The Testing Effect and Emotions

Investigating Potentially Influencing Factors on the Mnemonic Benefits of Testing

Kumulative Inaugural-Dissertation zur Erlangung der Doktorwürde
der Fakultät für Humanwissenschaften
der Universität Regensburg

vorgelegt von

KATHRIN JOHANNA EMMERDINGER

aus Deggendorf

Regensburg 2020

Erstgutachter: Prof. Dr. Christof Kuhbandner

Zweitgutachter: Prof. Dr. Karl-Heinz T. Bäuml

DANKSAGUNG

An dieser Stelle möchte ich denjenigen Menschen danken, die mich während der Entstehung dieser Arbeit begleitet und unterstützt haben.

Zunächst gilt mein Dank Prof. Dr. Christof Kuhbandner für die Betreuung dieser Arbeit, die wertvollen fachlichen Anregungen und Diskussionen sowie den stets hilfreichen Rat bei allen Fragen und Bedenken. Weiterhin danke ich dem gesamten Team des Lehrstuhls für Pädagogische Psychologie für die schöne und inspirierende gemeinsame Zeit am Lehrstuhl. Den studentischen Hilfskräften danke ich für die Unterstützung bei der Datenerhebung. Zudem möchte ich allen Personen danken, die an den präsentierten Studien teilgenommen haben.

Ein großer Dank geht an Aleix für seine Ermutigung, seine Unterstützung und sein Verständnis während meiner gesamten Promotionszeit. Besonders danken möchte ich auch meiner Familie. Meiner Schwester danke ich für ihr offenes Ohr und ihre klare Sichtweise, die mir so manches Mal weitergeholfen hat. Mein ganz besonderer Dank gilt schließlich meinen Eltern, denen ich diese Arbeit widmen möchte. Ihr habt immer an mich geglaubt, seid hinter mir gestanden und für mich da gewesen. Dafür danke ich euch aus ganzem Herzen.

PREFACE

In educational contexts as well as in basic cognitive research, tests are regarded mainly as a tool to assess the success of previous learning. However, this view does not adequately reflect the role tests play in knowledge acquisition, as the event of testing itself contributes significantly to the formation of durable memory representations. Numerous studies show that retrieving previously studied information from memory benefits later memory more than re-exposure to that information (Roediger & Butler, 2011; Roediger & Karpicke, 2006a). To date, a large body of research has been devoted to the investigation of influencing factors and boundary conditions of this so-called "testing effect" (Carrier & Pashler, 1992). Based on this research, extensive claims have been made regarding educational applications (e.g., Nunes & Karpicke, 2015). Yet, little research has examined whether emotional factors may affect the testing effect, although emotions are ubiquitous in real-life educational settings (Pekrun, Goetz, Titz, & Perry, 2002), and impact all stages of memory processing (for a review, see Fiedler & Hütter, 2014; Holland & Kensinger, 2010; Kensinger, 2009). In this context, both the emotional significance of memories itself, and the emotions that learners experience during studying and testing, may play a role. In addition, one special type of emotionally significant memory representations is memory of personally experienced events (i.e., autobiographical memories). To date, it is unknown whether the testing effect persists in autobiographical memory.

This cumulative thesis presents three studies that address the above-mentioned potentially moderating variables of the testing effect. These studies have been published in peer-reviewed journals over the last three years. A short overview of the studies is given on page 6, and the contributions of the co-authors are shown on page 7. In the first part of the thesis, the theoretical and empirical background underlying the three studies is presented. For this purpose, an overview of research on the testing effect and on the influence of emotions on cognitive processing is provided. In the second and main part of the thesis, the three studies are reproduced in their accepted pre-print versions with the following changes. First, the references for all three studies are combined into one bibliography at the end of the thesis. Second, section headings are standardized, and the numbering of the figures and tables is adjusted to allow consecutive numbering throughout the thesis. Third, Study 2's supplementary material (originally provided online) is included at the end of the article. The third and last part of the thesis provides a comprehensive discussion of the three studies' results and outlines how they contribute to the existing testing-effect literature and may spark new research.

CONTENTS

ABSTRACT

The finding that testing previously presented information promotes later memory more than presenting the information again (i.e., the testing effect) is a well-documented phenomenon, and numerous studies have identified influencing factors and boundary conditions. The present cumulative thesis aimed to clarify the role of factors that have received only marginal attention in previous testing-effect research: the potential influences of emotions. **STUDY 1** examined whether the effects of testing differ for emotionally significant memories compared to neutral memories. For this purpose, a standard cued recall testing-effect paradigm was employed. Participants first studied emotionally negative or neutral cue-target pairs, and then repeatedly retrieved or restudied the pairs. In a 1-week delayed memory test, both an emotional enhancement effect and the typical pattern of the testing effect, with higher memory performance for previously tested than for previously restudied memories, were observed. However, emotionally negative and neutral memories did not differ in the effects of testing and restudying. **STUDY 2** investigated whether the testing effect extends to emotional and unemotional autobiographical memories. Participants initially described memories of personally experienced events in response to emotionally negative, emotionally positive, and neutral cue words, and were subsequently tested on their previously described memories or restudied them. In two delayed memory tests at a 2-week and 13-week retention interval, regardless of the emotional significance of the memories, no advantage of testing compared to restudying autobiographical memories emerged. In **STUDY 3**, across two experiments, the effect of the affective state experienced during studying and testing was investigated. An established testing-effect paradigm using educationally relevant text material was employed, and affective states were manipulated before participants initially studied the texts (Experiment 1) or before they were tested on the texts or restudied them (Experiment 2). In both experiments, in a 1-week delayed memory test, the typical pattern of the testing effect emerged, with superior performance for previously tested compared to previously restudied contents. However, in none of the experiments did affect influence the occurrence or magnitude of the testing effect. In summary, Study 1 and Study 3 indicate that the testing effect may be relatively robust against emotional influences. This finding is of special relevance regarding potential educational applications. Study 2, on the other hand, shows a potential boundary condition: The advantage of testing compared to repeated re-exposure may not persist in autobiographical memory. The present cumulative thesis concludes with a comprehensive discussion of the three studies' results, which outlines their contributions to existing literature, and their implications for future testing-effect research.

CONTRIBUTIONS

STUDY 1

Testing Emotional Memories: Does Negative Emotional Significance Influence the Benefit Received From Testing?

Study idea	Christof Kuhbandner
Study design	Franziska Berchtold, Kathrin J. Emmerdinger, Christof Kuhbandner
Data analysis	Kathrin J. Emmerdinger, Christof Kuhbandner
Manuscript writing	Kathrin J. Emmerdinger
Manuscript revision	Kathrin J. Emmerdinger, Christof Kuhbandner

STUDY 2

Testing Memories of Personally Experienced Events: The Testing Effect Seems Not to Persist in Autobiographical Memory

Study idea	Christof Kuhbandner
Study design	Kathrin J. Emmerdinger, Christof Kuhbandner
Data analysis	Kathrin J. Emmerdinger, Christof Kuhbandner
Manuscript writing	Kathrin J. Emmerdinger
Manuscript revision	Kathrin J. Emmerdinger, Christof Kuhbandner

STUDY 3

Tests Improve Memory – No Matter if you Feel Good or Bad While Taking Them

Study idea	Christof Kuhbandner
Study design	Kathrin J. Emmerdinger, Christof Kuhbandner
Data analysis	Kathrin J. Emmerdinger, Christof Kuhbandner
Manuscript writing	Kathrin J. Emmerdinger
Manuscript revision	Kathrin J. Emmerdinger

PART I

INTRODUCTION

The Testing Effect

The act of retrieving information from memory in a test can provide a powerful means to enhance later memory for that information. While these beneficial consequences of testing previously learned information have already been demonstrated in empirical studies dating back about 100 years (Abott, 1909; Gates, 1917; Jones, 1923-1924; Spitzer, 1939), and were sporadically revisited by cognitive researchers from the 1960s to the 1990s (e.g., Allen, Mahler, & Estes, 1969; Bartlett & Tulving, 1974; Carrier & Pashler, 1992; Darley & Murdock, 1971; Glover, 1989; Hogan & Kintsch, 1971; Landauer & Eldridge, 1967; Madigan & McCabe, 1971; Tulving, 1967), increased interest in the testing effect (Carrier & Pashler, 1992) emerged in the 2000s (Roediger & Karpicke, 2018), resulting in a boost of further research.

