

NEURAL CORRELATES OF FEEDBACK ENHANCED LEARNING

Vom Fachbereich Sozialwissenschaften
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor rerum naturalium (Dr. rer.nat.)
genehmigte

D i s s e r t a t i o n

vorgelegt von
Petra Ludowicy

Tag der Disputation:	Kaiserslautern, 24. Mai, 2022
Dekanin:	Prof. Dr. Michael Fröhlich
Vorsitzende/r:	Prof. Dr. Florian Böller
Gutachter/in:	1. Prof. Dr. Daniela Czernochowski 2. Prof. Dr. Thomas Lachmann

D 386

Juni 2022

Contents

List of Figures	iv
List of Tables	v
Abbreviations	vi
Chapter 1: Introduction	1
<i>Literature Review</i>	3
The human memory system	3
The explicit memory system	4
Memory consolidation	5
The testing effect.....	7
Feedback in explicit memory research: Test-potentiated encoding.....	10
Theories regarding the testing effect	11
The implicit memory system	12
Feedback in implicit memory research	12
Potential performance feedback effects on TPE.....	14
<i>Empirical methods in memory research</i>	16
Behavioral measures.....	16
Functional magnetic resonance imaging	17
Electroencephalography.....	18
<i>Research Objectives</i>	20
Chapter 2: Timing matters: Immediate performance feedback enhances test-potentiated encoding	23
Abstract.....	24
Introduction	25
Methods	28
Results.....	33
Discussion.....	36
Conclusion	39

Chapter 3: Performance feedback enhances test-potentiated encoding	41
<i>Chapter 3a: Performance feedback enhances test-potentiated encoding: an event-related potential (ERP) study on the testing effect.....</i>	42
Abstract.....	42
Introduction	43
Methods	47
Results.....	53
Discussion.....	61
Conclusion	66
<i>Chapter 3b: Performance feedback enhances test-potentiated encoding: oscillatory correlates of the testing effect</i>	69
Abstract.....	70
Introduction	71
Methods	74
Results.....	77
Discussion.....	80
Conclusion	85
Chapter 4: Functional underpinnings of feedback assisted test-potentiated encoding	87
Abstract.....	88
Introduction	89
Methods	92
Results.....	100
Discussion.....	110
Conclusion	115
Chapter 5: General Discussion.....	117
Comparison of behavioral results of the three studies	117
Influences of performance feedback on theoretical accounts of the testing effect.....	120
Neural correlates of the testing effect and feedback assisted TPE	121
EEG correlates collected in the repetition phase and the final test	122
fMRI results collected on the final test	124
Benefits of multiple methods	125

Toward a multimodal account.....	126
Future perspectives.....	127
Chapter 6: Conclusion	129
References.....	131
Appendix.....	147
Acknowledgements.....	149
Declaration.....	150
Curriculum Vitae.....	151

List of Figures

Figure 1. Schematic procedure of the experiment (Study 1)	30
Figure 2. Accuracy and RT results in the final test (Study 1).....	34
Figure 3. Accuracy and RT results in the final test as a function of retrieval history (Study 1).....	35
Figure 4. Retrieval accuracy and RT results in the repetition phase (Study 1)	36
Figure 5. Schematic procedure of the experiment (Study 2)	48
Figure 6. Accuracy and RT in the repetition phase and final test (Study 2)	55
Figure 7. ERPs during TPE in repetition cycle 1 (Study 2)	58
Figure 8. ERPs during TPE in repetition (Study 2)	60
Figure 9. Oscillatory activity in the final test (Study 2)	78
Figure 10. Oscillatory activity in the repetition phase (Study 2)	80
Figure 11. Schematic procedure of the experiment (Study 3)	94
Figure 12. Accuracy and RT in the repetition phase and final test (Study 3)	102
Figure 13. Brain rendering and axial slice selection (Study 3).....	105
Figure 14. ROI analyses (Study 3).....	106
Figure 15. Pairwise functional connectivity analyses (Study 3)	109

List of Tables

Table 1: Summary of final test results (oscillatory activity, Study 2)	79
Table 2: Summary of the repetition phase results (oscillatory activity, Study 2).....	80
Table 3: Summary of ROI Results (Study 3)	107
Table 4: Summary of functional connectivity results (Study 3)	110

Abbreviations

4AFC	four-alternative forced choice
ACC	anterior cingulate cortex
ANOVA	Analyses of Variance
BOLD	blood-oxygen-level-dependent
BSG	backward strength
CSD	current source density
dB	decibel
DLPFC	dorsolateral prefrontal cortex
DTI	diffusion tensor imaging
EEG	electroencephalography
EOG	electrooculogram
EPSP	excitatory post-synaptic potential
ERO	event-related oscillations
ERP	event-related potential
FA	flip angle
FB	feedback
FDR	false-discovery rate
FFT	fast fourier transformation
fMRI	functional magnetic resonance imaging
FoV	field of view
FRN	feedback-related negativity
FSG	forward strength
FWE	Familywise error rate
FWHM	full-width half-maximum
GLM	general linear model
HRF	hemodynamic response function
Hz	hertz
ICA	Independent Component Analysis
IFG	inferior frontal gyrus
IPL	inferior parietal lobe

IPSP	inhibitory post-synaptic potential
kΩ	kiloohm
LPN	late posterior negativity
LTD	long-term depression
LTM	long-term memory
LTP	long-term potentiation
M	mean
MB	Multiband
MFG	middle frontal gyrus
MNI	Montreal Neurological Institute
MPRAGE	Magnetization Prepared Rapid Acquisition with Gradient Echoes
MRI	Magnetic resonance imaging
ms	millisecond
MSG	mediated strength
MTG	middle temporal gyrus
MTL	medial temporal lobe
MTT	multiple trace theory
OSG	overlapping strength
PFC	prefrontal cortex
RF	radio frequency
rm-ANOVA	repeated measures Analyses of Variance
ROI	region of interest
RPE	retrieval practice effect
RT	reaction time
s	second
SD	standard deviation
SE	standard error
SME	subsequent memory effect
SMG	supramarginal gyrus
SPL	superior parietal lobe
SPM	Statistical parametric mapping
STC	standard theory of memory consolidation
STG	superior temporal gyrus

STM	short-term memory
T	Tesla
tDCS	transcranial direct current stimulation
TE	time-to-echo
TPE	test-potentiated encoding
TR	time-to-repetition
TTT	trace transformation theory
VLPFC	ventrolateral prefrontal cortex
VS	ventral striatum
μV	microvolt

Chapter 1

Introduction

“Has it ever struck you ...

that life is all memory, except for the one present moment

that goes by you so quickly you hardly catch it going?

It’s really all memory ... except for each passing moment.”

(Eric Kandel)

Every day, while awake or asleep, people perceive thousands of experiences and impressions through their senses. All these pieces of information are constantly recorded, filtered, sorted and stored by our brain, all of which is accomplished via the memory system (Atkinson & Shiffrin, 1968). While some of the experiences and impressions are memorized, others are immediately lost. As the memory system is a central part of human being, many research projects have examined the neural basis of memory formation (i.e., encoding), and memory access, (i.e., retrieval) (Schacter, Norman, & Koutstaal, 1998; for review see Schacter & Addis, 2007). Previously, such studies were mostly limited to animals (e.g., Kandel & Tauc, 1965; Lømo, 1966, 2003) or humans with lesions or diseases (e.g., Scoville & Milner, 1957) until neuroimaging techniques have been developed such as functional magnetic resonance imaging (fMRI, see Ogawa, Lee, Nayak, & Glynn, 1990) and electroencephalography (EEG, see Berger, 1930). Such methods provide new insights into structural and temporal aspects of the neural processes underlying the memory system in healthy humans. Therefore, to examine the neural processes of long-term memory formation and retrieval, this PhD thesis investigates behavioral as well as neural correlates collected with EEG or fMRI.

Before elaborating on these studies, the following sections provide a brief introduction to the human memory system and its two sub systems: the explicit and implicit memory systems (Tulving, 1972, 1985). Since the implicit memory system was proposed to affect the

explicit system (e.g., Miendlarzewska, Bavelier, & Schwartz, 2016), the present PhD thesis aimed at investigating their relationship by combining characteristics from both systems. More precisely, performance feedback generally assigned to the implicit memory system is provided in an explicit memory paradigm that induces the testing effect. Hence, the processes underlying explicit memory formation and retrieval are summarized with a detailed description of the testing effect. Furthermore, the implicit memory system is outlined and insights about performance feedback are presented. Afterwards research methods typically used to investigate the explicit and implicit memory systems are outlined.

Literature Review

The human memory system

The human memory system enables individuals to adapt their behavior to environmental changes by learning, maintaining and retrieving information from previous experiences. While some information is only maintained for a short period, other information persists for long time intervals up to several decades, which is referred to as long-term memory (LTM; see Atkinson & Shiffrin, 1968). The information maintained in LTM can be further classified into two sub-systems based on conscious or unconscious retrieval: explicit and implicit memory (see Tulving, 1972, 1985). While the explicit (or declarative) memory system is related to predominantly conscious and intentional retrieval of either previous experiences (episodic memory) or factual information, meanings and concepts (semantic memory), the implicit (or non-declarative) memory system is mainly acquired and retrieved unconsciously and includes perceptual priming, procedural memory and conditioning (Tulving, 1972, 1985).

Historically, the explicit and implicit memory systems were explored separately as they differ in several factors, such as the involvement of brain structures (e.g., Poldrack et al., 2001, Poldrack & Foerde, 2008). However, recent research has discovered that implicit learning can modulate explicit memory formation, suggesting that explicit and implicit memory systems share some commonalities (e.g., Ergo, De Loof, Janssens, & Verguts, 2019; Miendlarzewska, Bavelier, & Schwartz, 2016). For instance, learning from positive and negative feedback has traditionally been categorized as part of the implicit memory system and modulates expectations regarding upcoming events (see Sutton & Barto, 1998). However recently, several studies have observed that performance feedback learning affects episodic memory formation and retrieval (e.g., Ergo et al., 2019; Miendlarzewska et al., 2016; Sinclair & Barense, 2019) and hence supports the notion that concepts of implicit memory affect explicit memory formation. Therefore, further research is needed to examine in detail the relationship between explicit and implicit memory.

As feedback supports the updating of predictions and promotes subsequent learning (e.g., Nieuwenhuis, Holroyd, Mol, & Coles, 2004), four studies that constitute the present research aimed at exploring the effects and underlying neural mechanisms of performance feedback (traditionally examined in implicit memory research) on learning from subsequent correct answer feedback (traditionally examined in explicit memory research) in an episodic memory task. A well-established phenomenon that serves to investigate memory formation and retrieval, as well as learning from correct answer feedback, is the testing effect (for review see Roediger & Butler, 2011). Hence, the typical paradigm structure of research investigating the testing effect provides a solid basis from which to examine the present research question. Accordingly, following a retrieval test, we provide positive or negative feedback indicating retrieval performance. Immediately after the performance feedback, correct answer feedback is presented. In the next sections, the explicit memory system and the testing effect will be described more specifically.

The explicit memory system

Atkinson and Shiffrin (1968) have proposed the multi-store model, according to which the human memory system is comprised of sensory memory, short-term memory (STM) and LTM. In brief, all perceptual input is stored for a very short time in sensory memory. Based on attentional processes, information is either transferred from sensory memory to STM or it is lost. Next, information in STM may be transferred to LTM for storage purposes but only if the information undergoes encoding processes (e.g., through rehearsal). From LTM, the information can be retrieved at some point, which means that it is again available in STM (Shiffrin, & Atkinson, 1969).

The multi-store model by Atkinson and Shiffrin (1968) is a simplification of the memory processes and does not provide information regarding the involvement of different brain regions. Research conducted over the following years explored the mechanisms underlying the “transfer” of memory representations from STM to LTM, suggesting that the hippocampus highly contributes to memory formation (see Eichenbaum, 2004; Squire, Wixted, & Clark,

2007). In detail, each episode that is supposed to be encoded in LTM involves various pieces of information and concepts that already exist in memory, mostly in the cortical system. These already existing representations are activated and then linked to each other, primarily by hippocampal involvement (Moscovitch, 1992; Moscovitch et al., 2005). The hippocampus links the contextual attributes and stores new information related to the episodic context by creating a register. This register provides the opportunity to later access the information and concepts, which can then be combined in the mind to mentally reproduce the episode (i.e., memory retrieval; see Nadel & Hardt, 2011; Nadel, Hubbach, Gomez, & Newman-Smith, 2012).

Memory consolidation

Even before the multi-store model was suggested (Atkinson & Shiffrin, 1968), Müller and Pilzecker (1900) re-discovered that memory formation takes time to stabilize and undergo specific processes, which they termed “consolidation”. However, more recent research has distinguished between synaptic and system consolidation processes (e.g., Asok, Leroy, Rayman, & Kandel, 2019; Dudai, 2004). Synaptic consolidation comprises the neurobiological processes occurring within the first few hours of memory formation (see Asok, Leroy, Rayman & Kandel 2019). Due to repetitive activation of neurons, protein synthesis and membrane potentials are adjusted by several intracellular transduction processes. These trigger a change in gene expression followed by a strengthening (long-term potentiation, LTP) or weakening (long-term depression, LTD) of synaptic transmission. In contrast, system consolidation refers to the subsequent processes in the brain causing the slow reinstatement of memories in the neocortex (see Dudai, 2004; Squire, et al., 2007). These processes were proposed to occur due to hippocampal activity, which repeatedly replays the new information interleaved with reactivation of already existing knowledge in the neocortex.

However, two theories have been introduced that consider the duration of the hippocampal involvement in system consolidation: the standard theory of memory consolidation (STC) and the multiple trace theory (MTT), (see Moscovitch et al., 2005, Nadel & Moscovitch, 1997; Squire, 1986). According to the STC, the involvement of the hippocampus

in memory formation is limited in terms of time (Dudai, 2004; McGaugh, 2000; Squire & Alvarez, 1995). The hippocampus contributes to the generation of associations, which are needed for later retrieval of the episode and to maintain it for a relatively short period. After some time, the memory representations are reorganized, or consolidated, in the cortex. Once this reorganizational stage is completed, retrieval can occur without the involvement of the hippocampus (Squire, 1986; Squire & Alvarez, 1995; Squire, Genzel, Wixted, & Morris, 2015).

Alternatively, the MTT suggests that the memory formation process differs for episodic compared to semantic memory (e.g., Nadel & Moscovitch, 1997; Moscovitch et al., 2005). For episodic memory, the hippocampus is consistently needed to form memories as well as to perform later retrieval. Each time hippocampus-based memories are retrieved, a new memory trace is added to the initial memory representation, which leads to the reinforcement and strengthening of the memory representation. After repeated retrieval iterations, regularities among several memory traces are integrated with preexisting knowledge (semanticization; Moscovitch et al., 2005), which leads to memory representations that are stored without contextual information and are therefore accessible without hippocampal involvement (semantic memory). Hence, the hippocampus is involved in early stages for semantic memory formation, but its role decreases until the memory representation is accessible without hippocampal involvement. In contrast, for episodic memory retrieval, hippocampal involvement continues and is not time-limited (Nadel & Hardt, 2011; Nadel et al., 2012).

Lately, the MTT was refined and extended into the trace transformation theory (TTT), which suggests that the hippocampal involvement during memory retrieval differs based on various factors such as memory strength or retrieval circumstances (Sekeres, Moscovitch, & Winocur, 2017; Winocur, Moscovitch, & Bontempi, 2010; Winocur, Moscovitch, & Sekeres 2007). In line with this theory, prior studies have revealed that factors such as sleep highly affect successful memory formation (sleep consolidation; for review see Klinzing, Niethard, & Born, 2019). Sleep consolidation refers to the memory-related processes during sleep, which result in transformation as well as integration of memory representations in the neocortex by hippocampus-related neural replay (e.g., Dudai, 2012; Klinzing et al., 2019).

In contrast to the previously presented theories about memory consolidation, research has proposed that memory representations are not only stabilized throughout the “lifetime” of a memory. Instead, theories of reconsolidation suggest that memories are based on dynamic processes, resulting in either the strengthening or weakening of these memories (see Nader & Hardt, 2009). In detail, memory retrieval causes the previously fixed memory representation to enter a labile state, which means that interferences might affect upcoming retrieval attempts as memory representations are consolidated once more (see Lee, Nader, & Schiller, 2017).

In the human brain, the explicit memory system engages several areas in the temporal, frontal and parietal cortex, with the left hemisphere more dominantly involved in encoding as well as retrieval processes (Spaniol et al., 2009; Kim, 2016). Particularly relevant regions are the medial temporal lobe (MTL) including the hippocampus in the temporal cortex, the dorsolateral and ventrolateral prefrontal cortex (DLPFC and VLPFC, respectively) in the frontal cortex and superior and inferior parietal lobe (SPL and IPL, respectively) in the parietal cortex.

In summary, explicit memories first undergo encoding and consolidation processes and can later be retrieved. The hippocampus is highly involved in performing these processes. The retrieval additionally leads to reconsolidation processes, which can update or disrupt the memory representations. A memory effect including all these processes is the testing effect, which is further explained in the following section.

The testing effect

To assess levels of knowledge acquisition, tests are regularly performed in educational systems (see Larsen, 2018; Roediger & Karpicke, 2006a). For example, when students study a foreign language, they must learn vocabulary, which comprises several stages. First, students associate a word that does not exist in their native language to a word and its meaning in their native language. Next, they encode this information to LTM for later recall (e.g. Baddeley, Gathercole & Papagno, 1998). Memory recall is examined at a later point in time through vocabulary tests to rate the student’s learning success at, for example, translating a

word from their own language into the foreign language. Hence, students are expected to intentionally recall the translation of the word from memory (see Carpenter & Olson, 2012).

Research has revealed that such testing has several advantages. While the intentional retrieval of information from LTM on a test provides the possibility to rate a student's learning level, testing also improves later recall success (e.g., Larsen, 2018; Roediger & Butler, 2011; Roediger & Karpicke, 2006a, b). Hence, practicing with memory tests can serve as a learning strategy to facilitate knowledge acquisition. The finding that memory tests compared to studying increases later retrieval success is known as the testing effect (or retrieval practice effect, e.g., Karpicke & Aue, 2015; Roediger & Butler, 2011). This effect was discovered and reported earlier than the 1900s (see Abott, 1909) and is highly replicable, as revealed by numerous reviews and meta-analyses (e.g., Adesope, Trevisan, & Sundararajan, 2017; Karpicke & Aue, 2015; Rowland, 2014; Schwieren, Barenberg, & Dutke, 2017). Furthermore, the testing effect is not limited to the laboratory environment but instead can be replicated in group experiments in schools or universities (e.g., Rohrer, Taylor, & Sholar, 2010; Vojdanoska, Cranney, & Newell, 2010).

In general, research investigating the testing effect is commonly structured in three phases: an initial study phase, followed by a practice phase and, lastly, a final test phase (see Roediger & Butler, 2011; Rowland, 2014). In the initial study phase, the learner is presented with the learning material for the first time and is expected to study and encode the material. In the example of vocabulary learning, the learner is provided with all words in his or her native language together with their translations into the foreign language. Next is a practice phase, in which the learner practices the material through either studying or testing. In regard to vocabulary learning, the learner is either presented with the words and their translations again and is expected to study these. In contrast, if the learner is tested, then only the word is presented in the native language without its translation and the learner is supposed to retrieve the translation from memory. Lastly, a final test is performed independently of repetition practice (testing or studying) with only the word being presented without the translation. Typically, the testing effect describes the difference in retrieval success on the final test based

on prior practicing by testing compared to studying (Roediger & Butler, 2011). Moreover, other beneficial effects such as improved metacognitive monitoring assessed with judgments of learning (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Dunlosky & Nelson, 1992) or increases in retrieval speed measured by reaction times (RT) have been reported as testing effects as well.

Several factors have been cited as affecting the magnitude of the testing effect, which is the retrieval benefit as a result of prior testing minus the benefit due to prior studying (see Rowland, 2014). Examples of these modulators are retrieval tasks, retention intervals, variability in learning material or the number of repetitions. Hence, while designing an experiment that aims at investigating the testing effect, these factors should be considered:

- Regarding the retrieval task, prior studies have employed various tasks such as cued recall (e.g., Kornell, Bjork, Garcia, 2011; Pyc & Rawson 2010, 2011), free recall (Fritz, Morris, Acton, Voelkel, & Etkind, 2007; Karpicke & Blunt, 2011; McConnel & Hunt, 2007), and recognition tasks (e.g., Jacoby, Wahlheim, & Coane, 2010, Rohrer et al., 2010; Wartenweiler, 2011). Some studies even changed the retrieval task in the repetition phase compared to the final test (e.g., Mulligan, Buchin, & West, 2020; Zaromb & Roediger, 2010). However, meta-analyses have revealed increased testing effects for recall tasks compared to recognition tasks.
- The retention interval is defined as the delay between the repetition phase and the final test. While some studies have investigated immediate final test retrieval success or with only a short delay of several minutes (e.g., Carpenter, 2009; Toppino & Cohen, 2009; Vestergren & Nyberg, 2014), others increased the delay to one day or week (e.g., Keresztes, Kaiser, Kovács, & Racsmány, 2014; Racsmány, Szöllősi, & Bencze, 2018) up to several months (Eriksson, Kalpouzos, & Nyberg, 2011). However, the magnitude of the testing effect was reported to increase with longer delays (Rowland, 2014).
- The testing effect was examined with various types of learning materials. While some studies used paired associates such as word pairs or vocabulary, others used

word lists or prose texts (e.g., van Eersel & Verkoeijen, 2016). Notably, the testing effect was mostly studied on but not limited to verbal material, as revealed by studies using pictures (Carpenter & Olson, 2012; Nguyen & McDaniel, 2015). Research has shown that the testing effect can be measured for all types of materials mentioned above (Rowland, 2014).

- Another factor modulating the testing effect is the number of initial tests. While some studies have provided only one repetition cycle (e.g., Carpenter, 2009), others used for instance seven cycles (e.g., van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013). This difference mostly depends on the difficulty of the learning material: whereas highly associated word pairs (i.e., the cue to target association strength is high, e.g., “toast - bread”) can be learned with only one repetition (Carpenter, 2009), learning vocabulary is difficult and hence requires several repetition cycles (e.g., four study cycles and three test or study cycles in the study by van den Broek et al., 2013). Indeed, several repetition cycles led to increased retrieval success (Zaromb & Roediger, 2010).

Feedback in explicit memory research: Test-potentiated encoding

Correct answer feedback is another factor that modulates the magnitude of the testing effect (Butler, Godbole, & Marsh, 2013; Butler & Roediger, 2007, 2008; Pan, Gopal, & Rickard, 2016). In other words, the presentation of the learning material following a testing situation enhanced later retrieval success compared to only testing (Arnold & McDermott, 2013; Izawa, 1966, 1971; Rickard & Pan, 2018). This effect is called test-potentiated learning (e.g., Arnold & McDermott, 2013) or test-potentiated encoding (TPE; van den Broek et al., 2016). According to the literature, encoding abilities are enhanced following prior testing, possibly because the feedback can be used to correct errors and validate correct responses given with low confidence (Butler & Roediger, 2007, 2008). The presentation of the learning material is equal to a study opportunity; hence, alterations of study and test cycles have sometimes not been termed as feedback in prior literature even though it is conceptually the same (Rowland, 2014,

see Appendix 1). Especially in repetitive study and test cycles, correct answer feedback was provided mostly block-wise instead of immediately following the test.

Theories regarding the testing effect

Several theories have been proposed offering an explanation for the cognitive mechanisms underlying the testing effect. Some of these characterize the magnitude of the testing effect in detail, providing the possibility to create assumptions on the relatedness of various modulating factors, while other hypotheses suggest cognitive mechanisms underlying the testing effect (see Rowland, 2014; van den Broek et al., 2016). These mainly focus on two accounts: the (semantic) elaboration account and the search-set restriction account.

The elaboration account indicates that through practicing by retrieval tests (but not by studying), the learner actively searches for the correct response in memory (see Carpenter, 2009; Carpenter & DeLosh, 2006). Until the correct response is retrieved, this memory search leads to the activation of similar and/or related concepts. Next, these concepts are connected to the concept of the correct response either by the creation of new connections or by the strengthening of already existing connections that developed during the first encoding, resulting in an elaboration of the mental representation of the learning material. These elaborations facilitate successful memory retrieval at a later time point since more retrieval routes are available. For example, when studying the word pair “animal - cat”, retrieval while testing (“animal - ?”) leads to the activation of related concepts such as “dog”, “lion” or “kitten” before retrieving the word “cat”. When aiming to recall the word “cat” at a later time, the other concepts might provide additional memory cues for the target word (see Rowland, 2014, for review).

The search-set restriction account also starts with an active retrieval attempt initiated by testing but not by studying (e.g. Karpicke & Zaromb, 2010; Thomas & McDaniel, 2013). Again, until the correct response is retrieved, similar and/or related concepts are activated resulting in a selective weakening of their relevant cue-target associations. In contrast, the activation of the correct response causes a selective strengthening of the cue-target

association. Hence, testing helps to refine the memory representation. Referring to the previous example, retrieval while testing (“animal -?”) leads to a selective weakening of other associations such as “animal - dog” or “animal - lion”, decreasing the probability that these concepts will be activated again in future tests. In contrast, successful retrieval of the association “animal - cat” from memory results in strengthening this association (see Rowland, 2014, for review).

The implicit memory system

As previously stated (page 3), the implicit memory system is acquired and retrieved unconsciously and includes perceptual priming, procedural memory or conditioning (see Butler & Berry, 2001). Research in this field has frequently investigated associative and non-associative learning. In associative learning, two stimuli (in classical conditioning) or several consequences following a specific behavior (in operant conditioning) are associated to another, whereas non-associative learning refers to a change in the strength of response to a single stimulus due to repeated exposure. A large portion of research on associative learning explores feedback processing (e.g., Dickinson, 2012; Nieuwenhuis et al., 2004). In this research conglomerate, feedback is mostly either positive or negative and hence, varies in content compared to other types of feedback such as correct answer feedback (see test-potentiated encoding, page 10).

Feedback in implicit memory research

Research investigating the underlying mechanisms of positive and negative feedback processing differ depending on the relationship between prior cues, the following behavior and the feedback. In more detail, several paradigms provide feedback independent of prior cues or performance, such as gambling tasks (e.g., Hajcak, Moser, Holroyd, & Simons, 2007; Marco-Pallares, Cucurell, Münte, Strien, & Rodriguez-Fornells, 2011; Peterburs, Suchan, & Bellebaum, 2013), which do not allow the participants to associate a cue or behavior to the feedback. Hence, the participants are unable to create expectations on feedback outcome. In

contrast, other studies have examined paradigms that provide a specific feedback based on the preceding behavior following a cue, termed performance feedback (e.g., Chase, Swainson, Durham, Benham, & Cools, 2011; Magosso, Forcelli, Garofalo, Di Pellegrino, & Ursino, 2015; Miltner, Braun, and Coles, 1997). Here, the participants can learn to associate the feedback with their behavior and generate expectations for upcoming events. Therefore, performance feedback offers the possibility to facilitate learning by providing external information on prior performance, which leads to the adaptation of future behavior.

However, it should be noted that these expectations rely on the specific relationship among the cue, the subsequent behavior and the resulting feedback. While some studies have always provided the same feedback for a specific behavior after a given cue (e.g., Ludowicy, Czernochowski, Weis, Haese, & Lachmann, 2019), others have presented a changing feedback for the same cue and behavior, which interrupts the behavior-feedback contingency. In probabilistic learning tasks, for example, a cue results in a behavior and causes a specific feedback in 75% and different feedback in 25% of the trials (e.g. Pfabigan, Alexopoulos, Bauer, & Sailer, 2011). Thus, the contingency between a behavior and a feedback outcome can be modified. Nonetheless, the present research exclusively investigates positive and negative feedback contingent to correct or incorrect performance, respectively.

Other research has revealed that positive feedback increases whereas negative feedback decreases relevant behavior (e.g., Sutton & Barto, 1998), but learning from feedback has recently been hypothesized to considerably rely on prior predictions (for review see Ullsperger, Fischer, Nigbur, & Endrass, 2014). If the feedback outcome matches feedback prediction, then learning does not occur, whereas if both do not correspond to each other, future predictions will be corrected based on the feedback. The difference between feedback outcome and prior expectation is referred to as a reward prediction error. As rewarding effects of performance feedback were detected even in the absence of rewards, such as monetary or social reward (e.g., Daniel & Pollmann, 2010; Miendlarzewska et al., 2016; see also review by Ferdinand & Czernochowski, 2018), knowledge gained in the context of reward-prediction errors provides beneficial insights relevant to the present research.

Reward learning and reward prediction error are modulated mostly by the dopaminergic system in the brain (e.g., Schultz, 1998; Shohamy & Adcock, 2010). The reward system includes areas such as the ventral tegmental area, the ventral and dorsal striatum, the cingulate cortex (mostly anterior portion), the substantia nigra of the basal ganglia, the insula, portions of the prefrontal cortex, the amygdala, the hippocampus and the thalamus among others (e.g., Berridge & Kringelbach, 2015; Miendlarzewska, 2016). The ventral striatum was especially cited to strongly contribute to reward learning ranging from simple stimulus-action-outcome learning to more complex learning tasks (see Daniel & Pollmann, 2014). On the biological level, successful learning of cue-outcome associations was proposed to depend on LTP at corticostriatal synapses (e.g., Waldschmidt & Ashby, 2011). Synaptic consolidation is based on LTP and LTD (for further information, see page 5).

Potential performance feedback effects on TPE

Recently, performance feedback was suggested to modulate more complex types of learning linked to complex cognitive processes, such as memory LTM formation (see Sinclair & Barense, 2019; Ergo et al., 2019; Ernst & Steinhauser, 2012), which are sensitive to internal and external cognitive factors (such as attention or motivation, see Miendlarzewska et al., 2016). Consequently, the present thesis aimed at investigating performance feedback effects on explicit LTM formation in more detail by providing performance feedback in a testing effect paradigm combined with correct answer feedback (TPE).

Performance feedback may influence the testing effect and TPE by modulating several factors. In detail, the testing effect has been proposed to depend on mainly two accounts: the semantic elaboration account (Carpenter, 2009; Carpenter & Delosh, 2006) and the search-set restriction account (Karpicke & Zaromb, 2010; Thomas & McDaniel, 2013). On the basis of the semantic elaboration account, the semantic representation could be extended by the representations of positive or negative feedback, which could then serve as additional cues to access the representation during the next retrieval attempts. In contrast, performance feedback either selectively strengthens or weakens specific associations (see reward

prediction error and reinforcement learning, see Sutton and Barto, 1998), which might also assist with search-set restriction as the underlying principles of feedback learning and search-set restriction have been thought to share some commonalities (e.g., Ergo et al., 2019) such as their neural basis. As previously stated, the reward system and the episodic memory system are built on hippocampal activation, which is regulated by dopaminergic activity (e.g., Miendlarzewska et al., 2016; Shohamy & Adcock, 2010). Hence, performance feedback may assist with search-set restriction by adjusting dopaminergic activity.

Empirical methods in memory research

To explore the mechanisms underlying learning and memory, several measures can be investigated. Behavioral measures including retrieval success or RT as well as neuroimaging techniques including functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), provide insights into research questions such as which brain structures are involved and when specific processes occur. The basic principles of these methods and examples are discussed in detail in the following sections.

Behavioral measures

Behaviorally, retrieval success and RT are typical measures to investigate learning outcome in episodic memory and reward learning research (e.g., Butler et al., 2013; Ludowicy et al., 2019; van den Broek, Segers, Takashima, & Verhoeven, 2014). Retrieval success measures the amount of information that is successfully retrieved from memory compared to unsuccessfully retrieved information (see Roediger & Butler, 2011; Rowland, 2014). In comparison, RTs, which are measured from cue onset until a response is provided, have been associated with the time needed to successfully retrieve the information from memory as an indicator of retrieval difficulty (van den Broek et al., 2014; Wixted & Rohrer, 1993). In the testing effect literature, retrieval success is the more prominent of these two measures and it has been revealed that prior practicing by testing compared to studying enhances retrieval success, especially with a delay of one day and more. Recent research has proposed that practicing by testing more than studying enhances RTs for successfully retrieved items on a delayed test as well (van den Broek et al., 2014). The decrease in retrieval speed was suggested to reflect improved cue-response associations.

Overall, such behavioral measures are relatively simple and low-cost tools compared to neuroimaging techniques. However, they only provide information about memory retrieval, but the neural processes cannot be examined in this way. In contrast, neuroimaging techniques such as fMRI and EEG can provide such a perspective.

Functional magnetic resonance imaging

Magnetic resonance imaging (MRI) is a noninvasive imaging technique that uses the magnetic properties of certain molecules to create images of specific tissues of organs in the body such as the brain (e.g., Hoult & Bhakar, 1997; Schenck, 1996). To generate these images, a strong magnetic field (in neuroimaging, mostly 3 T) is needed. Certain parts of molecules placed in this magnetic field can absorb radio frequency (RF) energy, which again results in specific spin polarizations that transmit an RF signal that can be detected with RF coils and then converted to images (Raichle, 1994).

In neuroscience, fMRI is used to rapidly generate images of the brain that can be used to quantify blood flow, which is related to brain activity (for review see Logothetis, 2008; Ogawa et al., 1990). Generally, neurons in the brain need glucose and oxygen to be able to transmit the electrical signal and these resources must be refilled quickly after activation. To accomplish a specific task, many neurons are collectively activated (i.e., transmit electrical signals). This activation causes a high demand for resources (e.g., oxygen) that are transported by the blood system to the previously activated brain areas. This leads to an increase in blood flow in the specific brain region, which causes a change in blood oxygenation, referred to as the hemodynamic response that can be observed by blood-oxygen-level-dependent (BOLD) contrasts (Fischl, Sereno, & Dale, 1999; see Logothetis, 2008, for review). BOLD contrasts measure the change in concentrations of oxyhemoglobin and deoxyhemoglobin due to their different magnetic properties. Oxygenated hemoglobin is diamagnetic and hence, only slightly interferes with the RF signal, indicating increased neural activity. Conversely, deoxygenated hemoglobin is paramagnetic and therefore creates an inhomogeneous field that strongly interferes with the RF, resulting in less signal detection. Research using fMRI has mapped BOLD changes to specific experimental events to detect activation patterns of particular brain regions depending on different experimental conditions. For example, in a testing effect study, increased activity in VLPFC was reported while practicing by testing compared to studying (van den Broek et al., 2013).

In addition to BOLD activation levels, recent research has proposed functional connectivity analysis to examine the temporal dynamics of BOLD activity (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Rissman, Gazzaley, & D'Esposito, 2004). Therein, the hemodynamic response function (HRF) modeling neurovascular coupling can be fitted to each experimental trial, resulting in statistical parameter estimates (beta values, for review see Buckner, Krienen, & Yeo, 2013; Mohanty, Sethares, Nair, & Prabhakaran, 2020). Next, these estimates can be used to produce condition-specific beta series for each voxel. The temporal level of coactivation between anatomically separated brain regions can differ and it has been theorized that high synchronization between different brain regions reflects improved communication among these regions, resulting in information integration (van den Heuvel & Pol, 2010). For example, increased coactivation of hippocampus and supramarginal gyrus (SMG) was detected in a testing effect study comparing tested and studied items (Wing, Marsh, & Cabeza, 2013).

Electroencephalography

Even though fMRI offers a great spatial resolution, its temporal resolution is relatively slow since it indirectly measures neural activity by measuring differences in blood flow which leads to a delay. Additionally, the acquisition of fMRI images is relatively slow, taking at least 1 s at present. In contrast, EEG directly captures changes in the electrical field on the scalp due to cortical activity with a high temporal resolution of for instance 500 Hz but with relatively low spatial resolution (e.g., Gevins, Leong, Smith, Le, & Du, 1995; Ward, 2003).

In more detail, several electrodes (mostly 32-, 64- or 128-channel systems) are placed on the participant's scalp to measure the electrical fluctuations of the cortical surface layer, comprising mostly pyramidal cells, to a reference electrode (e.g., Gevins et al., 1995; Ward, 2005). As mentioned previously, neural activation leads to synchronous firing patterns of neurons, resulting in simultaneous post-synaptic potentials, which are either excitatory or inhibitory (EPSP/IPSP, see Hölscher, Anwyl, & Rowan, 1997). Depending on the number of neurons activated simultaneously and the alignment of these neurons in the gyri and sulci of

the cortex, the EEG can capture the neural activation patterns but with low spatial resolution compared to MRI. Similar to fMRI, the EEG signal can be analyzed with respect to specific experimental events. To increase signal-to-noise ratio, averages of the activity following several trials are analyzed by investigating either event-related potentials (ERPs) or event-related oscillations (EROs) (see Cohen, 2014; Luck, 2005, 2014).

