity is usually reported. The 500-700 ms time window was selected to capture the left-parietal old/new effect. Additionally, the LPN is usually observed not before 700 ms after onset of a retrieval cue (Johansson & Mecklinger, 2003) and thus the 700 to 1000 ms time window was used to capture this effect.

Repeated measures ANOVAs with Subsequent Memory Performance condition and regions limited to a 3 × 3 grid of electrodes, including 3 Anterior-Posterior (frontal, central, parietal) and 3 Laterality (left, middle, right) were included as factors. Degrees of freedom were adjusted for the ANOVAs by incorporating the Greenhouse-Geisser correction for violations of sphericity when appropriate for both behavioral and ERP data.

Results

Behavioral Data

Figure 2.2 shows mean proportions of correct recall for the testing/restudy conditions on Day 1 and Day 2. An ANOVA with factors testing/restudy condition and time (Day 1, 2 recall) revealed a main effect of time, $F(1, 14) = 298.90$, $p < .01$ and an interaction between testing/restudy conditions and time, $F(1, 14) = 33.39$, $p < .01$. To follow up the interaction effect, we compared the amount of recalled items between testing and restudy conditions on Day 1 and Day 2 respectively. The result showed that on Day 1 more restudied items (M = .68, SD = .13) were recalled than tested items (M = .62, SD = .10), $t(14) = -2.31$, $p < .05$, while on Day 2 this difference was reversed. A marginally significant testing effect was found on Day 2 where participants were able to recall more tested items (M = .35, SD = .09) than restudied items (M = .32, SD = .11), $t(14) = 1.79$, $p = .10$. In addition, the difference in the amount of correctly recalled items from Day 1 to Day 2 is significantly smaller in testing (Mean difference from Day 1 to Day 2 = .27, SD = .06) than in the restudy condition (Mean difference from Day 1 to Day 2 = .36, SD = .09), $t(14) = 4.98$, $p < .01$. This suggests that once the tested items were successfully recalled in Phase 2, they were less likely to be forgotten on the Day 2 in comparison to merely restudied items. This benefit of testing from Day 1 to Day 2 recall licensed us to precede the ERP analysis to explore the neural underpinnings of this behavioral testing effect and its relevance on later memory performance as presented in the following.

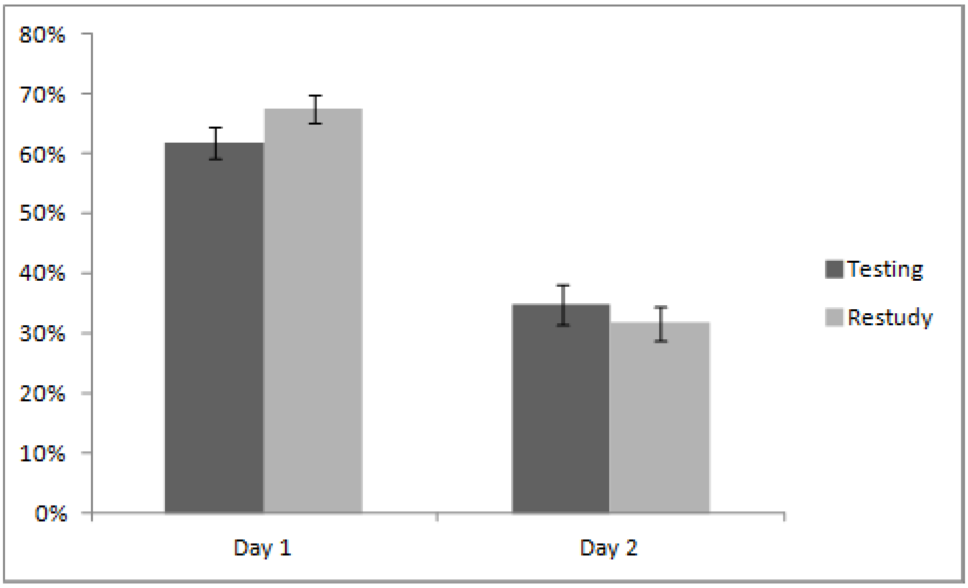


Figure 2.2 Percent correct for the tested items and the restudied items at cued recall on Day 1 and at final recall on Day 2. Error bars show standard error mean.

ERP Data

Restudy Condition

This analysis compared ERPs elicited by restudied items which were either remembered or forgotten on Day 2 recall (contrast (i): Restudy SME). As shown in Figure 2.3a, small differences from 300-500 ms at posterior sites were observed; however, a global ANOVA with the factors Subsequent Memory Performance (SR/SF: later remembered/later forgotten) $\times$ 3 AP $\times$ 3 Laterality in the three selected time windows did not reveal any main effect of Subsequent Memory Performance nor any interaction effect including this factor (See Table 2.2a). There were thus no significant ERP differences in the restudy condition of Phase 2 between items that were remembered or forgotten on the Day 2 recall test. Given this null effect and to make the remainder of the analyses more accessible, the two restudy conditions (SR/SF) were collapsed into one RS condition for the remainder of the relevant analyses.

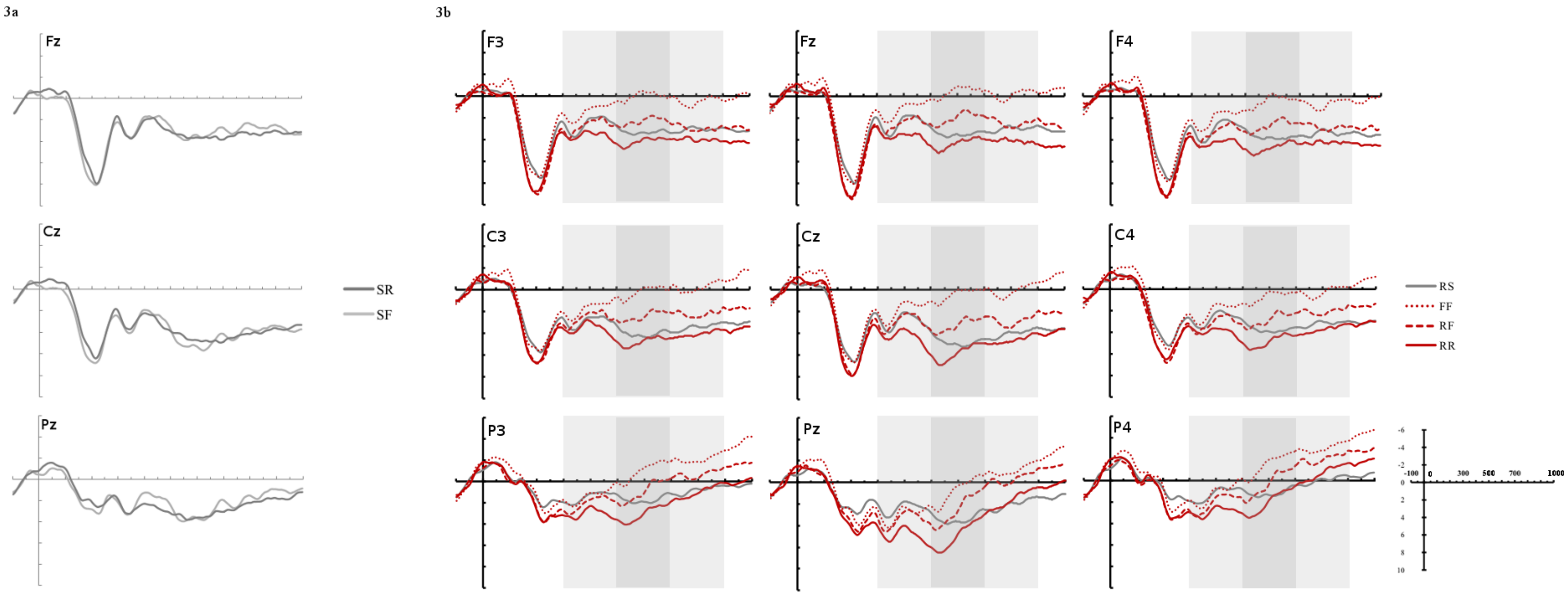


Figure 2.3 a) The ERP waveforms to restudied items which were later remember and forgotten (SR/SF) were not significantly different at any time windows of interest. b) ERP waveforms to all restudied items (RS) and tested items categories by Subsequent Memory Performance (RR, RF, FF). ERPs are plotted from 100 ms before stimulus onset to 1000 ms thereafter at frontal, central and posterior midline sites: Fz, Cz and Pz. Three time windows of interest are marked in grey. The waveforms were low-passed filtered at 12 Hz for illustration.

Testing Condition

Corresponding degrees of freedom, F-values and p-values for contrasts related to the items in the testing condition are reported in Table 2.2.

(ii) SME for tested items

As shown in Figure 2.3b, the ERPs to RR items start to diverge from ERPs to RF items around 300 ms post-stimulus, with a greater relative positivity for RR relative to RF items. This difference is widely distributed across the scalp (see 2.4 upper panel). A global ANOVA with factors of Subsequent Memory Performance (RR/RF) $\times$ 3 AP $\times$ 3 Laterality revealed a main effect of subsequent memory in all three time windows of interest from 300-500, 500-700 to 700-1000 ms. There was no interaction effect involving the Subsequent Memory Performance factor and the other factors found.

(iii) ERP correlates of immediate-retrieval

The waveforms to RF and FF in Figure 2.3b showed that the ERPs to later remembered items were more positive-going than to forgotten items. This effects starts around 500 ms and on the basis of visual inspection (Figure 2.4 lower panel), this effect appears to be more frontal-central than posterior. ANOVAs with factors of Subsequent Memory Performance (RF/FF) $\times$ 3 AP $\times$ 3 Laterality revealed a main effect of Subsequent Memory Performance in the 500-700 ms time window. In the 700-1000 ms time window, there was an interaction effect between Subsequent Memory Performance and AP. Follow-up tests with factors of Subsequent Memory Performance condition (RF/FF) by each of the 3 levels of AP (frontal, central or posterior) revealed that the main effect was significant at central ($F(1, 14) = 5.42$, $p < .05$) and marginally significant at frontal ($F(1,14) = 4.41$, $p = .05$), but not at posterior sites ($F(1,14) > 1.73$).

Comparing the SME and immediate-retrieval effect

Our main prediction was that if retrieval practice promotes learning by the recruitment of recollection-like processes, the SME for tested items should resemble the ERP correlate of immediate retrieval (as reflected in the RF/FF contrast). To directly test this, we examined whether the immediate-retrieval effect and the SME in the 500 to 700 ms time window differ in scalp topography, as would be expected if different neuronal circuitries have contributed to both effects. To improve the sensitivity of this contrast, all 58 recording sites were included in this analysis. An additional analysis was conducted on amplitude normalized mean values to ensure that any differences in scalp topography between the two conditions do not result from amplitude

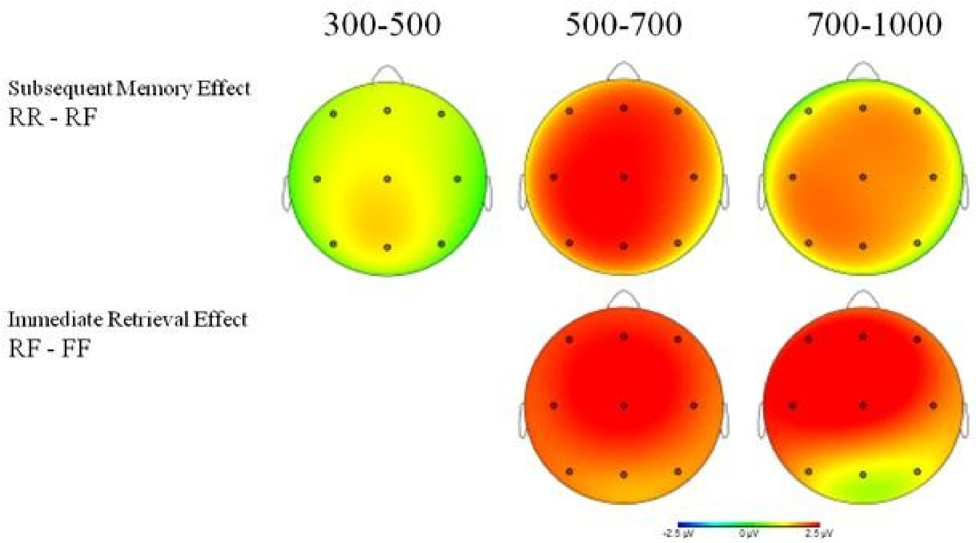


Figure 2.4 Topographical maps showing the scalp distributions of the subsequent memory effect (RR−RF) and the immediate retrieval effect (RF-FF). The subsequent memory effect started at an earlier time window (300-500 ms). Both effects are more comparable in the 500-700 ms time window and less resembling in the 700-1000 ms time window.

differences (McCarthy & Wood, 1985). The ANOVA with factors condition (RR−RF; RF−FF) and recording site did not reveal a significant interaction, (non-scaled data: $F(57, 789) = .66$, $p < .98$; scaled data: $F(57, 798) = .68$, $p < .97$, suggesting that highly similar brain circuitries were active in the immediate-retrieval processes and the 500 to 700 ms proportion of the SME contrast.