The term "testing effect" refers to the finding that retrieving previously studied information from memory in a test (also termed "retrieval practice") benefits long-term retention beyond the effect of control activities that imply the same amount of re-exposure to that information (Roediger & Butler, 2011; Roediger & Karpicke, 2006a). Hence, the typical testing-effect paradigm consists of three phases: an initial study phase, a subsequent restudy/ testing phase, where the previously presented material is either re-presented or tested, and a delayed memory test on the previously presented material. The testing effect has been replicated across various test formats, study materials, experimental designs, and sample characteristics (for meta-analyses, see Adesope, Trevisan, & Sundararajan, 2017; Rowland, 2014). Furthermore, variables modulating the testing effect have been examined in numerous studies, both in laboratory and applied classroom settings. The present chapter first provides an overview of contemporary theoretical accounts of the testing effect and then outlines the current state of research concerning influencing factors and boundary conditions of the testing effect.

Theoretical Accounts

To date, the mechanisms underlying the testing effect are not yet fully understood. Five theoretical accounts which have been prominently discussed in previous research are transfer-appropriate processing, retrieval effort, distribution-based bifurcation, elaborative processing, and episodic context (for reviews, see Karpicke, Lehman, & Aue, 2014; Roediger & Karpicke, 2006a; Rowland, 2014). It should be noted that these accounts are not necessarily mutually exclusive (see Rowland, 2014). The first of the above-mentioned theoretical accounts, transfer-appropriate processing, suggests that memory performance in a delayed test will be best if the required cognitive processes are similar or identical to the cognitive processes at encoding (Morris, Bransford, & Franks, 1977; Roediger & Karpicke, 2006a). The second theoretical account, retrieval effort, suggests that effortful processing yields more durable memory representations than easy processing (e.g., Bjork, 1994; Pyc & Rawson, 2009). The remaining three accounts, distribution-based bifurcation (e.g., Kornell, Bjork, & Garcia, 2011), elaborative processing (e.g., Carpenter, 2009, 2011; Kornell, Klein, & Rawson, 2015; Pyc & Rawson, 2010), and episodic context (e.g., Karpicke, Lehman et al., 2014; Rowland, Littrell-Baez, Sensenig, & DeLosh, 2014) are the most widely discussed accounts in recent research, and thus will be outlined in more detail below.

The distribution-based bifurcation model (Kornell et al., 2011; see also Halamish & Bjork, 2011), similar to the retrieval-effort account, originates from the assumption that successful retrieval provides a boost in memory strength. In addition, the model assumes that tests can bifurcate distributions of memory strengths across items because successfully retrieved items receive a large boost whereas non-retrieved items receive no boost in memory strength. By contrast, when restudying, all items are presented again, thus all restudied items receive a boost in memory strength, but to a lesser degree than successfully retrieved items. The bifurcation model can serve as a useful framework to describe typical patterns of results in testing-effect research, such as the common finding of a test-delay interaction, with the testing effect appearing or increasing after increased retention intervals (e.g., Roediger & Karpicke, 2006b; see Rowland, 2014, for a meta-analysis). According to the bifurcated distribution proposed by the model, at short retention intervals, restudying may seem more beneficial than testing because comparably more restudied than tested items received a boost in memory strength and thus lie above recall threshold. Over time, memory strength declines for all items, resulting in restudied items falling below the recall threshold, while formerly successfully retrieved items still possess enough memory strength to stay above the recall threshold even after a prolonged retention interval. In line with this assumption, the test-delay interaction seems to decrease or disappear if feedback is provided after initial retrieval or if all items are retrieved successfully, presumably because under these conditions, bifurcation occurs to a lesser degree (provision of feedback) or not at all (successful retrieval of all items) (Kornell et al., 2011). However, the bifurcation model does not aim to provide any explanations concerning specific underlying mechanisms (Kornell et al., 2011, p. 86); that is, the model presupposes that retrieval strengthens memories but does not specify why and how this strengthening of memories occurs (see Karpicke, Lehman, & Aue, 2014).

The next theoretical account, elaborative processing, attempts to specify why retrieving information in a test may strengthen memory more than restudying. This account posits that the act of retrieval initiates elaborative processing (Carpenter, 2009, 2011; Pyc & Rawson, 2010). Specifically, in search of the target information, related concepts are activated, which serves to update and strengthen the memory trace and creates multiple pathways by which the information can be accessed at retrieval. By contrast, restudying involves complete re-exposure to the target information, which renders extensive elaborative processing less likely. The elaborative processing account also aligns with the retrieval-effort account. More retrieval effort involves more extensive memory search until the correct target information is encountered. This results in more widespread activations of related memory contents, which may later serve as additional retrieval routes, compared to easy retrievals where the target information is readily accessible.

Another recent account of the testing effect that involves elaborative retrieval as the core component is the two-stage framework proposed by Kornell et al. (2015). According to this framework, both processing during the retrieval attempt (stage 1) and processing after the answer has become available (stage 2) are accountable for the benefit received from testing. Specifically, processing during the retrieval attempt involves the spreading of activations through the associative network of memory contents related to the presented cue or question. As soon as the correct answer is activated and successfully retrieved, the associations between the target information, related concepts, and the cue or question are updated and strengthened, which in turn facilitates subsequent retrieval attempts. By contrast, when restudying, the first stage of the process is omitted, and consequently less related concepts and associations are activated. According to this model, successful retrieval is not a prerequisite for the benefits of retrieval attempts or for subsequent answer processing to occur; unsuccessful retrieval attempts followed by correct answer feedback can yield similar memory benefits (Kornell, Hays, & Bjork, 2009; Kornell et al., 2015). The two-stage framework is also consistent with findings showing that factors influencing reconsolidation immediately after retrieval can modulate the magnitude of the testing effect (Finn & Roediger, 2011; Finn, Roediger, & Rosenzweig, 2012). These findings demonstrate that both processing during the retrieval attempt and after the answer has become available are crucial for the testing effect.

Besides semantic elaboration, the reactivation of contextual episodic information during testing may play an important role. The episodic context account (Karpicke, Lehman, & Aue, 2014; Rowland et al., 2014) presumes that during retrieval, the episodic context that was present during initial study is reinstated, and upon successful retrieval, memory traces are updated with additional features pertaining to the present episodic context. Those episodic contextual features can serve as additional retrieval cues in later tests, rendering previously retrieved memory traces more retrievable, compared to memory traces with non-updated context representations. This account also predicts that more effortful initial retrieval conditions should elicit more powerful testing effects. For example, greater lags between initial study and testing or between subsequent retrieval attempts, make context reinstatement more necessary (as the current context will have shifted more from the original learning context). Therefore, memory traces will benefit more from the enrichment with new, additional contextual features. Similarly, free recall tests may require reinstatement of prior contexts more than recognition tests. Finally, in cued recall tasks, context reinstatement may be more necessary if semantic associations between the cue and the target information are weak (see Carpenter, 2009).

Influencing Factors and Boundary Conditions

Numerous studies have sought to identify influencing factors and boundary conditions for the occurrence and magnitude of the testing effect. In the following section, previous findings related to the employed test conditions, study materials, research designs and settings, and the characteristics of the learners are reviewed.

Test Conditions

The testing effect has been replicated for a variety of different types of tests, including free recall tests (e.g., Roediger & Karpicke, 2006b; Smith, Floerke, & Thomas, 2016), short answer tests (e.g., Butler & Roediger, 2007; Nungester & Duchastel, 1982), cued recall tests (e.g., Carpenter, 2009, 2011; Karpicke, Blunt, & Smith, 2016), and recognition tests (e.g., McDaniel, Bugg, Liu, & Brick, 2015). Regarding influencing factors and boundary conditions related to the employed test materials and test conditions, three main factors have been identified as vital in many testing-effect studies.

First, the success of the initial retrieval is a crucial factor underlying the testing effect (e.g., Karpicke, Blunt, Sumeracki, & Karpicke, 2014). For example, in a recent meta-analysis (Rowland, 2014), reliable testing effects only emerged in studies where the initial retrieval success rate was at least 50% (unless feedback was provided, which will be discussed in more detail further below).