ERPs are averaged epochs of neural fluctuations time-locked to events in an experiment (e.g. Bae & Luck, 2019; Luck, 2005, 2014). These neural fluctuations consist of a series of positive and negative deflections, which can be characterized by their amplitudes and latencies. Previous research has related specific ERP components to distinct cognitive processes (e.g., Miltner et al., 1997; Polich, 2007). For example, every time a negative feedback is presented compared to a positive one, a negative deflection can be detected around 250 ms following feedback onset, referred to as the feedback-related negativity (FRN), which has been suggested to reflect cognitive processes related to feedback categorization as either subjectively positive or negative (e.g., Ludowicy et al., 2019; Miltner et al., 1997; Peterburs, Kobza, & Bellebaum, 2016).

In addition, ERPs consist of a combination of several deflections with different frequencies, referred to as oscillations (e.g., Cohen, 2014; Herrmann, Rach, Vosskuhl, & Strüber, 2014; Herrmann, Strüber, Helfrich, & Engel, 2016). With filtering methods such as bandpass filtering or fast Fourier transformation (FFT), the ERPs can be offline decomposed into specific frequencies mostly ranging around 1 - 90 Hz. Furthermore, temporal properties of these frequency decompositions can be analyzed by complex wavelet analysis (Bartnik, Blinowska, & Durka, 1992; Samar, Bopardikar, Rao, & Swartz, 1999). Similar to ERPs, specific frequency bands of the oscillatory activity measured following an event with a specific delay have been associated with various cognitive processes. For example, early theta activity was increased for remembered compared to forgotten stimuli and hence it was hypothesized to reflect memory processing (Klimesch, 1999; Klimesch, Freunberger, & Sauseng, 2010).

Research Objectives

The present research aims at investigating the effect of performance feedback on learning from correct answer feedback in a testing effect paradigm (which we refer to as feedback assisted TPE). Consequently, participants learned low-associated word pairs first in an initial study phase and then further practiced these word pairs in a repetition phase. While practicing, word pairs were either tested or studied, followed by correct answer feedback. For half of the tested trials, additional performance feedback was provided immediately following the test and before the presentation of correct answer feedback. After a delay of one day, participants performed a final test. Behavioral as well as neuroscientific methods were applied to examine the cognitive processes affected by feedback assisted TPE, in the repetition phase as well as in the final test.

This PhD thesis includes three experimental studies. The first experiment (referred to as Study 1) behaviorally investigates the previously outlined paradigm. Retrieval success and RTs in the repetition phase (Day 1) as well as the final test (Day 2) are analyzed to explore whether and at which point additional performance feedback affects learning from TPE (Chapter 2). In Study 2, behavioral and EEG data were collected throughout the entire experiment. The objective is to explore direct effects of performance feedback on learning from correct answer feedback by analyzing ERP correlates collected while correct answer feedback was provided in the repetition phase (Day 1; Chapter 3a). Moreover, direct and delayed effects of additional performance feedback on processes involved in memory retrieval are investigated by exploring oscillatory correlates gathered during the tests in the repetition phase (Day1) as well as the final test (Day 2; Chapter 3b). In the third experiment (Study 3), the paradigm is adapted to the specificities of fMRI research. In the final test (Day 2), fMRI data were collected to investigate delayed performance feedback effects modulating the final test with (1) whole-brain analyses, (2) regions of interest (ROI) analyses and (3) functional connectivity analyses (Chapter 4). Even though the paradigm only varies in minor details to adapt it to the specificities

of each method, all chapters provide the necessary theoretical background and thus can be read separately.

Study 1 and 2 (behavioral and EEG) have been carried out at the University of Kaiserslautern, Germany, whereas Study 3 (fMRI) has been conducted at the Basque Center on Cognition, Brain and Language located in San Sebastian, Spain. Furthermore, all studies are in preparation for publication as follows:

- Chapter 2: “Timing matters: Immediate performance feedback enhances test-potentiated encoding.” by Petra Ludowicy, Pedro M. Paz-Alonso, Thomas Lachmann, and Daniela Czernochowski
- Chapter 3a: “Performance feedback enhances test-potentiated encoding: an event-related potential (ERP) study on the testing effect.” by Petra Ludowicy, Pedro M. Paz-Alonso, Thomas Lachmann, and Daniela Czernochowski
- Chapter 3b: “Performance feedback enhances test-potentiated encoding: oscillatory correlates of the testing effect.” by Petra Ludowicy, Pedro M. Paz-Alonso, Thomas Lachmann, and Daniela Czernochowski
- Chapter 4: “Functional underpinnings of feedback-enhanced test-potentiated encoding.” by Petra Ludowicy, Daniela Czernochowski, Jaione Arnaez-Telleria, Kshipra Gurunandan, Thomas Lachmann, & Pedro M. Paz-Alonso

Chapter 2

Before investigating the neural mechanisms of feedback assisted TPE, the following chapter provides a detailed description of Study 1. In this study, we collected behavioral data while participants performed an episodic memory paradigm inducing the testing effect. Previously, many studies investigated the testing effect only by examining retrieval accuracy. However, recent research observed that the testing effect modulates RT as well, reporting faster RT for previously tested compared to studied material. Consequently, both, recall accuracy and RT, were investigated in the following data set. An initial learning phase and a repetition phase were performed during the first session and a final test was performed in a second session with a delay of 1 day between the first and the second session. Accordingly, we investigated retrieval accuracy and RTs of the repetition phase on Day 1 and the final test on Day 2.

Timing matters: Immediate performance feedback enhances test-potentiated encoding

Abstract

Long-term memory retention is enhanced after testing compared to studying (testing effect). Memory retrieval further improves when correct answer feedback is provided after the retrieval attempt (test-potentiated encoding - TPE). Before providing correct answer feedback, additional performance-contingent feedback was provided to evaluate whether explicit positive or negative reinforcement enhances memory performance beyond the beneficial effect of TPE. Thus, after an initial exposure to the full material, 40 healthy German speakers learned 210 weakly associated German cue-target word pairs by either studying or testing (two repetition cycles). Depending on the accuracy of the retrieval attempt, the tested word pairs were followed by positive or negative performance feedback (50%) or no feedback (50%). Irrespective of the type of repetition, all trials were followed by a study opportunity. Participants returned for a final cued-recall test on Day 2. Final test results replicated the testing effect, showing considerably better memory performance for tested compared to studied items. Furthermore, explicit performance feedback in addition to correct answer feedback increased recall performance, but only on Day 2. In addition, we used recall accuracy and reaction times during repetition cycles to assess the specific effects of learning history: Explicit feedback improved retrieval for material successfully encoded in the initial study phase (i.e., associated with consistent positive feedback) as well as material learned during the repetition phase (i.e., associated with mixed positive and negative feedback). Hence, performance feedback improves learning beyond the effects of retrieval practice and correct answer feedback, suggesting that it strengthens memory representations and promotes re-encoding of the material.

Introduction

Learning and memory are two separate research fields even though similar content is explored. Research investigating learning historically focuses on the occurrence and timing of stimuli (classical conditioning) and their influence on later behavior based on reinforcement or punishment (operant conditioning), thus providing insights into the biological mechanisms. However, more complex types of learning tied to higher cognitive functioning, such as episodic long-term memory, can be more susceptible to the influence of other internal and external cognitive factors (e.g., attention, monitoring and motivation). Just recently, a few investigations integrated both topics, for instance by implementing conditioning into memory tasks (for review see Miendlarzewska, et al., 2016) and thus, they have emphasized their relation and mutual influence. In line with these studies, the present study was aimed at investigating the role of positive and negative performance-contingent feedback for episodic learning by implementing principles of reinforcement into a cued recall paradigm.

Episodic memory paradigms investigating the *testing effect* provide a good basis to implement feedback. More precisely, memory tests are a powerful, although not widely used, technique to improve later memory recall (see Roediger & Butler, 2011). Generally, in testing effect paradigms an (1) initial study phase is followed by a (2) practice phase, in which participants practice the material by either testing or studying it, and then a (3) final memory test takes place. Increased memory performance is observed after testing compared to studying (i.e., testing effect, see Roediger & Butler, 2011; Rowland, 2014). This effect was replicated many times with different material (e.g., Butler et al., 2013; Karpicke & Roediger, 2008; Pastötter & Bäuml, 2016; Weinstein, McDermott, & Szpunar, 2011) and in different settings (e.g., Rohrer et al., 2010; Vojdanoska et al., 2010).

While testing enhances memory performance, receiving correct answer feedback after a test further boosts the mnemonic advantage of the testing effect (e.g., Carrier & Pashler, 1992; Cull, 2000). Another presentation of the full material (i.e., study opportunity) serves as correct answer feedback (i.e., *test-potentiated encoding* (TPE), (Arnold & McDermott, 2013; van den Broek et al., 2016). For instance, Butler, Karpicke, and Roediger (2007) reported

improved memory recall when correct answer feedback was provided after a test, compared to no study opportunity. Thus, a retrieval attempt maximizes the benefit from a subsequent presentation of the full material (Arnold & McDermott, 2013), potentially because the learner can correct errors and validate correct responses given with low confidence (Butler & Roediger, 2007, 2008).

Prior research primarily suggests two possible mechanisms as the source of the TPE. As originally suggested by Roper (1977), the modulation of attentional resources might be a key factor for TPE. When provided with correct answer feedback, participants can shift their attention to selectively re-encode relevant material (Roper, 1977), specifically for instances in which the initial retrieval was incorrect. This process can be enhanced by providing external feedback (Ludowicy et al., 2019). The second idea proposes that memory representations are extended and enriched during the reactivation of the encoding and retrieval context (Vestergren & Nyberg, 2014). It is still open whether error-detection processes and/or the re-encoding process underlie the effect of TPE.

In order to evaluate the underlying mechanisms, we propose to dissociate error detection and re-encoding of the material, occurring at the same time during correct answer feedback. A temporal separation was implemented by providing explicit performance-contingent feedback before presenting correct answer feedback. More precisely, some responses in the practice phase were immediately followed by a performance-contingent positive or negative feedback. Hence, participants would not need to verify their behavior themselves based on the correct answer feedback and could focus on re-encoding the material during the subsequent presentation of correct answer feedback instead. Both mechanisms previously discussed as the origin of TPE might benefit from performance feedback. On the one hand, previous studies suggest that rewards can modulate attentional orientation (Miendlarzewska et al., 2016). Note that recent literature on performance feedback revealed its rewarding effects even without monetary gains (e.g., Daniel & Pollmann, 2010; see also review by Ferdinand & Czernochowski, 2018). On the other hand, performance feedback might enrich the memory representation since this additional memory can be integrated into the

already existing memory structure and thus lead to memory strengthening. As suggested by van den Broek et al. (2014), strong memory traces are thought to be retrieved more rapidly and with more confidence.

Some prior studies combined performance feedback and correct answer feedback, although indirectly, for instance by presenting the correct answer in green or red font (Butler et al., 2007; Ernst & Steinhauser, 2012). In contrast, only few studies considered the differences between these types of feedback (Pashler, Cepeda, Wixted, & Rohrer, 2005; Roper, 1977). For instance, Roper (1977) and Pashler et al. (2005) compared memory recall boosted by either performance feedback, correct answer feedback or no feedback. While Roper (1977) reported a beneficial effect of performance feedback, Pashler et al. (2005) did not. These mixed findings suggest that performance feedback might enhance episodic memory under certain conditions only: Both studies did not combine performance feedback with a study opportunity, preventing re-encoding of the material, and hence, one of the potential mechanisms possibly underlying TPE. To the best of our knowledge, only one study (Fazio, Huelser, Johnson, and Marsh, 2010) combined performance feedback and correct answer feedback by replicating the previously mentioned experiments (Pashler et al., 2005; Roper, 1977) while adding another review cycle. This review opportunity might constitute a block-wise and delayed correct answer feedback (Rowland, 2014). Fazio et al. (2010) reported a selective performance feedback benefit for low-confidence correct responses. In line with Pashler et al. (2005), they did not find a general memory enhancement due to performance feedback, although the reinforcing quality of performance feedback has been previously suggested to influence learning (Jou & Foreman, 2007). However, firm conclusions on potential effects of performance feedback seem premature at this point and may for instance depend on the number of repetition cycles and delay to final test.

One possible reason for the small magnitude of performance feedback effects in the study by Fazio et al. (2010) is that initial testing with performance feedback and repeated presentation of the full material was performed block-wise. This delay between performance feedback and correct answer feedback might prevent effective attention shifting and focusing.

Hence, the present study aimed at disentangling the effects of performance feedback and correct answer feedback, using only a small temporal delay. We propose that separating the processes of error-detection and re-encoding the material might help to dissociate the mechanisms of the test potentiated encoding. Furthermore, these mechanisms might provide insights into the relation between learning by reinforcement on more complex types of learning. We expect to (1) replicate the standard testing-effect with our paradigm and stimuli. In addition, we predict that (2) memory performance in the final test increases when additional performance feedback is provided. Performance feedback might help to shift attention and enrich the memory representation. Enriched memory representations may be particularly relevant for trials in which correct, but weak memory traces have been formed initially, whereas shifts of attention are specifically useful if no retrievable memory traces have been formed initially. Hence, we evaluated whether the history of retrieval success on Day 1 influences memory performance on Day 2. More specifically, we assessed if (3) the beneficial effect of performance feedback is larger for items with mixed retrieval history compared to items consistently retrieved correctly. Finally, we explored the temporal course of memory performance. More specifically, we assessed (4) if the feedback effect might increase over time.

Methods

Participants

Forty-four healthy German native speakers volunteered in this experiment, approved by the local ethics committee. All participants were right-handed, reported to have normal or corrected-to-normal vision and had no red-green deficiency. None of the participants reported any history of psychological or neurological diseases, or acute use of psychoactive substances. In addition, participants were screened for excessively slow reading rates. Three participants had to be excluded from data analyses due to low memory performance (2 SD lower than the mean) and one participant due to technical problems during data acquisition. Data from a final sample of 40 participants (25 females, age: 24.5, SD = 2.7) were analyzed.

All participants reported normal sleep patterns, including the night before testing. Following information about the procedure, participants gave written informed consent. All volunteers either received payment (24 €) or course credit for their participation.

Material

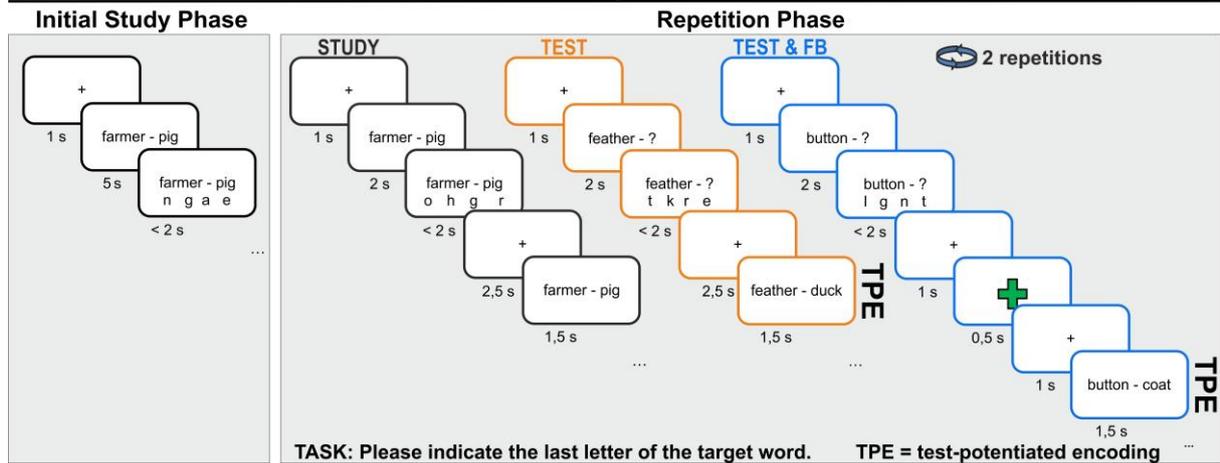
A total of 210 German translations of weakly associated cue-target word pairs (e.g., *feather-duck*; *towel-soap*) were selected from the Nelson, McEvoy, and Schreiber (2004) database. All word pairs were controlled for forward strength (FSG) ($>.04$), backward strength (BSG) ($>.04$), mediated strength (MSG) ($>.04$) and overlapping strength (OSG) ($>.05$). The assignment of word pairs to conditions was randomized across participants.

Procedure

The experiment was divided into two sessions spaced about 24 hours apart (see Figure 1). In the first session, participants were exposed to 210 word pairs in an initial study phase. Afterwards, 175 word pairs were repeated by either study or retrieval practice followed by another exposure to the full word pair (TPE). Half of the retrieval practice items were followed by a binary feedback contingent upon performance. In more detail, 35 word pairs were studied (*study* condition) and 140 word pairs were tested (70 followed by only TPE – *test* condition; 70 followed by binary performance feedback and TPE – *testFB* condition). As *control* condition, 35 word pairs were not presented in the practice phase.

One day later, participants returned to the lab for a final test. The experiment was conducted on a computer using Presentation (Neurobehavioral Systems, Berkeley, CA) to control stimulus presentation and timing.

Day 1



Day 2

Final Test

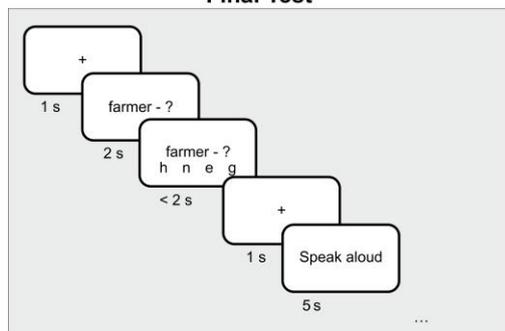


Figure 1. Schematic procedure of the experiment.

Session 1

In the Session 1, participants were instructed to intentionally learn as many word pairs as possible during the initial study phase, as well as afterwards during the repetition phase, in which they studied or retrieved the target words twice (repetition cycle 1 & 2). In the initial study phase, all word pairs (e.g., German: *Feder-Ente* / English: *feather-duck*) were presented individually for a maximum of 7 s at the center of the screen. Participants were informed that they would later be tested on the target words (e.g., *duck*) upon the presentation of the cue words (e.g., *feather*). After 5 s of stimulus presentation, four response letter options (e.g., "l r s o") appeared below the word pair. Participants were instructed to select the last letter of the target word from the four letters by pressing the corresponding key on a keypad with the left- or right-hand index or middle finger (e.g., "e" for *Ente (duck)*, i.e., 4-alternative forced choice recognition). We familiarized participants with this response procedure during the initial study

phase. Each trial was preceded by a 1s fixation cross. Short breaks of 60 s were carried out every 50 trials.

Following the initial study phase, a repetition phase was performed in which participants either retrieved or studied 175 word pairs with the instruction to memorize them. Word pairs were randomly assigned to one of the following three conditions: (I) In the *study* condition, participants were presented with 35 full word pairs (i.e., cue and target word together), one at a time at the center of the screen, for a maximum of 4 s. After 2 s, 4 letters appeared underneath the word pair. One of these letters represented the last letter of the target word. As in the initial learning phase, participants were asked to indicate this letter via button press. After the response, a fixation cross appeared for 2.5 s. (II) In the *test* condition, participants were only presented with the cue word of 70 word pairs and a question mark at the position of the target word, for a maximum of 4 s. Then, they were asked to retrieve the target word from memory. Next, the response options appeared after 2 s with the same response procedure. In this condition, a fixation cross appeared for 2.5 s after the participants gave a response. (III) The *testFB* condition was the same as the test condition, with additional performance-contingent positive or negative feedback for 500 ms, followed by a fixation cross of 1 s. For all conditions, the full word pair (i.e., cue and target word) was presented for another 1.5 s after each practice trial, and participants were instructed to study the word pair. Thus, as illustrated in Figure 1, word pairs were either studied, tested or tested with feedback. Critically, participants saw the correct answer for the same amount of time in all conditions to provide equal opportunity to re-encode the material.

After all word pairs were practiced in one of the 3 conditions (i.e., 1st repetition cycle), participants repeated these procedures in the 2nd repetition cycle to assess if the effect of feedback depends on a long consolidation phase or not. All words were assigned to the same conditions in both repetition cycles. Brief pauses of 60 s were carried out every 50 trials. Responses were scored correct only within 2 s after the response letters appeared. Before leaving the lab, participants were instructed not to rehearse the word pairs at home.

Session 2

After about 24 h ($M = 24\text{h}$; $SD = 1,3$), participants returned to the lab to perform the Session 2 in which all 210 word pairs were tested by presenting the cue word with a question mark for 2 s. The order of cue words was randomized, and rests of 60 s were carried out every 50 trials. Similar to the testing condition of the repetition phase, participants were asked to retrieve the target word from memory. The response procedure was the same as at the previous day (4-alternative forced choice of the last letter of the target word; response options appeared below the cue word after 2 s). Afterwards, another fixation cross was presented for 1 s. Finally, participants were requested to speak aloud the full word pairs (i.e., cued recall, pronouncing the cue word presented on the screen for 5 s and the target word retrieved from memory).

Analysis

The experiment was designed to investigate accuracy differences in the final test and repetition cycles. In addition, reaction time (RT) were assessed as potential index for memory strength. Recall accuracy in the final test was rated based on verbal responses. The analyses of the repetition phase were restricted to items correctly responded since there were not enough forgotten items for some conditions. In contrast to some previous studies on the testing effect, we used a within-participant manipulation in order to minimize variance caused by individual differences. Four questions were investigated: (1) Can the testing effect be replicated? (2) Does feedback further boost memory performance on Day 2? (3) Does positive and negative feedback during the repetition cycles differently influence memory performance in the final test? (4) Does feedback already influence memory performance on Day 1 (during the repetition cycles)?

First, in order to address research questions 1 and 2, recall accuracy and RT results from the final test (Day 2) were investigated as a function of the repetition *Conditions* (control, study, test, testFB) by means of a repeated measures Analyses of Variance (rm-ANOVA). Repeated within-subject contrasts were used to investigate different levels of main effects (control vs. study, study vs. test, test vs. testFB). These contrasts were planned beforehand,

according to the predictions and based on previous evidence (Fazio et al., 2010; Pashler et al., 2005; Racsomány et al., 2018; Roper, 1977, van den Broek et al., 2014).

Second, in order to directly assess the effects of retrieval history (Day 1) on final test performance (research question 3), results from the final test were conditionalized based on the retrieval success during the repetition cycles (only possible for the test and testFB conditions) and analyzed with a 2 x 2 rm-ANOVA with the factors *Condition* (test vs. testFB) and *Retrieval History* (early success vs. late success). Word pairs that were correctly remembered in both repetition cycles were correctly encoded during the initial study phase and hence, assigned to early retrieval success, whereas all others were assigned to late retrieval success. Note that trials receiving positive and negative feedback differ in both, original memory performance and type of feedback presented subsequently since feedback was always contingent on performance. Word pairs with less than two valid responses during the repetition phase could not be categorized and hence had to be excluded from this analysis (0.001 %).

Finally, in order to evaluate the temporal course of TPE (research question 4), correctly retrieved trials from the repetition phase (Day 1) were analyzed with a 2 (*Repetition cycle*: 1 vs. 2) x 2 (*Condition*: test vs. testFB) rm-ANOVA. For all analyses, Greenhouse-Geisser correction was employed where appropriate.

Results

Testing effect & feedback – Recall Accuracy and RT in the final test

An ANOVA with the factor *Condition* (control, study, test, testFB), $F(2.14, 83.29) = 513.10$, $p < .001$, $\eta_p^2 = .93$ (see Figure 2, left), with repeated contrasts revealed increased recall accuracy on Day 2 for tested ($M = 78\%$) compared to studied items ($M = 57\%$, $p < .001$, $\eta_p^2 = .77$). Likewise, memory performance was lower for items in the control condition ($M = 18\%$) compared to studied items ($p < .001$, $\eta_p^2 = .88$). In addition to these typical testing effect results, word pairs tested with performance feedback on Day 1 showed an increased memory performance on Day 2 ($M = 80\%$, $p < .05$, $\eta_p^2 = .12$) compared to tested items.

A similar pattern was found for RTs, $F(1.89, 71.76) = 42.50, p < .001, \eta_p^2 = .53$ (see Figure 2, right), with considerably faster responses for word pairs in the testing condition ($M = 687$ ms) compared to previously studied items ($M = 779$ ms, study vs. test: $p < .001$) or not repeated at all ($M = 961$ ms, control vs. study: $p < .001$). No RT difference was found for items that were tested with performance feedback ($M = 685$ ms) compared to the ones tested without feedback.

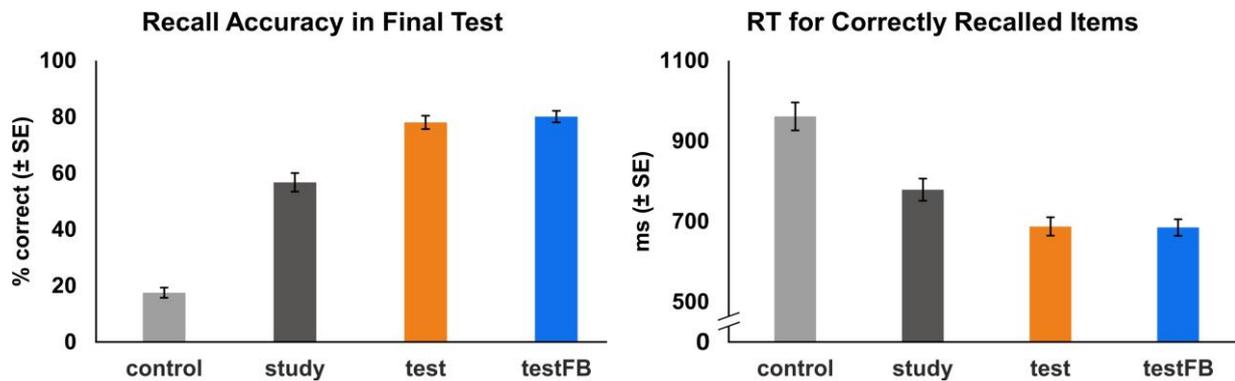


Figure 2. Recall accuracy and RT results (± 1 SE) in the final test as a function of repetition condition. Note that the control condition is associated with both, lowest recall accuracy and highest RTs. The intervention benefit is visible as increased accuracies and decreased RTs.

Effects of early vs late retrieval success on final test performance – Recall Accuracy & RTs as a function of initial response accuracy

A 2 x 2 ANOVA with the factors *Condition* (test vs. testFB) and *Retrieval History* (early vs. late success) revealed a main effect of *Retrieval History*, $F(1, 39) = 155.73, p < .001, \eta_p^2 = .80$ (see Figure 3, left), reflecting enhanced memory performance in the final test following early retrieval success during the repetition phase (early: $M = 89\%$, late: $M = 67\%$). In addition, a main effect of *Condition* indicated an improvement in memory performance for items previously tested with performance feedback ($M = 79\%$) compared to those without ($M = 77\%$), $F(1, 39) = 4.69, p < .05, \eta_p^2 = .11$.

The same pattern of results was found for RT: a main effect of *Retrieval History*, $F(1, 39) = 31.22, p < .001, \eta_p^2 = .45$ (see Figure 3, right) revealed faster responses in the final test if word pairs were retrieved correctly in both retrieval attempts (early: $M = 665$ ms, late: $M =$

752 ms). Moreover, a main effect for Condition indicated faster responses if additional performance feedback was provided previously ($F(1, 39) = 4.25, p < .05, \eta_p^2 = .10$; test: $M = 719$ ms, testFB: $M = 697$ ms).

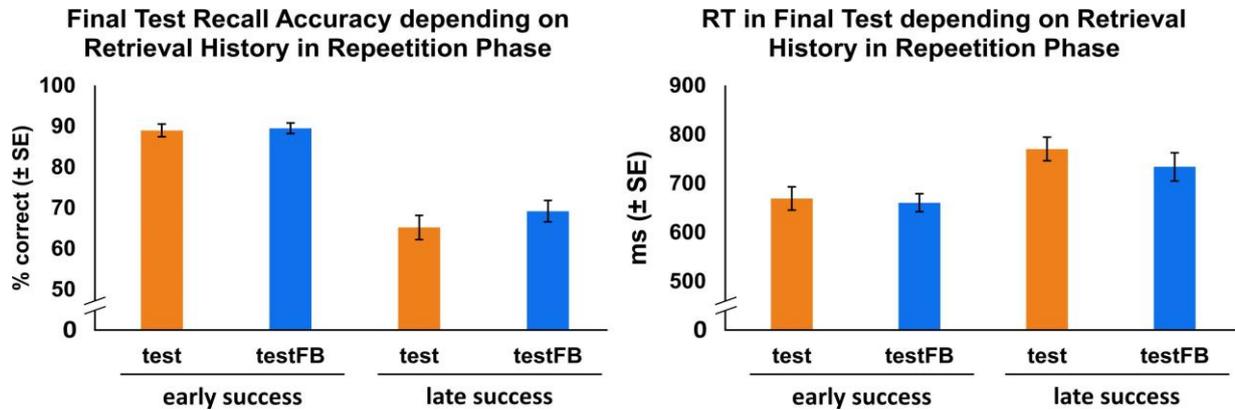


Figure 3. Accuracy and RT results (± 1 SE) in the final test as a function of retrieval history (at Day 1) (early vs. late success) and repetition practice (test vs testFB).

Effects of performance feedback and repeated exposure during the repetition phase – Accuracy & RT

A 2 x 2 ANOVA investigating retrieval accuracy in the repetition phase (i.e., 4AFC responses) with the factors *Condition* (test, testFB) and *Repetition Cycle* (1st cycle vs. 2nd cycle) revealed a main effect of *Repetition Cycle*, with increased memory performance after repeated exposure to the material, $F(1, 39) = 377.45, p < .001, \eta_p^2 = .91$ (see Figure 4, left). Memory performance in the second repetition cycle ($M = 84$ %) was increased compared to the first test in the first cycle ($M = 63$ %). In line with prior literature, no effect of performance feedback was found.

Analyzing RT of correctly retrieved items revealed a main effect of *Repetition Cycle*, $F(1, 39) = 59.53, p < .001, \eta_p^2 = .60$, (see Figure 4, right) confirming faster responses in the second compared to the first repetition cycle (cycle 1: $M = 807$ ms; cycle 2: $M = 665$ ms). Hence, independent of the presence of performance feedback, memory performance increased with repetition while RTs decreased.

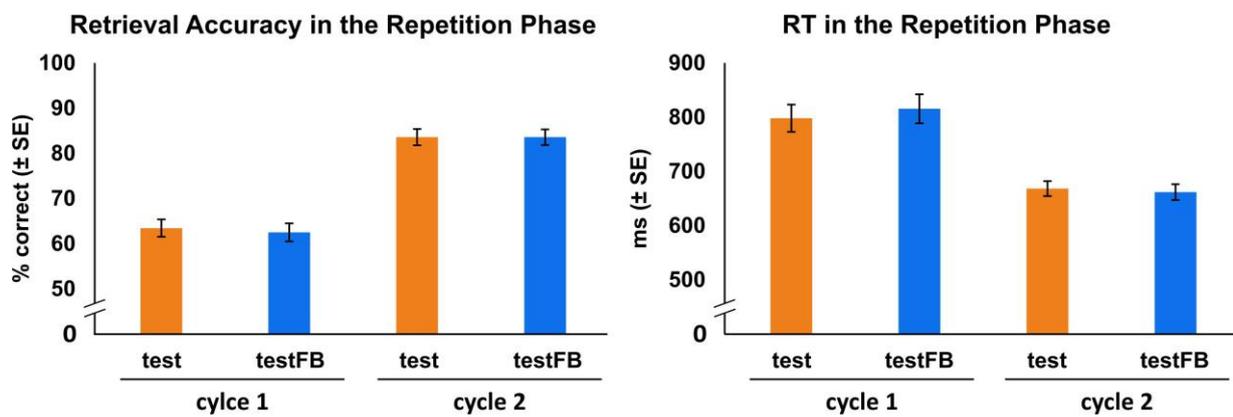


Figure 4. Retrieval accuracy and RT results for correctly retrieved material (± 1 SE) during repetition cycle 1 and 2 as a function of repetition practice (test vs testFB).

Discussion

The present study assessed whether performance-contingent feedback in combination with subsequent correct answer feedback modulates the testing effect, by comparing accuracy and RTs for the repetition phase (Day 1) and the final test (Day 2). Our results are in line with previous findings in terms of enhanced memory following retrieval practice (i.e., testing effect) and a benefit of correct answer feedback (i.e., TPE). When opportunities to re-encode are kept constant (TPE, c.f. van den Broek et al., 2016; Rowland, 2014), retrieval practice enhances later memory performance (Butler, Karpicke, & Roediger, 2007; Roediger & Karpicke, 2006a, b; Rowland, 2014). Specifically, prior testing compared to studying resulted in a performance increase from 57% to 79%, compared to 18% accuracy after a single learning cycle (control condition). We extend previous findings by demonstrating that explicit performance feedback further enhances TPE (by 2 %) after a delay of one day. In addition, analyzing final test recall and RT depending on Day 1 performance suggests a beneficial effect of performance feedback independent of early or late retrieval success on Day 1. Hence, we provide evidence that principles of reinforcement learning can modulate episodic memory performance. In the following, we will discuss how our findings clarify previously inconsistent results regarding the role of performance feedback and how previously neglected RT data can provide additional insight into the underlying mechanisms of feedback enhanced learning.

While our findings are in line with the literature in terms of TPE, the potential role of performance feedback may have been underestimated in prior studies. One factor concerns

the timing of the experimental paradigm. The cognitive processes associated with performance feedback and a study opportunity may depend on temporal contiguity of both factors, which is lost when a delay between performance feedback and correct answer feedback is introduced. A block-wise retrieval practice followed by a block-wise study opportunity presumably decreases the influence of performance feedback on TPE. In contrast to the study by Fazio et al. (2010), we addressed this point by presenting the correct answer feedback immediately after the performance feedback. A second important factor addresses the delay between initial learning and final test. A larger testing effect is observed when a delay of at least 24 hours is introduced between the repetition phase and the final test (Roediger & Karpicke, 2006a, b). Hence, shorter delays might underestimate the testing effect along with the factors that can modulate it. In our data, we can compare the beneficial effect of performance feedback on TPE in the end of Day 1 with the results in the final test on Day 2. In line with prior results (Fazio et al., 2010), we did not observe any differences between items tested with or without feedback after the second practice cycle on Day 1. Hence, recall benefits due to the combination of performance feedback and correct answer feedback appeared only after a delay of 24 hours.

Most studies investigating the testing effect and TPE have examined recall accuracy (e.g., Butler et al., 2007; Roediger & Karpicke, 2006a, b). In contrast, only very few studies considered RTs in addition to memory performance, as lower memory strength may lead to slower RTs since the recall is more effortful (Shiffrin, 1970; van den Broek et al., 2014; Wixted & Rohrer, 1993). In line with the results reported by van den Broek et al. (2014), we found faster RT for previously tested compared to studied items and in addition, studying improved retrieval speed compared to a single learning opportunity. Performance feedback modulated RT when retrieval success on Day 1 was included in the analysis. Two underlying mechanisms could explain this effect. On the one hand, recall mechanisms may become increasingly efficient with retrieval practice. In line with Racsomány et al. (2018), faster responses in the second compared to the first repetition cycle for correctly remembered words support this idea. On the other hand, an improvement in monitoring processes may lead to enhanced re-encoding during TPE as participants already received information about the correctness of

their response. In line with this thought, we found faster RT for both, material learned in the initial learning phase as well as material learned in the repetition cycles receiving additional support through performance feedback. In two instances, this type of support may be particularly helpful: (a) for items initially retrieved incorrectly or not at all and (b) for items retrieved correctly but with low confidence. Accordingly, positive and negative feedback may affect learning in different ways, which need to be disentangled in future studies. Due to its excellent temporal resolution, EEG may provide further insight into the processing of positive and negative feedback and the mechanisms underlying enhanced memory encoding (e.g. Pastötter & Bäuml, 2016). Finally, in the present study response options appeared 2 seconds after item presentation. This lower limit on RT data may mask stronger effects of memory strength that could be explored in future studies.