Comparison of Restudy and Testing Condition

(iv) All restudied items RS vs. one tested condition RR or RF or FF

As reported in Table 2.2b, in this set of contrasts we explored whether and how mnemonic processing in the testing condition is reflected in the LPN, a late onsetting ERP component elicited by retrieval cues when memory contents are searched and retrieved. We first contrasted the LPN in the study condition (RS; collapsed across later forgotten (SF) and later remembered (SR) trials) separately with the three testing conditions RR, RF and FF using ANOVAs with factors testing/restudy condition × 3 AP × 3 Laterality in the 700 to 1000 ms time window. For the RS vs. RR contrast there was a testing/restudy x AP interaction. Follow-up ANOVAs with 3 AP × 3 Laterality separately for RS and RR condition revealed different topographical distribution for the RS and RR conditions. In the RS condition, there was a main effect of AP, $F(1.24, 17.36) = 14.86$, $p < .01$ and interaction effect between AP and LAT, F(2.11, 19.52) = 10.95, p < .01. However, there was no significant effect of LAT on the amplitude in the restudy condition, p = .10. Whilst in the RR condition, there was a main effect of AP, F(1.30, 18.22) = 13.32, p < .01; main effect of LAT, F(2,28) = 3.39, p < . 05 and an interaction effect between AP and LAT, F(4, 56) = 7.95, p < .01. This finding suggests that the processes engaged in testing condition are not qualitatively comparable as the ones engaged in the restudy condition.

For the RS/RF contrast no effects involving the testing/restudy condition were obtained, whereas the RS/FF contrast revealed a main effect of testing/restudy condition, reflecting the broadly distributed LPN in the FF as compared to the RS condition. In a final contrast, we explored whether the LPN within the testing conditions is modulated by the ease with which information can be recovered, by contrasting tested items which were not retrieved at Practice or Day 2 (FF) with those that were retrieved during practice and at Day 2 (RR). This analysis revealed a main effect of testing/restudy condition, interactions between the condition factor and the two other factors, AP and LAT, as well as a three-way interaction. Tested separately for each of the Laterality by AP combinations, larger LPN for the FF than the RR condition were obtained at all nine electrode sites. Post-hoc analyses estimating the effect size using Cohen's d values revealed that the LPN is most pronounced at left middle-posterior C3 and P3 electrode sites (d > 0.9) and also middle-right central Cz and C4 (d > 0.8) electrodes.

Table 2.2 *ANOVA table for a) Restudy SME, Test SME and Immediate Retrieval Effect b) LPN analyses*

a)

Contrast	Effect	300-500	500-700	700-1000
(i) Restudy Subsequent Memory Effect				
SR/SF	Condition	n.s.	n.s.	n.s.
	... × AP	n.s.	n.s.	n.s.
	... × LAT	n.s.	n.s.	n.s.
	... × AP × LAT	n.s.	n.s.	n.s.
(ii) Test Subsequent Memory Effect				
RR/RF	Condition	F(1,14) = 5.35, p < .05	F(1,14) = 14.26, p < .01	F(1,14) = 6.21, p < .05
	... × AP	n.s.	n.s.	n.s.
	... × LAT	n.s.	n.s.	n.s.
	... × AP × LAT	n.s.	n.s.	n.s.
(iii) Immediate Retrieval Effect				
RF/FF	Condition	n.s.	F(1,14) = 6.74, p < .05	n.s.
	... × AP	n.s.	n.s.	F(1.31,18.27) = 5.95, p < .05
	... × LAT	n.s.	n.s.	n.s.
	... × AP × LAT	n.s.	n.s.	n.s.

b)

Contrast	Effect	700-1000
(iv) LPN		
RS/RR	Condition	n.s.
	... × AP	F(1.26,17.61) = 4.33, p < .05
	... × LAT	n.s.
	... × AP × LAT	n.s.
RS/RF	Condition	n.s.
	... × AP	n.s.
	... × LAT	n.s.
	... × AP × LAT	n.s.
RS/FF	Condition	F(1,14) = 6.67, p < .05
	... × AP	n.s.
	... × LAT	n.s.
	... × AP × LAT	n.s.
RR/FF	Condition	F(1,14) = 9.11, p < .01
	... × AP	F(1.36,19.09) = 5.27, p < .05
	... × LAT	F(1.31,18.38) = 4.19, p < .05
	... × AP × LAT	F(4,56) = 3.69, p < .01

Note. Degrees of freedom, F- and P- values are listed only for significant results (p < .05). Anterior-posterior (AP), laterality (LAT). SR: studied remembered; SF: studied forgotten; RR: remembered; RF: later forgotten; FF: immediate forgotten. Non-significant is abbreviated as n.s.

Discussion

Many studies have demonstrated that testing during learning enhances later memory performance. The episodic context account is one model of the underlying mechanisms thought to drive the testing effect (Lehman, Smith, & Karpicke, 2014). The core concept of this episodic context account suggested that retrieval practice encourages a retrieval process which leads the learners to reconstruct the learnt episode; consequently the episodic context was re-instantiated. Additionally, the dual process model of recognition memory indicated that a successful recollection requires actively retrieving the encoded item and the associated context in details. The current study integrated the two frameworks into one account and used ERP data to show that the well-known behavioral testing effect can be related to an active recollection-like process. To our knowledge, however, there is to date no evidence directly showing that the retrieval act at testing in the learning phase is related to successful memory performance at Day 2. In the present study, the benefit of testing was revealed by the finding that for materials which were tested on Day 1, the forgetting rate from Day 1 to Day 2 was lower than for restudied materials. We will now turn to the analyses of the ERP data to explore the neural underpinnings of this effect.

ERP results

Restudy

Although subsequent memory effects were expected in both restudy and testing condition, no such effect was observed in the restudy condition. We speculate that this is likely to be because of the inclusion of three learning blocks prior to the restudy condition in Phase 2 of the current design. The processes which predict later memory performance for the restudied items and which are typically seen in ERP SME contrasts (i.e. Paller & Wagner, 2002) could have occurred during any of the preceding learning block, rendering them unobservable during Phase 2. There are three possibilities concerning the timing of a word pairs being learnt: some words pairs might have been encoded successfully during any of the three learning blocks, or during the fourth learning chance at Phase 2, or was not encoded successfully at all. The possible jittering of encoding onset might have diluted the hypothesized SME in the restudy condition. Other alternative explanations could be derived from studies which demonstrated a reversed SME in the ERPs to non-words in contrast to the SME found in words condition (e.g. Otten, Sveen, & Quayle, 2007) or no SME found for non-semantic (spatial memory) in comparison to semantic condition (Mecklinger & Müller, 1996). Following these findings, it is reasonable to assume that repeatedly seeing a restudied word pair for the fourth time initiates less mnemonic processing or semantic processing in comparison to the tested items.

Testing

The first contrast between items in the testing condition, revealed ERPs to the later-remembered items (RR) that were more positive-going than ERPs to items that were later-forgotten (RF). This effect was widely distributed across the scalp at all time windows from 300 up to 1000 ms after stimuli onset. This SME has an earlier onset than the predicted time window where the neural correlations of recollection were often observed. There is a general consensus that cued recall relies on recollection or remembering both the cue and the target which is not the case for recognition task where familiarity or knowing the cue could arise a response (Lindsay & Kelley, 1996; Nobel & Shiffrin, 2001).

The contrast between RF and FF was presumed to reflect immediate-retrieval processes. This contrast was significant in the time window from 500-700 ms which is often associated with the neural correlates of recollection. Although most ERP effects associated with recollection have been reported to focus principally at left-parietal sites (Vilberg & Rugg, 2008), we find only a main effect of condition in this specific time window. Although the ERP correlate of recollection has been reported to focus principally at parietal recording sites (Vilberg & Rugg, 2008), we find only a main effect of condition in this specific time window. Previous studies have also found the parietal old/new effect to be larger and more widely distributed in free recall task than in recognition task (Paller, McCarthy & Wood, 1988). In addition, it is also conceivable that cues in a foreign language evoke processes additional to the recollection processes and this may have rendered the effect more widely distributed across the scalp. Given that this is the first ERP study which directly links the testing effect on subsequent memory performance and retrieval processes at testing, there is currently no suitable comparison in the literature to determine why this might be. One speculation might be that foreign cue words evoke processes additional to the recollection process and thus the effect is more widely distributed across the scalp.

Taken together, the SME and immediate-retrieval effect share a qualitatively similar neural circuitry which can be identified as an engagement of a recollection-like processes. We are inclined to make this conclusion because we have found that the two effects do not differ in their scalp topographies at the crucial time window which a recollection process was often reported between 500 and 700 ms even when using a highly conservative statistical measure for topographical differences including 58 scalp electrodes. That is, the testing effect which enhances later memory performance is related to the involvement of a recollection process during testing. The current findings supports the episodic context account which assumes that the testing effect is likely to be driven by an engagement of successful recollection at Testing condition which is absent in the restudy condition.

Our findings support the account that testing leads to the reinstatement of episodic context which is corresponding to the previous fMRI studies exploring the testing effect using SME paradigm (van den Broek et al., 2013; Wing et al., 2013). In both studies, they have found that testing engages activity in the hippocampus or MTL regions which are associated with binding associates with the temporal context information into an episode.

A second but not mutually exclusive possibility is that the amplitude of ERPs for tested items was modulated by the amount of memory strength. From visual inspection, a gradient change in amplitude from positive to negative was observed among RR, RF and FF (Figure 2.3b). An additional post-hoc analysis comparing the amplitude differences across the three testing conditions were conducted at the crucial time windows 500-700 ms after stimuli onset. An ANOVA with three testing conditions × 3 AP × 3 Laterality revealed a main effect of testing condition, $F(1.43, 19.99) = 13.18$, $p < .01$. The ERPs to RR was more positive-going than RF and the ERPs to RF was more positive-going than to FF. One interpretation is that the gradient main effect correlates with memory strength (Vilberg, Moosavi, & Rugg, 2006). The more information was retrieved at Phase 2, the higher memory strength it contains which leads to better recallability on Day 2.

LPN is associated with post-retrieval memory search

In the late 700-1000 ms time window, ERPs to tested items that were recalled correctly at Phase 2 and at Day 2 (RR) and tested items that could not be retrieved at either day (FF) elicited more negative going ERPs than items in the restudy condition (RS). While the former effect was smaller in amplitude and topographically bounded to the right posterior region, the latter LPN effect was more pronounced and showed a broad and posteriorly accentuated topography. In addition, within the testing conditions the LPN was most pronounced for items that could neither be retrieved at Day1 or Day 2. Overall this pattern of results is consistent with the view that the LPN reflects the search for and retrieval of memory bound information and is modulated by the specificity with which memory is searched (Mecklinger et al., in prep). In one illustrative source memory study, participants were required to discriminate either between performed and watched actions or between performed and interrupted actions. Consistent with the view that the LPN reflects the specificity with which memory is searched, there was a large LPN when discriminating the actions which were performed and interrupted, where there are only a few specific and diagnostic features that allow one to discriminate between both sources (Leynes & Kakadia, 2013).

Following this framework, a cued recall test as the one used in the current retrieval practice condition may constitute a paradigmatic case for a highly specific memory search, as it requires one to discriminate between highly overlapping words in order to identify the one originally paired with the cue word. The highly overlapping features of the German-target words may have lowered memory strength and may have given rise to extended retrieval processing as reflected in the LPN, in particular in situations in which retrieval processing was unsuccessful as in the FF condition.

Despite the fact that a robust benefit of testing is found when comparing the amount of forgetting from Day 1 to Day 2 between testing and restudy conditions, we have only found a marginally significant testing effect on Day 2. This could arise from a number of aspects of the current design. Unlike many other studies which manipulated learning times for each items (Karpicke & Roediger, 2008; van den Broek et al., 2013), in the current study, participants learned all word pairs three times before the critical manipulation was introduced. Given the large number of test items required to provide sufficient ERP trials, the relatively high task difficulty may have meant some items were not encoded sufficiently within the initial three learning blocks. For those items which were not learned during Phase 1, once they were assigned to the testing condition, there was a low-likelihood that the items would be recovered. In contrast, the unlearned items from Phase 1 would receive a fourth learning opportunity once they were assigned into the restudy condition (Bahrick & Hall, 1991; Jang, Wixted, Pecher, Zeelenberg, & Huber, 2012). This restriction of the experimental design conveys a disadvantage for tested items over restudied items on Day 1 (Toppino & Cohen, 2009), because the latter are shown again during Phase 2.

In summary, the current study provides direct link between the neural correlates of subsequent memory effect and of immediate retrieval at the time point when testing occurs in comparison to restudy condition. Our findings support the episodic context account that testing as practice engages an active recollect of the learnt word pairs which is then driving a better memory performance in a later cued recall task. A second possible explanation is that the higher memory strength it is, the higher recallability on a later recall task it leads to. In addition, we also show that the LPN is related to an active post-retrieval search independent from memory formation at test.

Chapter 3

Metacognitive judgment:

the delayed JOL effect

Introduction

Main messages of this chapter

1. Introduce metacognitive accuracy on predicting later memory performance
2. Showing the processes of metacognitive judgment cannot be reduced to a mnemonic process by ERP data

Rationale, Research question & Hypothesis

The ERP study

Method and results

Discussion

Introduction

The current study incorporates the behavioral and EEG technique to examine to assess the JOLs at two time points in order to better understand the basis of JOLs and its relation with memory retrieval process. One time point will be immediately after the presentation of the study material (immediate JOL) whereas for the remaining items the JOL will be given at a later time point after some delay from the learning phase (delayed JOL). EEG signals will be recorded at the study phase time locked to the onset of JOLs. The prediction is that items in the immediate JOL condition will be less accurate in predicting later memory performance than items in the delayed JOL condition (Dunlosky & Nelson, 1992). For the ERP result, the hypothesis is that if the delayed JOL effect relies on a reinstatement of cue-information, a recollection signature should be observed in the delayed condition between the high-confidence versus low-confidence JOL items. A successful recollection at this time point might enhance the likelihood that the item is also remembered at a later time point. The second effect investigated in this project is the judgment of learning (JOL) on learning. Participants first learn word pairs and are then asked "How confident are you that you will be able to recall the second word when prompted with the first?" (adopted from Nelson & Dunlosky, 1991, p. 268) using a scale from very confident to not at all confident. Studies have shown that a JOL made after a delayed interval between learning and JOL provides a more accurate prediction of later memory performance than immediate JOL ratings (JOLs) (Keleman & Weaver, 1997; Nelson and Dunlosky, 1992). Judgments of learning (JOLs) are taken as reference by learners to monitor their own performance and control their learning schedule (Son & Kornell, 2009; Veenman, Hout-Wolters & Afflerbach, 2006). Understanding the underlying mechanism of the JOLs may be helpful for learners to develop a learning schedule.