Second, more robust and larger testing effects generally result from conditions that render the initial retrieval more difficult. For example, testing effects are generally larger for initial tests involving productive recalls, such as free recall and cued recall tests, than for initial tests involving recognition (e.g., Butler & Roediger, 2007; Carpenter & DeLosh, 2006; Glover, 1989; Greving & Richter, 2018; McDaniel, Anderson, Derbish, & Morrisette, 2007; for a meta-analysis see Rowland, 2014). It is important to note that this seems to be the case regardless of the format of the final memory assessment (Carpenter & DeLosh, 2006; Glover, 1989; Kang, McDermott, & Roediger, 2007; McDaniel et al., 2007; for a meta-analysis, see Rowland, 2014), which contradicts the transfer-appropriate processing account. More potent testing effects are generally also produced by other conditions that render initial retrieval more effortful, such as providing less cue support (e.g., Carpenter & DeLosh, 2006), using weakly associated rather than strongly associated cue-target pairs (Carpenter, 2009), or extending the lag between initial studying and initial testing (e.g., Karpicke & Roediger, 2007a, 2007b; Pyc & Rawson, 2009). Similarly, additional benefits relative to a single test have mostly been found when several tests were administered spaced apart, but not always when several tests were administered in a massed schedule (e.g., Glover, 1989; Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2007a; Karpicke & Roediger, 2010; Lyle, Bego, Hopkins, Hieb, & Ralston, 2019; Roediger & Karpicke, 2006b). These findings demonstrating more potent testing effects following more difficult initial retrieval conditions support the theoretical accounts that emphasize the prominent role of effortful processing during initial retrieval, due to either more extensive elaborative processing or more extensive shifts in episodic contexts. In summary, the testing effect seems to require the initial retrieval to be difficult enough to exert effortful processing, but, on the other hand, not too difficult to jeopardize sufficiently high initial retrieval success.

Third, the length of the retention interval between the initial acquisition (i.e., study and restudy/testing phase) and the final memory assessment seems to be another important boundary condition for the testing effect. Usually, reliable testing effects have been found after retention intervals of several days up to a week (for a meta-analysis, see Rowland, 2014) or even longer (e.g., 1 month, Butler & Roediger, 2007; 9 months, Carpenter, Pashler, & Cepeda, 2009). However, the picture is less clear concerning short retention intervals. Typically, a test-delay interaction can be observed with no testing effect or even a restudy benefit emerging at short retention intervals of several minutes, and the pattern reversing when retention intervals extend to several days (e.g., Congleton & Rajaram, 2012; Mulligan & Peterson, 2015; Roediger & Karpicke, 2006b; Toppino & Cohen, 2009; Wheeler, Ewers, & Buonanno, 2003). Yet, there are also several studies in which potent testing effects have been demonstrated already after short retention intervals (e.g., Carpenter, 2009; Carpenter & DeLosh, 2006; Karpicke & Zaromb, 2010; Rowland & DeLosh, 2015). According to the distribution-based bifurcation model (Kornell et al., 2011) and two-stage model (Kornell et al., 2015), moderating factors may be exceedingly high initial retrieval rates or the provision of feedback after unsuccessful retrieval attempts (e.g., Kornell et al., 2011, 2015; Rowland & DeLosh, 2015). These factors may lead to an equal amount of item exposure for both restudy and testing conditions.

Beyond the three main boundary conditions described so far, providing test feedback and testing transfer of knowledge are further factors that can influence the testing effect. Concerning the role of feedback, in a recent meta-analysis (Rowland, 2014), regardless of the initial retrieval success, the largest testing effects were observed for studies in which test feedback was provided, either by presenting correct test answers or by re-presenting study material at least once after initial testing. Indeed, other studies demonstrate that the presentation of the correct answer after an unsuccessful retrieval attempt can yield a testing effect (Hays, Kornell, & Bjork, 2013; Kornell et al., 2009) and even lead to equal benefits in delayed memory performance as a successful retrieval (Kornell et al., 2015). When comparing the benefit received from testing with and without feedback, most studies reveal an advantage of testing with feedback (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Butler, Karpicke, & Roediger, 2007; Butler & Roediger, 2008; Kang et al., 2007; Karpicke & Roediger, 2007b; 2010; McDaniel & Fisher, 1991; Zaromb & Roediger, 2010). However, it is important to note that when providing feedback or additional study opportunities following testing, additional factors beyond the direct effects of testing itself come into play. For example, it has been shown that tests can render subsequent learning more effective (test-potentiated learning; Arnold & McDermott, 2013; Wissman & Rawson, 2018), even when new information is presented that is unrelated to the test contents (forward effect of testing; Pastötter & Bäuml, 2014, 2019; for reviews, see Chan, Meissner, & Davis, 2018; Yang, Potts, & Shanks, 2018). Thus, it is not clear whether observed benefits of tests followed by feedback stem from direct effects of testing itself, indirect effects of testing on subsequent learning, or the interplay of both mechanisms. Indeed, for this reason, providing feedback in testing-effect studies has been criticized because it is difficult to draw conclusions about the mnemonic benefits of retrieval (Karpicke, Lehman, & Aue, 2014).

An important issue regarding potential educational implications of the testing effect is to what extent testing boosts the transfer and application of previously acquired contents across different types of knowledge assessments. However, findings concerning the transfer of learning from tests are mixed (for a review, see Pan & Rickard, 2018). There is some evidence that testing compared to restudying improves knowledge transfer within the same or a different knowledge domain (e.g., Butler, 2010; Butler, Black-Maier, Raley, & Marsh, 2017; Cho & Powers, 2019; Dirkx, Kester, & Kirschner, 2014; Jacoby, Wahlheim, & Coane, 2010; Jensen, McDaniel, Woodard, & Kummer, 2014; Kang, McDaniel, & Pashler, 2011; Rohrer, Taylor, & Sholar, 2010). However, other studies failed to observe testing effects in knowledge transfer assessments (e.g., Hinze & Wiley, 2011; Pan, Gopal, & Rickard, 2016; Pan & Rickard, 2017; Wooldridge, Bugg, McDaniel, & Liu, 2014), such as deductive inferences (Tran, Rohrer, & Pashler, 2015), transfer to a new perspective in spatial memory (Brunyé et al., 2019), or when the initial type of test questions (factual vs. higher-order) mismatched the type of test questions in the final test (Agarwal, 2019).

Study Materials

The testing effect has been replicated using a wide range of study materials, including word lists (e.g., Carpenter & DeLosh, 2006; Karpicke et al., 2016; Smith et al., 2016; Zaromb & Roediger, 2010), paired associates (e.g., Carpenter, 2009, 2011; Carpenter, Pashler, & Vul, 2006; Carrier & Pashler, 1992), pictorial information (e.g., Kang, 2010; Lipowski, Pyc, Dunlosky, & Rawson, 2014; Smith et al., 2016), and more educationally relevant materials such as vocabulary pairs (e.g., Carrier & Pashler, 1992; Kuhbandner, Aslan, Emmerdinger, & Murayama, 2016; Pyc & Rawson, 2009, 2012; Tse & Pu, 2012), text materials (e.g., Agarwal et al., 2008; Butler, 2010; Dobson & Linderholm, 2015; Karpicke & Blunt, 2011; Roediger & Karpicke, 2006b; Weinstein, McDermott, & Roediger, 2010), mathematical functions (Kang et al., 2011), skills learning materials (Kromann, Jensen, & Ringsted, 2009), spatial learning materials (Brunyé et al., 2019; Carpenter & Kelly, 2012; Carpenter & Pashler, 2007; Rohrer et al., 2010), and videotaped or online lectures (Butler & Roediger, 2007; Heitmann, Grund, Berthold, Fries, & Roelle, 2018).

Evidence is less clear concerning learning from complex reading materials. Although the testing effect usually remains robust for text materials such as prose passages (e.g., Agarwal et al., 2008; Dobson & Linderholm, 2015; Karpicke & Blunt, 2011; Roediger & Karpicke, 2006b; Weinstein et al., 2010), recent studies failed to replicate the testing effect for highly interrelated reading materials (in contrast to learning materials in which each fact can theoretically be learned separately) (for a review, see Van Gog & Sweller, 2015; but see Karpicke & Aue, 2015; Rawson, 2015, for contrary positions). For instance, De Jonge, Tabbers, and Rikers (2015), in a 1-week delayed memory test, found no benefit for testing compared to restudying a complex science text. However, when the same text was dissected into a list of facts, which reduced the text complexity by disrupting the text coherence, the typical testing effect did emerge. Furthermore, several studies showed that repeatedly restudying worked examples of problem solutions may be equally or even more beneficial for later problem solving than various kinds of retrieval practices after studying worked examples (Leahy, Hanham, & Sweller, 2015; Van Gog & Kester, 2012; Van Gog et al., 2015; Yeo & Fazio, 2019).

Research Designs and Settings

The testing effect seems to remain robust across different experimental designs, such as manipulating testing and restudy conditions within subjects (e.g., Butler, 2010; Roediger & Karpicke, 2006b) or between subjects (e.g., Hinze & Rapp, 2014; Karpicke & Zaromb, 2010; Pyc & Rawson, 2010), or presenting restudy and test trials consecutively in blocked lists (e.g., Carpenter et al., 2006; Karpicke & Zaromb, 2010) or intermingled in mixed lists (e.g., Carpenter, Pashler, Wixted, & Vul, 2008; Karpicke & Zaromb, 2010, for a direct comparison, see Abel & Roediger, 2017; Rowland et al., 2014).