While the beneficial effects of TPE have been robustly demonstrated, the underlying mechanisms are still under debate. We predicted that performance feedback could either promote attention shifts toward the most relevant – previously incorrect – information (Ludowicy et al. 2019) or reinforce specific correct behavior and thus further enhance memory representations for initially correctly memorized items (Ernst & Steinhauser, 2012). In line with this idea, we analyzed final test results as a function of retrieval success during the repetition phase (Day 1), assuming that enhanced memory recall in the final test could vary depending on the timing of successful encoding. Material learned during the initial study phase benefits from consistent positive feedback during the repetition phase, whereas material learned during the repetition phase might take advantages of a combination of both, negative and positive feedback causing an improvement in learning as compared to correct answer feedback. In favor of this assumption, later memory recall and RT results of the present study implicate a beneficial effect of performance feedback for both, material learned at an early stage as well as material learned during the repetition cycles. Consequently, the present results imply that both mechanisms may be involved in TPE. However, behavioral measures only provide limited knowledge about specific mechanisms and the results highly depend on the design of the experiment. Neuroimaging techniques such as electroencephalography (EEG) or functional

magnetic resonance imaging (fMRI) offer additional insight into the temporal and structural neuronal processes underlying the testing effect (e.g., Ernst & Steinhauser, 2012; Pastötter & Bäuml, 2016; van den Broek et al., 2014, 2016).

Conclusion

In order to uncover the mechanisms underlying TPE, we assessed whether combining performance-contingent feedback with correct answer feedback further improves final memory recall. Overall, our results support this claim, providing evidence that principles of reinforcement learning can modulate episodic memory performance. In addition, the combination of recall accuracy and RT results demonstrates that both negative and positive feedback contribute to the beneficial feedback effect. Hence, performance-contingent feedback may be a helpful tool to promote episodic learning, as long as it is provided in close temporal proximity to the initial retrieval attempt and a subsequent study opportunity.

Chapter 3

Performance feedback enhances test-potentiated encoding

In Chapter 2, only behavioral measures of memory retrieval were investigated. In contrast, the paradigm previously investigated in Chapter 2 was adapted to EEG research in Chapter 3. EEG was collected throughout the entire experiment, including initial study phase, repetition phase as well as the final test. Two research questions exploring temporal aspects of the neural processes underlying feedback assisted TPE were addressed with this study:

- First, direct effects of performance feedback on learning from correct answer feedback were explored by analyzing ERP correlates collected during the presentation of the correct answer feedback in the repetition phase.
- Second, delayed effects of additional performance feedback on processes involved in memory retrieval were investigated by exploring oscillatory correlates collected during the memory tests in the repetition phase as well as the final test.

The first research question will be addressed in Chapter 3a whereas the second one will be addressed in Chapter 3b. Therefore, the sub chapters “*Participants*”, “*Material*”, “*Procedure*”, “*Behavioral Data analysis*” and “*Electrophysiological recordings*” as well as the “*Behavioral results*” will be provided only in Chapter 3a, but not in Chapter 3b to omit repetitions. Thus, Chapter 3b focuses on the oscillatory analyses and results.

Chapter 3a

Performance feedback enhances test-potentiated encoding: an event-related potential (ERP) study on the testing effect

Abstract

Long-term memory retention is enhanced when correct answer feedback (i.e., an opportunity to study the material) is offered following a testing situation (test-potentiated encoding, TPE). The goal of the present study was to explore the neural mechanisms underlying performance feedback assisted learning from correct answer feedback. After a first exposure to a list of 180 weakly associated cue-target word pairs, 25 German native speakers performed 2 repetition cycles by either studying, testing or testing with subsequent performance feedback. Correct answer feedback was offered directly afterwards independent of repetition practice. After a delay of one day, results of the cued retrieval test indicate increased performance for tested compared to studied items and performance feedback further boosted memory performance. ERPs collected during the presentation of correct answer feedback revealed that testing with additional performance feedback modulates early deflections (100 - 300 ms) related to attention orientation and reward learning as well as late components (700 - 1000 ms) related to elaborative encoding processes. Furthermore, increased ERPs were detected in the time interval from 300 - 500 ms if no additional performance feedback was provided, possibly reflecting early memory updating processes. Hence, the mechanism underlying the beneficial effect of explicit performance feedback beyond TPE appears to rely on improved attention orientation as well as reward learning principles and delayed encoding.

Introduction

Providing feedback while learning new information has long been recognized to efficiently enhance learning outcome (see Thurlings, Vermeulen, Bastiaens, & Stijnen, 2013). Nevertheless, feedback can provide different information depending on the study context (for review see Shute, 2008). Research investigating reinforcement learning present positive or negative performance feedback indicating correct or incorrect response to reinforce prior behavior by strengthening or weakening stimulus-response associations (e.g., Hajcak, Moser, Holroyd, & Simons, 2006; Holroyd & Coles, 2002). For instance, participants estimate the duration of one second and subsequently receive feedback about the accuracy of their time estimation (e.g., Miltner et al., 1997). In contrast, episodic memory research employs learning paradigms in which feedback following a memory test provides the correct answer, which offers just indirect performance feedback but the opportunity for studying as the learner is expected to re-encode the material for later memory recall (e.g., Butler et al., 2007; Rowland, 2014). Only few studies so far examined how different feedback types influence each other as prior research in both fields mostly focused on one specific type of feedback (see Miendlarzewska et al., 2016). However, investigating their relation by combining several types of feedback and the knowledge of both fields provides an opportunity to improve our understanding of feedback processing (Ergo et al., 2019; Miendlarzewska et al., 2016). Accordingly, recent literature proposed that reward learning influences declarative memory (see Miendlarzewska et al., 2016; Sinclair & Barense, 2019), but less is known about the underlying neural basis or the temporal course of this effect in humans. Thus, the present study aimed at investigating the behavioral as well as the neural correlates of positive and negative feedback in a declarative memory paradigm by providing performance feedback in addition to correct answer feedback.

Experiments designed to investigate the testing effect and test-potentiated encoding (TPE) are well-established declarative memory paradigms that provide the opportunity to include explicit performance feedback with only small alterations to the standard memory paradigms (e.g., Arnold & McDermott, 2013; Butler et al., 2007; Roediger & Butler, 2011). Retrieval tests are mostly used to assess learning outcome even though it can also serve as

a learning strategy to improve long-term memory retention (Butler et al., 2007; Roediger & Butler, 2011). Prior research revealed that practicing memory recall by testing enhances later recall success as compared to re-reading the study material again (studying), (see Roediger & Butler, 2011). This phenomenon is known as the testing effect. In addition to this effect, providing another study opportunity following a testing situation further enhances later memory recall, presumably due to a test-related enhancement of encoding processes (TPE, Arnold & McDermott, 2013; van den Broek et al., 2016).

When investigating TPE, correct answer feedback is presented after a testing situation by providing the material in the same way as in the initial study condition (Butler et al., 2007; Rowland, 2014). For instance, when participants were asked to learn vocabulary, each word is presented paired with its translation in the study condition as well as the correct answer feedback. Consequently, participants can implicitly infer their prior performance from the correct answer feedback. However, only few studies added explicit information on prior test performance to the correct answer feedback, for example by adding the word “correct” or “incorrect” (Ergo et al., 2019) or by coloring the word pair in green or red (Ernst & Steinhauser, 2012) to indicate correct or incorrect recall success. This immediate performance feedback can only be provided in experimental designs that contain a task to assess memory recall, such as choosing the correct response from multiple choices (Ernst & Steinhauser, 2012). Even though participants are typically able to implicitly rate their prior test performance themselves based on the correct answer feedback, providing explicit performance feedback along with the correct answer feedback might boost long-term learning, possibly due to reinforcement of prior behavior, guiding attention or enriching memory presentations (Ernst & Steinhauser, 2012; Miendlarzewska et al., 2016, see also Study 1, Chapter 2).

Accordingly, we provided additional performance feedback indicating prior recall success immediately before the correct answer feedback was presented (Study 1, Chapter 2). Results revealed a beneficial effect of performance feedback on recall success after a delay of 1 day, whereas immediate recall success was not increased. Several mechanisms might cause these results. While performance feedback can reinforce prior behavior and generally

increase motivation (Miendlarzewska et al. 2016), separating the information on prior test performance and correct answer feedback might enable participants to guide attention towards more relevant information, such as the correct answer feedback if the prior recall was not successful (Ernst & Steinhauser, 2012). In addition, positive and negative cues could assist with memory processing, for instance by elaborating the memory representation. Investigating the neural correlates of explicit performance feedback immediately prior to the studying opportunity during TPE may expose the mechanisms underlying this effect, which is why EEG was recorded in the present study.

The high temporal resolution of electroencephalography (EEG) provides the possibility to explore the neural mechanisms underlying feedback processing in detail (e.g., Luck, 2005, 2014). Besides exploring oscillatory activity (e.g., Ferreira, Maraver, Hanslmayr, & Bajo, 2019; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011), prior research examining the testing effect investigated event-related potentials (ERPs) collected in the repetition phase (e.g., Bai, Bridger, Zimmer, & Mecklinger, 2015; Liu, Tan, & Reder, 2018) or the final test (e.g., Liu et al., 2018; Rosburg, Johansson, Weigl, & Mecklinger, 2015). Mostly, indirect comparisons were used such as subsequent memory effects (i.e., later remembered vs. later forgotten items) or old-new effects (Rosburg et al., 2015) for instance, possibly due to their familiar ERP correlates. Several time windows reaching from 300 - 1000 ms following test/study onset were proposed to reflect modulations by the testing effect (e.g., Bai et al., 2015; Liu et al. 2018). While Gao et al. (2016) reported early ERP amplitudes (300 to 500 ms) to correlate with testing effect benefits, the findings of Rosburg et al. (2015) revealed increased left parietal old-new effects (500 - 900 ms), potentially reflecting recollection processes. In line with these results, Liu et al. (2018) suggested that learning from a prior retrieval attempt prompted by testing is based on two separate memory processes, first, an early process related to retrieval (400 - 700 ms) and second, a later process related to re-encoding (700 - 1000 ms). In accordance with these prior results, the present study investigated ERPs in several time intervals from ranging from 300 - 1000 ms.

At present, only few studies investigated the neural correlates of correct answer feedback processing in a declarative memory task employing the testing effect using EEG (Ergo et al., 2019; Ernst & Steinhauser, 2012). For instance, Ernst & Steinhauser (2012) investigated ERPs to predict learning from corrective feedback in a multiple-choice test. In their study, participants were asked choose translations of Swahili words from 4 alternatives and received the correct translation colored in green or red depending on prior correct or incorrect performance, respectively. Assessing the feedback-related negativity (FRN) and P3 typically known from reinforcement learning research (e.g., Holroyd & Coles, 2002; Ludowicy et al., 2019; Pfabigan, Sailer, & Lamm, 2015) implied that feedback assists memory processing and attention orienting (Ernst & Steinhauser, 2012). Furthermore, prior research focusing on feedback-related ERP components indicated that the anterior N1 (peaking around 135 ms) is modulated by performance feedback as well, possibly reflecting attention orientation (Ludowicy et al. 2019; Martin & Potts, 2009).

In brief, EEG provides the opportunity to uncover the mechanisms underlying correct answer feedback processing boosted by prior performance feedback (feedback assisted TPE) as the three previously suggested approaches (i.e., reinforcement, attention allocation and improved memory encoding) modulate ERPs at specific points in time (Ernst & Steinhauser, 2012; Ludowicy et al., 2019; Rosburg et al., 2015). Therefore, the present study aimed at investigating the neural processes underlying learning from studying correct answer feedback boosted by prior performance feedback. Consequently, the combination of ERP components typically known from the reinforcement learning literature (anterior N1, FRN and P3) and episodic memory research on the testing effect (from 300 - 1000 ms) were investigated. First, we expected to replicate the findings of Study 1 (Chapter 2). Second, regarding the EEG data, we predicted increased amplitudes for earlier components (N1, FRN, P3) correlated to feedback processing if additional performance feedback was presented just before the correct answer feedback. Third, later components related to memory encoding were predicted to differ depending on prior recall success and could be influenced by the presence of additional performance feedback as well (300 - 1000 ms following feedback onset). Fourth, in order to

explore if neural processing of feedback information changes over time, all previously suggested ERP components were compared between study-test cycles 1 and 2.

Methods

Participants

A group of 26 German native speakers (14 females) between 19 and 30 years ($M = 23.5$ yrs.) volunteered for this experiment. All participants were right-handed, reported normal or corrected-to-normal vision and spoke German as their native language. No participant reported any history of psychological or neurological disease. All participants were recruited at the University of Kaiserslautern. Data from one participant were excluded from analyses due to low memory performance (more than 2 SD lower than the mean). As compensation, participants either received 30 € or course credits. This study was conducted in line with the Declaration of Helsinki and approved by the local ethics committee.

Material

A total of 180 German translations of weakly associated cue-target word pairs (e.g., *feather-duck*) were selected from the Nelson et al. (2004) database. Word pairs were chosen based on several parameters describing their association strength: forward strength ($> .04$), backward strength ($> .04$), mediated strength ($> .04$) and overlapping strength ($> .05$) (see Nelson et al., 2004, for more information). All word pairs were randomly assigned to one of the three experimental conditions for each participant.

Procedure

The experiment consisted of three phases: an initial study phase, a repetition phase including 2 cycles and a final test. The final test was performed in a second session that took place with a delay of about 24 hours (see Figure 1). Stimulus presentation was conducted on a PC using Presentation Software (Neurobehavioral Systems, Berkeley, CA). Before the beginning of each phase, participants were instructed and provided with 5 practice trials.

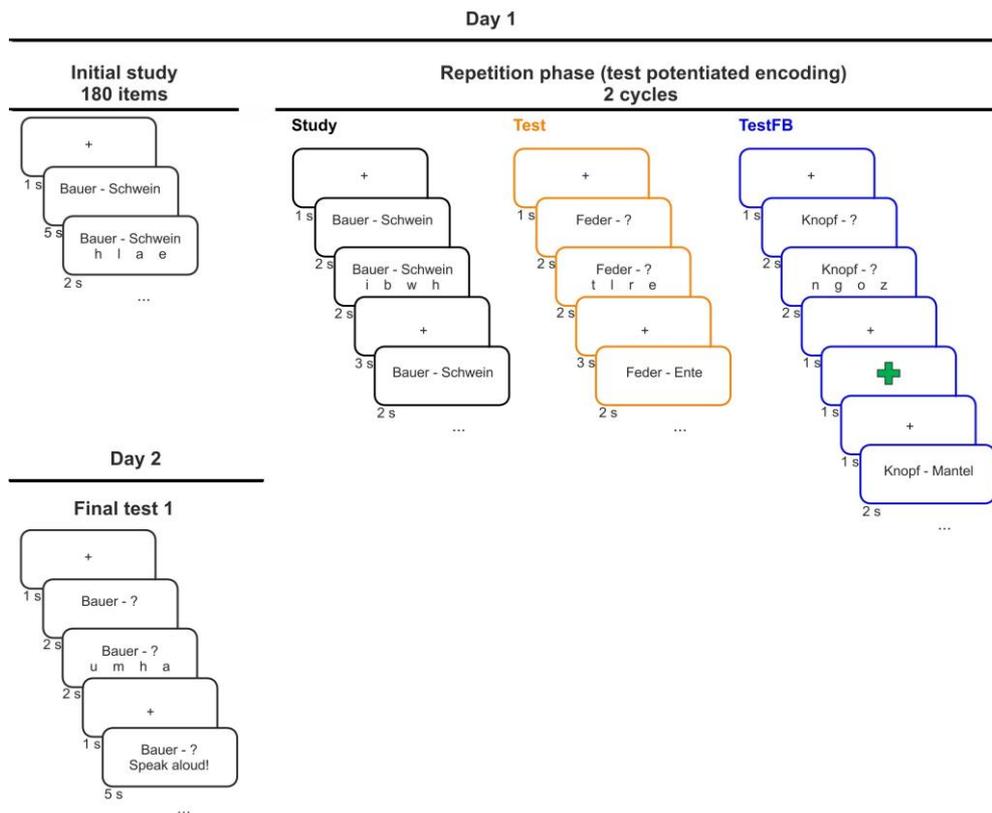


Figure 5. Schematic procedure of the experiment.

Session 1

In the first session, the initial study phase and the repetition phase (consisting of 2 repetition cycles) were conducted. During the initial study phase, participants were presented with all 180 word pairs for the first time. Their task was to intentionally study these word pairs that if they would see the word on the left side of the screen (cue word), they could recall the word presented on the right side of the screen (target word). Furthermore, participants were informed about the subsequent test and instructed not to use any particular strategy or mnemonic technique. Each word pair was presented on the screen for 7 s. During the last 2 s, four response options appeared on the screen below the word pair. Participants were asked to indicate the third letter of the target word by pressing the corresponding button on a response pad. This response procedure was chosen as the initial or last letter may serve as retrieval cues for the correct response, and introduced already during the initial study phase in order to familiarize participants with the response procedure. Each word pair was followed by a fixation cross for 1 s. Brief breaks of 30 s were carried out after every 45 trials.

The repetition phase was performed immediately after the initial study phase and was split into two equal repetition cycles. In each cycle, the word pairs were either practiced by studying the full word pair or by testing, which means that only the cue word was presented. Word pairs or cue words with question mark were presented on the screen for 4 s, one at a time. Consistent with the initial study phase, in all conditions of the repetition cycle four letter options appeared below the word pair on the screen for the last 2 s, and again, participants were asked to indicate the third letter of the target word. Finally, each word pair (i.e., cue and target word) was presented again for 2 s to provide another study opportunity. The correct answer (TPE) was presented for the same amount of time in all conditions to provide an equal opportunity to re-encode the material.

During the *study* condition, participants saw 60 word pairs. Participants were asked to study each word pair and to indicate the third letter of the target word when the response options appeared. Afterwards, a fixation cross was presented for 3 s, followed by the 2 s presentation of the full word pair.

During the *testing* condition (120 items), only the cue word was presented with a question mark next to it. Participants were asked to recall the target word and to indicate the third letter of the target word when the response options appeared. Afterwards, participants saw a fixation cross for 3 s for 60 items (*test* condition). For the remaining 60 items (*testFB* condition), they saw a fixation cross for 1 s, then received performance feedback for 1 s and then saw the fixation cross again for 1 s. Performance feedback consisted of either a green plus sign indicating a correct response or a red minus sign indicating an incorrect response. The feedback always matched the response accuracy and thus, only accurate information was provided (performance-contingent feedback). Independent of whether feedback was provided or not, participants saw the full word pair again for 2 s (TPE).

The assignment of each word pair to either the study, test or testFB condition were kept constant for both repetition cycles. Within each repetition cycle, trials were presented in random order. A fixation cross of 1 s was presented between the trials. Again, after 45 trials

participants were asked to rest for 30 s. In the end of the first session, participants were asked not to repeat the word pairs at home or to discuss them with anyone until the next session.

Session 2

On Day 2 (mean delay $M = 24.3$ h; $SD = 1.3$), participants returned to the lab and performed a final test. In this final test, all cue words were presented with a question mark instead of the target word for 4 s. Consistent with Day 1, 4 letter options appeared on the screen for the last 2 s. Participants were instructed to recall the target word and indicate the third letter of the target word. Next, a fixation cross appeared on the screen for 1 s. Finally, the cue word was presented again for 5 s, and participants were asked to speak aloud the full word pair (i.e., cue and target word) to precisely assess recall accuracy in the final test. Responses were only counted as correct if given within 5 s. A fixation cross of 1 s was presented between two consecutive trials and rests of 30 s were performed after 45 trials.

Behavioral data analysis

Performance of the repetition phase and the final test was investigated by analyzing recall accuracy as well as reaction times (RT). Recall accuracy in the final test was classified as correct or incorrect based on the verbal responses whereas recall accuracy in the repetition cycles was based on responses provided by button press. RT were defined as the time window following the onset of the response options until participants responded by button press.

First, response accuracy in the repetition phase was analyzed as a function of *Retrieval Practice* (test vs. testFB) and *Repetition Cycle* (cycle 1 vs. cycle 2) using a repeated measures (rm) Analysis of Variance (ANOVA). RTs were investigated separately for correct and incorrect responses resulting a $2 \times 2 \times 2$ rmANOVA with the factors *Retrieval Practice* (test vs. testFB), *Repetition Cycle* (cycle 1 vs. cycle 2) and *Retrieval Accuracy* (correct vs. incorrect). As retrieval success of previously studied items could not be classified as correct or incorrect based on the 4AFC response since the target word was presented with the cue word indicating the correct response, data from this condition was excluded from this analysis.

Next, final test recall accuracy and RTs were examined as a function of *Retrieval Practice* (study vs. test vs. testFB). In a second step, final test recall accuracy was further

grouped as a function of response history during the repetition phase. As participants responded correctly or incorrectly in two cycles, results were grouped into the following categories: correctly retrieved in both cycles (CC), correctly retrieved in the first, but incorrectly retrieved in the second cycle (CI), incorrectly retrieved in the first, but correctly retrieved in the second cycle (IC), and incorrectly retrieved in both cycles (II). Results were analyzed using a 2x4 rmANOVA with the factors *Retrieval Practice* (test vs. testFB) and *Response History* (CC vs. CI vs. IC vs. II).

Electrophysiological recordings

Throughout the experiment, the electroencephalogram (EEG) was recorded from each participant using the 32-channel active electrode system EEG-actiCap® (Brain Products, Gilching, Germany). Scalp voltages were recorded in the extended 10-20 system on positions Fp1, Fp2, F3, F4, F7, F8, Fz, FC1, FC2, FC5, FC6, C3, C4, Cz, CP1, CP2, CP5, CP6, P3, P4, P7, P8, Pz, T7, T8, Oz, as well as the left and right mastoid, with a reference electrode on position FCz and a ground electrode on position AFz. The electrode-skin impedance was kept below 5 k Ω and electrode FCz served as reference electrode. Vertical and horizontal electrooculogram (EOG) was recorded from 4 electrodes above and below the right eye and on the outer canthi of both eyes. A BrainAmp Standard amplifier (Brain Products GmbH, Gilching, Germany) was used with 16bit A/D conversion and EEG and EOG were continuously recorded at a sampling rate of 500 Hz.

Electrophysiological data analysis

EEG data were preprocessed and analyzed using Brain Vision Analyzer 2.1 (Brain Products GmbH, Gilching, Germany). The EEG recordings were re-referenced offline to the average mastoid reference (M1 & M2) and the online reference electrode FCz was restored. Next, the data were band-pass filtered (Butterworth) from 0.5 - 30 Hz (12 dB/ oct). Eye movements and blinks were corrected by using the Independent Component Analysis (ICA) based correction, implemented into the Brain Vision Analyzer 2.2. All trials were visually inspected for remaining artifacts. Even though EEG was collected throughout the entire experiment, the present analysis aimed at examining feedback processing and thus, epochs

ranging from 200 ms before to 1000 ms following the correct answer feedback (TPE) onset in repetition cycle 1 and 2 were extracted.

For each participant and condition, artifact free segments were averaged based on testing with or without feedback as well as prior correct or incorrect response accuracy for each of the two repetition cycles. Hence, the present analysis led to 4 conditions: test correct, test incorrect, testFB correct and testFB incorrect. Due to successful learning of the material over time, participant's memory performance in repetition cycle 2 increased, leading to fewer incorrectly retrieved items. Hence, for the analysis of repetition cycle 2, five participants with less than 5 trials in one of the conditions were excluded from the analysis to increase the signal-to-noise ratio. Thus, EEG analyses were based on the following trial numbers: Repetition cycle 1: test correct: $M = 31$, $SE = 1.7$ (range: 19 - 46); test incorrect: $M = 26$, $SE = 1.6$ (range: 11 - 40); testFB correct: $M = 30$, $SE = 1.9$ (range: 17 - 48); testFB incorrect: $M = 26$, $SE = 1.8$ (range: 7 - 37); Repetition cycle 2: test correct: $M = 40$, $SE = 1.6$ (range: 28 - 57); test incorrect: $M = 18$, $SE = 1.4$ (range: 5 - 29); testFB correct: $M = 40$, $SE = 1.6$ (range: 28 - 57); testFB incorrect: $M = 16$, $SE = 1.4$ (range: 5 - 28). TPE-locked ERPs were analyzed in several components reaching from 100 - 1000 ms.

The ERP components anterior N1 and FRN, as typically investigated in reinforcement learning literature, peak in the latency intervals I (Ludowicy et al., 2019; Martin & Potts, 2009) and II, respectively (Ludowicy et al., 2019; Pfabigan, Seidel et al., 2015; Smillie, Cooper, & Pickering, 2011). Both components can be detected mostly at fronto-central electrodes and hence, a region of interest consisting of the three electrodes Fz, FCz and Cz was evaluated for these latency intervals (Ernst & Steinhauser, 2012; Ludowicy et al., 2019; Pfabigan, Sailer et al., 2015). It should be noted that the FRN is measured after explicit feedback is provided. Hence, the present study investigated an ERP component matching the standard analysis window and location of the FRN when only implicit performance feedback is provided by the correct answer feedback. Hence, this fact is abbreviated here by adding quotation marks: "FRN". Mean amplitudes of these ERP components were subjected to a 2 x 2 rmANOVA with the factors *Retrieval Practice* (test vs. testFB) and *Retrieval Accuracy* (correct vs. incorrect).

In line with prior research investigating the testing effect using ERPs (Bai et al., 2015; Gao et al., 2016; Liu, Rosburg, Gao, Weber, & Guo, 2017; Liu et al., 2018; Rosburg et al., 2015), the three later latency intervals (III) 300 - 500 ms, (IV) 500 - 700 ms and (V) 700 - 1000 ms were analyzed as indicators of early and late memory processes. The analyses were restricted to a frontal (F3, Fz, F4) and a parietal electrode cluster (P3, Pz, P4; Liu et al., 2018). Mean amplitudes in these latency intervals and clusters were examined using a 2 x 2 x 2 rmANOVA with the factors *ROI* (anterior vs. posterior), *Retrieval Practice* (test vs. testFB) and *Retrieval Accuracy* (correct vs. incorrect).

In short, the following five latency intervals were investigated: (I) 100 - 150 ms corresponding to the N1, (II) 200 - 300 ms corresponding to the “FRN”, (III) 300 - 500 ms corresponding to the P3, (IV) 500 - 700 ms and (V) 700 - 1000 ms reflecting early and late memory processes. If applicable, sphericity violations were corrected using the Greenhouse-Geisser correction and the corrected p values are provided in the result section. Furthermore, post-hoc t-test were corrected using Bonferroni correction. Effect sizes are listed as η_p^2 . Statistical tests were computed with the SPSS software package 27.0 (IBM, USA).

Results

Behavioral data in the repetition phase and the final test

First, accuracy rates and RT collected during the repetition phase were analyzed based on the 4AFC responses for items tested with or without feedback comparing cycle 1 and 2. A main effect of *Repetition Cycle* ($F(1, 24) = 177.26, p < .001, \eta_p^2 = .88$) indicated that response accuracy increased from cycle 1 ($M = 54\%$) to cycle 2 ($M = 75\%$) (see Figure 2, upper left). Extending these results, RT resulted in a main effect for *Retrieval Accuracy* ($F(1, 24) = 178.98, p < .001, \eta_p^2 = .88$) and *Repetition Cycle* ($F(1, 24) = 4.63, p < .05, \eta_p^2 = .16$) and an interaction for these two factors ($F(1, 24) = 69.74, p < .001, \eta_p^2 = .74$) with faster RT for correctly retrieved items that accelerated from repetition cycle 1 to 2 (correct cycle 1: $M = 1026$ ms; correct cycle 2: $M = 899$ ms, $p < .001$) and slower RT for incorrectly retrieved items that even further

decelerated from the first to the second retrieval practice (incorrect cycle 1: $M = 1261$ ms; incorrect cycle 2: $M = 1334$ ms, $p < .01$; see Figure 2, upper right).

Final test recall accuracy revealed improved memory performance ($F(2, 48) = 54.66$, $p < .001$, $\eta_p^2 = .70$) for previously tested ($M = 71$ %) compared to studied items ($M = 50$ %, $p < .001$), which is in line with the typical testing effect. Furthermore, additional performance feedback further increased memory ($M = 75$ %, $p < .05$) compared to only testing with correct answer feedback (see Figure 2, lower left). Furthermore, RT results revealed a similar pattern ($F(2, 48) = 15.23$, $p < .001$, $\eta_p^2 = .39$) with faster responses for items previously tested ($M = 869$ ms) compared to studied ($M = 957$ ms, $p < .001$) ones, but no difference between items previously tested with or without feedback (testFB: $M = 868$ ms; see Figure 2, lower right).

Next, items correctly recalled in the final test were grouped based on *Response History* (CC, CI, IC, II) during the repetition phase. Note that this was only possible for items previously tested with or without feedback since no test results on day 1 were collected for studied items. Only finding a main effect of *Response History* suggests that the difference between testing with or without feedback is independent of prior retrieval success ($F(3, 72) = 89.03$, $p < .001$, $\eta_p^2 = .79$; CC: $M = 39$ %, CI: $M = 4$ %, IC: $M = 23$ %, II: $M = 7$ %; CC vs. CI, CC vs. II, CF vs. IC, IC vs. II: $p < .001$; CC vs. IC: $p < .01$; CI vs. II: $p < .05$).

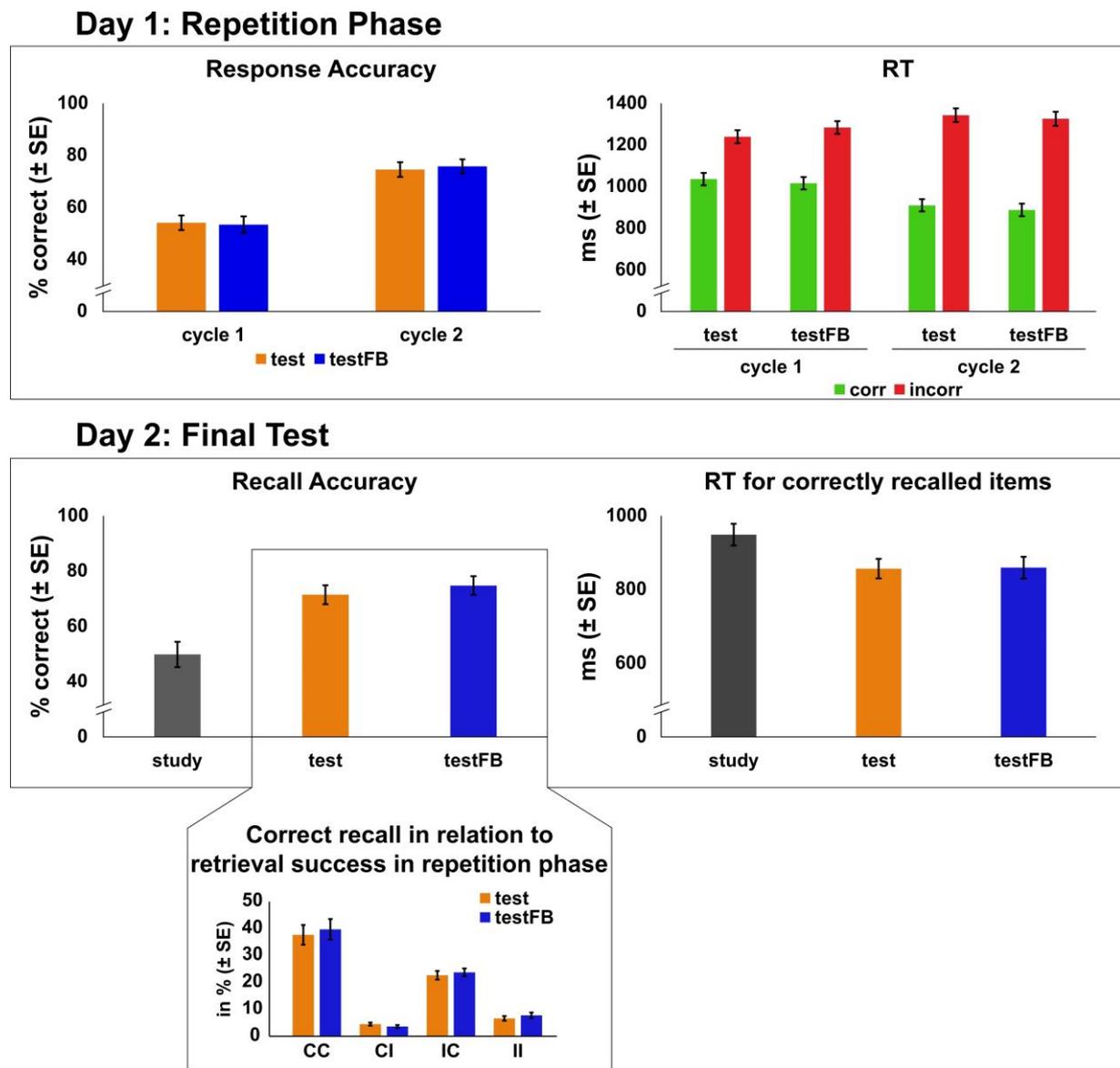


Figure 6. Accuracy and reaction time results (± 1 SE) in the repetition phase (cycle 1 and cycle 2) and final test as a function of *Retrieval Practice*. CC = correctly retrieved in repetition cycle 1 & 2; CI = correctly retrieved in repetition cycle 1 & incorrectly retrieved in repetition cycle 2; IC = incorrectly retrieved in repetition cycle 1 & correctly retrieved in repetition cycle 2; II = incorrectly retrieved in repetition cycle 1 & 2.

ERPs during Correct answer feedback in the repetition cycle 1

To investigate the neural basis of correct answer feedback processing with or without performance feedback, several ERP components were analyzed following the onset of the correct answer feedback (TPE). ERPs at selected electrode sides are depicted in Figure 3. In line with previous literature on the testing effect and feedback processing, we analyzed mean amplitudes in five latency intervals ranging from 100 - 1000 ms that correspond to the N1 (100

- 150 ms), FRN (200 - 300 ms), P3, (300 - 500 ms), and early and late memory processes (500 - 700 ms; (700 - 1000 ms).

The latency interval from 100 - 150 ms after TPE-onset at fronto-central electrodes (ROI: Fz, FCz, Cz), which corresponds to the N1, was investigated using a 2x2 ANOVA with the factors *Retrieval Practice* (test vs. testFB) and *Retrieval Accuracy* (correct vs. incorrect). Results revealed an Interaction ($F(1, 24) = 6.20, p < .05, \eta_p^2 = .21$) with increased amplitudes for previously incorrectly retrieved items if additional performance feedback was provided (test correct: $M = -1.99 \mu\text{V}, SE = 0.36$; test incorrect: $M = -1.40 \mu\text{V}, SE = 0.34$; testFB correct: $M = -1.85 \mu\text{V}, SE = 0.25$; testFB incorrect: $M = -2.08 \mu\text{V}, SE = 0.36$; test incorrect vs. testFB incorrect: $p < .05$).

Results of the "FRN" (latency interval 200 - 300 ms) peaking around 284 ms following TPE-onset at fronto-central electrodes (ROI: Fz, FCz, Cz) were analyzed with the same ANOVA as the first time interval and revealed a main effect of *Retrieval Practice* ($F(1, 24) = 4.52, p < .05, \eta_p^2 = .16$) and an interaction of *Retrieval Accuracy** *Retrieval Practice* ($F(1, 24) = 5.97, p < .05, \eta_p^2 = .20$). The analysis indicates decreased amplitudes for items previously tested with additional performance feedback as a function of initial accuracy (testFB correct: $M = 2.56 \mu\text{V}, SE = 0.44$; testFB incorrect: $M = 1.82 \mu\text{V}, SE = 0.49, p < .01$). Furthermore, amplitudes following incorrect responses decreased depending on the retrieval practice (test incorrect: $M = 2.75 \mu\text{V}, SE = 0.53; p < .05$), whereas no difference was found for correct responses (test correct: $M = 2.62 \mu\text{V}, SE = 0.48$).

The third latency interval (300 - 500 ms) was investigated at frontal (ROI: F3, Fz, F4) and parietal electrodes (ROI: P3, Pz, P4) using a 2x2x2 ANOVA with the factors *ROI* (anterior vs. posterior) x *Retrieval Practice* (test vs. testFB) and *Retrieval Accuracy* (correct vs. incorrect). We observed increased amplitudes at posterior electrodes ($F(1, 24) = 23.90, p < .001, \eta_p^2 = .50$; anterior: $M = 1.08 \mu\text{V}, SE = 0.39$; posterior: $M = 3.38 \mu\text{V}, SE = 0.35$) and decreased amplitudes if additional performance feedback was previously presented ($F(1, 24) = 16.40, p < .001, \eta_p^2 = .41$; test: $M = 2.60 \mu\text{V}, SE = 0.32$; testFB: $M = 1.86 \mu\text{V}, SE = 0.28$).