A JOL with delayed interval between studying and testing improves the predictability of the JOL on later memory performance. This is referred to as the delayed JOL effect (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). Spellman and Bjork (1992) proposed that the reason why a JOL made after some time interval (delayed JOL) is more accurate than a JOL made immediately after the learning phase (immediate JOL) is because the retention interval alters what is actually assessed during the JOL. A delayed JOL is presumably more related to the access of retrieval-ability of the learned information and it alters the accessibility of the encoded information so that a delayed JOL has a higher accuracy in predicting later performance (Koriat, 1997; Nelson and Dunlosky, 1992; Spellman and Bjork, 1992). Koriat (1997) suggested that the JOLs assessed at different time points rely on the different kinds of information available at the time point when a judgment is made. When making a JOL immediately after studying, the JOL is based on information which is related to item-characteristics, such as word frequency whereas when making a JOL with some delays after initial studying, the JOL is based on learners' assessment on whether they can remember the items at that point and this checking process induced by the JOLs can better predict the likelihood of which the items will be remembered or not.

Rationale

Although there are theoretical proposals on the basis of JOL, there is a lack of neuroimaging studies investigating the underlying mechanism of JOL. There were about two studies using ERP techniques to investigate the relation between JOLs and memory processes. However, both of them indirectly correlated the JOLs to the encoding or retrieval phase instead of directly time-locking the ERP onset at the moment when JOLs are made. Although a correlation between the JOL ratings at study and later recollection (Skavhaug, Wilding & Donaldson, 2013), it is not clear which information related to the study materials contributes to JOLs and how the temporal delay of JOLs from initial learning is related to JOL accuracy and memory performance.

Koriat (1997) explicitly suggested that with time, the basis of JOLs shift from the characteristics inherited from the study materials pre-experimentally, towards mnemonic-based heuristics including the retrieval accessibility of the study materials (Spellman & Bjork, 1992). By this logic, the accuracy of JOLs depends on a retrieval mechanism principally for items for which the JOL was delayed. The ERP signature of recollection, the left parietal old/ new effect will be used to capture the time course (Rugg & Curran, 2007) and to decompose the contributions of different cues as the basis of JOLs (Koriat, 1997). If the delayed JOLs are more related to participants' monitoring of the retrieval process, the ERPs to the delayed JOLs will resemble a recollection signature.

Research question and hypothesis

The aim of the current study is to examine the mechanisms which determine judgments of learning at these different time points. If the delayed JOL is based more on episodic recollection of the learned information than a direct access on the item-characteristics information, a differential pattern between high and low JOLs should be observed. More specifically, it is expected that the high-low JOLs contrast for delayed JOLs will resemble the left-parietal old/new effect which is related to recollection-based recognition (Rugg & Curran, 2007) than the high-low JOLs contrast in immediate JOLs. The hypothesis is that a successful recollection at this time point might enhance the likelihood that the item is also remembered at a later time point. Or if we do not find a recollection related process engaged during the delayed JOL, the increased likelihood of JOL predictive value is then driven by some processes which is not purely related to mnemonic process, or cannot be reduced to mnemonic process.

An ERP study

Method and results

Participants

30 native German speakers gave informed consent to participant and were all rewarded with cash (8€ per hour). Six of them were excluded due to a lower amount of JOL trials rated in one condition (< 10 %). Two other participants were excluded after artifact rejection because there were lower than 10 trials left in one of the condition. The remaining 22 participants (15 female) were on average 25±4 years old (range: 19~36). All participants were right-handed (Oldfield, 1971), reported no history of neurological disorders and had normal or corrected to normal vision.

Materials

Stimuli were 216 German-German semantically-related word pairs. 8 additional pairs were included in the practice session. All words were concrete nouns and semantically-related (Kriukova, Bridger and Mecklinger, 2013).

Design

The list was constructed as the following. There were six word lists created, which were composed by 36 pairs of words. The six word lists were matched by type of semantic relatedness (thematic or categorical), word length and amount of pairs which the syllable-initials is shared. The assignment of the experimental conditions was counterbalanced across the three lists. In the first study block, there is the 1st list of 36 pairs assigned into the immediate condition and a 2nd list of 36 pairs assigned into the delayed condition. And then, in the second study block, a 3rd list of 36 pairs into the immediate and a 4th list of 36 pairs into the delayed condition. In total, there were 144 words learned in two study blocks. The temporal delay between learning and judgment of learning is in average 55 trials, including 36 learning-only trials and in average 10 learning plus JOL for the immediate condition, which is about 5 minutes interval. The order of presentation within each study block was pseudorandomized with the constraints that no more than three consecutive trials being in the same condition; in addition, the initial letters of an adjacent trial is never identical. An additional 12 word pairs were selected with the same criteria to be served as the practice items.

Each participant saw four out of the six lists in either the immediate or the delayed condition in the study blocks. The remaining two lists were treated as new items and used in the later recognition test. The order of seeing word lists in block 1 or 2 and the assignment of the three conditions was counterbalanced across participants.

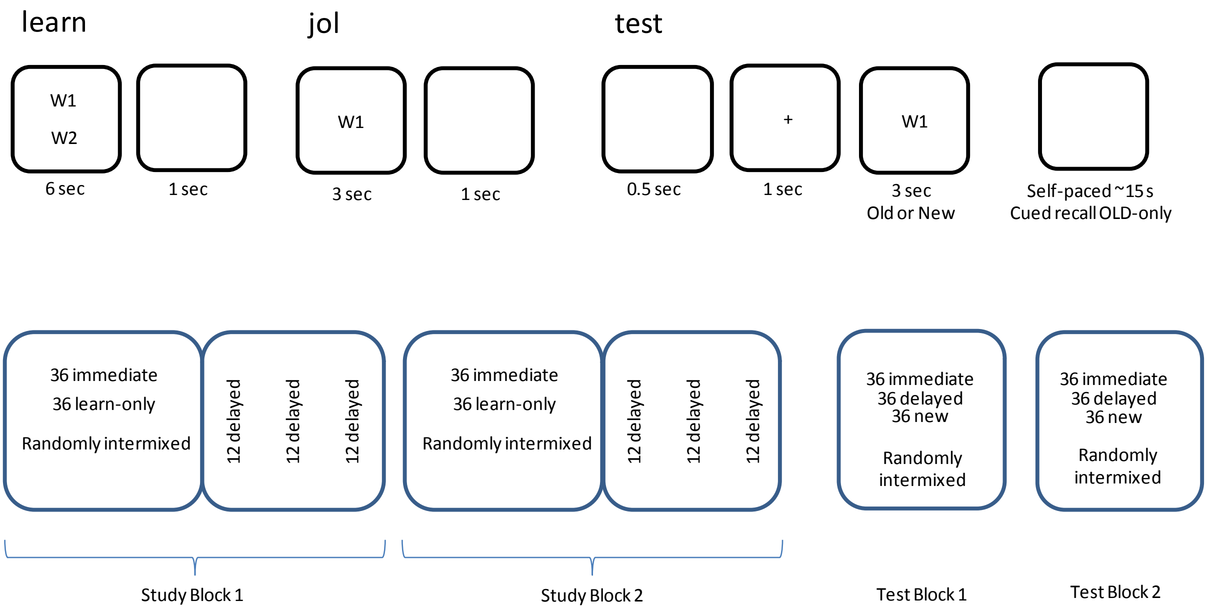


Figure 3.1 Procedure and design of this ERP study. During the learning blocks, participants learn the word pairs for 6 second. In the immediate condition, a 3-sec cue JOL trial follows. Otherwise, a next word pair follows. After 72 word pairs, 36 JOL trials continue as the delayed condition. The 1/3 word pairs which were learned in the first 1/3 were cued first for the JOL. In the test blocks, new words were added for the recognition memory test. Word 1 (W1) stands for cue word and Word 2 (W2) stands for target word. W1 is placed on top of W2 vertically on the display.

Procedure

The participants were informed that this is an incidental learning task where they were asked to remember the word pairs as well as possible. The stimuli were presented in Calibri font, with the cue word on top of the target word, followed by a 1-sec blank. A cue word together with a 5-point scale underneath was presented for 2 seconds as the cue for a JOL response followed by a 1-sec blank. After two study blocks, a distraction task was given where the participants performed automatic operational span task (Unsworth et al., 2005). In the learning phase, they first see a word pair presented on the screen for six seconds. From time to time, after the initial presentation and a 1-second blank screen, they see the German cue words from the learnt-pair and a 5-point-scale. They were instructed that when they saw the cue word, they had to make a judgment about the likelihood that they could remember the word pairs in a later memory performance: from definitely forgotten (1), probably forgotten (2), not sure (3), probably remember (4) to definitely remember (5).

The participants were instructed to use their right index finger to make the judgment. Each study block took about 12 minutes.

In the final test phase, participants were asked to perform a recognition memory task. They were instructed to make an old response even if they could not recall the word. When a cue word was not presented in the study block, they had to give a “new” response. The response buttons were counterbalanced across participants. The given trial terminated with response input. The cue word was displayed after a 1-second fixation-cross for 3 seconds. The ITI was 0.5 second. The test blocks took about 20 minutes.

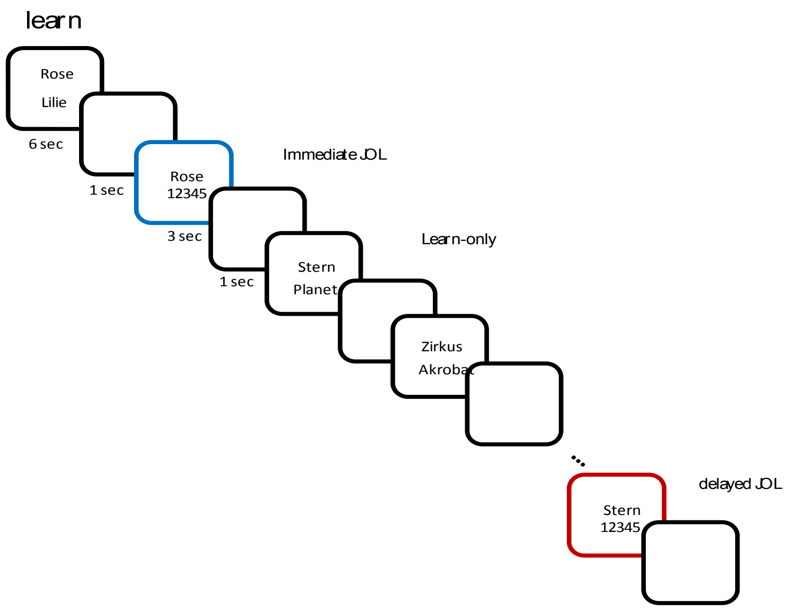


Figure 3.2 Procedure at the learning blocks. Blue box is immediate JOL trial right after the word learning. Red box is delayed JOL trial with temporal delay from the learn-only trial.

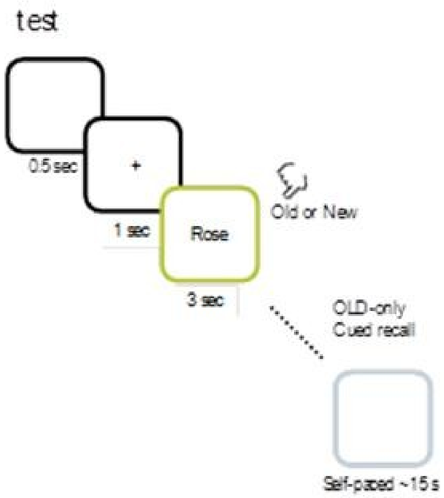


Figure 3.3 Procedure at the test blocks. In a test trial, after a 0.5-sec blank screen, a 1-sec fixation cross signals a coming cue word and asks for participants' attention on the location of the cross sign (+). Participants are asked to respond old or new since the onset of the cue word (green box in this illustration). The trial determines as soon as an old/ new response is given. The deadline of the old/new decision is up to 3 seconds. According to participants' responses, when they press OLD, a cued recall test follows. Otherwise, the trial determines.

EEG Acquisition and Analysis

58 Ag/AgCI electrodes were embedded in an elastic cap (Easycap, Herrsching, Germany) based on the extended international 10-20 system. The electroencephalogram (EEG) was recorded continuously with a sampling rate of 500 Hz. Two additional pairs of electrodes were used: one pair was placed on the outer canthi for horizontal EOG. Another two electrodes were placed above and below the right eye for vertical EOG respectively. An electrode placed anterior to Fz served as the ground. EEG was referenced online to the left mastoid. The impedances of the recording electrodes were kept below 5 kΩ. Data was recorded online and processed offline by commercial software Brain Vision Recorder and Analyzer (Brain Products). EEG signals were recorded with a digital bandpass filter (DC-70 Hz) at a rate of 500 Hz with an extra filter applied offline (0.03-30 Hz, 12 dB/ oct). Final epochs extended from 100 ms prestimulus until 1000 ms after stimulus presentation during Phase 2. Data were downsampled to 250 Hz and offline re-referenced to the average of the mastoid signals. Baseline correction started from 100 ms before stimulus onset to stimulus onset. A correction algorithm based on independent component analysis (ICA) was employed for EOG artifact rejection (Makeig et al. 2004). ERP waveforms were created for four conditions related to the JOL time condition (immediate, delayed) and the confidence level (high, low) at learning phase.