Regarding potential educational implications of the testing effect, an important question is whether the effect, as established in laboratory settings, remains stable when examined in field studies. Indeed, the effects of testing have been replicated in several classroom studies, mostly by applying short (multiple-choice or short-answer) quizzes on parts of the class material (e.g., Agarwal, 2019; Bjork, Little, & Storm, 2014; Carpenter et al., 2009; Cranney, Ahn, McKinnon, Morris, & Watts, 2009; Hopkins, Lyle, Hieb, & Ralston, 2016; Jensen et al., 2014; McDaniel et al., 2007; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014; Roediger, Agarwal, McDaniel, & McDermott, 2011; Trumbo, Leiting, McDaniel, & Hodge, 2016). However, regarding the actual importance of these findings for educational settings, two points should be kept in mind. First, the testing effects in these classroom studies might be an artificial effect due to the somewhat weak control conditions (e.g., doing nothing or simply reading the quizzed statements, which has been shown to be an inefficient study strategy; Callender & McDaniel, 2009; see also Kornell, Rabelo, & Klein, 2012). Second, the final criterion tests often consisted of similar or identical test questions and/or test formats so that other factors beyond retrieval benefits may have played a role (e.g., test practice; McCabe, Langer, Borod, & Bender, 2011). Generally speaking, when considering testing effects outside of laboratory contexts, observed benefits may actually reflect the effects of a number of other factors beyond the direct effects of testing, such as increased motivation for continuous, spaced study of contents, increased metacognitive awareness of current knowledge gaps, test expectancy, and increased test experience (e.g., Bangert-Drowns, Kulik, & Kulik, 1991; Finley & Benjamin, 2012; McCabe et al., 2011; Roediger & Karpicke, 2006a; Thomas & McDaniel, 2007).

Characteristics of the Learners

Although most testing-effect studies have been conducted with undergraduate students, there is some evidence that the testing benefits can generalize across age and education levels. Testing effects have been demonstrated with elementary or preschool children (e.g., Bouwmeester & Verkoeijen, 2011; Fritz, Morris, Acton, Voelkel, & Etkind, 2007; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Jones et al., 2016; Karpicke, 2016; Lipowski et al., 2014; Rohrer et al., 2010; for a review, see Fazio & Marsh, 2019), middle school students (e.g., McDermott et al., 2014; Roediger et al., 2011), high school students (e.g., Morris, Fritz, Jackson, Nichol, & Roberts, 2005), young adults who do not attend university (Meyer & Logan, 2013), and middle-aged to older adults (e.g., Bishara & Jacoby, 2008; Coane, 2013; Meyer & Logan, 2013; Tse, Balota, & Roediger, 2010).

Additionally, several studies have examined whether the magnitude of the testing effect is influenced by individual differences in cognitive abilities. Some studies show that the testing effect persists across different levels of reading comprehension, processing speed, and visual word decoding skills in children (Karpicke et al., 2016; Moreira, Pinto, Justi, & Jaeger, 2019), and for populations with memory dysfunctions (Sumowski et al., 2010). Other studies indicate that the testing effect remains robust (Bertilsson, Wiklund-Hörnqvist, Stenlund, & Jonsson, 2017; Minear, Coane, Boland, Cooney, & Albat, 2018; Wiklund-Hörnqvist, Jonsson, & Nyberg, 2014), or may even be more pronounced, for individuals with low working memory capacity (Agarwal, Finley, Rose, & Roediger, 2017; but see Tse & Pu, 2012 for a finding that testing effect and trait test anxiety may be negatively correlated in individuals with low working memory capacity) or low episodic memory skills (Brewer & Unsworth, 2012; Pan, Pashler, Potter, & Rickard, 2015; Robey, 2019). In addition, the testing effect may be larger for individuals who possess less effective learning strategies (Minear et al., 2018; Robey, 2019) and less prior topic knowledge (Cogliano, Kardash, & Bernacki, 2019). In summary, previous research indicates that individuals with lower memory abilities and ineffective learning strategies may benefit equally or even more from testing than others (but see Carpenter et al., 2016 for the contradictory finding of a larger testing effect in high-achieving students).

Beyond trait-like individual differences, another important issue is whether transitory psychological states, such as the learner's motivational or affective states, may influence the occurrence and magnitude of the testing effect. However, this issue has received only marginal attention in previous research. To date, only two studies have directly manipulated participants' motivational state during retrieval practice (Kang & Pashler, 2014; Kuhbandner et al., 2016). These studies, while still replicating the testing effect, did find motivational influences on the overall benefits received from testing. For example, Kuhbandner et al. (2016) demonstrated that inducing extrinsic motivation by providing monetary reward contingent on the initial retrieval success decreased correct recall, increased commission errors for previously tested contents, and even reduced correct recall for initially successfully retrieved memories, compared to a no-reward condition. Thus, extrinsic motivation may reduce the benefit received from successful retrieval.

A few studies examined the testing effect in contexts of performance pressure and acute stress. Hinze and Rapp (2014) induced performance-related affect during retrieval practice by using a peer-pressure method. In the pressure condition, researchers informed participants that they could earn themselves and a partner a monetary bonus if they scored higher than the university average on the retrieval task, and that their partner had already achieved the required score. In a delayed test, a testing effect emerged when comparing a no-pressure condition to a reread condition. However, the testing effect was absent, and for some question types even reversed, when comparing the pressure condition to the reread condition. Some classroom studies found that ungraded quizzes may reduce test anxiety with respect to the final assessment (Agarwal, D'Antonio, Roediger, McDermott, & McDaniel, 2014) and that ungraded quizzes yield better final exam performance compared to taking no quizzes (Khanna, 2015; Khanna & Cortese, 2016). In contrast, students reported more anxiety regarding a course when graded quizzes were employed, and the graded quizzes did not lead to improved final exam performance (Khanna, 2015; Khanna & Cortese, 2016). Finally, when acute stress was induced before the final memory assessment, the testing effect remained intact (Szőllősi et al., 2017). Some studies suggest that previous retrieval practice may even protect memory against stress-induced detrimental effects (Smith et al., 2016; Smith, Davis, & Thomas, 2018; Smith, Race, Davis, & Thomas, 2019).

Emotions and Memory

As reviewed in the previous chapter, numerous studies have been conducted to identify moderating variables and boundary conditions of the testing effect. However, one important factor that has received only marginal attention in testing-effect research is the potential influence of emotions. This is especially surprising given the fact that extensive implications for educational settings have been drawn on the basis of testing-effect research (e.g., Karpicke, 2016; Nunes & Karpicke, 2015). In real-life educational settings, emotions are ubiquitous

(Pekrun et al., 2002). At the same time, a wealth of research has demonstrated impacts of emotions on all levels of information processing (for reviews, see e.g., Fiedler & Hütter, 2014; Kensinger, 2009; Reisberg & Heuer, 2004). Thus, it is essential to investigate whether emotions may influence the effects of testing.

From a functional perspective, two general levels of emotional influences on memory can be distinguished (Kuhbandner & Pekrun, 2010, p. 37). First, numerous studies indicate that emotionally significant stimuli may be processed and stored qualitatively different than neutral stimuli (for reviews, see Kensinger, 2009; Levine & Edelstein, 2009). In this context, one special category of emotionally significant memory representations is memory of personally experienced events (i.e., autobiographical memories). Autobiographical memory representations, beyond being often emotionally significant, also exhibit other distinct characteristics (Conway, 2005) that may influence the way they are processed. Second, previous research shows that an individual's currently experienced affective state may influence cognitive processing, regardless of whether emotionally significant or neutral information is processed (for a review, see Clore & Huntsinger, 2007; Fiedler & Hütter, 2014). This chapter presents a brief review of research on the processing of emotional memories, memories for personally experienced events, and the effects of affective states on memory.

Emotions – a Short Conceptual Clarification

Although exact definitions of the term "emotion" diverge among different theoretical perspectives, there are a number of widely agreed-upon constitutional criteria of an emotion or emotional episode (Moors & Scherer, 2013). First, emotions are generally considered as temporarily restricted episodes that, albeit they may vary in duration, are usually short-lived. In contrast, more enduring, mildly intense affective states are often labelled as "moods" (Scherer, 2005). Second, emotional episodes are marked by changes in multiple organismic subsystems, including cognitive, motivational, physiological, and behavioral changes, as well as changes in subjective experience (often conceptualized as result of the perception of changes in the other subsystems, Moors, 2013). Different emotion theories diverge on how many of these components are necessary to speak of an emotion, whether certain components are causes, consequences, or constituents of an emotion, and the degree of synchronization along the changes in different subsystems. However, it is generally agreed upon that each of these changes and their interplay serve an adaptive function, namely, to evaluate and signal the relevance of stimuli for major goals of the organism, and prepare adaptive reactions (Moors, 2017; Moors & Scherer, 2013).