The fourth latency interval (500 - 700 ms) was analyzed with the same 2x2x2 ANOVA as the third time interval, revealing only increased amplitudes for the posterior ROI compared to the anterior one ($F(1, 24) = 72.92, p < .001, \eta_p^2 = .75$; anterior: $M = -0.34 \mu\text{V}, SE = 0.31$; posterior: $M = 3.65 \mu\text{V}, SE = 0.35$).

The fifth latency interval (700 - 1000 ms), investigated with the same 2x2x2 ANOVA as latency window II and III, indicated again an amplitude increase from anterior to posterior ($F(1, 24) = 81.72, p < .001, \eta_p^2 = .77$; anterior: $M = 0.31 \mu\text{V}, SE = 0.20$; posterior: $M = 2.51 \mu\text{V}, SE = 0.19$). Furthermore, an interaction was found for *ROI* Retrieval Practice* supporting the amplitude increase from anterior to posterior electrodes ($F(1, 24) = 9.68, p < .01, \eta_p^2 = .29$; test anterior: $M = 0.62 \mu\text{V}, SE = 0.21$; test posterior: $M = 2.31 \mu\text{V}, SE = 0.20$, test anterior vs. test posterior: $p < .001$; testFB anterior: $M = -0.01 \mu\text{V}, SE = 0.25$; testFB posterior: $M = 2.71 \mu\text{V}, SE = 0.21$, testFB anterior vs. testFB posterior: $p < .001$) and additionally revealing an amplitude decrease at anterior sites if no additional performance feedback was provided and a parietal amplitude increase if performance feedback was provided (test anterior vs. testFB anterior: $p < .01$; test posterior vs. testFB posterior: $p < .01$). Next, an interaction for *ROI*Retrieval Accuracy* was observed with increased amplitudes at posterior electrodes ($F(1, 24) = 11.63, p < .01, \eta_p^2 = .33$; anterior correct: $M = 0.40 \mu\text{V}, SE = 0.19$; posterior correct: $M = 2.22 \mu\text{V}, SE = 0.19, p < .001$; anterior incorrect: $M = 0.21 \mu\text{V}, SE = 0.26$; posterior incorrect: $M = 2.81 \mu\text{V}, SE = 0.22, p < .001$) and increased amplitudes for incorrect responses at posterior electrodes (posterior correct vs. posterior incorrect: $p < .01$).

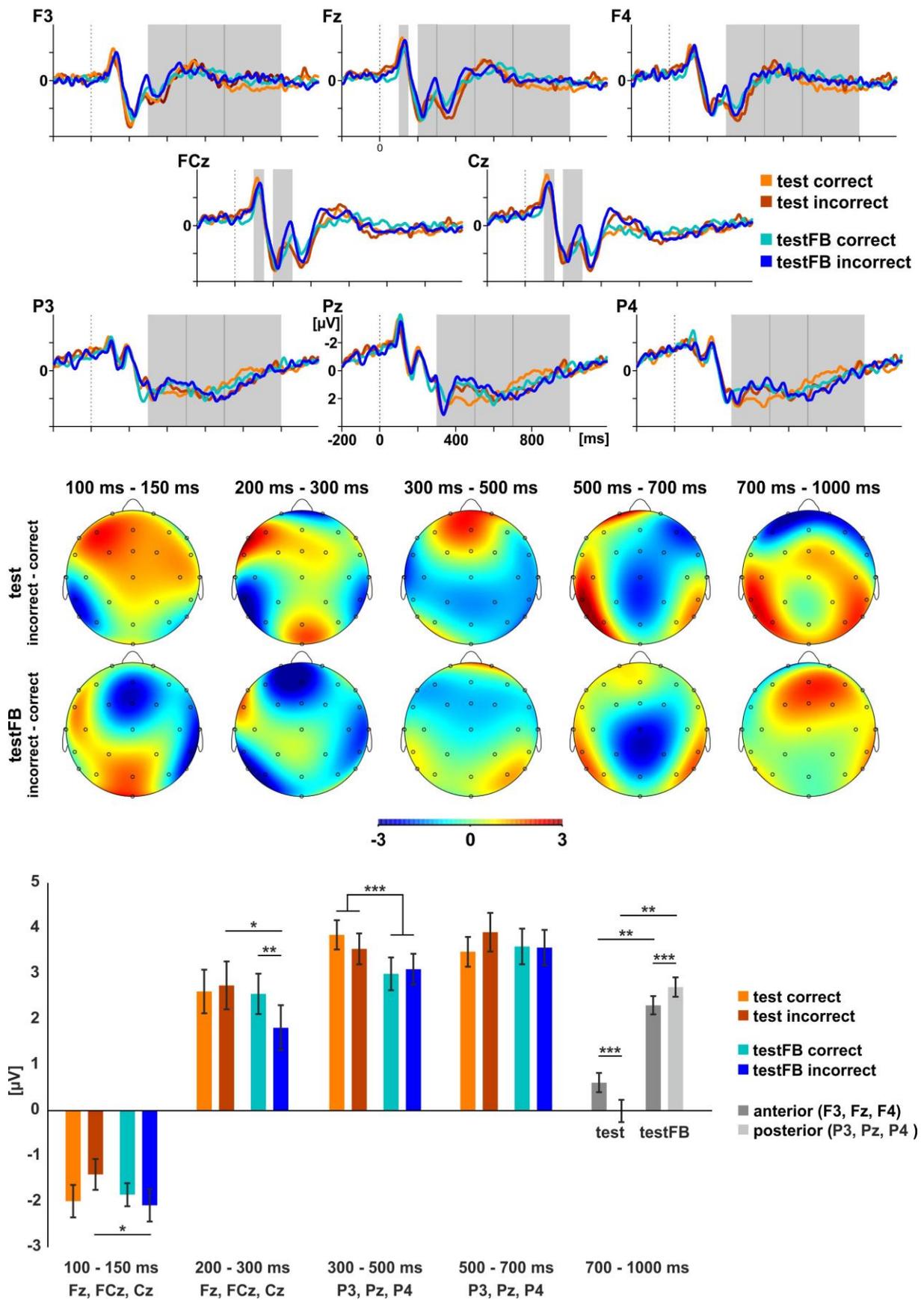


Figure 7. ERPs at frontal, central and parietal electrodes at TPE (correct answer feedback) during repetition cycle 1 as a function of *Retrieval Practice* (test vs testFB) and *Retrieval Accuracy* (correct vs. incorrect) (upper panel) and current source density

(CSD) maps for the respective time intervals displaying differences (incorrect-correct) as a function of *Retrieval Practice* (middel). Means (\pm SE) of each ROI as a function of *Retrieval Practice* (test vs testFB) and *Retrieval Accuracy* (correct vs. incorrect) (lower panel)

ERPs during correct answer feedback in the repetition cycle 2

In line with previous studies on the testing effect, we investigated the correct answer feedback not only during the first, but also during the second repetition cycle, in order to observe if the neural processes are stable over time or change depending on repetition. The same latency intervals and analyses were evaluated and the corresponding results are illustrated in Figure 4. Note that participants were excluded from analysis if less than 5 trials were available for the respective mean resulting in an analysis of 20 participants.

In contrast to the previous findings, analyses on the latency intervals 100 - 150 ms (I) and 200 - 300 ms (II) corresponding to N1 and “FRN” did not indicate any differences. Furthermore, analyzing the latency interval 300 - 500 ms, which corresponds to the P3, revealed only a main effect of *ROI* ($F(1, 19) = 25.64, p < .001, \eta_p^2 = .57$) with increased amplitudes for posterior electrode sites (anterior: $M = 0.37 \mu\text{V}, SE = 0.33$; posterior: $M = 3.36 \mu\text{V}, SE = 0.42$).

Mean amplitudes collected at the latency interval 500 - 700 ms (IV) revealed increased amplitudes for posterior compared to anterior electrode sites ($F(1, 19) = 63.40, p < .001, \eta_p^2 = .77$; anterior: $M = -0.41 \mu\text{V}, SE = 0.31$; posterior: $M = 3.45 \mu\text{V}, SE = 0.35$). In addition, we observed a main effect of *Retrieval Accuracy* ($F(1, 19) = 7.79, p < .05, \eta_p^2 = .29$) with increased amplitudes for previously incorrectly retrieved trials ($M = 1.90 \mu\text{V}, SE = 0.32$) compared to correctly retrieved ones ($M = 1.14 \mu\text{V}, SE = 0.18$).

The same pattern of results were found for the latency interval 700 - 1000 ms (V) with increased amplitudes at posterior electrode sites ($F(1, 19) = 57.33, p < .001, \eta_p^2 = .82$; anterior: $M = -0.32 \mu\text{V}, SE = 0.33$; posterior: $M = 3.54 \mu\text{V}, SE = 0.39$) and increased amplitudes for incorrectly retrieved items ($F(1, 19) = 6.71, p < .05, \eta_p^2 = .26$; correct: $M = 0.97 \mu\text{V}, SE = 0.13$; incorrect: $M = 1.50 \mu\text{V}, SE = 0.24$). In addition, an interaction was observed for *ROI*Retrieval*

Accuracy with increased amplitudes at posterior electrodes ($F(1, 19) = 5.26, p < .05, \eta_p^2 = .22$; anterior correct: $M = 0.25 \mu V, SE = 0.18$; posterior correct: $M = 1.70 \mu V, SE = 0.15$, anterior correct vs. posterior correct: $p < .001$; anterior incorrect: $M = 0.51 \mu V, SE = 0.32$; posterior incorrect: $M = 2.49 \mu V, SE = 0.24$, anterior incorrect vs. posterior incorrect: $p < .001$) and increased amplitudes for incorrect responses but only at posterior electrodes (posterior correct vs. posterior incorrect: $p < .01$).

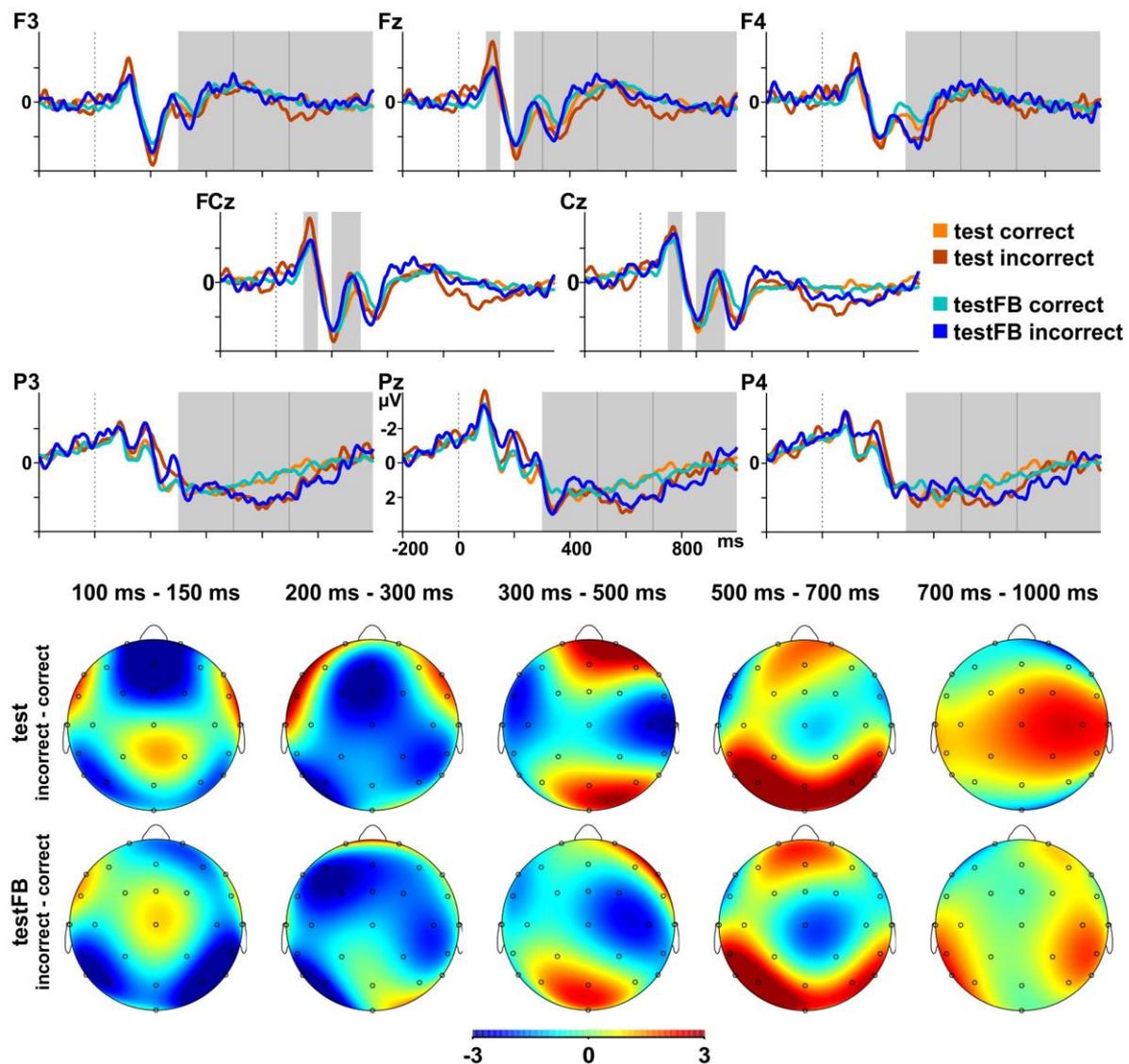


Figure 8. ERPs at frontal, central and parietal electrodes at TPE (correct answer feedback) during repetition cycle 2 as a function of *Retrieval Practice* (test vs testFB) and *Retrieval Accuracy* (correct vs. incorrect) (upper panel) and CSD maps for the respective time intervals displaying differences (incorrect-correct) as a function of *Retrieval Practice*.

Discussion

The present study assessed the effect of performance feedback on correct answer feedback in a declarative memory paradigm with a focus on the neural correlates investigated with EEG. While prior studies on reinforcement learning consider positive and negative feedback to promote specific behavior (Ernst & Steinhauser, 2012; Holroyd & Coles, 2002), feedback in declarative memory studies consist of the learning material and thus, represents a chance for re-encoding as well as elaboration of the memory representation (see Roediger & Butler, 2011; Rowland, 2014).

Behavioral results revealed the typical testing effect with increased recall success in the final test following practicing by testing compared to studying, whereas testing with additional performance feedback enhanced final recall success even further compared to only testing. ERP components typically investigated in research on reinforcement learning, including anterior N1, FRN and P3 (e.g., Ernst & Steinhauser, 2012; Ludowicy et al., 2019), were combined with later ERP's (300 - 1000 ms) reported in memory research focusing on the testing effect (e.g., Bai et al., 2015; Rosburg et al., 2015). ERP results in the repetition cycle 1 revealed that additional performance feedback influenced learning from correct answer feedback almost throughout the entire time interval analyzed in the present study whereas retrieval accuracy affected later processing (700 - 1000 ms). In addition, different processing of previously retrieved versus forgotten items due to additional performance feedback was detected in early processing (N1 and "FRN"). In contrast to repetition cycle 1, results in repetition cycle 2 were only modulated by recall success, but not due to the presentation of additional performance feedback. As outlined in the following paragraphs, performance feedback modulates learning from correct answer feedback (feedback assisted TPE) by search-set restriction and semantic elaboration, but potentially only once.

In line with prior literature investigating the testing effect (e.g., Roediger & Butler, 2011; Rowland, 2014; van den Broek et al., 2014), the present results revealed increased recall success as well as increased RTs in the final test for items previously practiced by testing compared to studying. In addition, providing performance feedback immediately after testing

and before correct answer feedback was presented enhanced final test recall success more than only providing correct answer feedback. However, RTs were not modulated. Hence, the present study replicated the findings of Study 1 (see Chapter 2) regardless of the minor modifications of the experimental design. Investigating behavioral results collected in the repetition cycles revealed that recall success increased while RT decreased due to repetition, whereas additional performance feedback did not affect results at this stage, which is again in line with the findings of Study 1 (see Chapter 2). Thus, additional performance feedback promoted learning from correct answer feedback.

EEG is due to its high temporal resolution a great method to investigate neural mechanisms and we expected correct answer feedback to elicit ERPs correlated with attention shifting and feedback valence processing on the one side and memory related processing on the other side. To the best of our knowledge, only few studies so far investigated learning from correct answer feedback using EEG (Ergo et al. 2019; Ernst & Steinhauser, 2012). In the study by Ernst and Steinhauser (2012), multiple choice testing was followed by correct answer feedback colored in green or red contingent to prior performance with green indicating a correct response and red indicating an incorrect response. ERP results collected while correct answer feedback was presented, suggested that learning from corrective feedback predominantly relies on enhanced encoding processes rather than reinforcement learning since FRN amplitudes only differed for positive and negative feedback whereas P3 amplitudes predicted successful learning (Ernst & Steinhauser, 2012). In line with these results, the present study investigated ERP components known from studies investigating reinforcement learning, including anterior N1 and "FRN", even though the correct answer feedback does not include explicit performance feedback. We expected to detect such ERP components in the present study since we presented performance feedback just before the correct answer feedback and participants were able to implicitly evaluate their performance based on the correct answer and.

Recent literature on reinforcement learning reported increased amplitudes following negative feedback onset compared to positive feedback for the anterior N1 and FRN (Ludowicy

et al., 2019; Pfabigan, Sailer et al., 2015). While the anterior N1 peaks around 120 ms following feedback onset at frontal electrodes reflecting early top-down attention shifting processing (Ludowicy et al. 2019; Luque, Morís, Rushby, & Le Pelley, 2015; Martin & Potts, 2009), the FRN emerges 200 ms following feedback onset (Walsh & Anderson, 2012) reaching its maximum around 250 ms at midline electrode sites (Holroyd & Coles, 2002; Pfabigan, Alexopoulos, Bauer, & Sailer, 2011) reflecting feedback valence and expectation. The present results revealed increased anterior N1 amplitudes for incorrectly retrieved items following testing with prior performance feedback. In contrast, increased “FRN” amplitudes were detected for incorrectly compared to correctly retrieved items tested previously with additional performance feedback as well as for incorrectly retrieved items tested with additional performance feedback compared to only correct answer feedback. Hence, the present results suggest that ERP components typically measured for performance feedback in reinforcement learning paradigms can be detected as well without explicitly providing positive or negative feedback. Furthermore, the findings of anterior N1 suggested that attention is modulated at a very early time point following the onset of correct answer feedback possibly reflecting attention shifts towards relevant information which was incorrectly retrieved previously.

Regarding the FRN, Ernst and Steinhauser (2012) reported increased amplitudes for negative compared to positive feedback as well as enhanced negative deflections for subsequently failed recall attempts. In line with these results, a negative deflection peaking between 200 and 300 ms was detected matching the typical result pattern of the FRN even though we did not present explicit positive or negative feedback. This result was detected only if performance feedback was provided preceding the correct answer feedback. Thus, even if performance feedback is not provided at the same point in time as the correct answer feedback, but instead 1 s before, it affects learning from correct answer feedback. While some researchers proposed that the FRN represents reinforcement related learning from FRN to enhance later behavior (Holroyd & Coles, 2002), others suggested that it reflects conflict monitoring instead with increased activation of the ACC due to high conflict (Gehring & Fencsik, 2001; Jia et al., 2007; Wang et al., 2017). Hence, providing additional performance

feedback might facilitate subsequent learning from correct answer feedback by either modulating reinforcement of prior behavior or by improving conflict monitoring.

Both, research investigating reinforcement learning (e.g., Bellebaum & Daum, 2008; Ludowicy et al, 2019; Pfabigan, Sailer et al., 2015) as well as research exploring the testing effect (e.g., Bai et al., 2015; Gao et al., 2016; Liu et al., 2017; Rosburg et al. 2015) analyzed ERPs around 300 - 500 ms following feedback or study onset. The P3 peaks around 400 ms at centro-parietal electrode sites (Ludowicy et al., 2019) and was suggested on the one hand to detect feedback valence (Bellebaum & Daum, 2008; Yeung & Sanfey, 2004) and on the other hand to reflect early memory related processes (e.g., Ernst & Steinhauser, 2012; Gao et al., 2016; Rosburg et al., 2015). Results of the present study revealed increased amplitudes for material tested without prior performance feedback at posterior electrodes. Thus, these findings suggest early memory related processes instead of reinforcement learning to modulate correct answer feedback processing in this time interval, potentially reflecting the retrieval mode (i.e., recalling past exposure to a specific cue; Gao et al. 2016) or processes related to retrieval depth (see Rosburg et al. 2015).

Since participants perform distinct cognitive processes when practicing by testing compared to studying, research exploring the neural mechanisms underlying the testing effect in the repetition phase mostly contrast studying and testing indirectly by comparing the magnitude of other effects (e.g., Bai et al. 2015; Rosburg et al., 2015). For instance, prior literature examined subsequent memory effects (i.e., later remembered vs. later forgotten items; Bai et al. 2015), source memory tasks or recognition memory tasks (Gao et al., 2016) or old-new effects (Rosburg et al., 2015). In contrast, the present study investigated correct answer feedback, which was presented independent of prior practice condition, and hence, the task does not differ and therefore, we directly compared testing with or without performance feedback. Hence, we evaluated ERPs in latency windows from 500 - 1000 ms following correct answer feedback similar to previous research.

The present results revealed enhanced amplitudes in the later phase (700 - 1000 ms) due to previously testing with additional feedback as well as due to prior unsuccessful retrieval,

predominantly at posterior sites, which was not modulated by retrieval accuracy. Previous research on episodic memory reported a late occurring component peaking at parietal sites starting around 800 ms, which was called late posterior negativity (LPN) (Johansson & Mecklinger, 2003). This component was assumed to indicate prolonged memory processing if recollection processes are based on weak memory representations and increased action monitoring is needed (Johansson & Mecklinger, 2003; Rosburg et al., 2015; Rosburg, Mecklinger, & Johansson, 2011). In line with the LPN, the present findings presumably reflected similar prolonged memory processes, potentially elaborating already existing memory representations if additional performance feedback was presented or by encoding critical information previously not recalled.

Along with Liu et al. (2018), the present study aimed at investigating the stability of the previously reported findings and hence, ERP results collected at the first repetition cycle were compared to the second cycle. Liu et al. (2018) investigated recall success in the two time windows 400 - 700 ms and 700 - 1000 ms reporting stable patterns of results for the early and changing patterns for the later time window. In contrast to their findings, the present study revealed no effect of prior additional performance feedback on correct answer learning in any of the time windows investigated in the second repetition cycle. Instead increased amplitudes for previously unsuccessfully retrieved compared to successfully retrieved items already starting at 500 ms up to 1000 ms at posterior sites were detected in repetition cycle 2. Liu et al. (2018) suggested overlearning as the reason for their changes in result patterns and that re-encoding processes become more irrelevant for items successfully retrieved several times. The present results support this idea as only incorrectly retrieved items led to increased encoding processes and additionally imply that prior performance feedback enhances learning of correct answer feedback only once. Taking this idea into account, future studies might investigate the effect of presenting additional performance feedback once, twice and more times, perhaps using complex learning material that needs several repetitions for successful learning such as acquiring vocabulary.

Over the last decades, many hypotheses have been provided to explain the underlying mechanisms of the testing effect, but most studies focused on two accounts (for review see Rowland, 2014; van den Broek et al., 2016). One is the elaboration account suggesting that testing provides an opportunity to strengthen memory by enriching the memory representation (Carpenter, 2009; Carpenter & Delosh, 2006; Roper, 1977). According to the second account, the search-set restriction account, testing facilitates later memory recall by strengthening correct and weakening competing representations (Karpicke & Smith, 2012; Karpicke & Zaromb, 2010; Thomas & McDaniel, 2013). As testing increases learning not only of previously studied material (see Roediger & Butler, 2011), but also affects learning from correct answer feedback following a test (Arnold & McDermott, 2013; van den Broek et al., 2016) as well as the next testing situation (known as the “forward testing effect”, e.g., Yang, Potts, & Shanks, 2018; Yang & Shanks, 2018), the same accounts might serve to explain learning from correct answer feedback as well. Enhanced learning from correct answer feedback following a test could either result from context reactivation elaborating or refining the memory representations (see Ernst & Steinhauser, 2012) or from improved performance evaluation and attention allocation to critical information (van den Broek et al., 2016; see as well Study 1, Chapter 2). The present results support both accounts. Rosburg et al. (2015) suggested based on their findings that multiple mechanisms instead of only one strategy affects the testing effect. In accordance with this suggestion, the present results support both accounts as well. Hence, learning from correct answer feedback potentially relies on a combination of attention allocation (N1) as well as search-set restriction and semantic elaboration of memory representations (FRN, P3 and 700 - 1000 ms). In contrast, the second presentation of correct answer feedback only supports further encoding processes (500 - 1000 ms).

Conclusion

In the present study, EEG was used to explore the temporal course of neural mechanisms underlying learning from correct answer feedback modulated by prior performance feedback (feedback assisted TPE). We hypothesized that additional performance

feedback could enhance learning from correct answer feedback by modulating attention shifting processes, facilitating search-set restriction processes and as well elaborating memory representations. ERP results collected while correct answer feedback was present suggest that prior performance feedback led to increased processing of feedback valence potentially facilitating search-set restriction ("FRN": 200 - 300 ms). Furthermore, enhanced memory related processing was detected in later time intervals (P3: 300 - 500 ms & LPN: 700 - 1000 ms) possibly further enriching memory representations. In addition, prior performance feedback modulated attention orientation as well (N1: 100 - 150 ms). Comparing the first and the second repetition cycle suggested that boosting learning from correct answer feedback by additional performance feedback might be limited. To conclude, these findings suggest that explicit performance feedback affects encoding of correct answer feedback by modulating attention orientation, conflict monitoring as well as memory processing, specifically during the first repetition cycle. Hence, performance feedback might assist learning from correct answer feedback only once.

Chapter 3b

While Chapter 2 focused on the behavioral aspects of memory retrieval, Chapter 3a investigated the neural processes underlying learning from correct answer feedback modulated by performance feedback (feedback assisted TPE) in the repetition phase. Behavioral results of both studies revealed increased retrieval success in the final test following material previously tested with additional performance feedback and correct answer feedback compared to only testing with correct answer feedback in the repetition phase. Hence, In contrast to Chapter 3a, Chapter 3b examines temporal aspects of the neural processes underlying successful retrieval in the final test as well as during the memory tests in the repetition phase. Furthermore, while Chapter 3a explored ERPs, Chapter 3b investigated oscillatory activity associated to retrieval processes.

Since Chapter 3a and Chapter 3b investigate different phases of the same study, the sub chapters "*Participants*", "*Material*", "*Procedure*", "*Behavioral Data analysis*" and "*Electrophysiological recordings*" as well as the "*Behavioral Results*" were already provided in Chapter 3a and will not be repeated in Chapter 3b.

Performance feedback enhances test-potentiated encoding: oscillatory correlates of the testing effect

Abstract

Compared to repeated presentation of study material, a retrieval attempt prompted by a memory test enhances later memory recall success (testing effect). In addition, providing correct answer feedback following a retrieval test further boosts memory performance since it provides a new chance to encode the material and implicitly offers information on prior test performance (test-potentiated encoding, TPE). Less is known about how reinforcement by explicit performance feedback affects TPE. Here, 25 native speakers learned 180 weakly associated word pairs once. Next, all word pairs were repeated twice by studying or testing followed by correct answer feedback. For half of the tested items, explicit performance feedback was presented after the retrieval attempt in order to facilitate error monitoring before the correct answer feedback is provided. One day later, a final test was conducted. Electroencephalography (EEG) was recorded throughout the entire experiment. Behaviorally, testing compared to studying on Day 1 increased memory performance on Day 2. Explicit feedback further enhanced retrieval success in the final test. Event-related oscillatory activity revealed increased slow theta as well as slow beta activity for tested compared to studied material reflecting increased memory processes, for instance by semantic elaboration. Furthermore, performance feedback led to increased fast theta as well as slow alpha activity potentially enhancing attentional processes as well as memory. In conclusion, results revealed that testing enhances memory retrieval by semantic elaboration as well as search-set restriction and that the type of feedback additionally provided modulates the intensity of these mechanisms.

Introduction

According to the “testing effect”, practicing by testing compared to studying facilitates later recall success (for review see Roediger & Butler, 2011; Rowland, 2014; also called “retrieval practice effect”, RPE). For example while learning vocabulary, trying to retrieve the translation of a word from memory (testing) instead of reading the word together with its translation (studying) increases the chance to correctly recall the translation at a later time point. In addition, if the word and its translation is presented following a test (i.e., correct answer feedback), then later retrieval success is increased compared to no additional exposure, which is known as test-potentiated encoding (TPE, Arnold & McDermott, 2013; van den Broek et al., 2016). Several theories exist that aim to explain the testing effect (see Rowland, 2014), but most studies focused on the suggestion that testing facilitates memory retrieval either by elaborating the semantic network (e.g., Carpenter, 2009; Carpenter & Delosh, 2006) or by improving selection processes (e.g., Karpicke & Zaromb, 2010; Thomas & McDaniel, 2013). Exploring the neural basis underlying these effects could assist with understanding the cognitive mechanisms. Therefore, the present study was aimed at investigating frequency modulations over time caused by the feedback enhanced testing effect using electroencephalography (EEG).

At present, most research investigating TPE provided correct answer feedback by presenting the material in the same way as in the study condition. In the example of vocabulary learning, the word was presented together with its translation (e.g., Butler, Karpicke, & Roediger, 2008; Butler & Roediger 2008; Rowland, 2014). Such correct answer feedback provides on the one hand an opportunity to encode the material and on the other hand, participants can implicitly self-evaluate their prior retrieval success. In contrast, only few studies additionally offered explicit information on prior test performance by e.g., presenting the learning material in green or red (Ernst & Steinhauser, 2012) or adding the words “correct” or “incorrect” to indicate correct or incorrect retrieval success respectively (Ergo et al., 2019). However, these studies did not investigate the difference between providing explicit

information on retrieval performance compared to implicit performance evaluation based on correct answer feedback.

Since performance feedback guides attention orientation, potentially towards relevant information and reinforces relevant behavior (Ernst & Steinhauser, 2012; Miendlarzewska et al., 2016, see also Chapter 2 and Chapter 3a), providing explicit performance feedback immediately before the presentation of the correct answer feedback instead of at the same time might enhance the benefits of additional performance feedback. To the best of our knowledge, only one study behaviorally investigated the effect of additional performance feedback provided before the correct answer feedback so far reporting increased recall success after a delay of one day for items previously tested with additional performance feedback before the correct answer feedback compared to only correct answer feedback (Study 1, Chapter 2). The present study aimed at replicating this study while additionally collecting EEG to investigate feedback-induced modulations of oscillatory activity during memory retrieval.

So far, several studies investigated the neural processes of memory retrieval and the testing effect for example by examining event-related potentials (ERPs) (e.g., Bai et al., 2015; Liu et al. 2018; Rosburg et al., 2015) or oscillatory activity (e.g., Ergo et al., 2019; Ferreira et al., 2019; Pastötter et al., 2011). For oscillatory activity, an increase or decrease in specific frequencies was reported to facilitate memory formation by synchronizing activation patterns of various regions in the brain (see Hanslmayr, Axmacher, & Inman, 2019). Multiple frequencies have been associated to the processes underlying memory retrieval, including alpha, beta and theta and hence, these were explored in the present study as well (e.g., Ergo et al., 2019; Pastötter & Bäuml, 2016; Pastötter et al., 2011).

Cortical theta activity, formerly defined from about 4 - 7.5 Hz (Klimesch, 1999), was suggested to be correlated with the potential to encode new information and to retrieve it at a later time point as well as the generation of associations (see also Herweg, Solomon, & Kahana, 2020). Recently, research proposed to dissociate “slow” and “fast” theta oscillations as the first is positively related and the second one is negatively related to retrieval success

measured by subsequent memory effects (SME, i.e., later remembered vs. later forgotten items; Lega, Jacobs, & Kahana, 2012). In line with this suggestion, Pastötter & Bäuml (2014) performed a testing effect study collecting EEG and results supported the differentiation between slow theta oscillations (~3 Hz) associated to recollection and conscious awareness (e.g., Bramão & Johansson, 2017; Gruber, Tsivilis, Giabbiconi, & Müller, 2008) and fast theta (~7 Hz) related to interference processing and cognitive control (see Hanslmayr, Staudigl, Aslan, & Bäuml, 2010). In contrast to theta activity, alpha desynchronization (7.5 - 12.5 Hz; Beardsley, Hernández-Leo, & Ramirez-Melendez, 2017; Klimesch, 1999) as well as lower beta oscillations (~15 Hz) were suggested to correlate with semantic long-term memory processing as well as deep encoding and elaboration of item information (Klimesch, 1999; Pastötter & Bäuml, 2016; Scholz, Schneider, & Rose, 2017). Additionally, alpha frequencies were as well suggested to reflect attention orientation processes. Furthermore, fast beta frequencies (15 - 25 Hz, Pastötter et al., 2011) were assumed to be differently modulated by retrieval compared to encoding as well (Guran, Herweg, & Bunzeck, 2019).

Research investigating oscillatory activity modulated by the testing effect mostly examined EEG that was collected in the practice phase (e.g., Beardsley et al., 2017; Ferreira et al., 2019; Ferreira, Marful, Staudigl, Bajo, & Hanslmayr, 2014; Guran et al., 2019). In these studies, either comparisons of remembered and forgotten items were investigated (Pastötter & Bäuml, 2014, 2016) or competitive and noncompetitive cues were contrasted (Ferreira et al., 2014) or learning of the first word list was compared to learning of last word list (Pastötter et al., 2011) for example. In contrast, Guran et al. (2019) compared EEG collected in the practice phase comparing an encoding task (indoor / outdoor categorization) to a retrieval task (old / new categorization). Hence, the direct comparison of testing and studying is not possible, potentially due to the different task demands.

To the best of our knowledge, Pastötter & Bäuml (2016) were the only ones investigating oscillatory activity collected in the final test modulated by previously testing compared to studying comparing remembered and forgotten material. Results revealed increased slow and decreased fast theta as well as decreased beta activity. Furthermore, they

provided correct answer feedback following a first test at day 2 and results revealed increased slow theta activity as well as increased alpha and slow beta oscillations for previously remembered compared to forgotten items. To the best of our knowledge, the present study is the first that aimed at investigating the testing effect in the final test as well as in the repetition phase comparing items previously tested with or without additional performance feedback to studied ones.

Based on previous research, the present study collected EEG data over the course of a testing effect paradigm to investigate oscillatory modulations due to studying, testing or testing with additional performance feedback. The present study aimed at investigating three main ideas: First, (I) we examined if prior testing compared to studying affect oscillatory activity of correctly recalled items in the final test. Since testing with additional performance feedback increases recall success in the final test (see Study 1, Chapter 2), (II) we expected modulations of oscillatory activity in the final test due to the additional performance feedback, possibly in slow theta as well as alpha / beta frequency bands (similar to the results reported for feedback processing by Pastötter & Bäuml, 2016). Lastly, we investigated oscillatory changes due to the presentation of additional performance feedback in the repetition phase (III) as oscillatory difference collected for items tested with compared to without additional performance feedback were assumed for the second but not the first repetition cycle. Additionally, Staudigl, Hanslmayr, and Bäuml (2010) reported a decrease in theta oscillations from repetition cycle 1 to 2 and hence, this comparison will be investigated as well.

Methods

Note: As the current chapter is based on the same experiment as Chapter 3a, the sub chapters of the Methods part in Chapter 3a “Participants”, “Material”, “Procedure” “Behavioral Data analysis” and “Electrophysiological recordings” are the same for the current Chapter 3b and therefore, are not repeated here (see page 47).

EEG Data analysis

EEG data were preprocessed and analyzed using Brain Vision Analyzer 2.2 (Brain Products GmbH, Gilching, Germany). First, EEG recordings of the repetition phase and final test were re-referenced offline to the average mastoid reference (M1 & M2) and the online reference electrode FCz was restored. Next, EEG data were band-pass filtered (Butterworth) from 0.5 - 50 Hz (12 dB/ oct) and Independent Component Analysis (ICA) based correction as implemented in the Brain Vision Analyzer 2.2 was used to correct artefacts due to eye movements. All trials were visually inspected for remaining artifacts and removed. Even though the EEG was collected throughout the entire experiment, the present analysis focused on recall processes modulated by repetition practice (either studying, testing or testing with additional performance feedback) in the repetition phase and final test. Hence, epochs ranging from 2000 ms before to 4000 ms after study or test onset were extracted for items successfully recalled in the final test.