Behavioural results

Global JOL rating distribution shows how participants use the JOL-scale

Firstly, participants assigned the JOL rating proportionally least to the "definitely forget" (M=.10, SD=.08). Secondly, "probably forget" (M=.19, SD=.07) and "unsure" (M=.18, SD=.07) received intermediate amount of ratings and the difference in amount of ratings were not significantly different (t=.55). Most ratings were given to the "probably remember" (M=.22, SD=.07) and "definitely remember" (M=.25, SD=.13) and the difference was also not significant between these two levels of remembering (t=-.84). Generally speaking, participants treated the 5-point scale as a 3-level scale and used it in a linear trend. In addition, participants were confident on their own performance.

JOL rating by JOL time condition

Globally speaking, participants used the JOL-scale distinctively depending on whether the JOL was assessed immediately or with delay. In the immediate condition, the first fourth levels of the confidence scale were used linearly ("definitely forget" (M=.07, SD=.05); "probably forget" (M=.19, SD=.08); "unsure" (M=.24, SD=.08); "probably remember" (M=.29, SD=.08)) while there is a drop of amount at the most-confident level (M=.18, SD=.12) (the t-values ranged from -7.21 to 2.85 for every two levels contrasted). In contrast, in the delayed condition, participants were more inclined to rate items as "definitely remember" (M=.33, SD=.16) than any other confidence levels. The two forgotten-related levels "definitely forget" (M=.13, SD=.13) and "probably forget" (M=.19, SD=.08) were not different in the amount of items assigned (t=-1.80).

Recognition memory task performance is modulated by the JOL time condition

The overall Pr-score was 0.64(.14) with the Br-score 0.39(.21). The hit rate was higher for items which were rated in the delayed condition (M=.81,SD=.13) in the recognition test than items which were rated in the immediate condition (M=.75,SD=.12), $t(21)=-3.12, p<.01$. The reaction time was longer for the immediate judged items (M=1392, SD=383) than for the delayed items (M=1374, SD=401). However, the difference was not significant.

Cued recall test performance does not depend on the JOL time conidition

There were 41% (SD=18%) of items which were rated immediately and 43% (SD=20%) of items which were rated with delay being recalled accurately in the cued recall test. Statistically, there was no effect of condition on cued recall accuracy, which replicated the previous results in Nelson and Dunlosky.

Table 3.1 Mean gamma values by JOL conditions. Standard deviation is shown in parentheses. *n=21 due to a 100% HIT rate in one participant; gamma cannot be calculated with the absence of MISS items.

	Recognition memory	Cued recall
Immediate JOL	0.19(0.28)	0.26(0.24)
Delayed JOL	0.35(0.26)*	0.58(0.18)
t-test	$t(20) = -2.65$, $p = .02$	$t(20) = -5.71$, $p < .01$ $t(21) = -5.57$, $p < .01$

JOL accuracy

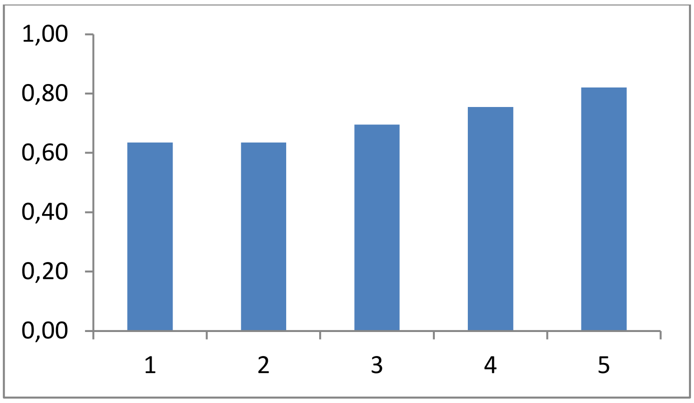
Goodman-Krusal gamma correlation coefficient is an ordinal correlation and a recommended measure of JOL accuracy in the previous studies demonstrating the delayed JOL effect. Gamma correlation ranges between -1 and +1. The rating in delayed JOL predicts better the Pr-score in the recognition memory than the rating in immediate JOL. Recall success in cued recall among the HIT items also showed this effect. That is, in both recognition and cued recall tests, the delayed JOL effect was replicated (See Table 3.1). This way of measuring JOL accuracy is termed "relative JOL" in the literature.

Another way to measure the JOL accuracy is to calculate the ratio between the amount of items rated at any five JOL levels and the amount of items which were then recognized or recalled in the later memory tests. Similar effect was found (data). This way of measuring JOL accuracy is termed "absolute JOL" in the literature.

Recognition memory performance depends on JOL confidence levels

An ANOVA with 2 time and 5 levels of JOL as factors revealed that there is a main effect of JOL confidence level on the JOL accuracy in predicting later recognition memory performance ($F(2.77,58.09)=7.17, p<.01$). There was no main effect of time found ($F=2.56$, $p=.073$).

a)



b)

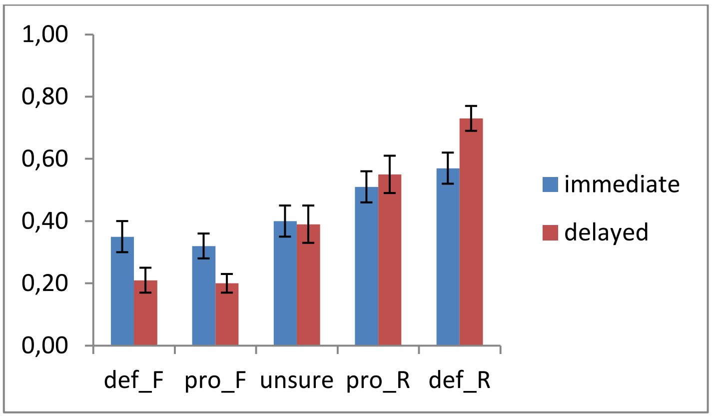


Figure 3.4 a) The mean percentages of correctly recognized items by JOL ratings collapsed across conditions. b) The mean percentages of correctly recognized items by JOL ratings by two JOL time conditions. X-axis from left to right: definitely forget, probably forget, unsure, probably remember, and definitely remember.

Cued recall and JOL

An ANOVA with 2 time and 5 levels of JOL as factors revealed that there is a main effect of JOL confidence level on the JOL accuracy in predicting later cued recall performance ($F(2.04,42.84)=48.13$, $p<.01$), a main effect of time found ($F(1,21)=5.43$, $p=.03$) and an interaction effect between time and JOL time and confidence level ($F(2.81,59,10)=10.18$, $p<.01$).

There are more items which received a low-confidence JOL were later remembered in the immediate condition than in the delayed condition. This finding showed that in the immediate condition, participants were less accurate in their judgment in comparison to the judgments made in the delayed condition. The significant difference for items which were rated with high confidence further supported this finding. Higher proportion of items were recalled received a "definitely remember" rating in the delayed condition ($M=.73, SD=.17$) than in the immediate condition ($M=.57, SD=.26$), $t(21)=-5.18, p<.01$.

ERP Results

Two contrasts were made between the items which were rated with high confidence (4 and 5) versus items with low confidence (1, 2 and 3) by JOL conditions (Figure 3.5).

In the immediate condition, the ERPs evoked by high-confidence JOL and low-confidence JOL started to diverge at around 900 ms after the onset of a JOL. The negative difference is visually more left frontal distribution at 900-1000 ms and then the effect seems to spread to frontal midline electrode sites widely at around 1200-1400 ms. In the late time window 1500-1600 ms, the effect was more left and mid-frontal distributed. However, the ERP amplitudes between the immediate high versus low confidence were not found statistically different in any time window of interest.

In the delayed condition, the ERPs evoked by high-confidence JOL and low-confidence JOL started at early time windows at around 200 ms after the onset of the JOL. The negative difference visually was more at frontal site until 1100 ms and then the effect became more widespread over the whole scalp until 1600 ms. The EPRs for the high-confidence JOL were significantly more negative-going than the ERPs for the low-confidence JOL at the 300-500 time window $F(1,21)=9.69, p=.005$. At a later time window 700-900 ms, this main effect was no longer present. Instead, marginally significant difference was found at the maximal at F3, $t(21)=-1.85$, $p=.08$ and Pz, $t(21)=-1.61$, $p=.07$. Then, the effect remained marginal-significantly maximum at F3, $t(21)=-1.87$, $p=.07$ at 900-1100 ms time window. Taken the findings together, the neural correlates of immediate JOL is distinct from the ones of delayed JOL from visual inspection. In addition, the confidence level in the delayed condition has a marginally significant effect on the ERP waves at F3 and Pz. This effect was not found in the immediate condition. We further conduct a test to examine whether the JOL time condition (immediate, delayed) is modulating the amplitude of the ERP waves.

A global ANOVA with 2 JOL time conditions (immediate, delayed) × 2 JOL confidence levels (low, high) × AP × LAT in six time windows of interest revealed that there is a main effect of JOL conditions on the ERP amplitude in the two early time windows 300-500, $F(1,21)=74,47$, $p=.000$, and 500-700 ms $F(1,21)=16.84$, $p=.001$. This finding points to a differential processes underlying between immediate and delayed JOLs.

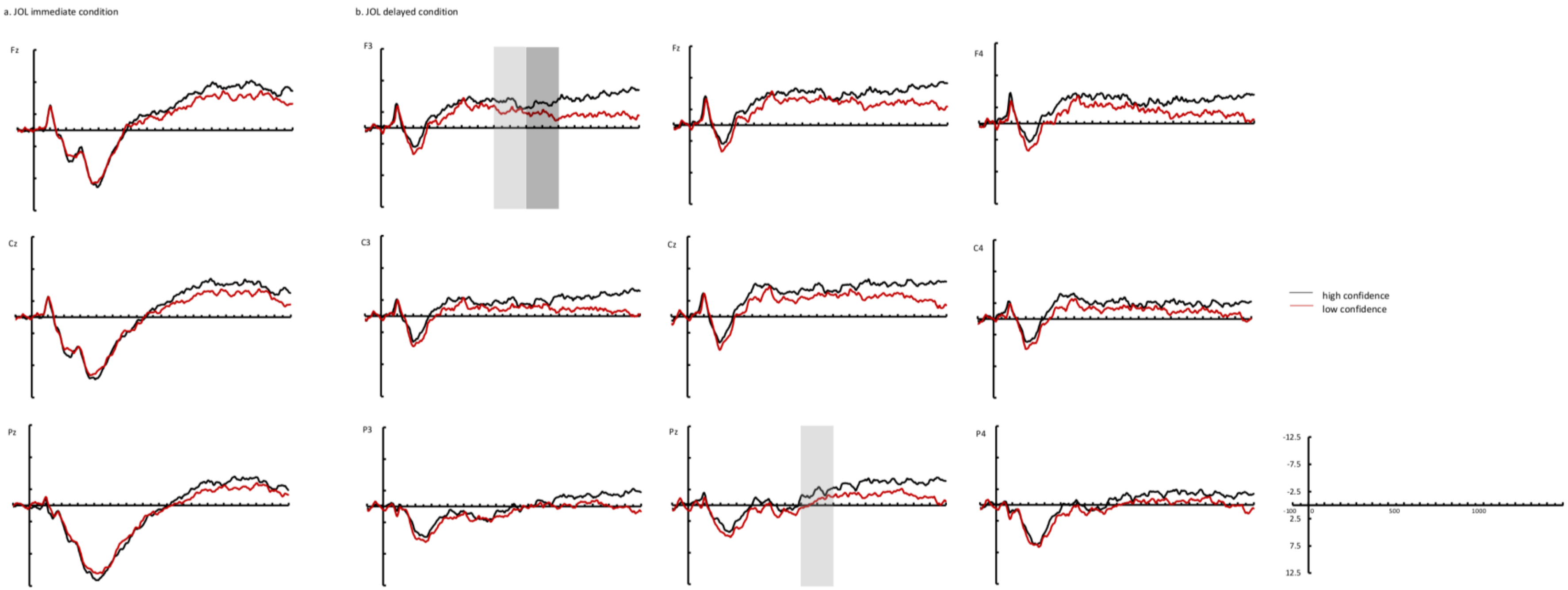


Figure 3.5 ERP waveforms elicited by the onset of a JOL trial during the learning blocks sorted by high versus low confidence ratings in the immediate (left panel a) and in the delayed condition (right panel b). There was no main effect of JOL confidence level on the amplitude in the immediate condition. Therefore, only three central electrodes Fz, Cz and Pz are plotted. All nine electrodes used for the statistical tests are plotted for the delayed JOL condition, including F3,z,4, C3,z,4 and P3,z,4. The epoch is from -100 to 1600 ms. Positive voltage plotted downwards. The shaded time windows mark the marginally significant effects found at F3 and Pz in the delayed condition.

Discussion

This current study uses German-German semantic associative word pair to assess metacognitive judgment on later memory performance. The metacognitive judgments were assessed at two timings: immediate after learning a word pairs “immediate condition” or with some temporal delayed “delayed condition”. Behaviourally, the delayed JOL effect was replicated. In addition, we provide new data comparing two different memory tasks: recognition or cued recall tests on the predictive value of the JOLs. Most importantly, the current study is the first study using ERP technique which time-locked at the onset on a JOL trial to investigate directly the underlying mechanism on metacognitive judgment. Despite the fact that we did not find the ERP signature of recollection in the delayed condition as hypothesized, this study shows that the ERPs at JOLs are modulated by the time of assessing JOL. The finding suggests that the mechanism which supports a delayed JOL might be non-mnemonic. Instead, the predictive accuracy in the delayed JOLs might be related to visual repetition effect or cognitive control or monitoring. These processes provide additional information on top of the mnemonic processes engaged during a metacognitive judgment which correspond to a later memory retrieval outcome better than in the immediate condition where these processes were less engaged.