A major distinction between different theoretical perspectives concerns the basic underlying structure of emotions. In this context, two broad classes of theories can be distinguished: categorical and dimensional theories (Gray & Watson, 2007; Moors, 2017). Categorical theories postulate the existence of discrete, mutually exclusive basic emotional categories or affect programs (e.g., Ekman, 1992, 1999; Izard, 1993, 2009). Dimensional theories, on the other hand, describe affective states as values in a multidimensional space. Concerning the number of underlying dimensions, two-dimensional models have proven the best empirical fit (e.g., Russell & Barrett, 1999; Watson, Wiese, Vaidya, & Tellegen, 1999). These dimensions are mostly interpreted as valence, with values varying between displeasure and pleasure, and arousal, with values varying between activation and sleep (e.g., Barrett & Russell, 1999; Russell, 1980; Russell & Barrett, 1999). More recent models strive for an integration of dimensional and categorical emotional theories. They assume that affect is a composite of current values on the dimensions of valence and arousal, and that the categorization of affective experience into acquired mental concepts forms the experience of discrete emotional episodes (such as anger or sadness) (Barrett, 2006a, 2006b; Russell, 2003, 2009).

Memory for Emotionally Significant Information

Emotional significance is one of the main factors influencing our memory about past events. A large body of research shows that emotionally significant information is better remembered than neutral information, which is referred to as emotional memory enhancement (for reviews see Kensinger, 2004; Reisberg & Heuer, 2004). However, the emotional enhancement effect does not necessarily occur across all elements of an emotional scene; typically, a memory trade-off at the expense of contextual information can be observed (for a review, see Kensinger, 2009).

Memory enhancement for emotionally significant stimuli has been demonstrated using a wide range of stimuli such as words, pictures, or narrated slide shows (for reviews, see Buchanan & Adolphs, 2002; Kensinger, 2004; Reisberg & Heuer, 2004). The advantage of emotional stimuli compared to neutral stimuli can be found across all stages of information processing, from the earliest pre-attentive stages such as iconic memory (Kuhbandner, Spitzer, & Pekrun, 2011) to attentional selection and prioritized encoding, to later stages such as memory consolidation and retrieval. For instance, ample evidence indicates that emotionally significant stimuli attract attention more readily than neutral stimuli and receive prioritized processing (for reviews, see Vuilleumier, 2005; Yiend, 2010). Emotionally significant stimuli are encountered more rapidly in visual search tasks (e.g., Lundqvist, Bruce, & Öhman, 2015;

Öhman, Flykt, & Esteves, 2001), are fixated on first and gazed at longer compared to neutral stimuli (e.g., Calvo & Lang, 2004; Nummenmaa, Hyönä, & Calvo, 2006), and are more likely to be processed under conditions of limited attentional resources than neutral stimuli (e.g., Anderson, 2005; Kensinger & Corkin, 2004; Talmi & McGarry, 2012).

Regarding subsequent stages of information processing, emotional significance seems to modulate memory consolidation such that emotionally significant memories are more likely to be consolidated into durable memory traces, leading, in turn, to an increased likelihood of memory retrieval. This is supported by findings of a more pronounced emotional enhancement effect after longer retention intervals (e.g., LaBar & Phelps, 1998; for reviews, see Kensinger, 2004; McGaugh, 2004; Tyng, Amin, Saad, & Malik, 2017). Furthermore, studies on sleep-mediated memory consolidation show that sleep provides a larger benefit to memories for emotionally significant stimuli than for non-emotional stimuli (e.g., Payne, Stickgold, Swanberg, & Kensinger, 2008; Wagner, Hallschmid, Rasch, & Born, 2006). However, the evidence regarding the influence of emotional significance on other factors that modulate memories after initial encoding is less conclusive. For example, forgetting as a consequence of intentionally attempting to forget previously encoded information (e.g., Barnier et al., 2007; Taylor, Quinlan, & Vullings, 2018; Wessel & Merckelbach, 2006, but see Payne & Corrigan, 2007, for a contradictory finding), forgetting in the think/no-think paradigm as a consequence of repeatedly suppressive thinking of previously encoded information (e.g., Murray, Muscatell, & Kensinger, 2011; Sakaki, Kuhbandner, Mather, & Pekrun, 2014), and retrieval-induced forgetting of previously encoded information as a consequence of repeatedly retrieving related information (e.g., Barber & Mather, 2012; Barnier, Hung, & Conway, 2004; Kuhbandner, Bäuml, & Stiedl, 2009), appear not to differ for memories of emotionally significant and neutral stimuli.

Furthermore, memory-enhancing effects of emotional significance do not occur across all elements of an emotional event. A prominent example is the so-called “weapon focus effect” where the weapon in a crime scene is remembered with great clarity, but at the expense of other crucial details of the scene (e.g., Stanny & Johnson, 2000). Typically, when a visual scene contains an emotionally significant stimulus, a memory trade-off is observed: enhanced memory for the emotionally significant stimulus, but no memory enhancement or even reduced memory for background information and peripheral details present in the same visual scene (e.g., Christianson & Loftus, 1991; Kensinger, Garoff-Eaton, & Schacter, 2007a; Waring & Kensinger, 2011; for reviews see Kensinger, 2009; Levine & Edelstein, 2009; Mather, 2007; Reisberg & Heuer, 2004). Similarly, when studying paired associates, item memory for an emotionally significant image or word is typically enhanced, but memory for the associate reduced, compared to paired associates consisting of two neutral stimuli (e.g., Bisby & Burgess, 2014; Madan, Caplan, Lau, & Fujiwara, 2012; Nashiro & Mather, 2011; but see Guillet & Arndt, 2009, for a contradictory finding). Importantly, several studies monitored participants' eye movements and suggested that selective visual attention towards emotion-eliciting elements of a scene cannot fully account for the memory trade-offs (Riggs, McQuiggan, Farb, Anderson, & Ryan, 2011; Steinmetz & Kensinger, 2013). In fact, as the formation of single-item memory and the formation of associations between stimuli seem to rely on distinct mechanisms (e.g., Bisby & Burgess, 2014; Davachi, 2006; Giovanello, Schnyer, & Verfaellie, 2004; Giovanello, Verfaellie, & Keane, 2003; Hannula et al., 2015), conditions that benefit the two types of memory formation do not necessarily coincide. Thus, while emotionally significant stimuli are preferentially stored in memory, this may come at the expense of weakened associations to other stimuli that were present in the same scene.

Memory for Personally Experienced Events

Besides being often emotionally significant, memories of personally experienced events (i.e., autobiographical memories) differ from other memory representations in several ways. Autobiographical memories exhibit a high degree of self-relevance and a high degree of connectivity in an associative network (Conway & Pleydell-Pearce, 2000; Holland & Kensinger, 2010). Self-relevance and personal involvement are important factors in forming and modulating memory. Many studies have shown that memory performance increases if information is encoded in reference to the self (e.g., "does the information describe yourself?") (e.g., Kelley et al., 2002; Kuiper & Rogers, 1979; Reeder, McCormick, & Esselman, 1987; Rogers, Kuiper, & Kirker, 1977; for a review, see Symons & Johnson, 1997). Personal involvement also seems to play a vital role in the retention and vividness of autobiographical memories. For instance, Muscatell, Addis, and Kensinger (2010) found that hockey players remembered events in which they were directly involved more vividly (e.g., scoring a goal compared to watching a goal from the bench).

Autobiographical memories are assumed to be strongly connected to personal goals, motives, and self-images (Conway, 2005; Conway & Williams, 2008). One of the main functions of autobiographical memory, in addition to guiding future behavior and facilitating social relationships, is to maintain a coherent sense of self over time (for a review, see Bluck, Alea, Habermas, & Rubin, 2005). Memories of specific events are thought to be embedded in a rich context of more abstract knowledge about oneself and the personal past and, as such, strongly interconnected in an associative network. More precisely, autobiographical memory is thought to be hierarchically organized, with major lifetime periods at the top level, temporally extended events (e.g., a vacation) and repeated categories of events (e.g., every Christmas) at the intermediate level, and memories of specific episodes at the bottom level. At retrieval, rather than depicting a perfect record of experienced events, memories are reconstructed from this rich database of knowledge about oneself (Conway & Pleydell-Pearce, 2000; Conway & Williams, 2008).