To examine evoked oscillatory activity in the time-frequency domain, data were averaged for each subject and across all electrodes depending on the repetition practice condition (either studied, tested or tested with additional performance feedback) in the repetition cycle 1 (Day 1), repetition cycle 2 (Day 1) and final test (Day 2) time-locked to the onset of either the test (i.e., presentation of cue word with "?") or the study opportunity (i.e., presentation of the cue and the target word in the repetition phase). Complex Morlet wavelet transformations with a length of 7 cycles were applied. Wavelet coefficients for frequencies ranging from 1 - 30 Hz with 40 frequency steps (logarithmically layered) were calculated. Analyses were based on relative power changes (in percent) relative to a baseline interval starting -750 to -250 ms prior to stimulus onset. Time-frequency plots focusing on an interval of -500 - 2000 ms around stimulus onset were used to depict relative power changes of all conditions and time points.

Statistical analysis

For statistical analyses, the present analysis focused on theta, alpha and beta frequencies split into five frequency bands of interest, which were selected based on recent

episodic memory literature: 2 - 4 Hz & 5 - 7 Hz (e.g., Ferreira et al., 2014; Hanslmayr et al., 2010; Lega et al., 2012; Pastötter & Bäuml, 2014); 8 - 10 Hz (Beardsley et al., 2017; Guran et al., 2019); 11 - 17 Hz (Guran et al., 2019; Pastötter & Bäuml, 2016) and 18 - 24 Hz (Pastötter et al., 2011). The slow theta frequency band (2 - 4 Hz) was investigated in a time interval from 200 - 800 ms post-stimulus onset (see Pastötter et al., 2016) and the other frequency bands were analyzed in an early and a later time interval ranging from 0 - 500 ms and from 500 - 1000 ms respectively.

Only items correctly recalled in the final test were examined. Mean power changes (in percent) of each frequency band and time interval collected in the final test were examined based on repetition by studying, testing or testing with additional feedback. For the 2 - 4 Hz frequency band, means were analyzed using a one-way ANOVA with the factor factors *Repetition practice* (study vs. test vs. testFB) whereas means of the other frequency bands were submitted to repeated-measures (rm) ANOVAs with the factors *Repetition practice* (study vs. test vs. testFB) and *Time Interval* (0 - 500 ms vs. 500 - 1000 ms).

In the repetition phase, participants performed dissimilar tasks for studying compared to testing: They were presented with the cue and the target word in the study condition whereas only the cue word was presented in the test condition and participants had to retrieve the target word from memory. Therefore only items tested either with or without additional performance feedback were examined for the repetition phase. Furthermore, frequency changes from repetition cycle 1 to cycle 2 were examined as well. Thus, means for the 2 - 4 Hz frequency band were submitted to repeated-measures (rm) ANOVAs with the factors *Repetition practice* (test vs. testFB) and *Repetition cycle* (cycle 1 vs. cycle 2). The factor *Time interval* (0 - 500 ms vs. 500 - 1000 ms) was added again to the analyses of the remaining frequency bands. All ANOVAs were followed up with post-hoc t-tests. Furthermore, ANOVAs and post-hoc t-tests reported in the results were corrected for multiple comparisons using false-discovery rates (FDR; Benjamini & Yekutieli, 2001) with a significance level of $q < 0.05$ (see e.g., Busch, Dubois, VanRullen, 2009).

Results

Note: As the current chapter is based on the same experiment as Chapter 3a, the behavioral results can be found in Chapter 3a, page 53.

Electrophysiological results in the final test

Time-frequency results of the oscillatory brain activity in the final test were analyzed in the frequency bands 2 - 4 Hz, 5 - 7 Hz, 8 - 10 Hz, 11 - 17 Hz and 18 - 24 Hz. Modulations of mean power changes due to prior repetition practice in these frequency bands were examined using ANOVAs with the factor *Repetition practice* (study vs. test vs. testFB). In Figure 1, time frequency-spectrograms of the 1 - 30 Hz frequency interval are depicted whereas statistical results can be found in Table 1. Results collected in the frequency band 2 - 4 Hz investigated in a time interval from 200 - 800 ms following test onset revealed increased power for items previously tested with or without feedback compared to studied ones (study vs. test: $q = 0.027$; study vs. testFB: $q < 0.045$; test vs. testFB: $q = 0.929$).

For the remaining frequency bands, the factor *Time Interval* (0 - 500 ms vs. 500 - 100 ms) was included to the analyses. Means collected in the fast theta frequency band (5 - 7 Hz) differed for items previously tested with or additional feedback compared to studied ones, while previously tested items were in between the other conditions in the 0 - 500 ms time interval (study vs. test: $q = 0.075$; study vs. testFB: $q = 0.015$; test vs. testFB: $q = 0.179$). In the 8 - 10 Hz frequency band items previously tested with additional performance feedback revealed increased power changes compared to tested or studied items in the 0 - 500 ms time interval (study vs. test: $q = 0.109$; study vs. testFB: $q = 0.011$; test vs. testFB: $q = 0.049$). Similar to the 2 - 4 Hz frequency band, power changes in the 11 - 17 Hz frequency band were increased following items previously tested either with or without feedback compared to studied items (study vs. test: $q = 0.038$; study vs. testFB: $q = 0.011$; test vs. testFB: $q = 0.764$). Lastly, beta power (i.e., 18 - 24 Hz) was increased for items previously tested compared to items tested with additional feedback and items previously studied resulted in power means in between the

two other conditions in the 0 - 1000 ms time interval (study vs. test: $q = 0.073$; study vs. testFB: $q = 0.175$; test vs. testFB: $q = 0.020$).

All analyses for the time interval from 500 - 1000 ms for the frequency bands 5 - 7, 8 - 10, 11 - 17 Hz for the comparisons study vs. test, study vs. testFB or test vs. testFB did not reach statistical significance ($qs \geq 0.345$) whereas all comparisons between first and second time interval in these frequencies were significant ($q < 0.000$).

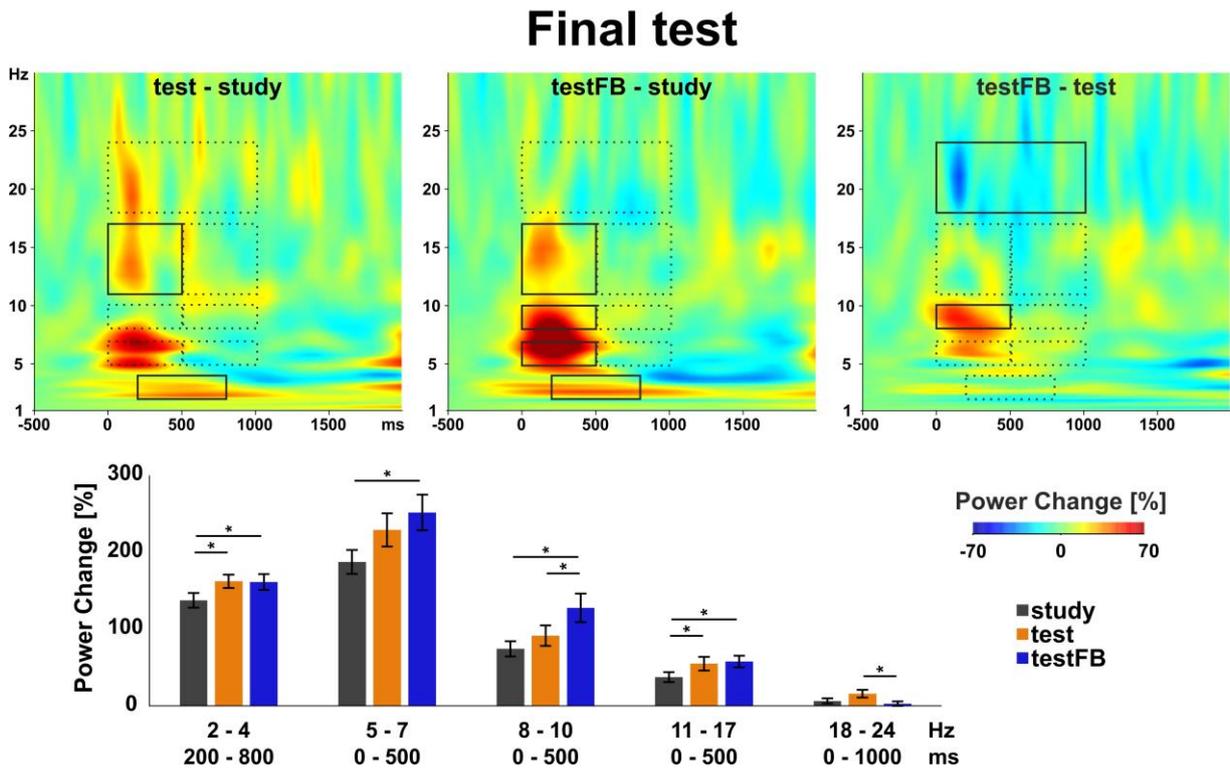


Figure 9. Upper panel: Time-frequency spectrograms of power differences averaged over all electrodes depicting the EEG analyses time-locked to test onset in the final test for the following contrasts: test - study; testFB - study; testFB - test; lower panel: Mean power changes in the time interval from 200 - 800 ms for 2 - 4 Hz or 0 - 500 Hz for 5 - 7, 8 - 10 and 11 - 17 Hz or 0 - 1000 ms for 18 - 24 Hz for each condition. (* $q < .05$).

Table 1: Summary of final test results

Frequency	Effect	F-values	FDR q -values	η_p^2 -values
2 - 4 Hz	<i>Repetition practice</i>	$F(2, 48) = 4.340$	0.024	0.153
5 - 7 Hz	<i>Repetition practice</i>	$F(2, 48) = 3.751$	0.036	0.135
	<i>Time Interval</i>	$F(1, 24) = 105.620$	0.000	0.815
	<i>Repetition practice X Time Interval</i>	$F(2, 48) = 8.292$	0.006	0.257
8 - 10 Hz	<i>Repetition practice</i>	$F(2, 44) = 6.684$	0.015	0.233
	<i>Time Interval</i>	$F(1, 22) = 62.389$	0.000	0.739
	<i>Repetition practice X Time Interval</i>	$F(2, 44) = 8.197$	0.002	0.271
11 - 17 Hz	<i>Time Interval</i>	$F(1, 23) = 45.794$	0.000	0.666
	<i>Repetition practice X Time Interval</i>	$F(2, 46) = 7.658$	0.003	0.250
18 - 24 Hz	<i>Repetition practice</i>	$F(2, 46) = 5.292$	0.024	0.187
	<i>Time Interval</i>	$F(1, 23) = 47.643$	0.000	0.674

Electrophysiological results in the repetition cycle 1 and 2:

To examine if testing modulates oscillatory power changes already during practicing on day 1, data collected in the repetition phase was investigated in the same frequency bands and time intervals as the final test data. Since participants performed dissimilar tasks for studying compared to testing, only tested items were analyzed comparing the conditions test and testFB. Results for the testFB - test contrast for repetition cycle 1 and 2 were depicted as time-frequency spectrograms in Figure 2. For the 2 - 4 Hz frequency band, a 2 x 2 rm ANOVA with the factors *Repetition practice* (test vs. testFB) and *Repetition cycle* (cycle 1 vs. cycle 2) was performed. A significant main effect of *Repetition cycle* revealed an increase in slow theta power from cycle 1 to cycle 2 ($F(1, 24) = 16.264$, $q = 0.003$, $\eta_p^2 = 0.404$; cycle 1: $M = 73.530$ % , $SE = 8.869$; cycle 2: $M = 96.906$ % , $SE = 9.711$). For the remaining frequency bands, the factor *Time interval* (0 - 500 ms vs. 500 - 1000 ms) was additionally examined. The significant main effect of *Time interval* revealed increased power in the 0 - 500 ms compared to the 500 - 1000 ms time interval for all frequency bands. Statistical results were reported in Table 2. All remaining main or interactive effects did not reveal statistical significance ($F_s \leq 4.072$, $q_s \geq 0.284$, $\eta_{ps}^2 \leq 0.145$).

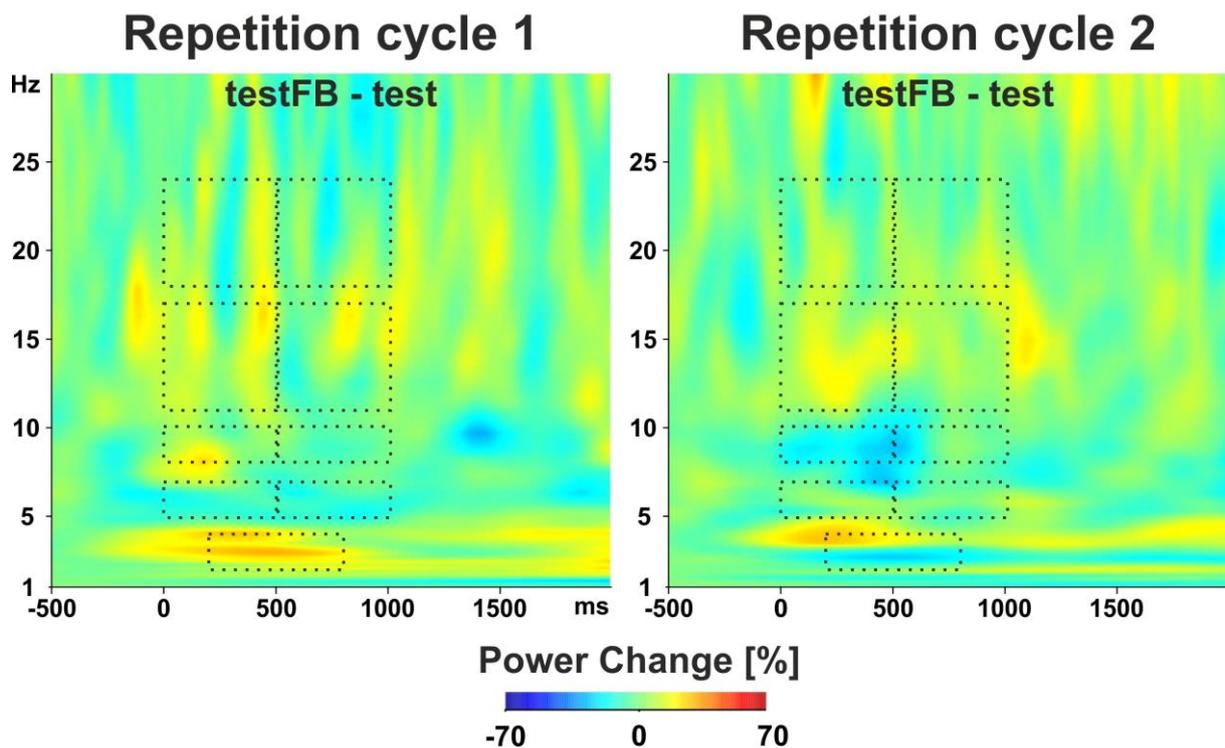


Figure 10. Time-frequency spectrograms of power differences averaged over all electrodes depicting the EEG analyses time-locked to test onset in the repetition cycle 1 (left) and cycle 2 (right) subtracting power changes for testFB - test.

Table 2: Summary of repetition phase results

Frequency	F-values	FDR q-values	η_p^2 -values	0 - 500 ms M (SE)	500 - 1000 ms M (SE)
5 - 7 Hz	F(1, 23) = 178.680	0.000	0.886	120.94 (7.62)	11.54 (3.81)
8 - 10 Hz	F(1, 22) = 87.558	0.000	0.799	71.87 (7.55)	0.84 (3.82)
11 - 17 Hz	F(1, 24) = 52.140	0.000	0.685	36.08 (4.59)	1.98 (2.46)
18 - 24 Hz	F(1, 22) = 22.561	0.001	0.506	10.66 (1.66)	2.20 (1.46)

Discussion

The present study investigated the effect of retrieval tests in combination with performance and correct answer feedback on later retrieval performance as well as oscillatory activity collected with EEG. Participants learned low-associated word pairs by either studying, testing or testing with additional performance feedback. After a delay of 1 day, a final recall test was performed revealing the typical testing effect with increased recall success for items previously tested compared to items previously studied. Furthermore, testing with additional performance feedback increased recall success more than only testing. Results of the retrieval related oscillatory activity suggested on the one hand that testing more than studying boosts

synchronization of brain regions by increasing slow theta as well as the slow beta activity and on the other hand additional performance feedback especially led to an increase in fast theta and alpha activity. Moreover, oscillatory activity differed due to testing or studying only in the final test, but not in the repetition phase.

Most studies investigating oscillatory power modulated by the testing effect examined EEG data collected during the repetition phase instead of the final test (e.g., Beardsley et al., 2017; Ferreira et al., 2014, Guran et al., 2019; Pastötter et al., 2011). Nonetheless, the present study aimed at exploring final test results since the beneficial effect of testing over studying was consistently reported to increase with a delay of one day and more (e.g., Roediger & Butler, 2011; Rowland, 2014), possibly due to memory consolidation processes (Antony, Ferreira, Norman, & Wimber, 2017). In line with this idea, Pastötter & Bäuml (2016) examined oscillatory activity in two final tests revealing increased fast theta power for previously tested compared to studied items in the second final test, but not the first. In contrast, results of the present study revealed increased slow theta as well as slow beta activity for items previously practiced by testing compared to studying. These differences might result from specificities in experimental design and analysis. For example, while Pastötter & Bäuml (2016) defined their analysis on subtracting activity collected for recalled and forgotten, we directly compared only successfully retrieved items previously tested or studied.

Research proposed mainly two hypotheses to explain the mechanisms underlying the testing effect (see Rowland, 2014; van den Broek et al., 2016). More precisely, testing enhances later retrieval success more than studying by improving either processes related to elaboration of the semantic network (e.g., Carpenter, 2009; Carpenter & Delosh, 2006) or search-set restriction processes (e.g., Karpicke & Zangrando, 2010; Thomas & McDaniel, 2013). In line with these ideas, prior studies associated alpha and slow beta power changes to semantic long-term memory processing, including deep encoding and elaboration of item information (e.g., Klimesch, 1999; Pastötter & Bäuml, 2016, Scholz et al., 2017). Hence, the present results support the semantic elaboration hypothesis by revealing an increase in slow beta activity due to prior testing. Additionally, an increase in lower theta oscillations was

recorded as well for practicing by testing compared to studying. Some studies suggested slow theta oscillations to reflect processes related to recollection and conscious awareness (e.g., Bramão & Johansson, 2017; Gruber et al., 2008). In contrast, increased slow theta activity was reported as well for processing of correct answer feedback in exact the same time interval and frequencies as in the present results (Pastötter & Bäuml, 2016). Since we provided correct answer feedback in the repetition phase for all conditions, the slow theta activity could also reflect retrieval processes related to the correct answer feedback. In order to clarify whether processes related either to recollection or correct answer feedback or both led to increased low theta power following prior testing, modulations of the level of retrieval depth during memory retrieval could be investigated in future studies.

Prior research revealed that performance feedback affects learning from correct answer feedback and thereby boosts later recall success possibly by guiding attention as well as reinforcement learning (Ernst & Steinhauser, 2012; Miendlarzewska et al., 2016, see also Chapters 2 & 3a). In line with this idea, the present study aimed at investigating the neural mechanisms underlying improved retrieval success following testing with additional performance feedback and subsequent correct answer feedback using oscillatory activity. Results in the final test revealed increased alpha frequencies for items successfully retrieved following testing with additional performance feedback compared to practicing by testing or studying. Alpha-band oscillations were suggested to be linked to attention that enables memory access by the suppression of task-irrelevant memories (Klimesch, 2012). Additionally, fast theta oscillations increased as well due to testing with additional performance feedback compared to prior studying possibly reflecting increased interference processing and cognitive control (Hanslmayr et al., 2010). Hence, in line with the two hypotheses previously proposed, the present results implied that performance feedback seemed to enhance memory selection processes more than to elaborate the semantic network.

Oscillations in the beta frequency range were investigated in testing effect studies as well (Guran et al., 2019; Pastötter et al., 2011) reporting mixed findings. Beta oscillations were reported to decrease due to practicing by testing compared to studying and thus, were

suggested to reflect a decrease in task demands (Guran et al., 2019). In contrast, increased beta oscillations were as well reported for final word list learning compared to first list, but no change due to testing compared to studying was found and therefore, beta oscillations were proposed to reflect shifts in learning strategy from rote rehearsal to nonvocalized forms of encoding (Pastötter et al., 2011). In contrast, results of the present study revealed decreased beta power for items previously practiced by testing with additional performance feedback followed by correct answer feedback compared to testing with only correct answer feedback.

The findings could differ on the one hand due to the delayed memory test since we investigated results collected in the final test, which was performed with a delay of one day whereas the other studies investigated the practice phase instead. Such a delay possibly provided more time to successfully consolidate the memories (see Antony et al., 2017). On the other hand, research investigating performance feedback processing for example in probabilistic learning tasks reported decreased beta oscillations to reflect reduced striatal involvement if feedback was delayed (e.g., Foerde & Shohamy, 2011; Marco-Pallarés, Münte, & Rodríguez-Fornells, 2015; Weismüller, Kullmann, Hoenen, & Bellebaum, 2019). Correspondingly, decreased beta activity in the present study could as well reflect striatal activity instead. These assumptions might be addressed in more detail with other neuroimaging techniques such as functional magnetic resonance imaging (fMRI).

In line with prior studies investigating the repetition phase (e.g., Ferreira et al., 2014, 2019; Guran et al., 2019), the present study aimed at examining if differences in oscillatory activity between items previously tested with or without additional performance feedback detected for the final test were present already during the repetition phase. Since the additional performance feedback was presented following a testing situation but before the correct answer feedback was presented, oscillatory activity collected during testing in repetition cycle 1 was not influenced by additional performance feedback. In contrast, learning from correct answer feedback was boosted by additional performance feedback right after the test. Therefore, it may have modulated the oscillatory activity of the following test that was performed in repetition cycle 2. Thus, differences between testing with or without additional

performance feedback were only expected for repetition cycle 2. Results of the present study did not reveal differences in oscillatory activity between testing with or without additional performance feedback, neither for repetition cycle 1 nor for cycle 2. These results suggested that the beneficial effect of additional performance feedback was only detected after a delay of one day and hence may be a result of modulated consolidation processes (Antony et al., 2017).

Furthermore, Staudigl et al. (2010) reported decreased theta oscillations in repetition cycle 2 compared to cycle 1, which we sought to replicate. Instead, the present results revealed an increase in theta oscillations from repetition cycle 1 to cycle 2, but only for slow theta power, which possibly indicated enhanced recollection processes in repetition cycle 2 (Bramão & Johansson, 2017; Gruber et al., 2008). These mixed findings might arise from the specific definitions in theta bands. While Staudigl et al. (2010) investigated theta activity ranging from 5 - 9 Hz, the present study evaluated activity for slow (2 - 4 Hz) and fast (5 - 7 Hz) separately as separate neural processes were suggested to modulate these frequencies (e.g., Lega et al., 2012). Hence future studies might mind such variations for slow and fast theta activity and further investigate their activity patterns.

In summary, slow beta activity for previously tested compared to studied word pairs was increased possibly reflecting semantic long-term memory processing as well as information elaboration (Pastötter & Bäuml, 2016; Scholz et al., 2017). Hence, these findings support the semantic elaboration hypothesis, which suggests that testing boosts later memory retrieval by enhancing elaboration of the semantic network (e.g., Carpenter, 2009; Carpenter & Delosh, 2006). In contrast, providing additional performance feedback led to an increase in fast theta as well as alpha activity, which was suggested to reflect attention modulation and increased cognitive control regulating interference processing (Hanslmayr et al., 2010). Thus, these results promote the search-set restriction account instead, which proposes that testing leads to improved selection processes at memory retrieval (e.g., Karpicke & Zangwill, 2010; Thomas & McDaniel, 2013). Consequently, results indicated that both mechanisms seem to evoke the testing effect and in addition as suggested by Rosburg et al. (2015), the type of

feedback that was provided following testing might modulate how intensely each mechanism is used.

Conclusion

The present study aimed at investigating the neural mechanisms of performance feedback enhanced learning from correct answer feedback in a testing effect paradigm by analyzing oscillatory activity in the repetition phase as well as final test. Results revealed increased slow theta as well as slow beta activity for tested compared to studied material supporting the semantic elaboration hypothesis. Furthermore, performance feedback resulted in an increase of fast theta as well as slow alpha activity probably reflecting enhanced memory selection as well as attention orientation processes instead. Detecting modulations of oscillatory activity in the final test but not in the repetition phase suggests, that consolidation processes might be modulated by the feedback type provided while practicing. Finally, successful memory retrieval due to testing with correct answer feedback can be modulated by additional performance feedback possibly enhancing semantic elaboration as well as search-set restriction.

Chapter 4

So far, the behavioral data in the repetition phase as well as the final test phase was investigated in Chapter 2 and 3a. Results revealed increased retrieval success after a delay of one day if performance feedback was provided following a memory test and before correct answer feedback was presented in the repetition phase (feedback assisted TPE). Furthermore, we investigated the temporal course of the neural processes underlying learning from correct answer feedback with EEG (Chapter 3a). In this study, results suggested that additional performance feedback enhances learning from correct answer feedback by improving attention orientation as well as search-set restriction and semantic elaboration. Examining oscillatory activity following the memory tests in the repetition phase as well as final test (Chapter 3b) revealed increased synchronization of the neural firing patterns for previously tested compared to studied material in the final test, which was even more enhanced due to the additional presentation of performance feedback. This effect was only observed after a delay of one day and hence, might be related to consolidation processes.

However, the studies presented so far did not provide detailed information about which areas in the brain modulate the enhanced retrieval following feedback assisted TPE. Therefore, we adapted the present research design and collected fMRI during the final test to examine structural aspects of the neural mechanisms underlying memory retrieval. This study is explained in the following paragraphs in detail.

Functional underpinnings of feedback assisted test-potentiated encoding

Abstract

The testing effect describes the finding that retrieval practice compared to study practice enhances memory performance. Prior evidence consistently demonstrates that this effect can be further boosted by providing feedback after retrieval attempts: Test-potentiated encoding (TPE). The present functional magnetic resonance imaging (fMRI) study was aimed at investigating the neural processes during memory retrieval underlying the beneficial effect of additional performance feedback beyond the benefits of only adding correct answer feedback. Twenty-seven participants learned 180 weakly associated cue-target word pairs either by studying, testing or testing followed by immediate positive or negative performance feedback. Afterwards, full word pairs were presented again. One day later, a cued retrieval test was performed inside the MRI scanner. Behavioral results confirmed the expected testing effect on long-term memory, and demonstrated that adding explicit performance feedback further enhanced memory performance relative to the study or standard test conditions. Stronger functional engagement while retrieving items previously studied were found in left prefrontal cortex (PFC), including Broca and left middle frontal gyrus (MFG), as well as left superior parietal lobe (SPL). In contrast, lateral temporoparietal areas including left supramarginal gyrus (SMG) and hippocampus were more strongly recruited while retrieving items previously tested. Performance feedback increased posterior hippocampal activation during final test and resulted in tighter functional coupling between hippocampus with ventral striatum (VS) and SMG as well as VS and SMG with middle temporal gyrus (MTG) and SPL. Our findings revealed that binary performance feedback combined with test-potentiated encoding enhances memory performance mostly by increasing the functional coupling between hippocampus, SMG and ventral striatum with lateral temporoparietal cortex.

Introduction

Learning by memory retrieval improves long-term memory compared to repetition by studying (see Roediger & Butler, 2011; Roediger & Karpicke, 2006a, b). For instance, when learning vocabulary, recalling the translation from memory with only reading a cue word (testing) is a more effective learning strategy compared to simple studying the word together with its translation. This highly reproducible and strong phenomenon is called “testing effect”. In addition, prior research revealed that providing a study opportunity after a testing situation can further increase long-term retention, termed “test-potentiated encoding” (TPE; see Arnold & McDermott, 2013; van den Broek et al., 2016). While learning vocabulary, a testing situation can boost learning of the correct answer, which is presented following the test as the word with its translation, probably due to an enhancement of encoding processes.

Several theories and hypotheses have been put forward about the underlying mechanisms of these effects. In fact, most studies focused either on the idea that testing helps elaborating the semantic network (e.g., Carpenter, 2009; Carpenter & DeLosh, 2006) or that testing reinforces the representation of information and thus leads to an improvement of the selection processes during memory retrieval (Karpicke & Zaromb, 2010; Thomas & McDaniel, 2013). In line with these theoretical accounts, research investigating the neural basis of the testing effect using functional magnetic resonance imaging (fMRI) reported increased activation in prefrontal cortex (PFC) areas, such as the middle frontal gyrus (MFG) and inferior frontal gyrus (IFG), during testing compared to studying (e.g., Rosner, Elman, & Shimamura, 2014; Vannest et al., 2012). The involvement of these PFC regions has been associated to higher attentional demands induced by testing as well as improved conflict monitoring (Rosner et al., 2014; van den Broek et al., 2013). In contrast, practicing by studying compared to testing has been associated with the engagement of temporoparietal regions, such as the inferior parietal lobe (IPL) and middle temporal gyrus (MTG), which were related to accessing semantic representations (van den Broek et al., 2013; Vannest et al., 2012; Wing et al., 2013).

Previous research focusing on final retrieval tests reported mixed findings ranging from no differences in neural activation for previously tested compared to studied items or vice versa

(Rosner et al., 2014; Wirebring et al., 2015) to a decrease in neural activation in frontoparietal networks following practice testing compared to studying (Keresztes et al., 2014; Wiklund-Hörnqvist, Stillesjö, Andersson, Jonsson, & Nyberg, 2021). This suggests that practicing by testing reduces the required neural activation to perform accurately in the final test (van den Broek et al., 2016). Moreover, recent research on the testing effect focused on the medial temporal lobe (MTL) reporting increased hippocampal engagement due to testing, with the anterior hippocampus being more strongly recruited after multiple testing whereas the posterior hippocampus being more engaged for items successfully retrieved once during practicing (Wiklund-Hörnqvist et al., 2021).

At present, only few neuroimaging studies have investigated the neural mechanisms underlying learning from correct answer feedback after practice testing (TPE), reporting also inconsistent results. Whereas Vestergren and Nyberg (2014) did not find differences in functional activation patterns comparing later remembered to forgotten information, Liu, Liang, Li, and Reder (2014) discovered a marginal increase in striatal activation during studying following an unsuccessful test leading to a subsequent successful retrieval and suggested a modulation in memory representations due to negative feedback. Supporting this finding, Wiklund-Hörnqvist, Andersson, Jonsson, and Nyberg (2017) reported increased activation of the ventral striatum during the first correct retrieval, which was decreased in following tests.

So far, prior studies exploring testing with additional correct answer feedback mostly presented the to-be-learned material in the same way as in the study condition (Rowland, 2014). In contrast, only few studies explicitly added information regarding the accuracy of the retrieval attempt, either by additionally presenting the words “correct” or “incorrect” (e.g., Jacoby et al., 2010; Ergo et al., 2019) or by coloring the font in green or red (e.g., Ernst & Steinhauser, 2012). Offering information on prior test performance diminishes internal processes related to self-performance evaluation following correct answer feedback. Only few behavioral studies to date provided performance feedback immediately after testing and preceding the correct answer feedback, even though separating these feedback information might support attention shifting towards the correct answer feedback and hence possibly

enhance encoding processes (see Study 1, Chapter 2). Memory benefits from such a preceding performance feedback could provide an opportunity to investigate the mechanisms of the testing effect as the reinforcing quality of positive and negative feedback might explicitly boost processes, which result either in the search-set restriction or in semantic elaboration. However, to the best of our knowledge, no fMRI studies to date have examined the functional underpinnings promoting the effects of positive and negative feedback on subsequent test-potentiated encoding (i.e., feedback assisted TPE).

Examining this assumption behaviorally, we designed a study in, which a positive or negative feedback indicating prior test performance was presented just before the study opportunity (see Study 1, Chapter 2 & Study 2, Chapter 3). Results revealed a modest but solid and consistent beneficial effect of additional performance feedback on retrieval performance in a final test performed one day after the initial learning, which was not present in the end of the practice phase. In line with prior studies (e.g., Vestergren & Nyberg, 2014), we suggested that relocating attentional resources and the reinforcing aspects of performance feedback in addition to elaborating the semantic network might cause this result (Study 1, Chapter 2).

The present fMRI study constitutes the first study investigating the neural basis of memory retrieval underlying feedback assisted TPE. To this end, participants were asked to learn word pairs with low association strength in an initial study phase followed by a practice phase in which the word pairs were either repeated by testing or by studying and subsequently followed by correct answer feedback. For half of the tested word pairs, participants received an additional performance feedback in between the retrieval attempt and the correct answer feedback. One day later, participants performed a final test on all previously learned word pairs inside the MRI scanner.

First, we expected to replicate the behavioral results observed in our previous behavioral paper (see Study 1, Chapter 2) as well as the fMRI results for tested compared to studied materials (e.g., Keresztes et al., 2014; Wiklund-Hörnqvist et al., 2021). Given previous evidence on successful retrieval (e.g., Spaniol et al., 2009), we expect to find increased

involvement of frontal and parietal areas following studying compared to testing (van den Broek et al., 2016; Wiklund-Hörnqvist et al., 2017) even though some other research reported contrasting findings such as an increase in frontal activity for tested compared to studied items (Wiklund-Hörnqvist et al., 2021). Moreover, we aimed at extending prior findings by comparing items tested with additional performance feedback to those with only correct answer feedback or studied items. Second, we sought to examine differences in anterior versus posterior hippocampus based on accounts suggesting differential involvement of hippocampal regions along the y-axis (e.g., Poppenk, Evensmoen, Moscovitch, & Nadel 2013; Poppenk & Moscovitch, 2011) with the posterior hippocampus being more involved in retrieval of detailed, pattern-separated representations, hence more involved in test than study (see Wiklund-Hörnqvist et al., 2021). Third, to specifically focus on the beneficial effect of performance feedback, we also examined the involvement of ventral striatum due to its well-known recruitment in feedback-related learning (e.g., O'Doherty et al., 2004; Wiklund-Hörnqvist et al., 2017). Fourth, we sought to examine functional coupling between left hippocampus, ventral striatum and other frontoparietal areas reported to be involved in the testing effect and extended to additional performance feedback (see Wing et al., 2013).

Methods

Participants

A total of 29 Spanish native speakers participated in this study. Data from two participants was excluded due to poor memory performance (i.e., 2 SD lower than the mean) on the final test after a 1-day retention interval. The final sample consisted of 27 right-handed participants (mean age = 24.2 years, SD = 4.4 years; 15 females) that reported to have normal or corrected-to-normal vision, no red-green deficiency and no history of major medical, neurological or psychiatric disorders. The study protocol was approved by the Ethics Committee of the Basque Center on Cognition, Brain and Language (BCBL) and was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving human participants. Prior to their inclusion in the study, all

participants provided informed written consent. Participants received monetary compensation for their participation.

Materials

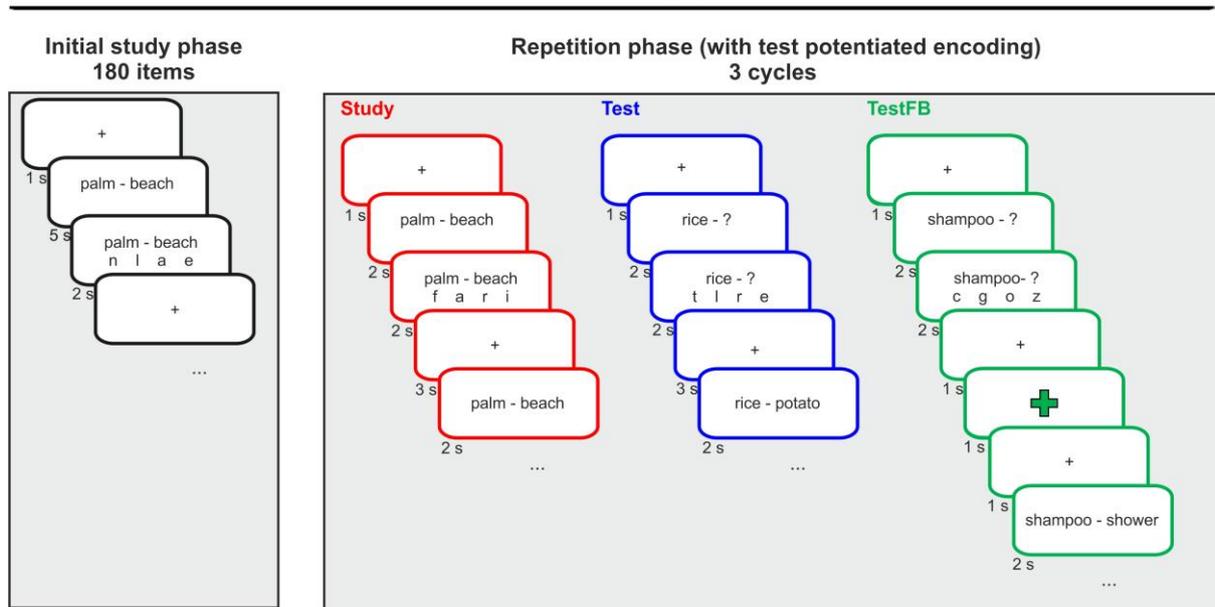
The stimulus material consisted of 180 Spanish weakly associated cue-target word pairs (e.g., *feather-duck*; *towel-soap*) taken from the database by Nelson, McEvoy, and Schreiber (2004). All word pairs were chosen based on their forward (FSG) (>.04), backward (BSG) (>.04), mediated (MSG) (>.04) and overlapping (OSG) (>.05) association probability to ensure that participants were unlikely to produce the correct response by guessing and that they would need to learn the associations to provide correct responses. The word pairs were randomly assigned to one of the repetition practice conditions for each participant.