Behavioural results

Participants were confident in general at the learning phase

The task given in the current study is a German-German word pair association incidental task. The materials included in the experiment are high frequent words and the word pairs are all semantically related. According to Koriat’s cue utilization view on the JOLs, in the immediate condition, it could be likely that participants assess immediately the item information from the German words. In this case, rating item by item, each word pairs are expected to be equally in their difficulty level objectively. Previous studies reported a normal distribution on a 5-point scale for rating JOLs. What we have found is that in the immediate condition, the participants were rather confident in their performance. After some delay, the participants were less confident in their later performance.

Memory performance

The current study employs a paradigm which has recognition memory task and then when a hit-response was given, a cued recall test follows in the same trial. Previous studies investigating the recognition memory usually used the cued recall test to be a secondary test which assesses the confidence level for the primary recognition task result. Or the original studies showing the delayed JOL effect often used cued recall test to assess the final memory performance. Our design mixed the two and we report the two memory tests independently. However, it is clear to us that there must be a carryover effect from the recognition memory test to the cued recall test. The results should be interpreted with caution in both behavioural and electrophysiological data.

The recognition memory test result confirmed the hypothesis that this current task is rather easy for the participants. As emphasized above, delayed JOL effect should be understood independently from the subsequent memory performance. Delayed JOL effect only suggests the predictive value instead of accuracy on later memory performance. The significantly better performance for items which were later rated in the delayed condition in comparison to the ones in the immediate condition can be better explained by the retrieval-practice effect or spacing effect or generation effect as highlighted in this dissertation work. The temporal delay between learning the word pairs for the first time to the delayed JOL trial where one has to activate some degree of retrieval attempt to make the metacognitive judgment. This retrieval attempt might be related to an increased performance in a later recognition task.

Interestingly, this delayed-JOL-benefit-on-later-memory-performance effect did not transfer to the cued recall test. This result is not surprising because it truthfully replicated the finding in Nelson and Dunlosky. However, it makes the speculation above less convincing. As discussed in chapter 1 and 2, we learnt that retrieval practice is most useful for free recall, than cued recall and least on recognition memory. If the JOL time effect observed in the recognition memory test can be explained by the retrieval practice, we should be able to find an even stronger effect in the cued recall test. Alternatively, the transfer-appropriate processing might be a way to account for the different pattern found in the two types of memory tests here. If we re-visit the task participants encountered at JOL, they see a cue word and they had to judge within a short time window, i.e. seconds, how likely they can recall the target word in a later test. Note that all the words in the delayed JOL trials are old items. From the behavioural result, we could not be sure about whether the process at the delayed JOL trial can go very deep in the level of processing. Perhaps there was only an attempt to retrieve the target word, or it could be a mere judgment basing on the signal seeing the cue word. It could be that the process of a JOL at the delayed condition is closer to the one engaged in a recognition memory than in a cued recall test. This is one way to explain why there is an effect of JOL time in the recognition memory test but absent in the cued recall task.

ERP results

There was no effect of JOL confidence on the amplitude in the immediate condition, but in there was an effect found in the delayed condition. From visual inspection, the two JOL time conditions elicited rather different waveforms. There is a N1 and then two positive-going peaks in the immediate condition whilst there is only positive-going peak and this peak is attenuated in comparison to the P2 observed in the immediate condition.

The main prediction is that if the JOL in the delayed condition is supported by a process which is similar to retrieval attempt from an episodic memory formation, we should observe a recollection like ERP signature at around 500-700 ms. The hypothesis is based on an assumption that a recollection-like process is engaged in the delayed condition. Once the information is reinstated, there is a higher likelihood that the participants rate the item with high confidence. Therefore, we hypothesized that the ERPs associated with the high-confident items would be more positive-going than the low-confident items. The finding did not support this hypothesis. Instead, the main effect of JOL confidence level on the ERP amplitude in the delayed condition was found in an early time window 300-500 ms. And then, at a later time window 700-900 ms, a marginally significant effect was observed at left frontal and mid-occipital areas. The left frontal effect remains also at 900-1100 ms time window.

First, we speculate on the reasons why findings do not fit with the hypothesis. Second, explanations for the significant effects found in the delayed condition will be discussed. Third, we outline the major ERP findings which contribute to the understanding of a metacognitive judgment.

A parietal old/ new effect is often obvious in a recognition memory task. In a recognition memory task, participants have to identify if an item was studied before or not. The process which leads to the identification of an old item relies on a recollection process. From the behavioral result, we learn that the memory performance in the recognition memory test, not in the cued recall test is modulated by the JOL time condition. The behavioral result points to the direction that a delayed JOL might share similar processing with the one engaged in the later recognition memory test. However, this shared-processing was not observed in the ERP result.

We also recorded the ERPs elicited by the onset of the cue in a recognition memory test. There we did not find a recollection effect either (Figure 3.6). So in the current study, it is hard to determine whether the processes engaged during the JOL delayed trails is similar to the ones engaged in a recognition memory test or not. The reason why an ERP recollection effect is not observed in the recognition test could be due to the design. When the participants responded old or new, they also had to prepare for the coming cued recall test. The processing was then not comparable to a task where the participants had to only perform an old/ new judgment.

What we observed instead was an early effect at 300-500 ms after cue onset in the delayed JOL condition. The ERPs to the low-confidence items were more positive-going than the ones to the high-confidence items. This effect was observed over the scalp. At this early time window, the process might be related to lexical information. At the crucial time window 500-700 ms, none of the effect was observed. There was a marginally significant main effect of JOL confidence level found ($F=3.19$). It could be the case that the effect is carried over by a later effect where the ERPs elicited by the high-confidence items starting to develop to the negative polarity.

At a later time window 700-900 ms and 900-1100 ms, the ERPs evoked by items rated with high confidence were more negative-going than by those ones with low confidence. This effect was significant at left frontal area. The frontal lobe is associated with metacognition (Fernandez-Duque et al., 2000) and source monitoring (Johnson, Hashtroudi & Linsey, 1993). Left prefrontal cortex was found to be activated in the context where the source information was retrieved (Nolde, Johnson and D'Esposito, 1998; Rugg, Fletcher, Chua and Dolan, 1999). Shimamura and Squire (1986) were able to dissociate metacognition from other memory capacities by showing that metacognitive deficit in Korsakoff patients is not present in other form of amnesics.

The ERPs elicited by the items rated with high confidence were more negative-going than by those ones with low confidence. This effect was marginally significant at left frontal region and mid-posterior region at 700-900 ms. Activities observed at the left-frontal region are associated with cognitive control or decision making. Whilst the activities observed at the mid-posterior region could be related to visual replay occurred in correspondence to an onset of source monitoring or a specific type of metacognitive process, such as the late posterior negativity. A LPN usually is observed in a source monitoring task and has an onset after a recognition decision is made. In addition, the LPN is most pronounced when the extended retrieval processing is required. This interpretation appears to be counterintuitive at the first glance to the current findings. As it was hypothesized, the high JOLs should be easier to recover than the low JOLs. Then, we need to answer the question "Why is it then the case that the EPRs elicited by the high-confidence JOLs were more negative-going than the ERPs by the low-confidence JOLs?" One speculation is that LPN can only search for memory attributes from an available set of encoded information. The high JOL items which were rated with high confidence developed more information at the learning phase than the low JOLs. At the 700~900 ms time window at mid-posterior, the LPN-like component reflects a process where there are more information to be searched and then recovered for the high-confidence than the low-confidence

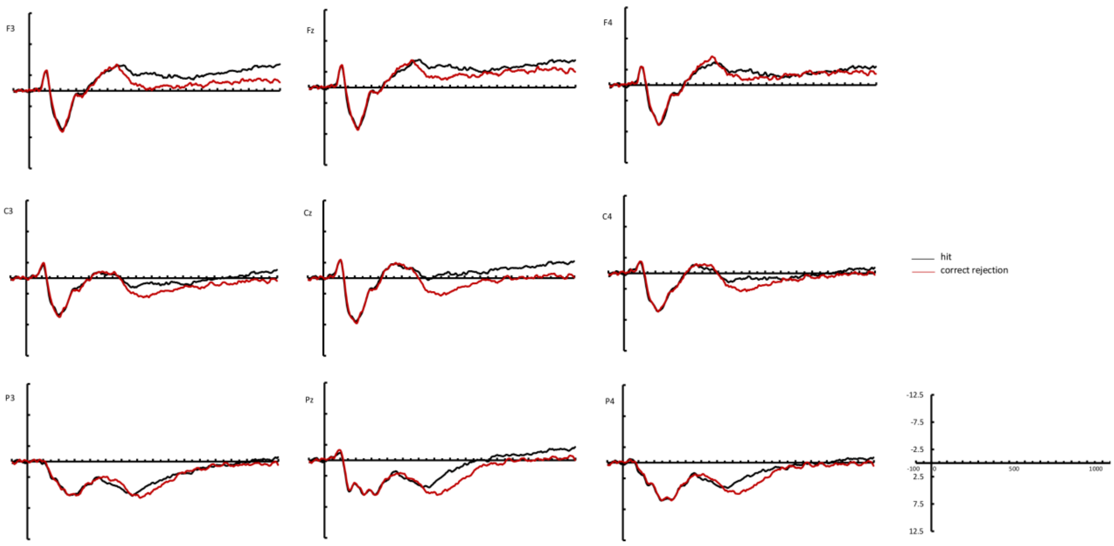


Figure 3.6 ERP waveforms elicited by the onset of a test trial by recognition memory performance: recognized correctly as old (hit) versus correct rejected as new items (correct rejection). The epoch is from -100 to 1600 ms. Positive voltage plotted downwards.

items. Or alternatively, one can argue that a metamemory task encapsulates a memory task. In each JOL trial, participants had to perform a JOL on top of a memory search process. The current finding might reflect an overlapping process which combines a meta-judgment together with a retrieval/recall process. As it has been emphasizing throughout the study, the delayed JOL effect only suggests the accuracy in the metamemory, rather than the accuracy in the memory recall performance. Therefore, the finding here where the high JOLs are less readily to be recovered and thus need continuous search in combination with an initial meta-judgment resulting into a more negative-going waveform than the low JOLs at the 700-900 ms maximal at the mid-posterior site.

In summary, there were no recollection-like processes found at the predicted time window evoked by the JOL rating event. The reason why the prediction was not confirmed might be due to a novel and complex paradigm used: a combined recognition memory task followed by a cued recall test directly. Instead, the findings showed that the monitoring process at the frontal area is greater for the high-confidence than in the low-confidence items which might be the crucial reason why the delayed high JOLs received the most accurate JOL in predicting later memory performance. The higher the monitoring process is engaged, the more accuracy to the judgment of learning is then observed. In addition, the greater activities at the central posterior were greater in delayed high JOLs than in delayed low JOLs. Whether this finding is indicating the replaying of visual information during memory search is a potential explanation which acquires future studies before drawing the conclusion.

Chapter 4

General Discussion

Summary

The current studies have demonstrated that the retrieval processes engaged by testing enhances subsequent declarative memory performance. Testing or retrieval-practice is a good way to acquire information for longer retention. On the other hand, the JOL rating does not encourage an active retrieval process as predicted. Instead, the results pointed to the non-mnemonic monitoring processes engaged by JOL enhance the accuracy of predicting declarative memory.

Responses to previous findings

The testing effect is conditional. Unlike what has been presented in many studies over decades using different materials or type of memory tests, a series of pilot studies and also the actual study failed to find a robust testing effect. This shows that the testing effect is sensitive to specific learning and practice conditions. What is new in this case is to clarify that the testing effect is conditional. Only when the word pairs are learnt properly before the practice (testing/restudy) Phase 2, the testing can enhance or prevent the memory trace to be lost. In addition, the type of task also matters. A testing effect is more often found in a free recall test because the information of list order is not important in free recall test.

Metacognition cannot be reduced to the cognitive processes of learning. In line with previous behavioural study, the delayed JOL effect is only relevant to the predictability of later memory performance; in the contrary the delayed JOL effect is not related to the memory performance itself. For me, this result was counter-intuitive with the testing effect. When a JOL trial is given, participants had to use the “cue” to decide how likely they can remember the target in a later memory test. I hypothesized that this spacing of time between learning and judging the word pair should lead to some kind of retrieval-effect at JOL trial. I failed to observe such a retrieval-related ERP at JOL trials. Given that a null result was found, it is still not certain that there is no retrieval-attempt during a JOL trial which supports the encoding of the word pairs. It could be that the task goal is so specific to judge later recall that this judgment is unaware or irrelevant to later memory performance. What I can suggest here is only that the JOL is a metacognitive process which is more related to monitoring reflected by activities observed in left-frontal areas than related to a parietal retrieval related process. And it remains unclear why a monitoring process does not enhance the word pair memory.

Limitation of the current studies and suggestions for future studies

The limitation of the testing paradigm used in the current project (Chapter 2). Although a robust benefit of testing was demonstrated when the amount of forgetting from Day 1 to Day 2 between testing and restudy conditions were compared, only a marginally significant difference between testing and restudy on Day 2 was observed. This could arise from a number of aspects of the current design. Unlike studies which controlled for learning success before retrieval practice took place (Karpicke & Roediger, 2008; van den Broek et al., 2013), in the current study, participants learned all word pairs three times before the critical manipulation was introduced. Given the large number of test items required to provide sufficient ERP trials, the relatively high task difficulty may have meant some items were not encoded sufficiently within the initial three learning blocks. For those items not learned during Phase 1, once they were assigned to the testing condition, there was a low-likelihood that the items would be recovered. In contrast, the unlearned items from Phase 1 would receive a fourth learning opportunity once they were assigned into the restudy condition (Bahrick & Hall, 1991; Jang, Wixted, Pecher, Zeelenberg, & Huber, 2012). This restriction of the experimental design conveys a disadvantage for tested items over restudied items on Day 1 (Toppino & Cohen, 2009), because the latter are shown again during Phase 2.