The emotional enhancement effect described in the previous section is also evident in autobiographical memory. Emotionally significant experiences are remembered more vividly and more durably (e.g., Rubin & Kozin, 1984; White, 2002; for a review, see Holland & Kensinger, 2010), and experienced as more personally relevant than neutral memories (e.g., Barnier et al., 2004, 2007; Walker, Skowronski, Gibbons, Vogl, & Ritchie, 2009). A functional explanation of emotional memory enhancement in autobiographical memory rests upon the assumption that emotional significance signals the importance of an experience for goals that are vital for the self (Levine & Edelstein, 2009). In this regard, it has also been observed that more memories of emotionally positive experiences than of emotionally negative experiences tend to be recalled, and that negative affect seems to fade faster over time than positive affect (e.g., Berntsen & Rubin, 2002; Thompson, Skowronski, Larsen, & Betz, 1996; Walker, Vogl, & Thompson, 1997; for reviews see Walker & Skowronski, 2009; Walker, Skowronski, & Thompson, 2003). Self-regulatory functions may cause negative affect to fade more rapidly, making negative autobiographical memories less accessible. Moreover, as people are generally motivated to maintain a positive self-concept, emotionally positive autobiographical memories may become integrated into people's self-concepts more easily (Holland & Kensinger, 2010).

Affective States and Memory

Besides the question to what extent memory for emotionally significant events differs from memory for neutral events, another crucial question regarding affective influences on memory is whether an individual's current affective state influences memory processes, regardless of the emotional significance of the information being processed.

Two frequently observed effects of affective states on memory are mood-congruent memory and mood-dependent memory. Mood-congruent memory refers to the phenomenon that emotional contents matching the current affective state are more easily processed. Mood-dependent memory means that (neutral) information is more easily retrieved if the affective state at retrieval matches the affective state at encoding (for a review of research on both phenomena, see Eich & Forgas, 2003). These phenomena have been explained by means of a network-theory, suggesting that humans encode their affective experience in an associative network together with the elements of the situation in which the affect has been experienced (Bower, 1981; Bower & Forgas, 2000). Thus, because of spreading activation through the network, information that is linked to an affective experience is more likely to be recalled when the same affect is experienced. Conversely, affective experiences are more likely to be reinstated when associated information is activated. Interestingly, some studies also demonstrate the opposite effect, mood-incongruent memory, which refers to better recall of memories with opposite emotional valence than the currently experienced affective state. Mood-incongruent memory is observed predominantly in negative affective states. Therefore, it might be a self-regulatory mood-repair mechanism that helps to avoid prolonged negative affective states (for reviews, see Holland & Kensinger, 2010; Morris & Reilly, 1987).

Another prominent assumption concerning the effects of affective states on cognition is that, dependent on current affective state, different cognitive processing styles are adopted. More specifically, it is assumed that negative affect fosters a predominantly item-specific processing style, focusing on details and distinctive attributes of an item, whereas positive affect fosters a predominantly relational processing style, focusing on associations between items in relation to other concepts in memory (for reviews, see Clore & Huntsinger, 2007; 2009; Fiedler & Hütter, 2014). Research has shown affective influences on processing styles from the earliest pre-attentive stages of information processing to several higher stages such as attentional selection, feature binding in working memory, and activation of associative knowledge structures.

Regarding attention, positive affect seems to broaden the scope of attention, and negative affect to narrow it (for a review, see Vanlessen, De Raedt, Koster, & Pourtois, 2016). For instance, positive affect has been shown to enhance the amount of information temporarily available in iconic memory by reducing spatial biases (Kuhbandner, Lichtenfeld, & Pekrun, 2011), to enhance access to remote semantic associations (Rowe, Hirsh, & Anderson, 2007), to induce a global bias in visual perception (Fredrickson & Branigan, 2005), but to impair visual selective attention (Rowe et al., 2007). Regarding feature binding and storage of visual information, positive affect seems to foster the storage of integrated object representations and negative affect seems to foster the storage of independent feature representations (Spachtholz & Kuhbandner, 2017; Spachtholz, Kuhbandner, & Pekrun, 2016). Finally, regarding processing in relation to existing knowledge structures, positive affect tends to promote the activation of schemata and the spreading of activation through associative networks in memory, whereas negative affect tends to inhibit both. For example, studies demonstrate that when experiencing positive affect, broader thought-action-repertoires (Fredrickson & Branigan, 2005) and more flexible problem-solving strategies (Haager, Kuhbandner, & Pekrun, 2014) are accessible, and intentionally forgotten memories are reactivated (Bäuml & Kuhbandner, 2009). By contrary, when experiencing negative affect, stereotypes are less likely activated (Huntsinger, Sinclair, & Clore, 2009), cognitive priming is inhibited (Storbeck & Clore, 2008), and false memories are reduced (Storbeck, 2013).

On a functional level, these effects of affective states on processing styles are supposed to adjust the cognitive system according to the respective demands of the current situation. In particular, positive affect signals a benign environment, allowing the individual to stick to a heuristic, schema-driven processing style (e.g., Schwarz, 2012) that frees cognitive resources and permits a momentary broadening of information processing, thus contributing to an adaptive expansion of cognitive and action repertoires (e.g., Fredrickson, 2001; 2004). Negative affect, on the other hand, acts as a metaphorical stop sign, signaling a problematic environment in which the default mode of heuristic processing based on inner knowledge structures and schemes may not be adaptive, and more careful processing of the environment and its details is necessary (e.g., Clore & Huntsinger, 2007). These opposing effects of affective states on cognitive processing have also been described in terms of two complementary adaptive functions: assimilation and accommodation. Assimilation refers to the interpretation of environmental information based on existing internal schemes and knowledge structures, whereas accommodation refers to the updating of internal knowledge structures according to novel environmental information (e.g., Bless & Fiedler, 2006; Fiedler & Hütter, 2014; see Piaget, 1954).

The Present Studies

Albeit previous research has shown a wide range of emotional influences on memory (for reviews, see e.g., Fiedler & Hütter, 2014; Kensinger, 2009; Reisberg & Heuer, 2004), the question whether emotions may potentially influence the mnemonic benefits of testing has been largely left unattended. The present research set out to fill this research gap. For this purpose, three studies were conducted.

STUDY 1 addressed the question whether the effect of testing, compared to restudying, may differ for emotionally significant stimuli compared to neutral stimuli. Numerous studies indicate that emotionally significant stimuli may be preferentially stored in memory (for a review, see Kensinger, 2009), but some previous studies found largely similar effects on emotionally significant and neutral stimuli for factors known to influence memories after initial encoding (e.g., Kuhbandner et al., 2009; Sakaki et al., 2014; Wessel & Merckelbach, 2006). Concerning post-encoding factors, to date only factors with detrimental consequences for later memory have been investigated and it is unknown whether emotional significance may influence factors that are beneficial for later memory, such as the testing of previously encoded information. In order to examine whether the testing effect is modulated by emotional significance, a standard cued recall testing-effect paradigm was employed. Participants studied either emotionally negative or neutral target stimuli together with neutral retrieval cues, and were subsequently tested via the retrieval cues, or restudied the cue-target pairs. After 1 week, in order to measure both the ability to access the target stimuli via the cue-target association and the availability of the targets in memory independently of the cue, a two-stage memory test was employed. As memory enhancements for emotionally significant stimuli often come at the expense of weakened associations to other stimuli present in the same scene (for a review, see Kensinger, 2009; Mather, 2007), it was expected that, despite increased storage strength of negative memories, the benefit received from testing would be similar or even decreased for the negative stimuli compared to the neutral ones.

STUDY 2 examined whether the testing effect extends to emotional and neutral memories of personally experienced events. Autobiographical memories, besides being often emotionally significant and personally relevant, stand out due to a high degree of complexity and interconnectedness in an associative network (Conway, 2005; Conway & Pleydell-Pearce, 2000; Holland & Kensinger, 2010). Due to these characteristics, re-exposure and retrieval practice may induce the same benefit to autobiographical memory recall and the typical pattern of the testing effect may be attenuated, or if at all only appear after longer retention intervals (see Kornell et al., 2011, 2015; Symons & Johnson, 1997; Van Gog & Sweller, 2015). To examine the testing effect for autobiographical memories, an experimental procedure (adapted from Barnier et al., 2004) was employed where participants initially collected memories for personally experienced events in response to emotionally negative, emotionally positive, or neutral cue words. On the following day, participants repeatedly restudied or repeatedly retrieved their memories. Memory for all initially collected events was assessed 2 weeks and 13 weeks after restudying/testing. It was hypothesized that the testing effect may disappear for memories of personally experienced events, or, if at all, only emerge after the longer retention interval of 13 weeks.