Procedure

The experiment was divided into two sessions spaced about 24 hours apart. Figure 1 provides an overview of the experimental design. On day 1, participants started with intentionally studying all word pairs followed by three repetition cycles. On Day 2, participants underwent MRI scanning and performed a final test inside the magnet, and then they performed a second final test in a behavioral cabin where they had to verbalize all word pairs they were able to retrieve inside MR scanner. This design was chosen to obtain around 50% retrievable items for all repetition practice conditions.

During all cycles, short breaks of 30 s were added after blocks of 45 word pairs and a fixation cross of 1 s preceded every stimulus presentation. Stimuli were presented using Presentation software (Neurobehavioral Systems, Berkeley, CA).

Session 1 (Day 1)



Session 2 (Day 2)

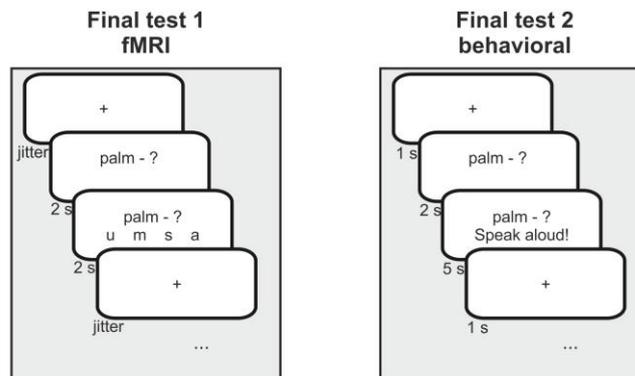


Figure 11. Depiction of the experimental procedure.

Session 1: initial study phase (behavioral)

In the first phase, participants studied all 180 items, one pair at a time. More specifically, each word pair was displayed in random order for a total of 7 s and participants were instructed to learn the word pairs as well as possible for a later test. Furthermore, participants were instructed that the word on the left side was the cue word (e.g., “*palm*”) whereas the word on the right was the target word (e.g., “*beach*”), which they would need to retrieve in a later test when presented with the cue word (e.g., “*palm*”). After 5 s, four letters appeared below the stimulus and participants were instructed to report the third letter of the target word by pressing the corresponding key (e.g., “a” is the third letter of the word “beach”). Participants were instructed that the response had to be given within the 2 s that the cue word was still presented

on the screen. This task was added to ensure that participants paid attention to the word pairs and to familiarize them with the response procedure for the next phases.

Session 1: repetition phase (behavioral)

In the second part of Session 1, participants went through three repetition cycles wherein one third of the word pairs (i.e., 60 items) were trained by studying, which means that the cue and target were presented again for 4 s (condition: “study”). The other two-thirds (i.e., 120 items) were repeated by testing, where only the cue word was presented with a “?” instead of the target word for 4 s). For all repetition practice conditions, four response options appeared on the screen during the last 2 s of word presentation and participants were requested again to indicate the third letter of the target word as in the initial study phase. If the participants did not know the target word when tested, they were instructed to provide their best guess.

After half of the tested word pairs (i.e., 60 items), performance feedback (i.e., positive or negative) was provided for 1 s (condition: “testFB”), and the other half was followed by a fixation cross not providing information about the accuracy of the response (condition: “test”). For the testFB condition, positive feedback consisting of a green plus sign was presented if the response was correct and given within the time limit of 2 s. Conversely, negative feedback was provided by a red minus sign for incorrect responses given within the time limit of 2 s. Last, independent of how the items were repeated previously, all cues and its target words were fully presented again for 2 s to provide the opportunity to re-encode the items (i.e., test-potentiated encoding).

This second part of Session 1 was repeated three times (i.e., three cycles) to boost later memory performance. All word pairs were randomly assigned to one of the three repetition practice conditions (study, test, or testFB) for each participant, with stimuli remaining in the same conditions not changing from one repetition cycle to the next one.

Session 2: final test 1 (fMRI) and test 2 (behavioral)

Around 24 hours after the Session 1, participants returned to the lab for the final test. The final test was split into two parts: the first one was performed inside the MR scanner (i.e.,

final test I) and the second one was conducted outside of the scanner and it was a behavioral test (i.e., final test II).

In the final test 1, only the cue words were presented for 4 s with a “?” on the right side. After 2 s, 4 letter options were displayed on the screen and participants were asked to indicate the third letter of the target word via button press. No feedback was provided afterwards. Between consecutive trials, a fixation cross was shown, with a jittered duration. Optimal sequencing software (OptSeq2, <https://surfer.nmr.mgh.harvard.edu/optseq/>; see Dale, 1999) was employed to determine the variable duration of the jitter fixation (0-5000ms) and the order of trial types in each of the repetition practice conditions within each functional run to optimize efficient recovery of the blood-oxygen level dependent (BOLD) response (Fischl et al., 1999). This part was split into two runs with 90 word pairs each.

The final test 2 was performed outside of the scanner as a confirmatory test for the in-scanner results. Again, only the cue word was presented for 2 s. Afterwards, participants were asked to verbalize the full word pair while the cue word was still presented and their voice was recorded for 5 s. Participants were again informed about the time limitation to respond which was only until the cue word disappeared. Responses were rated as correct if the participants correctly recalled the target word within the 5 s time window. A fixation cross of 1 s was displayed between trials.

fMRI data acquisition

Whole-brain MRI data acquisition was conducted on a 3-T Siemens Prisma Fit whole-body MRI scanner (Siemens Medical Solutions) using a 64-channel whole-head coil. Foam pillows were provided to help stabilize head position and scanner noise was reduced with earplugs and padded headphones. The MRI acquisition included T1-weighted structural images and functional T2* images. High-resolution MPRAGE T1-weighted structural images were collected with the following parameters: time-to-repetition (TR) = 2530 ms, time-to-echo (TE) = 2.36 ms, flip angle (FA) = 7°, field of view (FoV) = 256 mm, voxel resolution = 1 mm³, 176 slices. Functional images were acquired in two consecutive functional runs using a single gradient-echo echo-planar multiband pulse sequence with the following acquisition

parameters: TR = 1000 ms; TE = 35 ms; MB acceleration factor = 5; 65 axial slices with a 2.4 mm³ voxel resolution; no inter-slice gap; FA = 56°; FoV = 210 mm; 486 volumes. The first twelve functional volumes in each run were discarded to allow T1-equilibration effects.

fMRI data preprocessing

Standard SPM12 (The Wellcome Department of Cognitive Neurology, London, UK) preprocessing routines and analysis methods were employed. Images were first corrected for differences in timing of slice acquisition and then realigned to the first and mean volumes using rigid-body registration. Each subject's functional volumes were spatially smoothed with a 4-mm full-width half-maximum (FWHM) Gaussian kernel. Motion parameters extracted from the realignment process were used to identify outlier volumes with sudden scan-to-scan motion exceeding 0.5 mm and volumes whose global intensity fluctuate more than 1.3% and corrected them via interpolation between the nearest non-repaired scans (ArtRepair version 5b; Mazaika, Hoefft, Glover, & Reiss, 2009). After volume repair, functional volumes were co-registered to the T1 images using 12-parameter affine transformation and spatially normalized to the MNI space by applying non-linear transforms estimated by deforming the MNI template to each individual's structural volume. During normalization, the volumes were sampled to 3-mm cubic voxels. The resulting volumes were then spatially smoothed with a 7-mm FWHM Gaussian kernel. Due to the quadratic relation between separate smoothing operations, the total smoothing applied to the functional data was approximately equivalent to smoothing with an 8-mm FWHM Gaussian kernel. Finally, time series were temporally filtered to eliminate contamination from slow frequency drift (high-pass filter with cut-off period of 128 s).

fMRI data analyses

Statistical analyses were performed on individual participant's data using the general linear model (GLM). The motion parameters for translation (x, y, z) and rotation (yaw, pitch, roll) were also included as covariates of noninterest in this GLM. fMRI time series data were modeled by a series of events convolved with a canonical hemodynamic response function (HRF). Two different GLM models were conducted. The first model (i.e., analysis I), included 3 regressors of interest related to their prior repetition practice condition during the repetition

phase: *study*, *test* and *test feedback* (i.e., testFB). These regressors were time-locked to the onset of the stimulus presentation during the final in-scanner memory test. Trials incorrectly retrieved in the final test were modeled separately and excluded from the main analysis. For the second model (i.e., analysis II) we conditionalized final test results time-locked to the onset of the stimulus presentation during the final in-scanner memory test for items previously tested with and without feedback in the repetition phase depending on retrieval performance in repetition cycle 1. Hence, analysis II included 2 regressors of interest for the conditions test and testFB and 2 regressors of interest for correct and incorrect memory retrieval in repetition cycle 1, respectively. In analysis II, trials incorrectly retrieved in the final in-scanner memory test as well as trials belonging to the study condition (i.e., previously studied at Session 1) were modeled separately and excluded from further analyses.

SPM12 FAST was used for temporal autocorrelation modeling in these GLMs due to its optimal performance in terms of removing residual autocorrelated noise in first-level analyses (Olszowy, Aston, Rua, & Williams, 2019). The least-squares parameter estimates of the height of the best-fitting canonical HRF for each condition were used in pairwise contrasts. Contrast images computed on a participant-by-participant basis were submitted to group analysis. At the group level, whole-brain contrasts between repetition practice conditions were computed by performing one-sample t-tests on these images, treating participants as a random effect. For analysis I, a general All > Null (fixation baseline) contrast was computed to identify all the brain voxels involved in this fMRI experimental design across repetition practice conditions of interest for all the in-scanner correct responses. In addition, also for analysis I, we performed contrasts across all participants for the main conditions of interest: study vs. test, study vs. testFB, test vs. testFB. Familywise error rate (FWE) correction at the cluster level set at $p < 0.05$, using a voxel-extent threshold of $p < 0.001$, was applied to whole-brain maps involving all participants. Brain coordinates throughout the text, as well as in tables and figures, are reported in MNI atlas space (Cocosco, Kollokian, Kwan, & Evans, 1997).

The MARSBAR toolbox (Brett, Anton, Valabregue, & Poline, 2002) was used for ROI analyses. Given that previous meta-analyses on successful memory retrieval (Spaniol et al.,

2009; Neurosynth.org) do not include all the relevant regions that previous studies on the testing effect reported so far (van den Broek et al., 2016), we functionally defined left-lateralized regions of interest (ROIs) based on the overlap of the general activation produced by our design (All vs. Null) and anatomical masks of regions found as relevant in previous testing effect studies, including MFG (center of mass= -38 26 34; volume = 13488 mm³; e.g., Vannest et al., 2012; Wing et al., 2013), Broca (i.e., *pars opercularis* and *pars triangularis*; center of mass= -45 21 16; volume = 21288 mm³; e.g., van den Broek et al., 2013), SPL (center of mass= -24 -61 57; volume = 15128 mm³; e.g., Liu & Reder, 2016; Rosner et al., 2014), SMG (center of mass= -55 -26 32; volume = 3080 mm³; e.g., Nelson, Arnold, Gilmore, & McDermott 2013; Vestergren & Nyberg, 2014) and MTG (center of mass= -48 -59 3; volume = 4560 mm³; e.g., Wiklund-Hörnqvist et al., 2021). The left ventral striatum was also included as well as previous studies highlighted its importance for feedback evaluation (center of mass= -25 -4 -6; volume = 1280 mm³; see O'Doherty et al., 2004; Wiklund-Hörnqvist, et al., 2017). In addition, prior studies on the testing effect investigated the involvement of left and right anterior and posterior Hippocampus (Wiklund-Hörnqvist, et al., 2021) which we will focus on as well. In contrast to the other ROIs, the hippocampal ROIs were defined structurally due to its specific size and shape.

Functional connectivity analyses were conducted using the beta-series correlation method (Rissman et al., 2004) implemented in SPM12 with custom MATLAB scripts. The canonical HRF in SPM was fitted to each occurrence of each repetition practice condition and the resulting parameter estimates (beta values) were sorted according to the study conditions to produce a condition-specific beta series for each voxel. Pairwise functional connectivity analyses were performed for left hippocampus, left ventral striatum and left SMG with each of the remaining ROIs for each participant and repetition practice condition. Since the correlation coefficient ranges from -1 to +1, an arc-hyperbolic tangent transform (Fisher, 1921) was applied to these beta-series correlation values (*r* values) to make its null hypothesis sampling distribution approach that of the normal distribution. The normally distributed Fisher's *Z* values for each ROI pairs were submitted to statistical analyses (see below).

Statistical Analysis

Day 1 results including memory performance and reaction times (RT) collected in the three repetition cycles on Day 1 were analyzed using a 2 x 3 repeated measures ANOVA with the factors *Condition* (test vs. testFB) and *Repetition Cycle* (cycle 1-3).

Day 2 results including memory performance, RT, functional activation, and connectivity results were analyzed in 2 separate analyses: analysis I and II. For *analysis I*, final test results, i.e., memory performance, RT, functional activation and connectivity results, were examined based on the repetition practice conditions (i.e., study, test or testFB) and then submitted to a one-way ANOVA with post-hoc t-tests. For *analysis II*, items previously tested with or without feedback were conditionalized based on participants' retrieval performance in repetition cycle 1 using a 2 x 2 repeated-measures ANOVA with the factors *Condition* (test vs. testFB) and *Performance in repetition cycle 1* (correct vs. incorrect retrieval). Items belonging to the study condition during the repetition phase could not be investigated compared to items previously tested as participants were always presented with the correct response. Additionally, analysis I and II involving the hippocampus were extended by the two factors *Laterality* (left vs. right) and *Y-axis* (anterior vs. posterior) based on our hypotheses.

Post-hoc t-tests of the behavioral results were corrected using Bonferroni correction. For the analyses of the fMRI data, all ANOVAs and post-hoc t-tests reported in the results were corrected for multiple comparisons using false-discovery rates (FDR) with a significance level of $q < 0.05$. For functional connectivity analyses, significant ANOVAs were corrected using FDR as well.

Results

Behavioral results

On Day 1, all participants first studied and then practiced the word pairs three times by either studying or testing or testing with additional performance feedback followed by correct answer feedback. We examined retrieval performance in the repetition phase for items tested either with or without additional feedback with a 3 (*Repetition Cycle*: cycle 1 - 3) X 2 (*Condition*:

test or testFB) repeated measures ANOVA. For this analysis, trials belonging to the study condition were not analyzed since the target word was always presented next to the cue word and hence, retrieval performance was not assessed in this condition. Results revealed the statistically significant main effects of *Repetition Cycle* ($F(2, 52) = 152.91, p < 0.001, \eta_p^2 = 0.86$), showing that retrieval performance increased from the first to the second and third repetition phase (cycle 1 vs. cycle 2: $p < 0.001$; cycle 1 vs. cycle 3: $p < 0.001$; cycle 2 vs. cycle 3: $p < 0.001$) (Figure 2, panel A); and *Condition* (test vs. testFB) ($F(1, 26) = 5.08, p < 0.05, \eta_p^2 = 0.16$), which revealed an increase in retrieval performance for items tested with additional performance feedback. No interaction effect emerged in this analysis ($F < 1, p = 0.843, \eta_p^2 = 0.007$). We also assessed RT for correct responses. Again, a statistically significant main effect of *Repetition Cycle* revealed a decrease in RT from cycle 1 to 3 ($F(2, 52) = 89.06, p < 0.001, \eta_p^2 = 0.77$; cycle 1 vs. cycle 2: $p < 0.001$; cycle 1 vs. cycle 3: $p < 0.001$; cycle 2 vs. cycle 3: $p < 0.001$) whereas neither the main effect of *Condition* ($F < 1, p = 0.464, \eta_p^2 = 0.021$) nor the interaction ($F = 1.78, p = 0.180, \eta_p^2 = 0.064$) reached statistical significance.

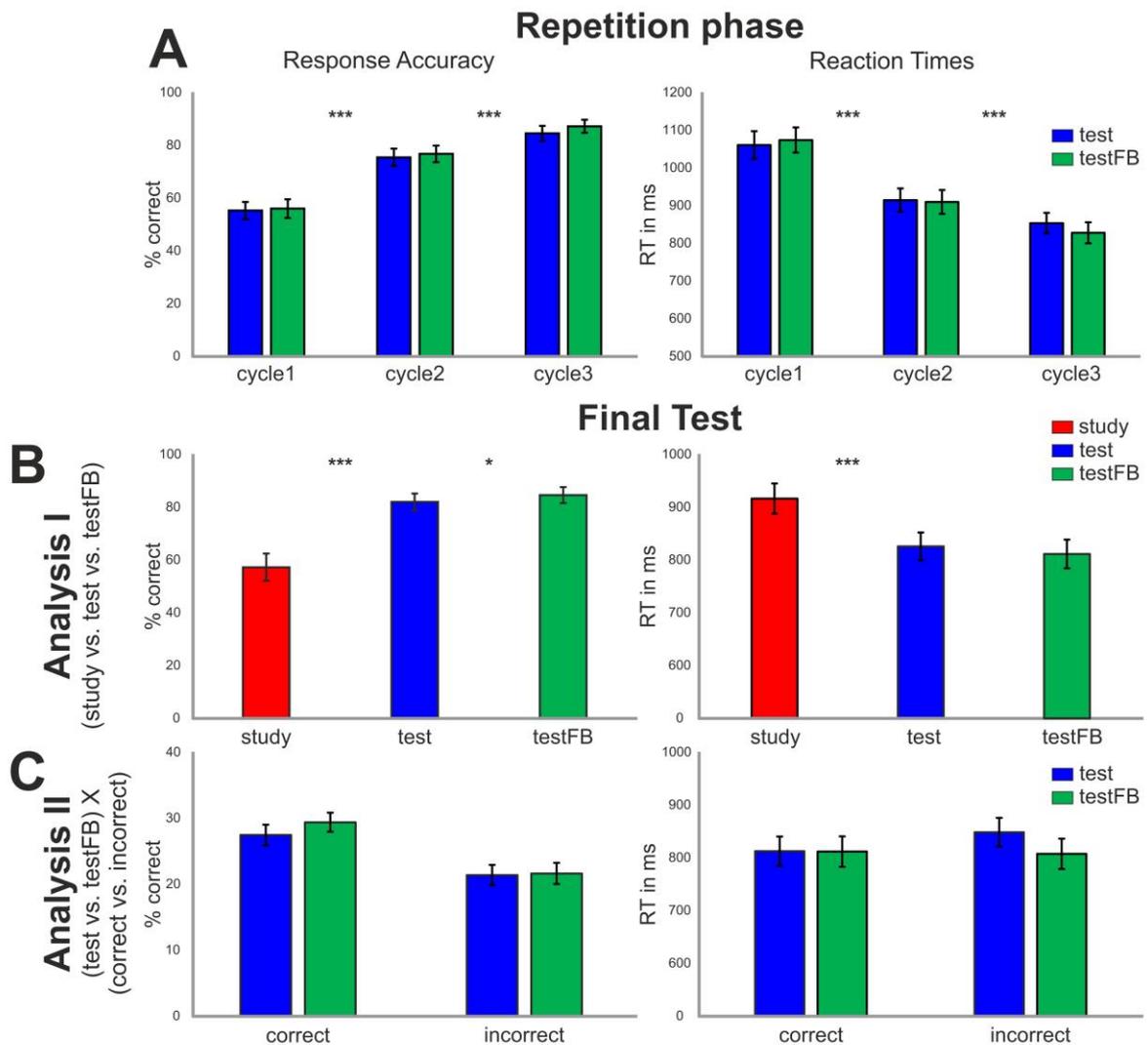


Figure 12. Mean percent response accuracy and reaction time results (A) in the repetition cycles 1-3 comparing trials tested with or without performance feedback (test vs. testFB). (B) Analysis I in the final test comparing the main repetition practice conditions, and (C) Analysis II in the final test of items tested with or without performance feedback conditionalized to retrieval performance in repetition cycle 1. Error bars show the standard error and asterisks denote statistically significant effects: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Analysis I. On Day 2, all word pairs were tested in a final test separated in two runs, one in the MRI collecting responses by button press and a second one afterwards in which participants were asked to say aloud all word pairs. Participants retrieved 75% ($SE = 4$) of all word pairs. The experiment was designed to allow participants to reach high levels of performance to have sufficient trials for fMRI analyses. A one-way ANOVA comparing final test retrieval performance for previously studied compared to tested or tested with feedback (Figure

2, panel B) revealed a statistically significant main effect of *Condition* indicating that participants retrieved 57 % of word pairs previously studied during the intervention phase, whereas they reached 82 % for items previously tested and 85 % if additional performance feedback was provided after the test in the repetition phase ($F(2, 52) = 74.01, p < 0.001, \eta_p^2 = 0.74$; study vs. test: $p < 0.001$; study vs. testFB: $p < 0.001$; test vs. testFB: $p < 0.05$). In line with prior findings on the testing effect in RTs, a statistically significant main effect of *Condition* revealed faster responses in the final test for items previously tested compared to items studied in the repetition phase, whereas additional performance feedback did not influence final test RT ($F(2, 52) = 29.32, p < 0.001, \eta_p^2 = 0.53$), study vs. test: $p < 0.001$; study vs. testFB: $p < 0.001$; test vs. testFB: $p = 0.519$).

Analysis II. To investigate if retrieval performance in Day 1 influenced final test results, we conditionalized final test results of items tested either with or without additional performance feedback during the repetition phase depending on retrieval performance on repetition cycle 1. The rm-ANOVA with the Factors *Condition* (test vs. testFB) and *Performance in repetition cycle 1* (correct vs. incorrect) revealed increased recall success if additional performance feedback was provided during the repetition phase ($F(1, 26) = 4.49, p < .05, \eta_p^2 = .15$). Furthermore, increased recall success was detected for word pairs correctly recalled in repetition cycle 1 ($F(1, 26) = 5.57, p < .05, \eta_p^2 = .18$), whereas the interaction ($F = 1.10, p = 0.304, \eta_p^2 = 0.041$) did not reach statistical significance. The same analyses was performed for RT, but no differences were detected in this analysis (*Condition* ($F = 2.71, p = 0.112, \eta_p^2 = 0.094$); *Performance in repetition cycle 1* ($F = 2.32, p = 0.139, \eta_p^2 = 0.082$); *Condition X Performance in repetition cycle 1* ($F = 2.12, p = 0.157, \eta_p^2 = 0.075$)).

fMRI results

Whole-brain contrasts

Analysis I. First, we sought to identify brain regions relevant for memory retrieval by contrasting All > Null or fixation baseline. Consistent with previous evidence, the results obtained in the current study revealed activation mostly of the left, but also on the right hemisphere (see Figure 3). Next, fMRI data were examined based on the repetition practice

on Day 1 in order to dissociate brain regions specifically activated for items previously either studied or tested or tested with additional performance feedback. All analyses were restricted to items correctly retrieved in the final test. A one-way ANOVA investigating the activation pattern for the three repetition practice conditions revealed no significant differences in neural activation due to testing with or without feedback and therefore, the results of the whole-brain analysis were examined together for test and testFB conditions. Brain areas more activated for studied compared to both tested items included left-lateralized regions, including *pars triangularis*, *pars orbitalis*, and middle frontal gyrus in the prefrontal cortex (PFC), precentral gyrus, inferior temporal gyrus, superior parietal cortex and inferior and middle occipital gyrus extending to fusiform gyrus. In contrast, correct retrieval of items previously tested either with or without performance feedback resulted in increased activation in bilateral temporoparietal regions, including superior temporal gyrus (STG), rolandic operculum, precentral gyrus and postcentral gyrus as well as left supramarginal gyrus (SMG) and right temporal pole (Figure 3, Appendix 2).

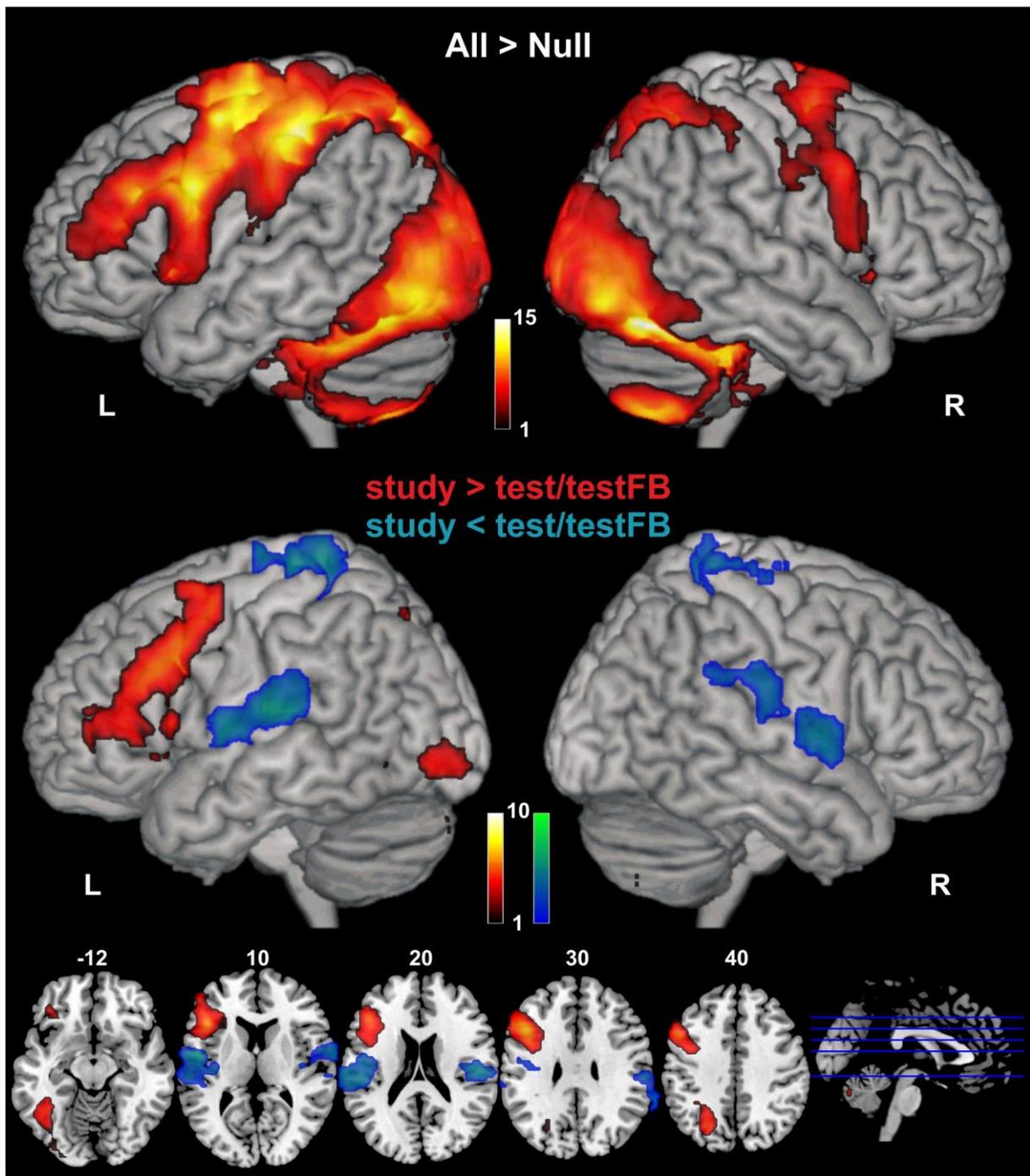


Figure 13. Brain rendering and axial slice selection showing activations for the All > Null whole-brain contrast (left) and for the study > test/testFB contrast (right) across all subjects using a voxel-level significance threshold set at $p < 0.001$, and an FWE-corrected cluster level significance threshold set at $p < 0.05$.

ROI analyses

Analysis 1. To evaluate regional activation patterns based on learning style (i.e., study, test, testFB), we conducted ROI analyses on regions reported in previous studies and meta-

analyses. These regions included left hemisphere MFG, Broca, SPL, SMG, MTG and ventral striatum (Figure 4 A). The fMRI parameter estimates were extracted for each repetition practice condition of interest against baseline for each ROIs and percent signal change values for each ROI were submitted separately to one-way ANOVAs with *Condition* (study vs. test vs. testFB) as the within-subject factor. Results revealed increased activation for previously studied compared to tested word pairs in left MFG, left Broca and left SPL (Table 1, Figure 4 A). In contrast, left SMG was more strongly engaged for previously tested items relative to studied items. No differences were found due to additional performance feedback. Results of the ventral striatum ($F < 1$, $q = 0.519$, $\eta_p^2 = 0.025$) and MTG ($F = 1.675$, $q = 0.225$, $\eta_p^2 = 0.061$) did not revealed statistically significant differences.

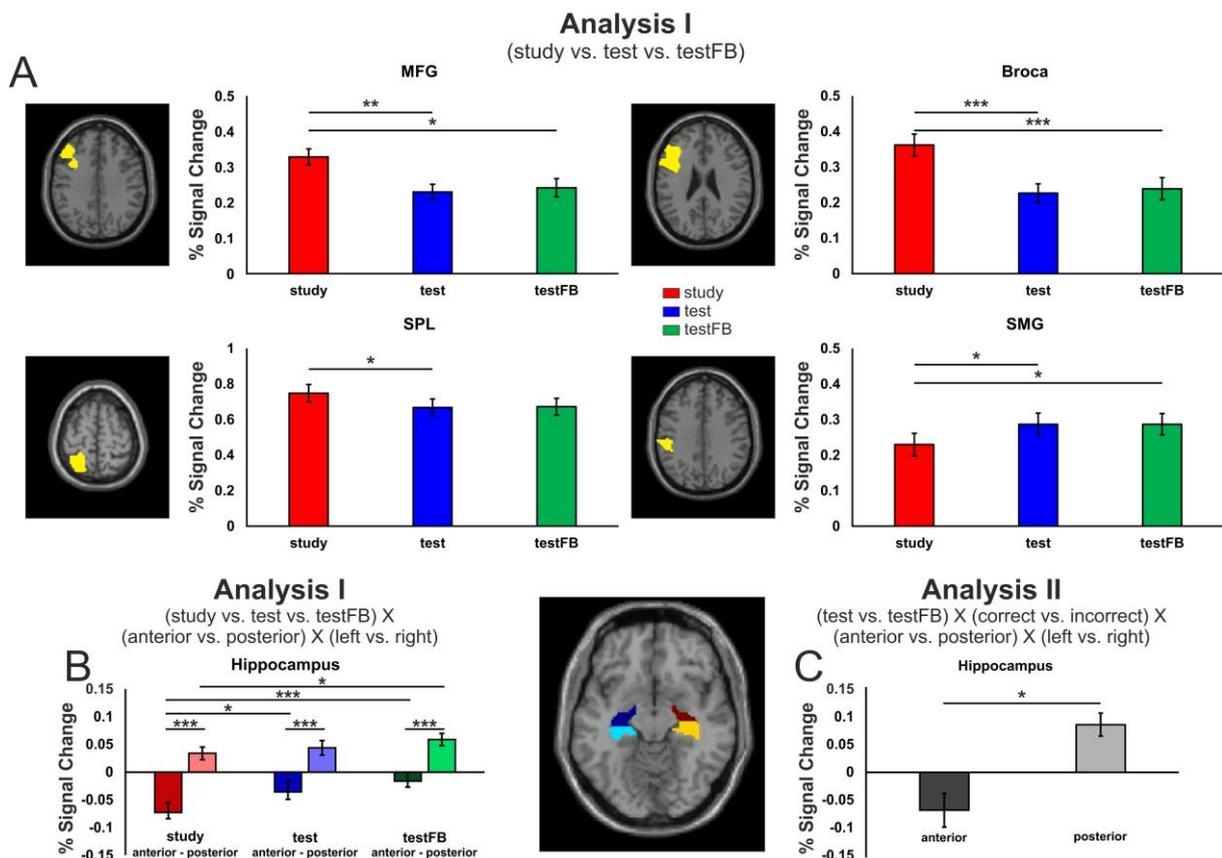


Figure 14. ROI analyses for Analysis I (A & B) and II (C). (A) Regions in analysis I revealing increased % signal change for study > test/testFB: MFG, Broca & SPL and for test/testFB > study: SMG (B) Hippocampal activation pattern in analysis I revealing increased % signal change for testFB > study. (C) Hippocampal activation pattern in analysis II showing increased % signal change in posterior (left: light blue; right: yellow)

vs. anterior hippocampus (left: dark blue; right: red) based on retrieval performance in repetition cycle 1. Asterisks indicate statistically significant effects: * $q < 0.05$, ** $q < 0.01$ *** $q < 0.001$. MFG = middle frontal gyrus; SPL = superior parietal lobe; SMG = supramarginal gyrus.

Table 3: Summary of ROI Results for Analysis I (study vs. test vs. testFB)

ROI	F-values	η_p^2	FDR q-values	Post-hoc tests (FDR q-values)		
				study vs. test	study vs. testFB	test vs. testFB
L. MFG	$F(2, 50) = 9.25$	0.27	0.003	0.003	0.010	0.501
L. Broca	$F(2, 52) = 21.76$	0.46	0.000	0.000	0.000	0.555
L. SPL	$F(2, 52) = 4.88$	0.16	0.031	0.036	0.056	0.835
L. SMG	$F(2, 52) = 4.73$	0.15	0.024	0.028	0.027	0.995

In line with recent literature (Wiklund-Hörnqvist et al. 2021) providing evidence that the testing effect benefits memory by differently activating the left and right anterior and posterior hippocampus, we investigated the left and right anterior and posterior hippocampus as well in the present study. A 3 x 2 x 2 repeated measures ANOVA with the factors *Condition* (study vs. test vs. testFB), *Laterality* (left vs. right) and *Y-axis* (anterior vs. posterior) revealed the main effects of *Condition* ($F(2, 48) = 5.64$, $q = 0.011$, $\eta_p^2 = 0.19$), *Laterality* ($F(1, 24) = 16.49$, $q = 0.001$, $\eta_p^2 = 0.41$), and *Y-axis* ($F(1, 24) = 35.95$, $q = 0.000$, $\eta_p^2 = 0.60$). Also the interaction of *Condition* X *Y-axis* resulted statistically significant ($F(2, 48) = 9.28$, $q = 0.000$, $\eta_p^2 = 0.28$) (see Figure 4 B) (study anterior vs. study posterior: $q = 0.000$; test anterior vs. test posterior: $q = 0.000$; testFB anterior vs. testFB posterior: $q = 0.000$; study anterior vs. test anterior: $q = 0.048$; study anterior vs. testFB anterior: $q = 0.000$; test anterior vs. testFB anterior: $q = 0.178$; study posterior vs. test posterior: $q = 0.452$; study posterior vs. testFB posterior: $q = 0.043$; test posterior vs. testFB posterior: $q = 0.178$). The remaining interactive effects did not reach statistical significance ($F_s \leq 1.502$, $q_s \geq 0.306$, $\eta_{ps}^2 \leq 0.059$).

Analysis II. Next, ROI results for word pairs either tested with or without additional performance feedback were conditionalized to the participant's performance during the first repetition cycle on Day 1. Similar to the behavioral final test results using a 2 x 2 repeated measures ANOVA with the factors *Condition* (test vs. testFB) and *Performance in repetition*

cycle 1 (correct vs. incorrect), here all previously described ROIs were analyzed, but none revealed significant differences ($F_s \leq 2.53$, $q_s \geq 0.511$, $\eta_{ps}^2 \leq 0.089$). For the hippocampus, a $2 \times 2 \times 2 \times 2$ repeated measures ANOVA with the factors *Condition* (test vs. testFB), *Performance in repetition cycle 1* (correct vs. incorrect), *Laterality* (left vs. right) and *Y-axis* (anterior vs. posterior) was conducted revealing only a significant main effect for *Y-axis* ($F(1, 23) = 29.61$, $q = 0.000$, $\eta_p^2 = 0.56$) (see Figure 4 C).