In order to make the design more elegant, I would suggest removing the Day 1 recall to make Day 2 recall the sole recall opportunity for the restudying items. In addition, I would also remove the JOL rating at the first learning block and equate the learning time for the three learning blocks.

One has to be cautious when interpreting the ERP testing effects as SME due to the item selection problem of the testing paradigm. When we investigate the testing effect, all effects influencing ERPs during memory retrieval can be mistaken as an encoding effect caused by testing itself. It is principally not possible to decide whether these ERP effects are related to processes that cause different memory strengths or are consequences of different memory strengths caused by former encoding. For example, it is possible that one sees differences in item difficulties. RR items could be easy items, and FF items the most difficult ones. If this happened the ERP effects were then an item selection bias. Due to the high number of items that we used we do not believe that a selection bias was at work but this issue should be more directly addressed in further studies.

The item-selection effect is an intrinsic problem for the testing paradigm. One way to examine whether this effect influences a given study is to observe how each word is recalled across participants/ different learners. As for the ERP of the task effect which contrasting between restudy and testing, this is a controversial problem. The brain waves can be sensitive to many factors. Although I have controlled for the visual complexity in the testing trials by presenting six question markers (as the average word length for the German words in the restudying condition), it is still the case that the visual and semantic information is richer in the restudy condition (Swahili word-German word) than in the testing condition (Swahili word-??????). This is of course not a problem for interpreting the testing effect because it is a design which benefits the restudy trials by providing more information on the visual form level which is against the main effect of interest. However, the two conditions are virtually depending on distinct inputs and task demands. I understand that the difference in task demand is precisely what this study aims to investigate. It is in this logic not a good idea to directly compare the ERPs evoked by the word stimuli at practice Phase 2 in the testing versus the restudy conditions. How we did in our study, to compare three types of testing by memory performance is, therefore, an elegant way to study the testing-provokes-reinstatement hypothesis.

To disentangle the two processes which provide similar outputs to the observers

The testing effect study in Chapter 2 provides a direct link between the neural correlates of the subsequent memory effect and of immediate retrieval at the time point when testing occurs in comparison to restudy condition. Our findings support the episodic context account that testing engages recollection and that enhanced recollective processes improve retention in a later cued recall task. A second possible explanation is that the higher memory strength is, the higher is recallability on a later recall task and this retrieval practice further enhances later memory on a delayed test. Future studies should develop experimental paradigm which can disentangle the two alternatives.

The limitation of the JOL paradigm

Participants use the JOL scale in a different ways. There were two type of JOL scales used in the current project, one is to use a 0, 20, 40, 60, 80 and 100% which has even number of points whilst another is to use a 5-point scale "definitely forget, probably forget, unsure, probably remember, definitely remember." I analysed the JOL rating tendency in the immediate trials to observe rating on individual level. I found that participants use the scale differently. Some tend to assign the rating evenly, or they choose a small set of point to rate, such as using majorly three out of the 6-point scale or 5-point scale.

JOL in delayed trials were given in a different environment from JOL in immediate trials. In the former case, participants kept seeing cue words and had to continuously rate for all the JOL trials.

Each JOL trial was in total 4 seconds, including 3-second cue-word presentation as JOL probe and 1 second ITI. In contrast, in the immediate condition, it was less predictable whether the JOL trial comes after the word pair or not. The total time for a trial in the immediate condition is 11 seconds which includes a 6-sec word-pair presentation, 1-sec blank and 4-second JOL trial. That is, the JOL trials alternated with new word pair learning only in the immediate condition. From the behavioural response point of view, the advantage for the delayed condition over the immediate condition was that there is a fluency and pure processing cost for the JOLs. There was no switch cost between learning and JOL ratings. In the contrary, more motor bias or response bias were likely to be established in the delayed condition than in the immediate condition. In the delayed condition, the experimenters observed that participants needed some time to get into the JOL-only mode in the delayed condition. Therefore, I suggest that there should be filler trials in the beginning of the delayed trials to set participants into the JOL-only mode. More studies are needed to find out an appropriate number of trials as filler.

From the processing point of view, the switch cost is higher in the immediate trial from learning mode to judging mode. There might be interplay between the mnemonic and non-mnemonic system within a short time window. It is still not clear to me whether there is an involvement of mnemonic process during delayed JOL trials. If so, perhaps the temporal intensity from one trial to another trial (3 second JOL presentation, 1 second ITI, continuously). A circle for an immediate trial is longer from learning to JOL (6-second learning, 1 sec ITI, 3-second JOL, 1-sec ITI). If participants in the learning block anticipate a JOL at the time of learning the word pairs, in the immediate trial, they expectations were reached; thus, the processing for a JOL is less effortful. In contrast, in the learn-only trial, their expectation fails and they had to readapt to a learning mode. In the delayed JOL trial itself, there is only one single mode of processing. The single task demand is to keep rating the JOLs in the delayed condition. In this case, the switch cost is low in terms of type of activities. However, it might a more demanding process in the view that the processing time for a JOL is shorter. That is, the JOL decisions have to come faster and then switch to the next cue word within every 3~4 seconds.

A possible problem for the ERP JOL study (Chapter 3) at the test phase was that the task demand was too complicated. The cued recall test after the recognition memory task might lead participant to some strategic position to resolve two stages of memory success in different manners as for one single type of memory test was given. In this case, our finding at the test phase is less easy to be interpreted. I would be interesting in conducting a between-group experiment with one recognition memory test or cued recall test to examine the contribution of JOLs in different types of memory task demand. This might be a way to understand if and how mnemonic process is involved in accessing metacognitive judgments on learning.

Implications

Should testing and quizzes be the best method for teaching in schools?

The robust testing effect seems to suggest frequent testing and quizzes can be useful for students to learn. As we all know that testing and any kind of assessment and evaluation are effective for learning from both an external and intrinsic point of view. Assessment and evaluation can help educators to track learning progress of the learners; meanwhile, they are useful as feedback to the students. On the other hand, over-extensive amount of testing create unnecessary pressure which can hinder the motivation of the learners and it also distorts the purpose of learning. How to develop an effective learning schedule for individual learners still remains to be a question.

Cramming and extensive testing system in East Asian countries

Taiwan, South Korea, and Japan all share one common trait in their educational systems which is testing and cramming. The purpose of education during school years is consequently highly distorted. After one stage of school, there are exams to take and the test scores determine the next school one can attend. The test scores became the most and nearly the only expecting outcome of schooling under this structure. Top grades and often nothing else are viewed as the entrance ticket to professional success. Students tend to learn for the exams rather than learn for the knowledge itself. In Taiwan, Baixiban (cram school) is where most of the pupils and school children spend their after-school time. In Baixiban, the major testing topics are provided, such as English, mathematics, language and literature. Most of the cram schools are situated in the capital city Taipei and about 3,000 cram schools are in the area in the South size of Taipei Main train Station (Chou and Yuan, 2011). At Baixiban, students take lessons and they are provided with repetitive exams. The major reasons why students go to Baixiban after school are mostly related to achieving higher testing scores. In South Korea, a very similar cramming culture is also reported, also known as shadow education. As an analog of Baixiban in Taiwan, the after-hours tutoring academy in South Korea is called hagwons (Ripley, 2011). In 2010, 74% of the students participated in some form of private education. The manifestion of the private tutoring is called juku in Japan. The history of ‘juku’ was analysed in some previous studies where the authors suggested that the ‘juku booms’ might be related to an increasing demand on continuing secondary education (Dawon, 2010; Komiyama, 2000).

Self-paced learning and one-on-one feedback wins against standardized testing

On the contrary to cramming, qualitative teaching was rewarded by reciprocity teaching and learning endeavor. Nancie Atwell is the founder of the Center for Teaching and Learning where the educational aim is not test scores. This center is an independent nonprofit demonstration school located in Maine, USA. This year 2015, Nancie Atwell was rewarded by the Global Teacher Prize. She is an educator who does not believe in test or quizzes. Instead, the teaching method in this center is based on students' choices and one-on-one interaction between teachers and students. This school owns a large body of library. Students choose which books they are going to read. Every year, each student in average read 40 books in this school (Coughlan, 2015). Learning is self-paced. Students do not only choose the books they read, also the topic they want to write about. In addition to the famous writing workshops, the center provides a full curriculum to prepare students for further academic education. For instance, there is a science lab and hands-on learning in science and history.

She said that "I've found, consistently, kids know what's interesting and what's valuable if we let them have some say in it." The role of the teachers under this framework is more like fellow writers and readers. "Teachers are being essentially asked to be technicians, to read a script, and the script is not valid," Atwell said. "[Test scores] are all that counts right now. It's all data analysis, metrics and accountability. It's a business model that has no business being applied to the craft of teaching or the science of learning."

Atwell disagrees with the politically contentious common core educational standards, which she said focus too much on test scores, rather than lessons learned, or books read, as a mark of achievement. Students all learn at different paces and levels, and the common core standards steamrolls individuality and forces everyone to be quite literally on the same page, she said. […] "I think the one thing we had in common, and it was really powerful to see this, was that none of us talked about test scores," Atwell said. "We were talking about making meaningful changes in kids' lives. I am so proud to be a part of a group of people who are professionals in every sense of the word. You just feel proud to be a teacher who was chosen to represent the profession." (Gambino, 2015)

Conclusion

Although the results in the current study showing a beneficial effect of testing, from the examples above, it is clear that education is not about test scores or cramming. Instead, education and learning require high motivation and etiquette feedback and resources. It is thus important to emphasize that the testing effect does not equate to giving more tests or quizzes to students only. In fact, the testing effect proposed in the experimental setting should not be understood as standardized testing or numerically evaluating students' performance. Instead, taken together with the neuroimaging results provided in this dissertation work, one should understand the testing effect as a method of which learners spend time retrieving learnt information. The reinstatement of learnt information is the key mechanism which can aid learning. On top of it, the metacognitive level of processes, including self-monitoring and motivation on learning are as crucial as the retrieval practice itself. The students have to be aware of their learning motivation, agenda and be pro-active in error-correction and receiving feedback. To spend time on the learning materials is also essential given that information needs time to be consolidated and episodic formation also requires time. To constantly reflect on one's own learning behaviour and to actively interact with the learning resources, such as recalling and applying the knowledge into practice, can lead to better memory and effective learning.

References

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. Neuron,65(4), 550-562.

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. Journal of experimental psychology, 81(1), 126.

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. Trends in cognitive sciences, 5(12), 533-539.

Bahrick, H., & Hall, L. (1991). Preventive and corrective maintenance of access to knowledge. Applied Cognitive Psychology, 5(1), 1–18. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/acp.2350050102/abstract

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium (pp. 123–144). Lawrence Erlbaum. Retrieved from http://philpapers.org/rec/BJORAA

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. European Journal of Cognitive Psychology, 19(4-5), 514–527. http://doi.org/10.1080/09541440701326097

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. Nature Reviews. Neuroscience, 9(8), 613–625. doi:10.1038/nrn2459

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. Journal of Experimental Psychology. Learning, Memory, and Cognition, 35(6), 1563–9. http://doi.org/10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. Journal of Experimental Psychology. Learning, Memory, and Cognition, 37(6), 1547–52. http://doi.org/10.1037/a0024140

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. Memory & Cognition, 34(2), 268–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16752591

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. Memory & Cognition, 36(2), 438–48. http://doi.org/10.3758/MC.36.2.438

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

Clayton, N. S., Griffiths, D. P., Emery, N. J., & Dickinson, A. (2001). Elements of episodic–like memory in animals. Philosophical Transactions of the Royal Society B: Biological Sciences, 356(1413), 1483-1491.

Conway, M. A., Pleydell-Pearce, C. W., & Whitecross, S. E. (2001). The neuroanatomy of autobiographical memory: A slow cortical potential study of autobiographical memory retrieval. Journal of Memory and Language, 45(3), 493-524.

Coughlan, S. (2015, March 15). Global teacher prize winner to give away $1m. BBC News. Retrieved from http://www.bbc.com/news/education-31861022

Cycowicz, Y.M., Friedman, D., Snodgrass, J.G., 2001. Remembering the color of objects: an ERP investigation of source memory. Cerebral Cortex 11, 322–334.

D'Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. Consciousness and cognition, 13(4), 844-858.

Davachi, L., Maril, A., & Wagner, A. (2001). When keeping in mind supports later bringing to mind: neural markers of phonological rehearsal predict subsequent remembering. Journal of Cognitive Neuroscience, 13(8), 1059–70. http://doi.org/10.1162/089892901753294356

Dawson, W. (2010). Private tutoring and mass schooling in East Asia: Reflections of inequality in Japan, South Korea, and Cambodia. Asia Pacific Education Review, 11(1), 14-24.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends in Cognitive Sciences, 11(9), 379–86. http://doi.org/10.1016/j.tics.2007.08.001

Donaldson, D.I., Rugg, M.D., 1999. Event-related potential studies of associative recognition and recall: electrophysiological evidence for context dependent retrieval processes. Brain Research, Cognitive Brain Research 8, 1–16.