Finally, **STUDY 3** addressed the question whether the affective state that learners experience during studying and testing of neutral information may modulate the testing effect. As described previously, research has shown that an individual's currently experienced affective state can influence the adopted processing styles (for a review, see Clore & Huntsinger, 2007; Fiedler & Hütter, 2014). Differences in adopted processing style may in turn modulate the benefit received from testing. Elaborative processing during retrieval is assumed to underlie the beneficial effects of testing (e.g., Carpenter, 2009, 2011; Kornell et al., 2015; Pyc & Rawson, 2010). Therefore, negative affect, which promotes item-specific processing and attenuates activation, may decrease the testing effect. In contrast, positive affect, which promotes relational processing and enlarges the spreading of activation, may increase the testing effect. To examine this hypothesis, two experiments were conducted. Emotionally negative, emotionally positive, or neutral affective states were induced either already before initial studying (Experiment 1) or directly before restudying/testing (Experiment 2). It was expected that in a 1-week delayed free recall test the typical pattern of the testing effect would emerge in the neutral affect condition, and that compared to the neutral affect condition, the advantage of testing compared to restudying would be increased in the positive affect condition and decreased in the negative affect condition.

PART II

PEER-REVIEWED STUDIES

STUDY 1

Testing Emotional Memories: Does Negative Emotional Significance Influence the Benefit Received From Testing?

Kathrin J. Emmerdinger, Christof Kuhbandner and Franziska Berchtold

Abstract

A large body of research shows that emotionally significant stimuli are better stored in memory. One question that has received much less attention is how emotional memories are influenced by factors that influence memories *after* the initial encoding of stimuli. Intriguingly, several recent studies suggest that post-encoding factors do not differ in their effects on emotional and neutral memories. However, to date, only detrimental factors have been addressed. In the present study, we examined whether emotionally negative memories are differentially influenced by a well-known beneficial factor: the testing of memories. We employed a standard cued recall testing-effect paradigm where participants studied cue-target pairs for negative and neutral target pictures. In a subsequent post-encoding phase, one third of the cue-target pairs were tested and one third restudied; the remaining third served as control pairs. After 1 week, memory for all cue-target pairs was tested. While replicating both the testing effect and the emotional enhancement effect, no differences between negative and neutral memories in the benefits received from testing and restudying were observed. Thus, it seems to be true that post-encoding factors do not influence emotional memories in any other way than neutral memories, even when they are beneficial.

A large body of research shows that one of the main factors influencing our memory about past events is their emotional significance. Using a wide range of stimuli such as words, pictures or narrated slide shows, numerous studies have demonstrated that emotional stimuli are better remembered than neutral stimuli (for a review, see Reisberg & Heuer, 2004). In particular, an advantage of emotional over neutral stimuli occurs across all information processing stages, even down to earliest stages such as sensory-memory (Kuhbandner, Spitzer, & Pekrun, 2011), and up to later stages such as the consolidation and retrieval of information in long-term memory (for a review, see LaBar & Cabeza, 2006). Thus, emotional stimuli seem to be preferentially stored in memory.

One question that has received much less attention is how emotional memories are influenced by factors that are known to influence memories *after* the initial encoding of stimuli. Intriguingly, other than at the level of initial encoding, several studies suggest that negative and neutral memories do not differ in the effects of post-encoding factors. For instance, the effects of factors such as the attempt to intentionally forget memories (i.e., directed forgetting; Wessel & Merckelbach, 2006), the intentional exclusion of memories from awareness (i.e., think/no-think forgetting; Sakaki et al., 2014), or the retrieving of related distracting memories (i.e., retrieval-induced forgetting; Kuhbandner et al., 2009) are largely similar for neutral and negative memories.

However, one characteristic that is common to all of the above mentioned post-encoding factors is that the effects are detrimental for stored memories, which all seem to reflect the memory consequences of inhibition – inhibition of (1) outdated memories (directed forgetting; MacLeod, 1998), inhibition of memories to intentionally exclude them from current awareness (think/no-think forgetting; Anderson & Green, 2001), and inhibition of distracting memories (retrieval-induced forgetting; Anderson, 2003). Thus, it may be that effects of negative emotions are found when factors are examined which are beneficial for stored memories and are based on cognitive mechanisms other than inhibition. One post-encoding factor that has been shown to be especially beneficial is the testing of memories. As revealed by numerous studies, retrieving an event from memory in a test provides a powerful boost for long-term memory of that event, a benefit that is even larger than the benefit received from repeatedly restudying the event (i.e., the testing effect; Carrier & Pashler, 1992), and which is commonly attributed to the fact that testing involves a more elaborative and deeper processing that enhances long-term retention (for a review, see Roediger & Butler, 2011). Interestingly, although the beneficial effect of testing has been examined using a large variety of materials (e.g., wordlists, vocabulary, prose texts), to our knowledge, the effect of testing memories for emotionally negative stimuli has not been examined to date.

At first glance, it may seem clear that the beneficial effect of testing should be larger for negative than for neutral stimuli. Due to preferential encoding, negative stimuli may be more likely retrieved in a test, which may lead to an increased testing effect because memories can only benefit from testing when they are successfully retrieved (Kornell et al., 2011). Furthermore, one may additionally assume that the benefit received from successful retrieval may be larger for negative than for neutral stimuli, similar to the memory advantage of negative over neutral stimuli found during initial encoding. However, a closer look suggests that such an assumption may be premature. As shown in several studies, when a visual scene contains a negative stimulus, the preferential encoding of negative stimuli leads to high storage strength for the stimulus itself. However, this benefit in stimulus-related storage strength comes at the cost of weakened associations to memories for other stimuli that were present in the same visual scene (e.g., Bisby & Burgess, 2014; Madan et al., 2012; for a review, see Mather, 2007). Regarding the beneficial effects of testing, this may have two detrimental consequences.

First, in order to be able to test one's memory about the presence of a stimulus in a past scene, other stimuli that were present in the same scene have to be used as retrieval cues. Thus, for the successful retrieval of memories in a test, high storage strength of the to-be-retrieved memory is not enough. Rather, successful retrieval additionally presupposes that the association between retrieval cues and to-be-retrieved memories is strong enough so that the memories can be accessed via the cues. Accordingly, despite the fact that negative stimuli are preferentially stored in memory, this does not necessarily mean that they are more likely retrieved in a test where memory is accessed via retrieval cues. Second, in case of successful retrieval, it may be that even if the test-induced benefit in storage strength is larger for negative than for neutral stimuli, the benefit at the level of associations may be smaller due to the attention-grabbing nature of emotional stimuli, so that overall no difference in the benefit received from testing is observed.

The aim of the present study was to investigate the effect of testing negative memories on long-term memory for those stimuli. We employed a standard cued recall testing-effect paradigm and varied the emotional significance of the to-be-studied stimuli. Participants first studied pairs consisting of a neutral or negative real-life picture (target) and a picture-independent visual cue. In a subsequent phase, for one third of the cue-target pairs, only the cues were presented, and participants were asked to retrieve the associated targets (testing condition). Another third of the cue-target pairs was completely presented, and participants were asked to restudy the pairs (restudy condition). The remaining third did not appear in this phase and served as a baseline for the benefits gained from testing and restudying (control condition).

After 1 week, memory for all cue-target pairs was tested in a delayed memory test. In order to measure both the ability to access the target pictures via the association formed during study between the target picture and the cue and the availability of the target pictures in memory independently of the cue, a two-stage memory test was used. In an initial test, a cued recall test was administered where pictures had to be retrieved via provided cues. In a subsequent free recall test, participants were asked to additionally recall all of the target pictures they remembered independently of the cues (for an illustration, see Figure 1).