Functional connectivity analyses

Analysis I. Prior fMRI evidence examining the testing effect suggested that testing influences functional coupling between the hippocampus and other brain regions (Wing et al., 2013). Here we investigated functional connectivity patterns of the left hippocampus with the remaining regions of interest previously examined (i.e., MFG, Broca, MTG, SPL, SMG, ventral striatum). In addition, due to the fact that the SMG was the only region showing increased engagement for testing compared to studying in analysis I, functional connectivity between left SMG with the remaining regions of interest was analyzed as well. Finally, due to the central role of the ventral striatum for feedback processing and its well-known anatomical connections with the hippocampus (see Kahn & Shohamy, 2013; Lisman & Grace, 2005), functional coupling of the ventral striatum with the remaining regions of interests was examined too. Hence, for Analysis I, a one-way ANOVA with the factor *Condition* (study vs. test vs. testFB) was conducted using condition-specific Fisher's Z transformed beta-series values (see Figure 5A). Results revealed a main effect of *Condition* for the left hippocampal-ventral striatum functional coupling ($F(2, 52) = 4.36$, $q = 0.018$, $\eta_p^2 = 0.14$) due to a stronger coupling among these regions for testFB relative to the other conditions (study vs. test: $q = 0.790$.; study vs. testFB: $q = 0.047$; test vs. testFB: $q = 0.047$).

Analysis II. To investigate whether prior retrieval performance modulates functional coupling of the left hippocampus, left SMG and left ventral striatum with the remaining regions of interests, word pairs tested with or without additional performance feedback were conditionalized based on participant's retrieval performance in repetition cycle 1. A series of 2×2 repeated measures ANOVAs with the factors *Condition* (test vs. testFB) and *Performance*

in repetition cycle 1 (correct vs. incorrect) were conducted. Only main effects of either Condition or Performance in repetition cycle 1 resulted statistically significant, with no interactions reaching statistical significance. Results are reported in Table 2 (see Figure 5 B & C).

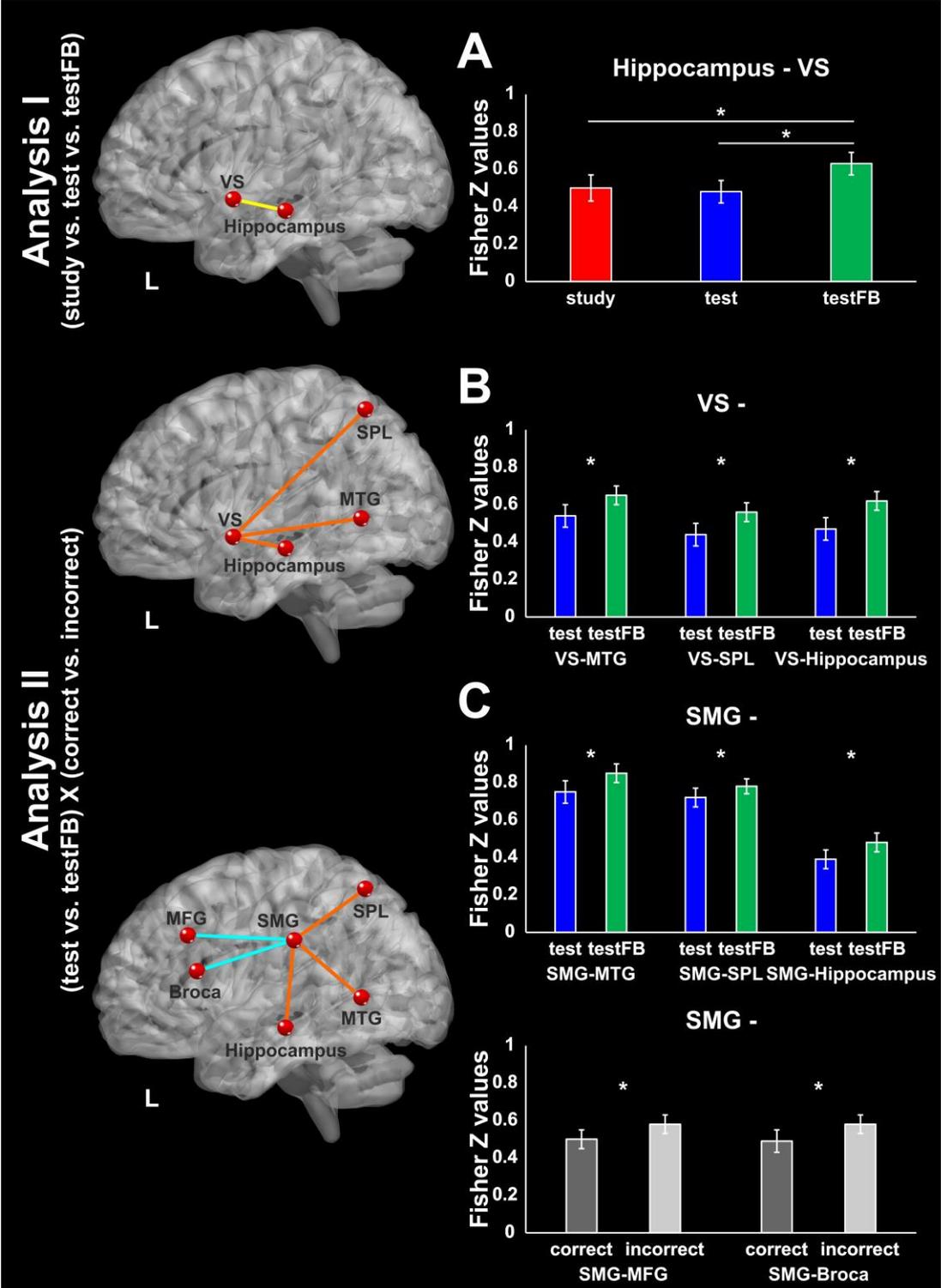


Figure 15. Pairwise functional connectivity analyses for analysis I (A) and analysis II (B & C). (A) Functional coupling of left hippocampus-ventral striatum (VS) was increased following items tested with additional performance feedback. (B) Increased VS-MTG, VS-SPL and VS-Hippocampus functional coupling for additional test performance feedback versus test. (C) Increased SMG-MTG, SMG-SPL, SMG-Hippocampus functional coupling for additional test performance feedback versus test; also, increased SMG-MFG and SMG-Broca functional coupling for incorrectly versus correctly retrieved items in repetition cycle 1. Asterisks indicate statistically significant effects: $*q < 0.05$. VS = ventral striatum; MFG = middle frontal gyrus; MTG = middle temporal gyrus; SPL = superior parietal lobe; SMG = supramarginal gyrus. Yellow edges in renderings = analysis I, *Condition* main effect; orange edges in renderings = analysis II, *Condition* main effect; cyan edges in renderings = analysis II, *Performance in repetition cycle 1* main effect.

Table 4: Summary of functional connectivity results for analysis II

<i>Condition</i> main effect	
	SMG
MTG	$F(1, 26) = 5.14, q = 0.046, \eta_p^2 = 0.17$
SPL	$F(1, 26) = 4.39, q = 0.046, \eta_p^2 = 0.14$
Left Hippocampus	$F(1, 26) = 5.84, q = 0.046, \eta_p^2 = 0.18$
Ventral Striatum	
MTG	$F(1, 26) = 4.59, q = 0.024, \eta_p^2 = 0.15$
SPL	$F(1, 26) = 6.60, q = 0.024, \eta_p^2 = 0.20$
Left Hippocampus	$F(1, 26) = 8.17, q = 0.042, \eta_p^2 = 0.24$
<i>Performance in repetition cycle 1</i> main effect	
	SMG
MFG	$F(1, 26) = 5.14, q = 0.046, \eta_p^2 = 0.17$
Broca	$F(1, 26) = 4.63, q = 0.046, \eta_p^2 = 0.15$

Discussion

The present study examined the neural correlates of performance feedback enhanced learning from TPE on successful memory retrieval. Specifically, we investigated whether explicit performance feedback after a retrieval attempt (1) further boosts performance relative

to the classic TPE effect and (2) modulates the neural network associated with successful memory retrieval. After an initial presentation of the material, participants repeated the word pairs three times by either testing or studying and correct answer feedback was presented afterwards. For half of the tested material, we additionally provided explicit feedback regarding the accuracy of the previous retrieval attempt immediately before (indirect) correct answer feedback was presented. On the following day, participants performed a final retrieval test while fMRI data were collected.

Behaviorally, retrieval performance increased whereas retrieval speed decreased for tested compared to studied items and additional performance feedback enhanced retrieval performance even more. At the neural level, while frontoparietal brain regions, including MFG, Broca and SPL were more active for previously studied compared to tested word pairs, increased activation in temporoparietal regions, in particular SMG and hippocampus, was observed for the reversed contrast. Moreover, additional performance feedback did not affect regional activation, but instead significantly enhanced functional coupling of ventral striatum as well as SMG with hippocampus and temporoparietal areas, including MTG and SPL.

Consistent with previous findings on the testing effect, our results confirmed an increase in retrieval performance and decreased RT for tested compared to studied items in the final memory test (see Roediger & Butler, 2011; Rowland, 2014; van den Broek, et al., 2014). Furthermore, the beneficial effect of learning with additional performance feedback on later retrieval performance reported in the Study 1 (Chapter 2) was replicated as well regardless of small changes in the experimental design to adapt the study to the specificities of fMRI. In contrast to Study 1 (Chapter 2), the increase in successful retrieval as a result of additional performance feedback was already measured during the repetition phases, potentially because an additional repetition cycle was provided to increase final test performance.

At present, only few neuroimaging studies have investigated the testing effect specifically in the final retrieval test instead in the repetition phase (Eriksson et al., 2011; Keresztes et al., 2014; Rosner, et al., 2014; Wirebring et al., 2015; Wiklund-Hörnqvist et al.,

2017, 2021). These studies have reported mixed findings, probably due to differences in the paradigms used to investigate the testing effect. As Wirebring et al. (2015) and Rosner et al. (2014) did not detect differences in neural activation contrasting previously tested / generated and studied items, Wirebring et al. (2015) investigated the neural patterns for remembered versus forgotten items and Rosner et al. (2014) compared successful recognition versus correct rejection reporting that successful retrieval was associated with increased activation in left frontoparietal regions and ACC. In contrast, Keresztes et al. (2014) and Wiklund-Hörnqvist et al. (2017) reported a decrease of neural activation in frontoparietal networks following testing compared to studying. The results of the present study are consistent with the later findings as we detected increased activation in frontoparietal brain regions, especially in MFG, Broca and SPL for items previously studied relative to previously tested ones. Hence, the present results support the notion, that the two learning strategies (i.e., studying vs. testing) differently affect successful memory retrieval at later stages, presumably by modulating cognitive control processes in frontal cortices and facilitating semantic memory related processes in the parietal cortex (Keresztes et al., 2014; van den Broek et al., 2016; Wiklund-Hörnqvist et al., 2017).

Furthermore, our results revealed an increase in neural activation in the final test in temporoparietal regions, including SMG, for correctly retrieved items previously practiced by testing compared to studying. As mentioned previously, Rosner et al. (2014) reported increased activation in several brain regions including SMG for successfully retrieved memories. Moreover, SMG has been implied in modulating the testing effect in prior literature, but so far only in studies investigating the repetition phase instead of the final test. While increased activation in SMG for subsequently retrieved compared to forgotten items was reported during practicing by testing (van den Broek et al., 2013; Vannest et al., 2012), other research reported enhanced activation in SMG while studying (Vestergren & Nyberg, 2014; Wing et al., 2013). Both SMG and angular gyrus were previously proposed as relevant brain areas for the storage of semantic information (see Binder, Desai, Graves, & Conant, 2009; van den Broek et al., 2016), but SMG has been also associated with the processing of phonological information (e.g., Braun, Hutzler, Ziegler, Dambacher, & Jacobs, 2009; Oberhuber et al.,

2016). The processing of phonological details specific to the material is sometimes required in order to solve the tasks used to investigate the testing effect. For example, in the studies by Rosner et al. (2014) or Vannest et al. (2012), participants had to generate associated words to target words with only several letters as cues (GARBAGE-W_ST_), which requires the processing of phonological information to guess the associated word. In the present study, participants were asked to indicate the third letter of the target word during the repetition phase, which promotes complex phonological processing of the target word as well. Thus, testing could possibly promote later retrieval performance by systematically strengthening only relevant connections between certain details of the representations that were specifically needed during practicing by testing such as the phonological representations. Alternatively, the parietal cortex was suggested to play a critical role in episodic memory processing with the ventral portion, including SMG, as modulator of bottom-up attention [see attention to memory (AtoM) model by Cabeza, Ciaramelli, Olson, & Moscovitch, 2008]. Further research is needed to properly disentangle these two possibilities with respect to the role of the SMG.

The goal of the present study was to investigate the neural processes underlying the beneficial effect of performance feedback in addition to correct answer feedback. Examining BOLD responses during the final retrieval test, we only detected changes in neural activation specific to items previously tested with or without feedback at hippocampal level. Prior neuroscientific studies on episodic memory highlighted the importance of hippocampal regions for retrieval enhanced learning, suggesting a complementary involvement with the default mode network (see Jonker, Dimsdale-Zucker, Clarke, & Ranganath 2018). In line with this research, Wiklund-Hörnqvist et al. (2021) performed a testing effect study collecting fMRI data in the final retrieval test to examine the role of hippocampal subregions, reporting increased activation for tested compared to studied items, especially in posterior hippocampus. In contrast to their findings, such activation patterns in posterior hippocampus were only observed in the present study when comparing items previously tested with additional performance feedback versus studied ones, but not for items tested without performance feedback. Results might differ due to the experimental designs, for example the retention

interval between the repetition phase and the final test was one day in the present study compared to 7 days in the study by Wiklund-Hörnqvist et al. (2021), and materials were practiced three times in the present study compared to six times in the study by Wiklund-Hörnqvist et al. (2021) due to the difficulty of learning the material. Nevertheless, additional performance feedback seems to boost posterior hippocampal activation, which led to a similar activation pattern as reported in the study by Wiklund-Hörnqvist et al. (2021). Future studies might dissociate further factors influencing learning that result in increased posterior hippocampal activation as it could relate to consolidation.

Besides examining activation levels to explore fMRI data, recent research examined functional connectivity to investigate whether specific behaviors are associated with the functional coupling between brain regions (see Noble, Scheinost, & Constable, 2019). To the best of our knowledge, only one study investigated the testing effect using functional connectivity analyses so far focusing on testing effect related coupling between hippocampus and other brain regions (Wing et al. 2013). Their results revealed enhanced hippocampal coupling with posterior cingulate cortex, medial PFC and left ventrolateral PFC for tested compared to studied items (Wing et al. 2013). Accordingly, the present study investigated functional coupling between hippocampus, ventral striatum and SMG with the remaining ROIs (MFG, Broca, MTG, SPL) revealing increased coupling between the hippocampus and ventral striatum for tested compared to studied items. Furthermore, increased functional connectivity was also observed for ventral striatum as well as SMG with MTG and SPL for items previously tested with additional performance feedback. To conclude, the present results suggest that additional performance feedback during the repetition phase enhances later retrieval performance by strengthening the connectivity between temporoparietal regions possibly using task-relevant regions, such as ventral striatum and SMG in the present study, as hubs facilitating semantic and phonological memory access.

Recent literature which aimed at explaining the mechanisms underlying the testing effect have emphasized two theoretical accounts. On the one hand, the elaboration account (e.g., Carpenter & DeLosh, 2006; Carpenter & Olson, 2011; Pyc & Rawson, 2009) suggested

that each testing situation helps to elaborate the semantic representation and hence increases the amount of cues which support later successful retrieval. On the other hand, the search-set restriction account (e.g., Karpicke & Smith, 2012; Thomas & McDaniel, 2013) proposed that retrieval attempts strengthen correct associations by reducing the connections to competing information and hence refining the cue-related search-set. In line with these theories, we hypothesized that testing with additional performance feedback would cause increased activation in areas related to feedback processing such as the ventral striatum and areas related to semantic memory such as the lateral temporal and parietal cortices. As the present results revealed increased activation in frontal-parietal brain regions for previously studied compared to tested items, and only in temporal regions for tested compared to studied material, the present study support the idea that testing reduces the neural activation (van den Broek et al., 2016) in anterior regions and instead helps to refine the search-set. Similarly, the results that additional performance feedback led to increased posterior hippocampal activation in the final test, but did not affect activation in ventral striatum might hint as well towards increased search-set restrictions instead of elaboration of the semantic representations. However, functional connectivity results revealing increased coupling between ventral striatum as well as SMG with temporal and parietal regions could support both theories as an increase in coupling could benefit semantic elaboration as well as restricted memory retrieval. Hence, future research might investigate the relation between functional coupling and specific mechanisms.

Conclusion

In conclusion, the present study confirmed previous results by demonstrating that practicing by testing compared to studying enhanced later memory retrieval and decreased retrieval pace. More interestingly, explicit performance feedback indicating prior retrieval performance provided just before the correct answer feedback enhanced retrieval performance even more and hence, might offer a useful tool to improve teaching outcome. Recording neural activation during the successful retrieval in the final test allowed us to dissociate the influence

of the different repetition conditions on the final memory representation. The fMRI results revealed that testing decreased frontoparietal activation in the final test. Furthermore, additional performance feedback on Day 1 led to increased posterior hippocampal activation on Day 2. With respect to the mechanisms underlying the testing effect, these results provide stronger support for the search-set restriction account compared to the elaboration account. In contrast, functional coupling of temporoparietal areas with hippocampus, ventral striatum and SMG following testing with additional performance feedback could support the elaboration account as well.

Chapter 5

General Discussion

The present research investigated the effect of performance-contingent feedback on the testing effect combined with correct answer feedback. Generally, the experimental structure applied to research the testing effect in combination with TPE consists of an initial study phase, a repetition phase and a final test. To answer the research question, performance feedback contingent on prior retrieval performance, was provided in the repetition phase following half of the retrieval tests. Next, correct answer feedback (TPE) was presented for all trials independent of previous studying, testing or testing with additional performance feedback. Three experimental studies were conducted to investigate the behavioral (Chapter 2) and neural correlates (Chapter 3 – EEG & 4 – fMRI) of the outlined experimental design. In this chapter, the findings of the three experimental studies are summarized and discussed with regard to the theoretical background and future research suggestions are provided.

Comparison of behavioral results of the three studies

Prior research examining the testing effect has predominantly investigated retrieval success in the final test, revealing an increase for previously tested compared to studied material (e.g., Butler & Roediger, 2007; Roediger & Karpicke, 2006a, b; for review see Roediger & Butler, 2011), particularly if the final test was delayed by about one day or more. In addition to this effect, providing correct answer feedback following the retrieval test further enhanced retrieval success at a later point in time (Arnold & McDermott, 2013; van den Broek et al., 2016). In line with these findings, all three studies revealed the same pattern of results on the final test (Day 2), which supports our expectation that the present experimental design was successful in capturing the testing effect. In addition to this effect, providing additional performance feedback in the repetition phase (Day 1) increased retrieval success on the final test (Day 2) by a moderate but consistent increase in retrieval success of 2 - 4 %. Hence, to the best of our knowledge, these studies (Chapter 2 - 4) are the first to discover and replicate

the benefit of feedback assisted TPE (i.e., testing followed by performance feedback and subsequent correct answer feedback) compared to only TPE (i.e., testing followed by correct answer feedback). In line with previous observations that reward learning and predictive coding affect LTM formation (see Sinclair & Barense, 2019, Miendlarzewska et al., 2016, Ergo et al., 2019), the present studies revealed that performance feedback learning (traditionally a part of the implicit memory system) affects LTM formation of explicit memories.

In addition to retrieval success, the testing effect has been suggested to affect RT as well, possibly reflecting retrieval effort (van den Broek et al., 2014; Racsomány et al., 2018; Wixted & Rohrer, 1993). Specifically, decreased memory strength has been proposed to result in more effortful retrieval attempts that cause RTs to increase and vice versa (see van den Broek et al., 2014). In line with prior literature, the final test results of all three studies support this claim by revealing reduced RTs for previously tested compared to studied material, but no additional benefit was detected for testing with compared to without additional performance feedback. The latter observation could possibly emerge from the precise details of the experimental design. To assess RTs, participants need to provide a response via button press and RTs can be measured from the onset of the response options until the response is provided by pressing the button. In the present studies, a retrieval test was performed by asking the participants to choose a letter specific to the target word from four options. These response options were presented with a delay to the cue word, which might have potentially masked modulations in RT as a result of prior performance feedback since participants probably retrieved the correct response before the response options were displayed. To explore this idea, future research could present the retrieval cue and the response options at the same time.

Along with analyzing final test results (Day 2) based on prior practice conditions on Day 1, results for material previously tested with or without additional performance feedback were examined depending on whether they were correctly or incorrectly retrieved in the repetition phase (similar to van den Broek et al., 2014). This comparison could only be investigated for items tested with or without additional performance feedback in the repetition phase since in

the study condition, the cue and the target word were presented at the same time and hence, retrieval success could not be assessed. In the first experiment (Chapter 2), word pairs correctly retrieved in repetition cycle 1 and 2 reflecting early retrieval success were contrasted with word pairs incorrectly retrieved in either cycles 1 or 2, reflecting late retrieval success. Results revealed increased retrieval success as well as decreased RT on the final test for material previously tested with additional performance feedback. This result did not differ due to early or late retrieval success in the repetition phase.

To further evaluate this finding of Study 1, Study 2 (Chapter 3) explored retrieval history more extensively by contrasting retrieval success on the final test (Day 2) for word pairs previously correctly retrieved twice (CC), correctly retrieved in the first but not second repetition cycle (CI), correctly retrieved in the second but not first repetition cycle (IC), or incorrectly retrieved twice (II) on Day 1. The results indicated no differences between materials tested with or without additional feedback. Furthermore, to assess if findings from Study 1 depend only on first retrieval success in the repetition phase, Study 3 (Chapter 4) investigated retrieval success and RT collected on the final test (Day 2) based on correct or incorrect retrieval in repetition cycle 1 (Day 1). In accordance with Study 1, recall success was increased if additional performance feedback was provided in the repetition phase, which could indicate that additional performance feedback mostly affects learning from correct answer feedback in the repetition cycle 1. However, one additional repetition cycle was provided in Study 3 compared to Study 1 and 2, in order to increase recall success in the final test to conduct fMRI analyses. Future studies could address these mixed findings by varying the number of repetition cycles and the material. Since it is more difficult to learn vocabulary (e.g. Jonsson, Wiklund-Hörnqvist, Stenlund, Andersson, & Nyberg, 2020; Toppino & Cohen, 2009) compared to low associated word-pairs as in the present studies, additional repetition cycles are needed to increase recall success in the final test. Furthermore, retrieval history effects have been claimed to change with test delays (e.g., Toppino & Cohen, 2009; van den Broek et al., 2014); thus, prolonged delay intervals of more than one day can be applied in future studies, as previously suggested.

The testing effect has been argued to provide a fast route to memory consolidation (Antony et al., 2017), perhaps due to effects related to reconsolidation (e.g., Finn & Roediger, 2011). Consequently, we hypothesized that if testing with additional performance feedback and correct answer feedback increased consolidation speed even more than only testing with correct answer feedback, then benefits due to additional performance feedback might possibly be detectable in the end of the repetition phase on Day 1. This result would suggest that performance feedback benefits arise from fast consolidation and/or reconsolidation. In contrast, if these benefits only emerge on the final test (Day 2), the effect might rather occur as a result of sleep consolidation processes (e.g., Petzka, Charest, Balanos, & Staresina, 2021). Results from Studies 1 and 2 (Chapters 2 and 3, respectively) indicated no difference between testing with, compared to without, additional performance feedback in the end of the repetition phase (Day 1) but only with a delay of at least one day. However, results from Study 3 (Chapter 4) already revealed performance feedback benefits in the end of the repetition phase. While two repetition cycles were provided in Studies 1 and 2, a third was added in Study 3 to enhance retrieval success on the final test inside the MR scanner. Therefore, either the mixed outcome could originate from the additional repetition cycle, or the number of repetition cycles modulates the benefits of fast consolidation and reconsolidation compared to sleep consolidation. Accordingly, this suggestion might provide a second argument to thoroughly explore the effect of additional repetition cycles.

Influences of performance feedback on theoretical accounts of the testing effect

The testing effect has been proposed to modulate the semantic network by either semantic elaboration or search-set restriction. In short, the semantic elaboration account claims that additional cues occur at each retrieval test that are integrated into the semantic representation. This elaboration of the representation facilitates next retrieval attempts by providing additional routes to access the correct memory representation (e.g., Carpenter & DeLosh, 2006; Carpenter & Olson, 2011; Pyc & Rawson, 2009). The search-set restriction account states that correct associations are strengthened, whereas the connection to

competing but incorrect associations is weakened by retrieval tests (e.g., Karpicke & Smith, 2012; Thomas & McDaniel, 2013). This restriction results in refined semantic representations, which eventually promote successful retrieval. Exploring the impact and relationship of both accounts on the testing effect is currently a main focus in related research (see Rowland, 2014; van den Broek, 2016).

Both previously outlined accounts offer explanations for the beneficial effect of additional performance feedback observed in the behavioral study (Chapter 2). On the one hand, the positive and negative feedback cues might be embedded into the semantic representation, as suggested by the semantic elaboration account. On the other hand, feedback is well known for strengthening (reinforcing) or weakening associations, especially if it does not correspond to the expectation (see: reward prediction error and reinforcement learning: e.g., Sutton and Barto, 1998). This type of learning is based on the dopaminergic system, including anterior cingulate cortex (ACC; e.g., Nieuwenhuis et al., 2004) and ventral striatum (see Daniel & Pollmann, 2014). More recently (e.g., Shohamy & Adcock, 2010), the dopaminergic system has been claimed to modulate hippocampal activity as well. Hence, performance feedback might assist with search-set restriction by modulating dopaminergic memory formation in the hippocampus. Consequently, providing additional performance feedback might affect both accounts, possibly to different degrees. In addition, performance feedback was noted to also modulate other factors that are not covered by the previous two accounts, particularly attention (see Buchin & Mulligan, 2017). In addition, early memory models (e.g., the multi-store model by Atkinson and Shiffrin, 1968) already have identified the importance of attentional processes in memory formation.

Neural correlates of the testing effect and feedback assisted TPE

The behavioral results (Chapter 2) presented thus far only provide limited insights into the previously suggested cognitive processes since retrieval success and RT only capture memory retrieval. In contrast, neuroscientific measures such as EEG and fMRI allow the possibility to observe temporal as well as structural aspects of the neural processes underlying

memory formation and retrieval. For both neuroscientific methods, previous literature has proposed distinct neural correlates to reflect processes specific to semantic elaboration, search-set restriction, potentially modulated by reinforcement learning, or attention modulation (e.g., Ernst & Steinhauser, 2014; Ludowicy et al., 2019; van den Broek et al., 2016). While EEG obtains insights into the temporal dynamics of these processes, fMRI offers detailed structural information. Hence, investigating the same paradigm with both neuroimaging techniques enables the examination of distinct features of the neural processes underlying learning from feedback assisted TPE. In the following sections, results of ERPs and oscillatory activity are discussed first, followed by a summary of fMRI results.

EEG correlates collected in the repetition phase and the final test

Principally, two research questions were investigated with EEG in Study 2 (Chapter 3). To begin, direct performance feedback effects on subsequent correct answer feedback processing in the repetition phase (Day 1) were investigated using ERPs (Chapter 3a). Then, oscillatory activity was applied to detect delayed performance feedback effects during successful memory retrieval in the repetition phase (Day 1) and the final test (Day 2; Chapter 3b). The findings of each study are discussed with respect to the previously outlined accounts and hypotheses.

In the repetition phase, ERPs in a time interval from 100 to 1000 ms following correct answer feedback onset were analyzed revealing components that corresponded with the time intervals and locations of N1, FRN, P3 and LPN (e.g., Ernst & Steinhauser, 2014; Ludowicy et al., 2019; Rosburg et al., 2015). The first negative deflection, N1, reflects attentional modulations due to prior performance feedback (Ludowicy et al., 2019) and results revealed increased N1 for incorrectly retrieved items previously tested with, compared to without, additional performance feedback. Thus, performance feedback modulates attention during correct answer feedback. Furthermore, even though we did not present positive and negative performance feedback while correct answer feedback was presented, a negative deflection was observed matching the properties of the FRN (e.g., Ernst & Steinhauser, 2014; Holroyd &

Coles, 2002; Pfabigan et al., 2011). These “FRN” results collected in Study 2 showed the typical contrast with more negative amplitudes following negative feedback compared to positive feedback but only if performance feedback was presented before correct answer feedback. The FRN was proposed to reflect selective strengthening or weakening of specific behavior (see reinforcement learning: e.g., Sutton and Barto, 1998), possibly modulating hippocampal activity assisting with search-set restriction. Furthermore, the P3 reflects modulations by feedback valence (e.g., Ludowicy et al., 2019) as well as explicit memory processing (Ernst & Steinhauser, 2012). Since the present results revealed increased amplitudes following testing without additional performance feedback, explicit memory processing seems to modulate ERPs from this time interval onward. In line with this observation, the LPN was assumed to reflect prolonged memory processing as well, revealing extended memory processing if performance feedback had been previously provided (see Rosburg et al., 2015). Nonetheless, later ERPs could reflect both, semantic elaboration and search-set restriction. Hence, we concluded that performance feedback modulates subsequent learning from correct answer feedback, first by modulating early attention orientation; second, by assisting with search-set restriction; and third, with adjusting the duration of memory processing, potentially facilitating semantic elaboration as well as search-set restriction.

In a next step, the effects of testing with additional performance feedback on delayed memory retrieval in the final test were examined by considering neural frequency alterations. Accordingly, final test data were analyzed depending on prior practice procedures in the repetition phase: studying, testing or testing with additional performance feedback. An increase in oscillatory activity from 2 to 4 Hz as well as 11 to 17 Hz has been detected for material previously tested either with or without performance feedback compared to studied word pairs. While the 2 to 4 Hz frequency interval was previously suggested to reflect recollection processes and conscious awareness (e.g., Gruber et al., 2008), the 11 to 17 Hz interval indicates semantic LTM processing and deep encoding (e.g., Klimesch, 1999; Pastötter & Bäuml, 2016, Scholz et al., 2017). Furthermore, increased activity was discovered for material

previously tested with performance feedback compared to studied items in the 5 to 7 Hz interval associated with interference processing and cognitive control (e.g., Hanslmayr et al., 2010) as well as in the 8 to 10 Hz interval, possibly reflecting either semantic elaboration or attentional processes suppressing task-irrelevant memories (e.g., Beardsley et al., 2017; Klimesch, 2012). Lastly, a decrease of 15 to 24 Hz frequencies for material tested previously with additional performance feedback compared to merely testing presumably reflects changing task demands (Guran et al., 2019), or activity changes in striatal areas (Foerde & Shohamy, 2011; Weismüller et al., 2019). It should be noted that, all results were detected on the final test whereas no differences were measurable during the repetition phase, hence indicating that performance feedback regulates consolidation processes.

Altogether, while the results covering frequencies from 11 to 17 Hz support the semantic elaboration account, oscillatory activity in the 5 to 7 Hz interval provides evidence for the search-set restriction account. Moreover, the results imply that testing with additional performance feedback potentially affects several processes underlying successful retrieval.

fMRI results collected on the final test

In addition to the examination of oscillatory activity, memory retrieval on the final test (Day 2) was investigated with fMRI with respect to prior testing with, or without, additional performance feedback or studying in the repetition phase (Chapter 4). For this purpose, we conducted whole-brain as well as ROI and functional connectivity analyses. Whole-brain and ROI analyses provide information specific to the involvement of different brain areas. Increased activation was detected in bilateral temporoparietal regions, including the superior temporal gyrus (STG), rolandic operculum, precentral gyrus and postcentral gyrus as well as in the left supramarginal gyrus (SMG) and right temporal pole for previously tested compared to studied material. Furthermore, prior testing with additional performance feedback resulted in the identical activation pattern as only testing. However, detailed ROI analyses revealed that the posterior hippocampus was more active for material previously tested with additional performance feedback compared to studied word pairs.

Along with whole-brain and ROI analyses, functional connectivity provides insights into the functional coupling of distinct brain areas regardless of their structural connection. Results showed increased coupling of the ventral striatum and SMG with the left hippocampus as well as MTG and SPL for material previously tested with, compared to testing without, performance feedback. As previously outlined, reward learning substantially depends on dopaminergic modulations in the ventral striatum (O'Doherty et al., 2004; Wiklund-Hörnqvist, et al., 2017). It is possible that these dopaminergic modulations additionally affect hippocampal activity, resulting in improved functional connectivity. Hence, these results may provide evidence for the search-set restriction account. In contrast, SMG was suggested to be involved in the storage of semantic information as well as the processing of phonological information (e.g., Oberhuber et al., 2016) and, as to modulate bottom-up attention (see attention to memory (AtoM) model by Cabeza et al., 2008). Thus, results linked to SMG could suggest that performance feedback related modulations trigger processes related to semantic elaboration or attention orientation. Overall, the obtained results indicate that performance feedback may facilitate delayed memory retrieval by adjusting search-set restriction processes as well as either semantic elaboration or memory-associated attention processes.

Benefits of multiple methods

In short, behavioral findings collected in the three studies including a total of 92 participants revealed that providing additional performance feedback following a retrieval test and before the presentation of correct answer feedback boosted retrieval success after a delay of one day. In addition to behavioral measures, neuroimaging techniques offer insights into temporal as well as structural aspects of the neural processes underlying feedback assisted TPE. As outlined in the introduction (pages 17 - 19) ERPs and oscillations capture neural activity directly by recording electrical fluctuations on the scalp whereas fMRI measures the neural activity indirectly by quantifying blood flow in the brain (see Logothetis, 2008; Luck, 2005). However, using multiple methods has several advantages: on the one hand, some research questions can only be addressed with a specific technique, and on the other hand, it

increases the validity and robustness of the findings (McCrudden, Marchand, & Schutz, 2019). Accordingly, regardless of the methodological variations, all results promoted the claim that both accounts, the semantic elaboration account and the search-set restriction account affect the testing effect.

Toward a multimodal account

In line with the suggestion that multiple mechanisms, instead of only one approach, affect the testing effect (Rosburg et al., 2015, Liu et al., 2018), results obtained in Studies 2 (EEG) and 3 (fMRI) provide support that semantic elaboration and search-set restriction jointly induce the testing effect. Moreover, additional factors including attention that are not considered in these accounts, were modulated by performance feedback and therefore, may essentially contribute to feedback assisted TPE and potentially the testing effect as well. More specifically, attention may assist with shifting the focus toward relevant memory representations (see van den Broek et al., 2016; Ernst & Steinhauser, 2012; Chapter 3a). Correspondingly, attention could potentially promote the transition between semantic elaboration and search-set restriction to improve memory retention, perhaps based on memory accessibility or other factors. In line with this idea, testing could assist with determining how accessible a memory representation is (see retrieval effort hypothesis, Karpicke & Roediger, 2007; Pyc & Rawson, 2009, or bifurcation model, Halamish & Bjork, 2011; Kornell et al., 2011) and then adjust attention orientation, which subsequently modulates if semantic elaboration or search-set restriction is applied based on the availability of external and internal memory cues. In accordance, performance feedback seems to facilitate attention orientation and assists with the modulation processes of either search-set restriction or semantic elaboration, possibly boosting search-set restriction. Thus, testing seems to regulate “semanticization”-processes (see Moscovitch et al., 2005; more details: page 6), which would also correspond to memory consolidation theories such as MTT and TTT (e.g., Moscovitch et al., 2005, Nadel & Moscovitch, 1997, Sekeres et al., 2017; Winocur et al., 2010; more details: page 6).

Future perspectives

To further evaluate the beneficial effects of performance feedback on the testing effect combined with correct answer feedback (TPE), future studies can adapt the present design by delaying the final test, modulating the availability of performance feedback or varying the stimulus interrelation (see Rowland, 2014). In detail, as the testing effect was revealed to increase due to the delay between the repetition phase and final test, future studies can collect final test results after increasing the delay interval from the current one day to several days (e.g., seven days or several month; e.g., Eriksson et al. 2011; Keresztes et al., 2014; Racsomány et al. 2018) to possibly detect an increase in the effect performance of feedback assisted TPE. Such adaptations may better reflect learning beyond the laboratory and therefore match actual learning in schools or universities for instance, since factors such as the availability of feedback and the delay between encoding and testing change in real life.