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. Frontiers in Human Neuroscience, 6, 69. doi:10.3389/fnhum.2012.00069

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The Use of Definite References Signals Declarative Memory Evidence From Patients With Hippocampal Amnesia. Psychological science, 22(5), 666-673.

Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. Nature neuroscience, 9(1), 140-146.

Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J. Dunloksy, and R. A. Bjork (eds), Handbook of memory and metamemory. Psychology Press. pp. 11-28.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. Memory & Cognition, 20(4), 374–80. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1495399

Efklides, A. (2008). Metacognition. European Psychologist, 13(4), 277–287. http://doi.org/10.1027/1016-9040.13.4.277

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: a parametric fMRI study of the testing effect. Neuroscience Letters, 505(1), 36–40. http://doi.org/10.1016/j.neulet.2011.08.061

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spitzer, M., & Walter, H. (2003). Emotional context modulates subsequent memory effect. Neuroimage, 18(2), 439-447.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. Consciousness and cognition, 9(2), 288-307.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34(10), 906–911. http://doi.org/10.1037//0003-066X.34.10.906

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. Nature Reviews Neuroscience, 6(2), 119-130.

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: a selective review. Microscopy Research and Technique, 51(1), 6–28. http://doi.org/10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.0.CO;2-R

Gambino, L. (2015, March 23). The world's best teacher lives in rural Maine and doesn't care about test scores. The Guardian. Retrieved from http://www.theguardian.com/education/2015/mar/23/worlds-best-teacher-nancie-atwell-maine-center-for-teaching-and-learning

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. British Journal of psychology,66(3), 325-331.

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16(6), 978.

Griffiths, D., Dickinson, A., & Clayton, N. (1999). Episodic memory: what can animals remember about their past? Trends in cognitive sciences, 3(2), 74-80.

Gupta, P. (2012). Word learning as the confluence of memory mechanisms: Computational and neural evidence. In M. Faust (Ed.), The Handbook of the Neurospychology of Language, Vol. 1 , pages 146-163. Chichester, England, Wiley-Blackwell, 1st edition.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. Trends in cognitive sciences, 11(7), 299-306.

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. The Journal of neuroscience, 27(52), 14365-14374.

Hutchinson, J. B., Uncapher, M. R., Weiner, K. S., Bressler, D. W., Silver, M. A., Preston, A. R., & Wagner, A. D. (2014). Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. Cerebral Cortex, 24(1), 49–66. doi:10.1093/cercor/bhs278

Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton.

Chou, C. P. & Yuan, J. K. S. (2011). Buxiban in Taiwan. The Newsletter IIAS, 56. Retrieved from http://www.iias.nl/the-newsletter/article/buxiban-taiwan

Jang, Y., Wixted, J. T., Pecher, D., Zeelenberg, R., & Huber, D. E. (2012). Decomposing the interaction between retention interval and study / test practice : The role of retrievability. The Quarterly Journal of Experimental Psychology A, 65(5), 962–975.

Johansson, M., & Mecklinger, A. (2003). The late posterior negativity in ERP studies of episodic memory: action monitoring and retrieval of attribute conjunctions. Biological Psychology, 64(1-2), 91–117. http://doi.org/10.1016/S0301-0511(03)00104-2

Johansson, M., Aslan, A., Bäuml, K.-H., Gäbel, A., & Mecklinger, A. (2007). When remembering causes forgetting: electrophysiological correlates of retrieval-induced forgetting. Cerebral Cortex (New York, N.Y. : 1991), 17(6), 1335–41. http://doi.org/10.1093/cercor/bhl044

Johnson, M. K. (1992). MEM: Mechanisms of recollection. Journal of Cognitive Neuroscience, 4(3), 268-280.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. Science, 319(5865), 966–8. http://doi.org/10.1126/science.1152408

Karpicke, J., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), Psychology of learning and motivation (Vol. 61, pp. 237–284). San Diego, CA: Elsevier Inc.

Kelemen, W. L., & Weaver, C. A. (1997). Enhanced metamemory at delays: why do judgments of learning improve over time? Journal of Experimental Psychology. Learning, Memory, and Cognition, 23(6), 1394–409. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9372607

King, J. F., Zechmeister, E. B., Shaughnessy, J. J., & College, H. (2013). Judgments of knowing : The influence of retrieval practice of Chicago, 93(2), 329–343.

Komiyama, H. (2000). Juku: Before the era of downsizing schools (Juku: Gakkô Surimuka Jidai wo Maeni). Tokyo: Iwanami Shoten.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. Psychological review, 100(4), 609.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. Journal of Experimental Psychology: General, 124(3), 311.

Koriat, A. (1997). Monitoring one's own knowledge during study : A cue-utilization approach to judgments of learning, 126(4), 349–370.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. Journal of Experimental Psychology. Learning, Memory, and Cognition, 40(4), 1–8. http://doi.org/10.1037/xlm0000012

Leynes, P. A., & Kakadia, B. (2013). Variations in retrieval monitoring during action memory judgments: Evidence from event-related potentials (ERPs). International Journal of Psychophysiology, 87(2), 189–199. http://doi.org/10.1016/j.ijpsycho.2013.01.004

Lindsay, D. S., & Kelley, C. M. (1996). Creating illusions of familiarity in a cued recall remember/know paradigm. Journal of Memory and Language, 35(2), 197–211. http://doi.org/10.1006/jmla.1996.0011

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. Contemporary Educational Psychology, 16(2), 192–201. http://doi.org/10.1016/0361-476X(91)90037-L

McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: the effects of cue-guided retrieval processing. Memory & Cognition, 17(4), 423–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2761400

McDaniel, M., & Masson, M. (1985). Altering memory representations through retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(2), 371–385. Retrieved from http://psycnet.apa.org/journals/xlm/11/2/371/

Mecklinger, A. (2000). Interfacing mind and brain: a neurocognitive model of recognition memory. Psychophysiology, 37(5), 565–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11037034

Mecklinger, A., & Müller, N. (1996). Dissociations in the processing of “what” and “where” information in working memory: an event-related potential analysis. Journal of Cognitive Neuroscience, 8(5), 453–73. http://doi.org/10.1162/jocn.1996.8.5.453

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: implications for Korsakoff amnesia. Psychological review, 100(1), 3.

Moscovitch, M., & Nadel, L. (1998). Consolidation and the hippocampal complex revisited: in defense of the multiple-trace model. Current opinion in neurobiology, 8(2), 297-300.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G. & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. Journal of anatomy, 207(1), 35-66.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. Current opinion in neurobiology, 7(2), 217-227.

Neath, I., & Surprenant, A. M. (2005). Mechanisms of memory. Handbook of cognition, 221-238.

Nelson, T. O. (1996). Gamma is a Measure of the Accuracy of Predicting Performance on One Item Relative to Another Item, not of the Absolute Performance on an Individual Item Comments on Schraw (1995). Applied Cognitive Psychology, 10(3), 257–260. http://doi.org/10.1002/(SICI)1099-0720(199606)10:3<257::AID-ACP400>3.0.CO;2-9

Nelson, T. O., & Dunlosky, J. (1991). When people’s judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “delayed-JOL effect.” Psychological Science, 2(4), 267–270. http://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Narens, L. (1994). "Why investigate metacognition?" In J. Metcalfe & A. P. Shimamura (Eds): Metacognition: Knowing about knowing. Cambridge, MA, US: The MIT Press, xiii, 334 pp. 1-25.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(2), 384–413. http://doi.org/10.1037//0278-7393.27.2.384

Nyberg, L., Kim, A. S., Habib, R., Levine, B., & Tulving, E. (2010). Consciousness of subjective time in the brain. Proceedings of the National Academy of Sciences, 107(51), 22356-22359.

O'reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus, 4(6), 661-682.

Otten, L. J., Sveen, J., & Quayle, A. H. (2007). Distinct patterns of neural activity during memory formation of nonwords versus words. Journal of Cognitive Neuroscience, 19(11), 1776–89. http://doi.org/10.1162/jocn.2007.19.11.1776

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. Trends in Cognitive Sciences, 6(2), 93–102. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15866193

Paller, K.A., McCarthy, G., Wood, C. C. (1988). ERPs predictive of subsequent recall and recognition performance. Biological Psychology, 26, 269–276.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? Journal of Memory and Language, 60(4), 437–447. http://doi.org/10.1016/j.jml.2009.01.004

Ripley, A. (2011, Sep 25). Teacher, leave those kids alone. Time. Retrieved from http://content.time.com/time/magazine/article/0,9171,2094427,00.html

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences, 15(1), 20–7. http://doi.org/10.1016/j.tics.2010.09.003

Roediger, H., & Karpicke, J. (2006). The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science, 1(3), 181–210. Retrieved from http://pps.sagepub.com/content/1/3/181.short

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. British Journal of Social Psychology, 31(2), 81-109.

Rosburg, T., Johansson, M., Weigl, M., & Mecklinger, A. (2015). How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. Cognitive, Affective & Behavioral Neuroscience. http://doi.org/10.3758/s13415-014-0310-y

Rosburg, T., Mecklinger, A., & Johansson, M. (2011). Electrophysiological correlates of retrieval orientation in reality monitoring. NeuroImage, 54(4), 3076–3084. http://doi.org/10.1016/j.neuroimage.2010.10.068

Rubin, R. D., Brown-Schmidt, S., Duff, M. C., Tranel, D., & Cohen, N. J. (2011). How do I remember that I know you know that I know?. Psychological science, 0956797611418245.

Rugg, M. D., & Allan, K. E. V. I. N. (2000). Memory retrieval: an electrophysiological perspective. The new cognitive neurosciences, 805-816.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. Trends in Cognitive Sciences, 11(6), 251–7. http://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. Trends in Cognitive Sciences, 4(3), 108–115. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10689345

Rugg, M. D., Otten, L. J., & Henson, R. N. (2002). The neural basis of episodic memory: evidence from functional neuroimaging. Philosophical Transactions of the Royal Society B: Biological Sciences, 357(1424), 1097-1110.

Sanquist, T., Rohrbaugh, J., Syndulko, K., & Lindsley, D. (1980). Electrocoritcal signs of levels of processing: perceptual analysis and recognition memory. Psychophysiology, 17(6), 568–576.

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. Consciousness and cognition, 9(2), 313-323.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1481), 773-786.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. Nature Reviews Neuroscience, 8(9), 657-661.

Skavhaug, I.-M., Wilding, E. L., & Donaldson, D. I. (2013). Immediate judgments of learning predict subsequent recollection: Evidence from event-related potentials. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(1), 159–66.

Son, L. K., & Kornell, N. (2009). Simultaneous decisions at study: time allocation, ordering, and spacing. Metacognition and Learning, 4(3), 237–248. http://doi.org/10.1007/s11409-009-9049-1

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of Learning may alter what they are intended to assess. Psychological Science, 3(5), 315–316.

Spitzer, B., Hanslmayr, S., Opitz, B., Mecklinger, A., & Bäuml, K.-H. (2009). Oscillatory correlates of retrieval-induced forgetting in recognition memory. Journal of Cognitive Neuroscience, 21(5), 976–90. http://doi.org/10.1162/jocn.2009.21072

Squire, L. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. Cognitive Neuroscience, Journal of, 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), 171–7. http://doi.org/10.1016/j.nlm.2004.06.005

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. Current opinion in neurobiology, 5(2), 169-177.

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: a meta-analysis. Neuropsychologia,44(12), 2189-2208.

Tomasello, Michael (2000). "First steps toward a usage-based theory of language acquisition". Cognitive Linguistics. 11.1/2: 61–82.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: questions and answers. Experimental Psychology, 56(4), 252–7. http://doi.org/10.1027/1618-3169.56.4.252

Tulving (1999), On the uniqueness of episodic memory, L. Nilsson, H.J. Markowitsch (Eds.), Cognitive Neuroscience of Memory, Ashland, OH, US: Hogrefe & Huber, pp. 11–42.

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

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. Hippocampus, 8(3).

Van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. NeuroImage, 78, 94–102. http://doi.org/10.1016/j.neuroimage.2013.03.071

van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. Trends in neurosciences, 35(4), 211-219.

Veenman, M. V. J., Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. Metacognition and Learning, 1(1), 3–14. http://doi.org/10.1007/s11409-006-6893-0

Vestergren, P., & Nyberg, L. (2013). Testing alters brain activity during subsequent restudy: Evidence for test-potentiated encoding. Trends in Neuroscience and Education, 3(2), 69–80. http://doi.org/10.1016/j.tine.2013.11.001

Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. Neuropsychologia, 46, 1787–1799. http://doi.org/10.1016/j.neuropsychologia.2008.01.004

Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. Brain Research, 1122(1), 161–70. http://doi.org/10.1016/j.brainres.2006.09.023

Wagner, A. D. (1998). Building Memories: Remembering and Forgetting of Verbal Experiences as Predicted by Brain Activity. Science, 281(5380), 1188–1191. http://doi.org/10.1126/science.281.5380.1188

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. Trends in Cognitive Sciences, 9(9), 445–53. http://doi.org/10.1016/j.tics.2005.07.001

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. Psychological Bulletin, 121(3), 331. Retrieved from http://psycnet.apa.org/psycinfo/1997-03609-001

Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. Neuropsychologia, 51(12), 2360–70. http://doi.org/10.1016/j.neuropsychologia.2013.04.004

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H.-J., & Düzel, E. (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. Neuron, 45(3), 459–67. http://doi.org/10.1016/j.neuron.2005.01.010

Yonelinas, A. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. Journal of Memory and Language, 46(3), 441–517. http://doi.org/10.1006/jmla.2002.2864

Appendices

Appendix A. Material list: 220 Swahili-German word pairs

Appendix B. Material list: 216 German-German semantically-related word pairs

Appendix A. Material list: 220 Swahili-German word pairs for experiment presented in Chapter 2

ID	cue	target	ID	cue	target	ID	cue	target	ID	cue	target
1	Afisa	Offizier	56	Kamata	Klinke	111	Meli	Flotte	166	Sahani	Teller
2	Alama	Fahne	57	Kanisa	Kloster	112	Meneja	Leiter	167	Saidi	Hausmeister
3	Askari	Polizist	58	Karakana	Stahlwerk	113	Mfuko	Paket	168	Sakafu	Etage
4	Aya	Absatz	59	Karatasi	Papier	114	Mfumo	Wolle	169	Saruji	Beton
5	Bahari	Ozean	60	Katibu	Sekretär	115	Mfupa	Knochen	170	Sebuleni	Wohnzimmer
6	Bahasha	Umschlag	61	Keki	Waffel	116	Mgodi	Grube	171	Shaba	Kupfer
7	Baiskeli	Fahrrad	62	Kiatu	Sattel	117	Mhimili	Achse	172	Shajara	Tagebuch
8	Bango	Plakat	63	Kiazi	Kartoffel	118	Mhudumu	Kellner	173	Shayiri	Roggen
9	Bata	Ente	64	Kibanda	Hütte	119	Miako	Flamme	174	Shujaa	Sieger
10	Biri	Zigarre	65	Kidini	Nonne	120	Milima	Gebirge	175	Siagi	Butter
11	Bomba	Pfeife	66	Kidole	Finger	121	Mishumaa	Kerze	176	Sikio	Hörer
12	Breki	Bremse	67	Kifaa	Apparat	122	Mitaro	Graben	177	Simama	Tribüne
13	Bunduki	Pistole	68	Kifaru	Panzer	123	Miti	Linde	178	Simba	Löwe
14	Bustani	Garten	69	Kijitabu	Broschüre	124	Mizigo	Gepäck	179	Simu	Telegramm
15	Chajio	Restaurant	70	Kikombe	Pokal	125	Mkazi	Einwohner	180	Sinema	Kino
16	Chama	Verband	71	Kilima	Hügel	126	Mkulima	Bauer	181	Skeli	Waage
17	Cheti	Urkunde	72	Kimbilio	Bunker	127	Mkurugenzi	Direktor	182	Sungura	Hase
18	Choma	Feuer	73	Kipando	Traktor	128	Moshi	Sauna	183	Sura	Fassade
19	Chuja	Stiefel	74	Kisiwa	Insel	129	Mtaalam	Experte	184	Suruali	Hose
20	Chuma	Eisen	75	Kitabu	Bibel	130	Mtoto	Baby	185	Suti	Anzug
21	Chumba	Schlafzimmer	76	Kiti	Sessel	131	Mtumishi	Diener	186	Taji	Krone
22	Chuo	Hochschule	77	Kiwanda	Fabrik	132	Mvua	Regen	187	Tariki	Fahrbahn
23	Chupa	Flasche	78	Kocha	Trainer	133	Mvulana	Junge	188	Tembe	Pille
24	Degaga	Brille	79	Kofia	Mütze	134	Mvuvi	Fischer	189	Teski	Taxi
25	Dereva	Reiter	80	Kombora	Bombe	135	Mwamba	Gestein	190	Tofali	Maurer

26	Duka	Laden	81	Kondakta	Dirigent	136	Mwandishi	Journalist	191	Tufaha	Apfel
27	Faharasa	Katalog	82	Koo	Kehle	137	Nafaka	Getreide	192	Tufani	Gewitter
28	Fataki	Rakete	83	Koti	Mantel	138	Ndege	Vogel	193	Tumbo	Magen
29	Fimbo	Hebel	84	Kucha	Nagel	139	Ndoo	Kübel	194	Ubao	Tafel
30	Flava	Musiker	85	Kulabu	Haken	140	Ngano	Weizen	195	Ubawa	Flügel
31	Forodhani	Flughafen	86	Kulisha	Futter	141	Ngazi	Treppe	196	Uga	Terrasse
32	Fulana	Weste	87	Kumbusho	Museum	142	Ngoma	Trommel	197	Ujumbe	Botschaft
33	Fundi	Handwerker	88	Kunya	Niederschlag	143	Ngozi	Leder	198	Ukanda	Korridor
34	Funguo	Schlüssel	89	Kupika	Köchin	144	Nguo	Uniform	199	Ukumbi	Halle
35	Gari	Motor	90	Kuruka	Fliege	145	Nguzo	Säule	200	Ukuta	Mauer
36	Gereji	Garage	91	Kuumia	Verletzung	146	Njia	Allee	201	Ulimi	Zunge
37	Gofu	Ruine	92	Lindi	Kanal	147	Njiwa	Taube	202	Unyasi	Rasen
38	Gumba	Daumen	93	Lori	Lastwagen	148	Nyota	Satellit	203	Uombi	Bewerbung
39	Habari	Radio	94	Mabao	Balkon	149	Nyundo	Hammer	204	Ushahidi	Zeugnis
40	Hati	Dokument	95	Madeski	Schreibtisch	150	Ofisa	Beamter	205	Ushairi	Dichter
41	Hatua	Stufe	96	Mafuta	Heizöl	151	Oga	Dusche	206	Uta	Bogen
42	Hekalu	Tempel	97	Mahewa	Klavier	152	Paka	Katze	207	Utenzi	Aktivist
43	Hospitali	Krankenhaus	98	Maji	Wasser	153	Pamba	Orden	208	Uwanja	Stadion
44	Irori	Benzin	99	Makaa	Kohle	154	Pazia	Vorhang	209	Vitunguu	Zwiebel
45	Jambazi	Räuber	100	Malkia	Königin	155	Picha	Fotograf	210	Wakodi	Mieter
46	Jangwa	Wüste	101	Mani	Wiese	156	Pikipiki	Motorrad	211	Walimu	Lehrerin
47	Jela	Gefängnis	102	Maua	Blume	157	Pombe	Alkohol	212	Waridi	Rose
48	Jengo	Gebäude	103	Mboga	Gemüse	158	Pua	Nase	213	Wasia	Testament
49	Jeraha	Wunde	104	Mbolea	Humus	159	Pumzikio	Villa	214	Wawindaji	Jäger
50	Jeti	Hubschrauber	105	Mbosho	Tasche	160	Pundamilia	Zebra	215	Wingu	Wolke
51	Jiko	Ofen	106	Mbunifu	Architekt	161	Punje	Hafer	216	Zahanati	Klinik

52	Joto	Heizung	107	Mchoraji	Maler	162	Rangi	Gemälde	217	Zana	Kamera
53	Jua	Sonne	108	Mchuma	Gewehr	163	Ridhe	Revolver	218	Zawadi	Geschenk
54	Jukwaa	Bühne	109	Mdomo	Lippe	164	Riwaya	Roman	219	Ziara	Gräber
55	Kadi	Einladung	110	Medali	Medaille	165	Rubani	Pilot	220	Ziwa	Schwimmbad

Appendix B. Material list: 216 German-German semantically-related word pairs for experiment presented in Chapter 3

ID	cue	target	ID	cue	target	ID	cue	target	ID	cue	target
1	Akrobat	Zirkus	55	Graffiti	Wand	109	Mascara	Nagellack	163	Schraube	Mechaniker
2	Anhänger	Ohrring	56	Gurke	Markt	110	Meer	Boot	164	Schrubber	Besen
3	Apfel	Korb	57	Hagel	Schnee	111	Messer	Fleischer	165	Schüssel	Becher
4	Aquarellfarbe	Tinte	58	Handschuh	Socke	112	Mikrofon	Sprachrohr	166	Seide	Leder
5	Armband	Geschenk	59	Handy	Rechnung	113	Mikroskop	Lupe	167	Seife	Schaum
6	Aschenbecher	Mülltonne	60	Hemd	Rock	114	Mikrowelle	Geschirr	168	Senf	Tube
7	Backofen	Toaster	61	Hocker	Stuhl	115	Mistgabel	Pflug	169	Sessel	Abend
8	Bagger	Panzer	62	Hose	Bügel	116	Mixtur	Pille	170	Shampoo	Zahnpasta
9	Ball	Sportler	63	Hügel	Berg	117	Mond	Rakete	171	Sitzbank	Holz
10	Banane	Geschäft	64	Huhn	Bauernhof	118	Mosaik	Künstler	172	Ski	Fahrrad
11	Bär	Höhle	65	Iglu	Schlafsack	119	Motte	Lampe	173	Sofa	Bett
12	Bart	Haar	66	Jacke	Pullover	120	Mücke	Spinne	174	Soldat	Matrose
13	Bauer	Fischer	67	Jäger	Hund	121	Mund	Medikament	175	Spatz	Taube
14	Beamter	Verkäufer	68	Jongleur	Clown	122	Muschel	Garnele	176	Speck	Frühstück
15	Beil	Werkstatt	69	Kaninchen	Hamster	123	Nase	Ohr	177	Spinat	Kohl
16	Bergwerk	Arbeiter	70	Kanone	Feuer	124	Notizheft	Schreibtisch	178	Spitzer	Büro
17	Biene	Käfer	71	Karpfen	Angel	125	Ölfarbe	Pinsel	179	Spüle	Handtuch
18	Birke	Vogel	72	Kartoffel	Karotte	126	Orange	Ananas	180	Spülmaschine	Kühlschrank
19	Birne	Aprikose	73	Karussell	Kind	127	Palast	Hofdame	181	Statue	Gemälde
20	Bleistift	Kreide	74	Kassette	Hülle	128	Palme	Schatten	182	Stein	Sand
21	Blinker	Ampel	75	Katalog	Broschüre	129	Papierkorb	Tisch	183	Stempel	Schublade
22	Blumentopf	Eimer	76	Ketchup	Essig	130	Pflaume	Erdbeere	184	Stereoanlage	Radio
23	Boden	Handfeger	77	Kino	Kasse	131	Pilot	Uniform	185	Stern	Planet
24	Bombe	Waffe	78	Kirsche	Plantage	132	Pinie	Bambus	186	Stiefel	Sportschuh
25	Brief	Postbote	79	Kissen	Nacht	133	Prärie	Sonne	187	Stirn	Kinn
26	Brosche	Mädchen	80	Klavier	Noten	134	Professor	Brille	188	Stoppel	Rasierklinge
27	Buchhalter	Formular	81	Klinik	Patient	135	Puder	Lidschatten	189	Tagebuch	Schreibblock
28	Burg	Schloss	82	Koffer	Handtasche	136	Radiergummi	Klebeband	190	Tänzer	Sänger
29	Diskette	Festplatte	83	Kommode	Regal	137	Rasenmäher	Hacke	191	Taschenrechner	Student
30	Dschungel	Affe	84	Kopierer	Computer	138	Rassel	Baby	192	Taxi	Gepäck
31	Dübel	Nagel	85	Krabbe	Strand	139	Rechen	Heu	193	Tee	Tasse

32	Dusche	Badewanne	86	Krähe	Nest	140	Regen	Himmel	194	Telegramm	Postkarte
33	Eichhörnchen	Wald	87	Krankenhaus	Apotheke	141	Regenwald	Gehölz	195	Teller	Suppe
34	Erbse	Garten	88	Krawatte	Hals	142	Ring	Kette	196	Teppich	Vorhang
35	Etage	Umzug	89	Kreisel	Puppe	143	Rollo	Zimmer	197	Theater	Museum
36	Fabrik	Mühle	90	Krug	Kanne	144	Rouge	Frau	198	Tiger	Käfig
37	Fackel	Kerze	91	Kugelschreiber	Schule	145	Rubin	Juwelier	199	Tomate	Küche
38	Fäustling	Finger	92	Lachs	Hering	146	Rücken	Bauch	200	Torte	Bonbon
39	Fernglas	Auge	93	Laken	Decke	147	Rucksack	Bahnhof	201	Traktor	Gelände
40	Fernseher	Kabel	94	Laterne	Dunkelheit	148	Rutsche	Schaukel	202	Treppe	Dach
41	Flasche	Milch	95	Lautsprecher	Konzert	149	Saft	Glas	203	Tulpe	Vase
42	Flöte	Geige	96	Lehm	Baustelle	150	Säge	Hammer	204	Wange	Pickel
43	Flugzeug	Bus	97	Lehrbuch	Binliothek	151	Salzfass	Zuckerdose	205	Wein	Kneipe
44	Fluss	Ozean	98	Lehrer	Ingenieur	152	Sandale	Fuß	206	Winkelmesser	Lineal
45	Foto	Skizze	99	Lenkrad	Motor	153	Saxophon	Cello	207	Wolle	Nadel
46	Fuchs	Wolf	100	Libelle	Blume	154	Schach	Kartenspiel	208	Wurst	Filet
47	Führerschein	Reisepass	101	Likör	Bier	155	Schaf	Schwein	209	Wüste	Wiese
48	Gangschaltung	Fahrer	102	Lilie	Rose	156	Schal	Halstuch	210	Zebra	Elefant
49	Gans	Ente	103	Limonade	Wasser	157	Schaufel	Scheune	211	Zeitschrift	Magazin
50	Gebäck	Konditorei	104	Lippenstift	Spiegel	158	Schauspieler	Bühne	212	Zeitung	Reporter
51	Gießkanne	Pflanze	105	Locher	Tacker	159	Schildkröte	Frosch	213	Zelt	Hütte
52	Giraffe	Zoo	106	Lotto	Gewinner	160	Schlitten	Winter	214	Ziege	Stall
53	Gitarre	Orchester	107	Löwe	Leopard	161	Schnecke	Aquarium	215	Zirkel	Architekt
54	Gletscher	Wanderer	108	Mantel	Garderobe	162	Schrank	Wäsche	216	Zwiebel	Paprika