Furthermore, as the EEG study (Chapter 3) revealed that performance feedback modulated ERP results collected in the first but not the second repetition cycle, future studies can vary how often correct answer feedback is combined with performance feedback. Presenting performance feedback once may be sufficient to generate feedback assisted TPE that increases delayed retrieval success. To assess this effect in depth, it may also be beneficial to conduct additional experiments with learning material that differs in stimulus interrelation (see Rowland, 2014), which refers to the relationship between the cue and the target information. For example, the present study used low associated word pairs. In contrast, other research on the testing effect has instead been performed with vocabulary, which requires more effort to acquire (e.g., van den Broek, 2013). Thus, the benefit of performance feedback may possibly increase with the effort needed for memory formation. In addition, the more effortful it is to learn the material, the more repetitions are needed to reach high retrieval success on the final test. Therefore, such learning material may provide an opportunity to focus in depth on the different effects of positive compared to negative performance feedback.

Lastly, since behavioral measures such as retrieval success and RT only provide limited insights into the processes underlying the testing effect as well as TPE and feedback

assisted TPE, studies using neuroimaging techniques such as fMRI and EEG should be conducted to further understand the underlying neural mechanisms in detail. Due to the great spatial resolution of fMRI and the high temporal resolution of EEG, future research may combine these two methods. However, several advantages and disadvantages of simultaneous EEG-fMRI should be considered beforehand (e.g., Scrivener, 2021). Hence, whether to record EEG and fMRI separately or simultaneously may depend on the detailed experimental design. Furthermore, other neuroimaging techniques, such as diffusion tensor imaging (DTI; Douet & Chang, 2015) or transcranial direct current stimulation (tDCS; Galli, Vadillo, Sirota, Feurra, & Medvedeva, 2019) may be considered as well.

Chapter 6

Conclusion

This PhD thesis investigated the relationship between the explicit and implicit memory systems by combining characteristics from both. Accordingly, the testing effect was examined in combination with TPE (part of the explicit memory system), and additionally, performance feedback was provided before the TPE, which is typically investigated in implicit memory research. Three studies were conducted to assess the beneficial effect of performance feedback on LTM with both, behavioral and neural methods.

First, Study 1 exclusively focused on behavioral measures in the repetition phase and the final test revealing increased retrieval success on the final test for material tested in the repetition phase with additional performance feedback and correct answer feedback. Next, Study 2 explored the repetition phase and the final test with EEG. On the one hand, learning from correct answer feedback in the repetition phase was examined with ERPs, suggesting that prior performance feedback modulates attention orientation as well as early and late memory processes. On the other hand, delayed effects of performance feedback on retrieval success in the final test were investigated by analyzing oscillatory activity. Results revealed increased synchronization of the neural activation patterns if performance feedback was provided in the repetition phase, but this effect was only observed in the final test and could therefore result from increased consolidation processes.

Lastly, successful retrieval in the final test was investigated with fMRI by analyzing activation levels as well as functional connectivity (Study 3). We observed the functional coupling of the ventral striatum and SMG with the hippocampus, MTG and SPL in the final test if additional performance feedback was provided in the repetition phase. Hence, the results suggest that performance feedback modifies learning from correct answer feedback by modulating attentional and memory related processes including semantic elaboration and search-set restriction.

Overall, these three studies provide both, behavioral and neuroimaging evidence that characteristics assigned to the implicit memory system affect explicit memory formation and retrieval success. The knowledge gained experimentally may help to improve teaching, for instance in schools and universities. Teachers should consider tests not only to assess a student's level of knowledge, but also to improve later recall success. Furthermore, they may provide combinations of various types of feedback following tests. However, future research is needed to validate the present results to more natural settings.

References

- Abott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, 11(1), 159.
- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659-701.
- Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger III, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(7), 861-876.
- Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in cognitive sciences*, 21(8), 573-576.
- Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 940.
- Asok, A., Leroy, F., Rayman, J. B., & Kandel, E. R. (2019). Molecular mechanisms of the memory trace. *Trends in neurosciences*, 42(1), 14-22.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, 105(1), 158.
- Bae, G. Y., & Luck, S. J. (2019). Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *NeuroImage*, 184, 242-255.
- Bai, C. H., Bridger, E. K., Zimmer, H. D., & Mecklinger, A. (2015). The beneficial effect of testing: an event-related potential study. *Frontiers in behavioral neuroscience*, 9, 248.
- Bartnik, E. A., Blinowska, K. J., & Durka, P. J. (1992). Single evoked potential reconstruction by means of wavelet transform. *Biological Cybernetics*, 67(2), 175-181.
- Beardsley, M., Hernández-Leo, D., & Ramirez-Melendez, R. (2018). Seeking reproducibility: Assessing a multimodal study of the testing effect. *Journal of Computer Assisted Learning*, 34(4), 378-386.
- Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, 27(7), 1823-1835.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188.

- Berger, H. (1930). Electroencephalogram of humans. *Journal fur Psychologie und Neurologie*, 40, 160-179.
- Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron*, 86(3), 646-664.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex*, 19(12), 2767-2796.
- Bishara, A. J., & Jacoby, L. L. (2008). Aging, spaced retrieval, and inflexible memory performance. *Psychonomic Bulletin & Review*, 15(1), 52-57.
- Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*, 34(4), 537-541.
- Bramão, I., & Johansson, M. (2017). Benefits and costs of context reinstatement in episodic memory: an ERP study. *Journal of Cognitive Neuroscience*, 29(1), 52-64.
- Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., & Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human brain mapping*, 30(7), 1977-1989.
- Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002, June). Region of interest analysis using an SPM toolbox. In *8th international conference on functional mapping of the human brain* (Vol. 16, No. 2, p. 497).
- Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language*, 66(3), 407-415.
- Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), 1934.
- Buckner, R. L., Krienen, F. M., & Yeo, B. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nature neuroscience*, 16(7), 832-837.
- Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of neuroscience*, 29(24), 7869-7876.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118.
- Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4-5), 514-527.
- Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & cognition*, 36(3), 604-616.
- Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, 105(2), 290.

- Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 918.
- Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13(4), 273.
- Butler, L. T., & Berry, D. C. (2001). Implicit memory: Intention and awareness revisited. *Trends in Cognitive Sciences*, 5(5), 192-197.
- 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.
- 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.
- 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), 268-276.
- Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of experimental psychology: learning, memory, and cognition*, 38(1), 92.
- Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14(3), 474-478.
- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test?. *Psychonomic bulletin & review*, 13(5), 826-830.
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438-448.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20(6), 633-642.
- Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of cognitive neuroscience*, 23(4), 936-946.
- Cocosco, C. A., Kollokian, V., Kwan, R. K. S., Pike, G. B., & Evans, A. C. (1997). Brainweb: Online interface to a 3D MRI simulated brain database. In *NeuroImage*.
- Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 14(3), 215-235.
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human brain mapping*, 8(2-3), 109-114.

- Daniel, R., & Pollmann, S. (2010). Comparing the neural basis of monetary reward and cognitive feedback during information-integration category learning. *Journal of Neuroscience*, *30*(1), 47-55.
- Daniel, R., & Pollmann, S. (2014). A universal role of the ventral striatum in reward-based learning: evidence from human studies. *Neurobiology of learning and memory*, *114*, 90-100.
- Douet, V., & Chang, L. (2015). Fornix as an imaging marker for episodic memory deficits in healthy aging and in various neurological disorders. *Frontiers in aging neuroscience*, *6*, 343.
- Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram?. *Annu. Rev. Psychol.*, *55*, 51-86.
- 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), 374-380.
- Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109-120.
- Ergo, K., De Loof, E., Janssens, C., & Verguts, T. (2019). Oscillatory signatures of reward prediction errors in declarative learning. *Neuroimage*, *186*, 137-145.
- 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.
- Ernst, B., & Steinhauser, M. (2012). Feedback-related brain activity predicts learning from feedback in multiple-choice testing. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(2), 323-336.
- Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory*, *18*(3), 335-350.
- Ferdinand, N. K., & Czernochowski, D. (2018). Motivational influences on performance monitoring and cognitive control across the adult lifespan. *Frontiers in Psychology*, *9*, 1018.
- Ferreira, C. S., Maraver, M. J., Hanslmayr, S., & Bajo, T. (2019). Theta oscillations show impaired interference detection in older adults during selective memory retrieval. *Scientific reports*, *9*(1), 1-11.
- Ferreira, C. S., Marful, A., Staudigl, T., Bajo, T., & Hanslmayr, S. (2014). Medial prefrontal theta oscillations track the time course of interference during selective memory retrieval. *Journal of Cognitive Neuroscience*, *26*(4), 777-791.
- Finley, J. R., Benjamin, A. S., Hays, M. J., Bjork, R. A., & Kornell, N. (2011). Benefits of accumulating versus diminishing cues in recall. *Journal of Memory and Language*, *64*(4), 289-298.
- Finn, B., & Roediger III, H. L. (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later recall. *Psychological science*, *22*(6), 781-786.

- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2), 195-207.
- Fisher, R. A. (1921). 014: On the " Probable Error" of a Coefficient of Correlation Deduced from a Small Sample.
- Foerde, K., & Shohamy, D. (2011). Feedback timing modulates brain systems for learning in humans. *Journal of Neuroscience*, 31(37), 13157-13167.
- Fritz, C. O., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology*, 21(4), 499-526.
- Galli, G., Vadillo, M. A., Sirota, M., Feurra, M., & Medvedeva, A. (2019). A systematic review and meta-analysis of the effects of transcranial direct current stimulation (tDCS) on episodic memory. *Brain stimulation*, 12(2), 231-241.
- Gao, C., Rosburg, T., Hou, M., Li, B., Xiao, X., & Guo, C. (2016). The role of retrieval mode and retrieval orientation in retrieval practice: insights from comparing recognition memory testing formats and restudying. *Cognitive, Affective, & Behavioral Neuroscience*, 16(6), 977-990.
- Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *Journal of Neuroscience*, 21(23), 9430-9437.
- Gevins, A., Leong, H., Smith, M. E., Le, J., & Du, R. (1995). Mapping cognitive brain function with modern high-resolution electroencephalography. *Trends in neurosciences*, 18(10), 429-436.
- Gruber, T., Tsvivilis, D., Giabbiconi, C. M., & Müller, M. M. (2008). Induced electroencephalogram oscillations during source memory: familiarity is reflected in the gamma band, recollection in the theta band. *Journal of cognitive neuroscience*, 20(6), 1043-1053.
- Guran, C. N. A., Herweg, N. A., & Bunzeck, N. (2019). Age-related decreases in the retrieval practice effect directly relate to changes in alpha-beta oscillations. *Journal of Neuroscience*, 39(22), 4344-4352.
- Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological psychology*, 71(2), 148-154.
- Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology*, 44(6), 905-912.
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 801.
- Hanslmayr, S., Axmacher, N., & Inman, C. S. (2019). Modulating human memory via entrainment of brain oscillations. *Trends in neurosciences*, 42(7), 485-499.

- Hanslmayr, S., Staudigl, T., Aslan, A., & Bäuml, K. H. (2010). Theta oscillations predict the detrimental effects of memory retrieval. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(3), 329-338.
- Herrmann, C. S., Rach, S., Vosskuhl, J., & Strüber, D. (2014). Time–frequency analysis of event-related potentials: a brief tutorial. *Brain topography*, *27*(4), 438-450.
- Herrmann, C. S., Strüber, D., Helfrich, R. F., & Engel, A. K. (2016). EEG oscillations: from correlation to causality. *International Journal of Psychophysiology*, *103*, 12-21.
- Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in cognitive sciences*, *24*(3), 208-227.
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *109*(4), 679.
- Hölscher, C., Anwyl, R., & Rowan, M. J. (1997). Stimulation on the positive phase of hippocampal theta rhythm induces long-term potentiation that can be depotentiated by stimulation on the negative phase in area CA1 in vivo. *Journal of Neuroscience*, *17*(16), 6470-6477.
- Hoult, D. I., & Bhakar, B. (1997). NMR signal reception: Virtual photons and coherent spontaneous emission. *Concepts in Magnetic Resonance: An Educational Journal*, *9*(5), 277-297.
- Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, *18*(3), 879-919.
- Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, *8*(2), 200-224.
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1441.
- Jia, S., Li, H., Luo, Y., Chen, A., Wang, B., & Zhou, X. (2007). Detecting perceptual conflict by the feedback-related negativity in brain potentials. *Neuroreport*, *18*(13), 1385-1388.
- 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.
- Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences*, *115*(43), 11084-11089.
- Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2020). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology*.

- Jou, J., & Foreman, J. (2007). Transfer of learning in avoiding false memory: The roles of warning, immediate feedback, and incentive. *Quarterly Journal of Experimental Psychology*, 60(6), 877-896.
- Kahn, I., & Shohamy, D. (2013). Intrinsic connectivity between the hippocampus, nucleus accumbens, and ventral tegmental area in humans. *Hippocampus*, 23(3), 187-192.
- Kandel, E. R., & Tauc, L. (1965). Heterosynaptic facilitation in neurones of the abdominal ganglion of *Aplysia depilans*. *The Journal of Physiology*, 181(1), 1.
- Kang, S. H. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*, 38(8), 1009-1017.
- Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4-5), 528-558.
- Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, 27(2), 317-326.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772-775.
- Karpicke, J. D., & Roediger III, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of experimental psychology: learning, memory, and cognition*, 33(4), 704.
- Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *science*, 319(5865), 966-968.
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, 67(1), 17-29.
- Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, 62(3), 227-239.
- Keresztes, A., Kaiser, D., Kovács, G., & Racsmany, M. (2014). Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. *Cerebral Cortex*, 24(11), 3025-3035.
- Kim, H. (2016). Default network activation during episodic and semantic memory retrieval: a selective meta-analytic comparison. *Neuropsychologia*, 80, 35-46.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3), 169-195.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, 16(12), 606-617.
- Klimesch, W., Freunberger, R., & Sauseng, P. (2010). Oscillatory mechanisms of process binding in memory. *Neuroscience & Biobehavioral Reviews*, 34(7), 1002-1014.
- Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature neuroscience*, 22(10), 1598-1610.

- Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, *17*(5), 493-501.
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85-97.
- LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, *67*(2), 259.
- Larsen, D. P. (2018, August). Planning education for long-term retention: the cognitive science and implementation of retrieval practice. In *Seminars in neurology* (Vol. 38, No. 04, pp. 449-456). Thieme Medical Publishers.
- Lee, J. L., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in cognitive sciences*, *21*(7), 531-545.
- Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*, *22*(4), 748-761.
- Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron*, *46*(5), 703-713.
- Liu, X. L., & Reder, L. M. (2016). fMRI exploration of pedagogical benefits of repeated testing: when more is not always better. *Brain and behavior*, *6*(7), e00476.
- Liu, X. L., Liang, P., Li, K., & Reder, L. M. (2014). Uncovering the neural mechanisms underlying learning from tests. *PLoS One*, *9*(3), e92025.
- Liu, X. L., Tan, D. H., & Reder, L. M. (2018). The two processes underlying the testing effect—Evidence from Event-Related Potentials (ERPs). *Neuropsychologia*, *112*, 77-85.
- Liu, Y., Rosburg, T., Gao, C., Weber, C., & Guo, C. (2017). Differentiation of subsequent memory effects between retrieval practice and elaborative study. *Biological psychology*, *127*, 134-147.
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869-878.
- Lømo, T. (1966, January). Frequency potentiation of excitatory synaptic activity in dentate area of hippocampal formation. In *Acta Physiologica Scandinavica* (p. 128). PO BOX 88, OSNEY MEAD, OXFORD OX2 0NE, OXON, ENGLAND: Blackwell Science Ltd.
- Lømo, T. (2003). The discovery of long-term potentiation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1432), 617-620.
- Luck, S. J. (2005). Ten simple rules for designing and interpreting ERP experiments. *Event-related potentials: A methods handbook*, *4*.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Ludowicy, P., Czernochowski, D., Weis, T., Haese, A., & Lachmann, T. (2019). Neural correlates of feedback processing during a sensory uncertain speech-nonspeech discrimination task. *Biological Psychology*, *144*, 103-114.

- Luque, D., Morís, J., Rushby, J. A., & Le Pelley, M. E. (2015). Goal-directed EEG activity evoked by discriminative stimuli in reinforcement learning. *Psychophysiology*, *52*(2), 238-248.
- Magosso, E., Forcelli, V., Garofalo, S., Di Pellegrino, G., & Ursino, M. (2015, August). Event-related brain potential signaling unexpected timing of feedback: A source localization analysis. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 618-621). IEEE.
- Marco-Pallares, J., Cucurell, D., Münte, T. F., Strien, N., & Rodríguez-Fornells, A. (2011). On the number of trials needed for a stable feedback-related negativity. *Psychophysiology*, *48*(6), 852-860.
- Marco-Pallarés, J., Münte, T. F., & Rodríguez-Fornells, A. (2015). The role of high-frequency oscillatory activity in reward processing and learning. *Neuroscience & Biobehavioral Reviews*, *49*, 1-7.
- Martin, L. E., & Potts, G. F. (2009). Impulsivity in decision-making: An event-related potential investigation. *Personality and Individual Differences*, *46*(3), 303-308.
- Mazaika, P. K., Hoefft, F., Glover, G. H., & Reiss, A. L. (2009). Methods and software for fMRI analysis of clinical subjects. *Neuroimage*, *47*(Suppl 1), S58.
- McConnell, M. D., & Reed Hunt, R. (2007). Can false memories be corrected by feedback in the DRM paradigm?. *Memory & Cognition*, *35*(5), 999-1006.
- McCrudden, M. T., Marchand, G., & Schutz, P. (2019). Mixed methods in educational psychology inquiry. *Contemporary Educational Psychology*, *57*, 1-8.
- McGaugh, J. L. (2000). Memory-a century of consolidation. *Science*, *287*(5451), 248-251.
- Miendlarzewska, E. A., Bavelier, D., & Schwartz, S. (2016). Influence of reward motivation on human declarative memory. *Neuroscience & Biobehavioral Reviews*, *61*, 156-176.
- Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, *9*(6), 788-798.
- Mohanty, R., Sethares, W. A., Nair, V. A., & Prabhakaran, V. (2020). Rethinking measures of functional connectivity via feature extraction. *Scientific reports*, *10*(1), 1-17.
- Morris, P. E., Fritz, C. O., Jackson, L., Nichol, E., & Roberts, E. (2005). Strategies for learning proper names: Expanding retrieval practice, meaning and imagery. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(6), 779-798.
- Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of cognitive neuroscience*, *4*(3), 257-267.
- Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., ... & 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.

- Müller, G. E. (1900). Pilzecker A ((1900)). *Experimentelle Beiträge zur Lehre vom Gedächtnis. Zeitschrift f. Psych, und Physiol. des Sinnesorgan., supplement.*
- Mulligan, N. W., Buchin, Z. L., & West, J. T. (2020). Assessing why the testing effect is moderated by experimental design. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(7), 1293.
- Nadel, L., & Hardt, O. (2011). Update on memory systems and processes. *Neuropsychopharmacology*, 36(1), 251-273.
- Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2), 217-227.
- Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews*, 36(7), 1640-1645.
- Nader, K., & Hardt, O. (2009). A single standard for memory: the case for reconsolidation. *Nature Reviews Neuroscience*, 10(3), 224-234.
- Nelson, S. M., Arnold, K. M., Gilmore, A. W., & McDermott, K. B. (2013). Neural signatures of test-potentiated learning in parietal cortex. *Journal of Neuroscience*, 33(29), 11754-11762.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 402-407.
- Nguyen, K., & McDaniel, M. A. (2015). The picture complexity effect: Another list composition paradox. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(4), 1026.
- Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, 28(4), 441-448.
- Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage*, 203, 116157.
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *science*, 304(5669), 452-454.
- Oberhuber, M., Hope, T. M. H., Seghier, M. L., Parker Jones, O., Prejawa, S., Green, D. W., & Price, C. J. (2016). Four functionally distinct regions in the left supramarginal gyrus support word processing. *Cerebral Cortex*, 26(11), 4212-4226.
- Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic resonance in medicine*, 14(1), 68-78.

- Olszowy, W., Aston, J., Rua, C., & Williams, G. B. (2019). Accurate autocorrelation modeling substantially improves fMRI reliability. *Nature communications*, *10*(1), 1-11.
- Pan, S. C., Gopal, A., & Rickard, T. C. (2016). Testing with feedback yields potent, but piecewise, learning of history and biology facts. *Journal of Educational Psychology*, *108*(4), 563.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3.
- Pastötter, B., & Bäuml, K. H. T. (2014). Distinct slow and fast cortical theta dynamics in episodic memory retrieval. *Neuroimage*, *94*, 155-161.
- Pastötter, B., & Bäuml, K. H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(3), 473-488.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K. H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 287.
- Peterburs, J., Kobza, S., & Bellebaum, C. (2016). Feedback delay gradually affects amplitude and valence specificity of the feedback-related negativity (FRN). *Psychophysiology*, *53*(2), 209-215.
- Peterburs, J., Suchan, B., & Bellebaum, C. (2013). You do the math: coding of bets and outcomes in a gambling task in the feedback-related negativity and P300 in healthy adults. *PloS one*, *8*(11), e81262.
- Peterson, D. J., & Mulligan, N. W. (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1287.
- Petzka, M., Charest, I., Balanos, G. M., & Staresina, B. P. (2021). Does sleep-dependent consolidation favour weak memories?. *Cortex*, *134*, 65-75.
- Pfabigan, D. M., Alexopoulos, J., Bauer, H., & Sailer, U. T. A. (2011). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology*, *48*(5), 656-664.
- Pfabigan, D. M., Sailer, U., & Lamm, C. (2015). Size does matter! Perceptual stimulus properties affect event-related potentials during feedback processing. *Psychophysiology*, *52*(9), 1238-1247.
- Pfabigan, D. M., Seidel, E. M., Paul, K., Grahl, A., Sailer, U., Lanzenberger, R., ... & Lamm, C. (2015). Context-sensitivity of the feedback-related negativity for zero-value feedback outcomes. *Biological Psychology*, *104*, 184-192.
- Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews*, *32*(2), 197-205.

- Poldrack, R. A., Clark, J., Paré-Blagoev, E. A., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546-550.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.
- Poppenk, J., & Moscovitch, M. (2011). A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. *Neuron*, *72*(6), 931-937.
- Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in cognitive sciences*, *17*(5), 230-240.
- Putnam, A. L., & Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect?. *Memory & Cognition*, *41*(1), 36-48.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*(6002), 335-335.
- Pyc, M. A., & Rawson, K. A. (2011). Costs and benefits of dropout schedules of test–restudy practice: Implications for student learning. *Applied Cognitive Psychology*, *25*(1), 87-95.
- 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.
- Racsomány, M., Szöllősi, Á., & Bencze, D. (2018). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 157.
- Raichle, M. E. (1994). Visualizing the mind. *Scientific American*, *270*(4), 58-64.
- Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*, *25*(3), 847-869.
- Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage*, *23*(2), 752-763.
- Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, *15*(1), 20-27.
- Roediger III, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255.
- Roediger III, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science*, *1*(3), 181-210.
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 233.
- Roper, W. J. (1977). Feedback in computer assisted instruction. *Programmed learning and educational technology*, *14*(1), 43-49.

- 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*, *15*(1), 195-210.
- Rosburg, T., Mecklinger, A., & Johansson, M. (2011). Electrophysiological correlates of retrieval orientation in reality monitoring. *NeuroImage*, *54*(4), 3076-3084.
- Rosner, Z. A., Elman, J. A., & Shimamura, A. P. (2013). The generation effect: Activating broad neural circuits during memory encoding. *cortex*, *49*(7), 1901-1909.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin*, *140*(6), 1432.
- Samar, V. J., Bopardikar, A., Rao, R., & Swartz, K. (1999). Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. *Brain and language*, *66*(1), 7-60.
- 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., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual review of psychology*, *49*(1), 289-318.
- Schenck, J. F. (1996). The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Medical physics*, *23*(6), 815-850.
- Scholz, S., Schneider, S. L., & Rose, M. (2017). Differential effects of ongoing EEG beta and theta power on memory formation. *PLoS one*, *12*(2), e0171913.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, *80*(1), 1-27.
- Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: A meta-analytic perspective. *Psychology Learning & Teaching*, *16*(2), 179-196.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11.
- Scrivener, C. L. (2021). When Is Simultaneous Recording Necessary? A Guide for Researchers Considering Combined EEG-fMRI. *Frontiers in Neuroscience*, *15*, 774.
- Sekeres, M. J., Moscovitch, M., & Winocur, G. (2017). Mechanisms of memory consolidation and transformation. In *Cognitive neuroscience of memory consolidation* (pp. 17-44). Springer, Cham.
- Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological review*, *76*(2), 179.
- Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in cognitive sciences*, *14*(10), 464-472.

- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: how incomplete reminders drive reconsolidation. *Trends in neurosciences*, 42(10), 727-739.
- Smillie, L. D., Cooper, A. J., & Pickering, A. D. (2011). Individual differences in reward–prediction–error: extraversion and feedback-related negativity. *Social Cognitive and Affective Neuroscience*, 6(5), 646-652.
- Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia*, 47(8-9), 1765-1779.
- Squire, L. R. (1986). Mechanisms of memory. *Science*, 232(4758), 1612-1619.
- Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current opinion in neurobiology*, 5(2), 169-177.
- Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor perspectives in biology*, 7(8), a021766.
- Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nature Reviews Neuroscience*, 8(11), 872-883.
- Staudigl, T., Hanslmayr, S., & Bäuml, K. H. T. (2010). Theta oscillations reflect the dynamics of interference in episodic memory retrieval. *Journal of Neuroscience*, 30(34), 11356-11362.
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: an introduction MIT Press. Cambridge, MA, 22447.
- Thomas, R. C., & McDaniel, M. A. (2013). Testing and feedback effects on front-end control over later retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 437.
- Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational Research Review*, 9, 1-15.
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental psychology*, 56(4), 252-257.
- Tulving, E. (1972). 12. Episodic and semantic memory. *Organization of memory/Eds E. Tulving, W. Donaldson, NY: Academic Press*, 381-403.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, 26(1), 1.
- Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in cognitive sciences*, 18(5), 259-267.

- van den Broek, G. S., Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, 22(7), 803-812.
- van den Broek, G. S., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage*, 78, 94-102.
- van den Broek, G. S., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the “testing effect”: A review. *Trends in Neuroscience and Education*, 5(2), 52-66.
- van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European neuropsychopharmacology*, 20(8), 519-534.
- van Eersel, G. G., Verkoeijen, P. P., Povilenaite, M., & Rikers, R. (2016). The testing effect and far transfer: The role of exposure to key information. *Frontiers in Psychology*, 7, 1977.
- Vannest, J., Eaton, K. P., Henkel, D., Siegel, M., Tsevat, R. K., Allendorfer, J. B., ... & Szaflarski, J. P. (2012). Cortical correlates of self-generation in verbal paired associate learning. *Brain research*, 1437, 104-114.
- Vestergren, P., & Nyberg, L. (2014). Testing alters brain activity during subsequent restudy: evidence for test-potentiated encoding. *Trends in Neuroscience and Education*, 3(2), 69-80.
- Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, 24(8), 1183-1195.
- Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews*, 36(8), 1870-1884.
- Wang, Y., Kuhlman, D. M., Roberts, K., Yuan, B., Zhang, Z., Zhang, W., & Simons, R. F. (2017). Social value orientation modulates the FRN and P300 in the chicken game. *Biological psychology*, 127, 89-98.
- Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in cognitive sciences*, 7(12), 553-559.
- Wartenweiler, D. (2011). Testing effect for visual-symbolic material: Enhancing the learning of Filipino children of low socio-economic status in the public school system. *International Journal of Research*, 6(1).
- Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face–name learning. *Psychonomic bulletin & review*, 18(3), 518-523.
- Weismüller, B., Kullmann, J., Hoenen, M., & Bellebaum, C. (2019). Effects of feedback delay and agency on feedback-locked beta and theta power during reinforcement learning. *Psychophysiology*, 56(10), e13428.

- Wiklund-Hörnqvist, C., Andersson, M., Jonsson, B., & Nyberg, L. (2017). Neural activations associated with feedback and retrieval success. *npj Science of Learning*, 2(1), 1-7.
- Wiklund-Hörnqvist, C., Stillesjö, S., Andersson, M., Jonsson, B., & Nyberg, L. (2021). Retrieval practice facilitates learning by strengthening processing in both the anterior and posterior hippocampus. *Brain and behavior*, 11(1), e01909.
- 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), 2360-2370.
- Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia*, 48(8), 2339-2356.
- Winocur, G., Moscovitch, M., & Sekeres, M. (2007). Memory consolidation or transformation: context manipulation and hippocampal representations of memory. *Nature neuroscience*, 10(5), 555-557.
- Wirebring, L. K., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *Journal of Neuroscience*, 35(26), 9595-9602.
- Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1024.
- Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(3), 485.
- Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: a review of the forward testing effect. *NPJ science of learning*, 3(1), 1-9.
- Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience*, 24(28), 6258-6264.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & cognition*, 38(8), 995-1008.

Appendix

Appendix 1. List of testing effect studies investigating correct answer feedback (see Rowland, 2014; PA: paired associates, SR: semantically related, O: other material, NS: nonsemantic, SW: single words, SU: semantically unrelated)

Author	Year	Behavioral measure	Material	Cue-target relation	Feedback
Bishara & Jacoby	2008	retrieval success	PA: word pairs	SR	Correct answer feedback
Brewer & Unsworth	2012	retrieval success	PA: word pairs	SR	Correct answer feedback
Butler	2010	retrieval success	prose	SR	Correct answer feedback
Carpenter & Pashler	2007	retrieval success	O: visuospatial material	NS, none	Correct answer feedback
Carpenter, Pashler, & Vul	2006	retrieval success	PA: word pairs	SR, none	Correct answer feedback
Carpenter, Pashler, Wixted, & Vul	2008	retrieval success	O, PA	SR, NS	Correct answer feedback
Carrier & Pashler	1992	retrieval success	PA: word pairs	NS	Correct answer feedback
Cull	2000	retrieval success	PA: word pairs	SU	Correct answer feedback
Finley Benjamin, Hays, Bjork, & Kornell	2011	retrieval success	PA: word pairs	NS	Correct answer feedback
Fritz Morris, Acton, Voelkel, & Etkind	2007	retrieval success	PA: word pairs	NS, none	Correct answer feedback
Jacoby, Wahlheim, & Coane	2010	retrieval success	PA: concepts	NS, same	Correct answer feedback + indication of prior test performance (correct/incorrect)
Kang	2010	retrieval success	PA	NS	Correct answer feedback
Kang, McDermott, & Roediger	2007	retrieval success	prose	same	Correct answer feedback
Karpicke & Blunt	2011	retrieval success	prose	none	Correct answer feedback
Kornell & Son	2009	retrieval success	PA: word pairs	SR	Correct answer feedback
Kornell, Bjork, & Garcia	2011	retrieval success	PA	SR	Correct answer feedback
LaPorte & Yoss	1975	retrieval success	prose	SR	Correct answer feedback
McConnell & Hunt	2007	retrieval success	SW	none	Correct answer feedback + self feedback
Morris, Fritz, Jackson, Nichol, & Roberts	2005	retrieval success	PA: names and faces	NS	Correct answer feedback
Peterson & Mulligan	2013	retrieval success	PA: word pairs	SU	Correct answer feedback
Putnam & Roediger	2013	retrieval success	PA: word pairs	SR	Correct answer feedback
Pyc & Rawson	2010	retrieval success	PA: word pairs	SU	Correct answer feedback
Pyc & Rawson	2011	retrieval success	PA: word pairs	SU	Correct answer feedback
Rohrer, Taylor, & Sholar	2010	retrieval success	O	same	Correct answer feedback
Thomas & McDaniel	2013	retrieval success	PA: word pairs	SR	Correct answer feedback
Wartenweiler	2011	retrieval success	PA: symbols and words	same	Correct answer feedback

Appendix 2. Peak activation in significant clusters for the whole brain contrast (study > test / testFB and study < test / testFB)

Cluster	Local maxima	Hem	x	y	z	Z-score	Voxels (k)
study > test / testFB							
1	Pars triangularis	L	-45	20	29	6.89	1313
	Pars orbitalis	L	-30	26	2	6.23	
	Middle frontal gyrus	L	-51	14	41	5.77	
	Precentral gyrus	L	-48	2	53	5.24	
	Middle frontal gyrus	L	-35	8	62	5.00	
	Pars orbitalis	L	-39	44	-1	3.97	
2	Inferior Temporal Gyrus	L	-42	-52	-7	5.32	280
	Inferior Occipital Gyrus	L	-36	-82	-7	4.47	
3	Cerebellum (Crus 2)	R	12	-79	-34	5.48	272
	Cerebellum (VIII)	R	30	-67	-49	4.92	
4	Middle Occipital Gyrus	L	-24	-67	38	5.26	237
	Superior Parietal Lobe	L	-27	-58	44	4.84	
study < test / testFB							
1	Superior temporal gyrus	L	-51	-7	5	6.7	661
	Superior temporal gyrus	L	-66	-19	14	5.87	
	Rolandic operculum	L	-60	-22	14	5.86	
	Rolandic operculum	L	-42	-10	23	4.83	
	Insula	L	-36	-16	8	4.72	
	Spramarginal gyrus	L	-60	-31	26	3.89	
2	Precentral gyrus	L	-27	-31	62	6.53	458
	Postcentral gyrus	L	-21	-40	65	6.00	
	Precuneus	L	-21	-49	65	4.7	
3	Rolandic operculum	R	54	-19	20	5.35	481
	Teporal Pole	R	66	-1	5	5.18	
	Superior temporal gyrus	R	54	-4	2	4.74	
	Insula	R	42	-7	-1	4.6	
4	Precuneus	R	12	-43	62	4.89	
	Postcentral	R	27	-40	71	4.77	
	Precentral	R	27	-19	68	3.86	

Acknowledgements

Danke, thank you, gracias y eskerrik asko!

Pursuing a PhD is truly a valuable and fascinating journey, and I would like to express my gratitude and appreciation to all those persons who supported me throughout my journey.

First, I thank my supervisors Prof. Dr. Daniela Czernochowski and Prof. Dr. Thomas Lachmann for giving me the opportunity to work on this exciting project and their insightful feedback which pushed me to sharpen my thinking. Your confidence and trust even beyond my research projects helped me to grow as a scientist as well as a person.

Furthermore, I would like to express my special thanks to Dr. Kepa Paz-Alonso for his invaluable guidance and advices during my research stay at the Basque Center on Cognition, Brain and Language in San Sebastian as well as throughout my entire PhD studies. Thank you for all the knowledge you shared with me, especially about the fMRI, and for your patient support during data collection and analysis. Moreover, I want to thank the LMC group and all colleagues from BCBL, especially Jaione Arnaez-Telleria and Kshipra Gurunandan for teaching me how enjoyable fMRI data collection can be.

I would like to acknowledge my colleagues from the Cognitive Neuroscience group and Cognitive and Developmental Psychology group as well as Charlotte Mock for our fruitful discussions, your encouraging words and your help in the lab.

Last, but not least, I would like to thank my family and friends, especially Maxi and Nora, for their wise counsel, their sympathetic ear and the happy distractions to rest my mind outside of my research.

Declaration

Hiermit versichere ich,

- dass ich die vorgelegte Arbeit selbst angefertigt und alle benutzten Hilfsmittel in der Arbeit angegeben habe,
- dass ich diese Dissertation nicht schon als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, und
- dass weder die gleiche noch eine andere Abhandlung der Dissertation bei einer anderen Universität oder einem anderen Fachbereich der Technischen Universität Kaiserslautern veröffentlicht wurde.

07. April 2022, Kaiserslautern

Petra Ludowicy

Curriculum Vitae

Name: Petra Ludowicy

Anschrift: Erwin-Schrödinger-Straße 57, 67653 Kaiserslautern

E-Mail: Petra.Ludowicy@sowi.uni-kl.de

wissenschaftlichen Werdegang

Seit 12/2015

Promotion (Dr. rer. nat.)

Technische Universität Kaiserslautern

10/2013 – 11/2015

Master of Science in Kognitionswissenschaft

Technische Universität Kaiserslautern

Titel der Masterarbeit: „Comparison of neural correlates after positive, negative and neutral feedback during speech – non-speech discrimination“

10/2010 – 09/2013

Bachelor of Science in Biowissenschaften

Technische Universität Kaiserslautern

Titel der Bachelorarbeit: „Methylation Pattern of the Genes MGMT, P15, P16 and Rb in Glioblastoma“