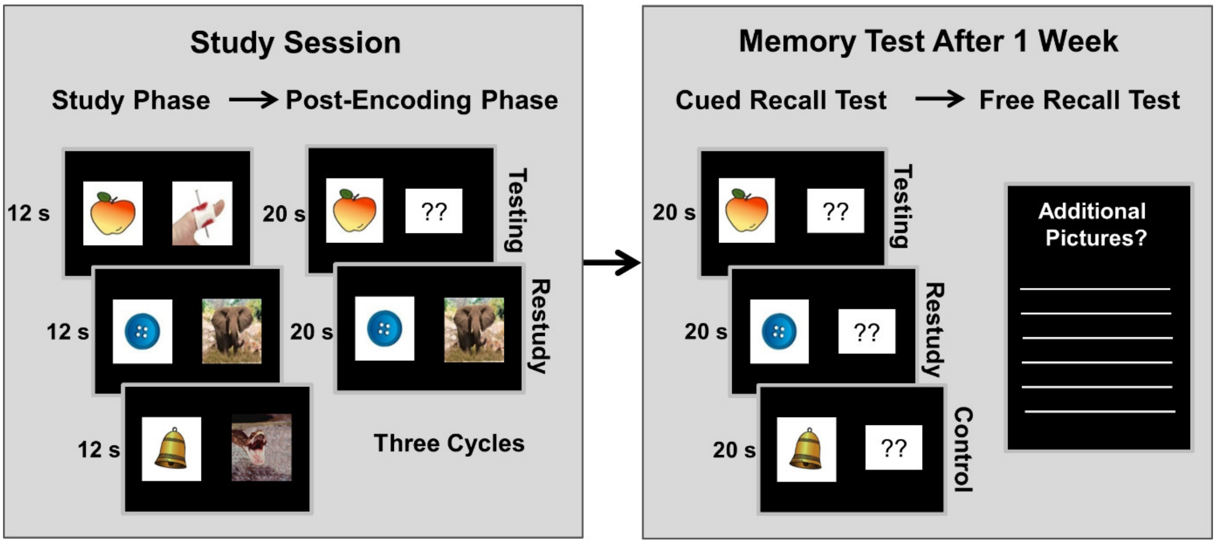

Figure 1. Procedure of the experiment. In an initial study phase, 24 negative and 24 neutral object-picture pairs were presented in random order. In a subsequent post-encoding phase, for one third of the pairs, the object was presented with the instruction to remember the associated picture (testing condition), and one third was completely presented again with the instruction to restudy the pairs (restudy condition); the remaining third did not appear in this phase (control condition). After 1 week, memory for all object-picture pairs was tested. In an initial cued recall test, all objects were presented with the instruction to recall the associated pictures. In a subsequent free recall test, participants were instructed to write down further pictures they could recall regardless of the associated object.

For neutral target stimuli, we expected to replicate the typical testing effect; that is, memory for the cue-target pairs in the delayed test after 1 week should be higher for tested than for restudied and control targets. Furthermore, we expected to replicate previous findings on memory for negative stimuli; that is, the availability of the target pictures in memory independently of the cue should be enhanced for negative compared to neutral targets, but the ability to access target pictures via the picture-independent cue should be decreased for negative compared to neutral targets due to the detrimental effect of negative emotional significance on associative memory. Thus, as a net effect, cued recall memory performance should be similar, or even decreased, for negative compared to neutral targets. As a consequence, given that the successful retrieval of a memory is a prerequisite for the occurrence of a testing benefit, we hypothesized that, despite the increased actual storage strength of negative memories, the benefit received from testing should be similar or even decreased for negative compared to neutral stimuli.

Method

Participants

A power analysis revealed that to achieve a power of .80 for detecting small to medium sized effects (dependent t-test, $d = 0.35$, $\alpha = .05$; G*Power 3.1.7; Faul, Erdfelder, Lang, & Buchner, 2007), a sample size of at least 52 would be required. Thus, we decided to recruit 60 undergraduates (39 females, 21 males; $M_{age} = 21.4$, $SD = 2.3$) who participated for course credit. All participants provided written informed consent. The study was approved by the Ethics Committee of the University of Regensburg, and conducted in accordance with the Helsinki declaration. All data exclusions, all manipulations, and all measures in the study are reported.

Material

The materials involved 48 negative and 48 neutral real-life target pictures. Each of the negative pictures was yoked with one neutral real-life picture that was visually similar but not arousing (taken from Mather & Nesmith, 2008). The resulting 48 negative/neutral picture pairs were grouped into three sets of 16 pairs each. Each set was assigned to one of three post-encoding conditions (testing, restudy, control); the assignment was counterbalanced across participants. Each participant was shown only one picture from each pair, resulting in eight negative and eight neutral target pictures in each post-encoding condition; the version that was shown was also counterbalanced across participants. The target pictures were randomly paired with 48 neutral objects selected from the revised colorful version of the Snodgrass and Vanderwart picture set (Rossion & Pourtois, 2004) that served as visual cues for the later memory tests on target picture memory. Stimuli and instructions were presented on a computer screen using E-Prime 2.0 (Psychology Software Tools, Sharpsburg, PA), and participants were provided with booklets in the post-encoding and the delayed memory test phases.

Design and Procedure

The experiment involved a study phase, a restudy/testing phase, and a delayed memory test phase that took place after 1 week. All participants were tested individually in separate cubicles. In the study phase, the 48 object-picture pairs were shown one by one for 12 s each. In order to control for order effects and prevent the induction of negative moods, the object-picture pairs were presented in random order, irrespective of emotional significance and post-encoding condition. Participants were instructed that they would see 48 object-picture pairs, and that they should memorize the pairs for a memory test 1 week later. They were told that it may be helpful to combine the objects and pictures in their imagination. It was not mentioned that only the pictures would be tested later via the objects as retrieval cues.

After the study phase, the restudy/testing phase followed. Participants were tested on one third of the studied object-picture pairs (testing condition) and restudied another third of the object-picture pairs (restudy condition). The remaining third of pairs did not appear in this phase and served as a baseline for the benefits gained from testing and restudying (control condition). The order of conditions was counterbalanced across participants. In the testing condition, the objects were shown one by one in random order, and participants were asked to recall the associated pictures. Participants did not receive any feedback. In the restudy condition, the object-picture pairs were completely presented in random order, and participants were asked to restudy the object-picture pairs. In both conditions, participants were instructed to write down a short description of the objects and pictures into their booklet within 20 s. The restudy/testing cycle was repeated three times. Afterwards, participants were thanked and asked to return to the laboratory 1 week later for a delayed memory test.

In the delayed memory test phase after 1 week, participants' memory for all target pictures was tested, irrespective of post-encoding conditions. For this purpose, all objects were presented one by one in random order. Participants were asked to recall the associated picture, and write a short description of both the object and the picture into their booklet. By pressing the space key, participants could move to the next object; after 20 s, the next object was presented automatically. Subsequent to the cued recall test, participants were asked if they could recall any further pictures without remembering the associated object, and instructed to write down a short description of these pictures without any time restrictions. Finally, all 48 previously shown target pictures were presented again in random order, and participants rated their valence (1 = extremely negative, 9 = extremely positive) and arousal (1 = low arousal, 9 = high arousal) using the affect grid (Russell, Weiss, & Mendelsohn, 1989), which was shown nearby the picture.

Scoring

All tests were scored independently by two raters. Target picture descriptions that could unambiguously be assigned to the corresponding target picture were scored as correctly remembered; missing target descriptions and vague, not interpretable descriptions were scored as not remembered. Cohen's Kappa (κ) was performed to determine consistency among raters and indicated high inter-rater reliability, $\kappa = .98$ ($p < .001$), 95% CI [0.979, 0.987].

Results

Emotional Ratings

Compared to neutral pictures, participants rated negative pictures as both more negative ($M_{\text{Neutral}} = 5.53$, $SD = 0.48$ vs. $M_{\text{Negative}} = 2.38$, $SD = 0.66$), $t(59) = 32.22$, $p < .001$, $g_{av} = 5.40$, and more arousing ($M_{\text{Neutral}} = 4.97$, $SD = 0.77$ vs. $M_{\text{Negative}} = 6.60$, $SD = 1.10$), $t(59) = -11.49$, $p < .001$, $g_{av} = 1.70$.

Immediate Memory Test

For the three immediate cued recall tests in the restudy/testing phase, a 3 (Test Number: test 1, test 2, test 3) × 2 (Emotion: negative, neutral) repeated-measure analysis of variance (ANOVA) was conducted on correct recall rates. Results showed that there were neither main effects of Test Number, Greenhouse-Geisser corrected $F(1.63, 96.16) = 0.07$, $p = .899$, $\eta_p^2 = .001$, nor of Emotion, $F(1, 59) = 2.08$, $p = .155$, $\eta_p^2 = .034$, nor a significant Test Number × Emotion interaction, Greenhouse-Geisser corrected $F(1.80, 106.33) = 0.50$, $p = .591$, $\eta_p^2 = .008$. Thus, correct recall rates stayed roughly the same throughout the three tests ($M = .77$, $SD = .20$ vs. $M = .77$, $SD = .19$ vs. $M = .77$, $SD = .19$), and did not significantly differ for negative targets compared to neutral targets ($M_{\text{Negative}} = .75$, $SD = .21$ vs. $M_{\text{Neutral}} = .79$, $SD = .22$, collapsed over the three tests).

Delayed Memory Test after 1 Week

Memory performance in the delayed memory test as a function of emotion (negative, neutral), post-encoding condition (control, restudy, testing), and type of memory test (cued recall test, free recall test) is shown in Figure 2.

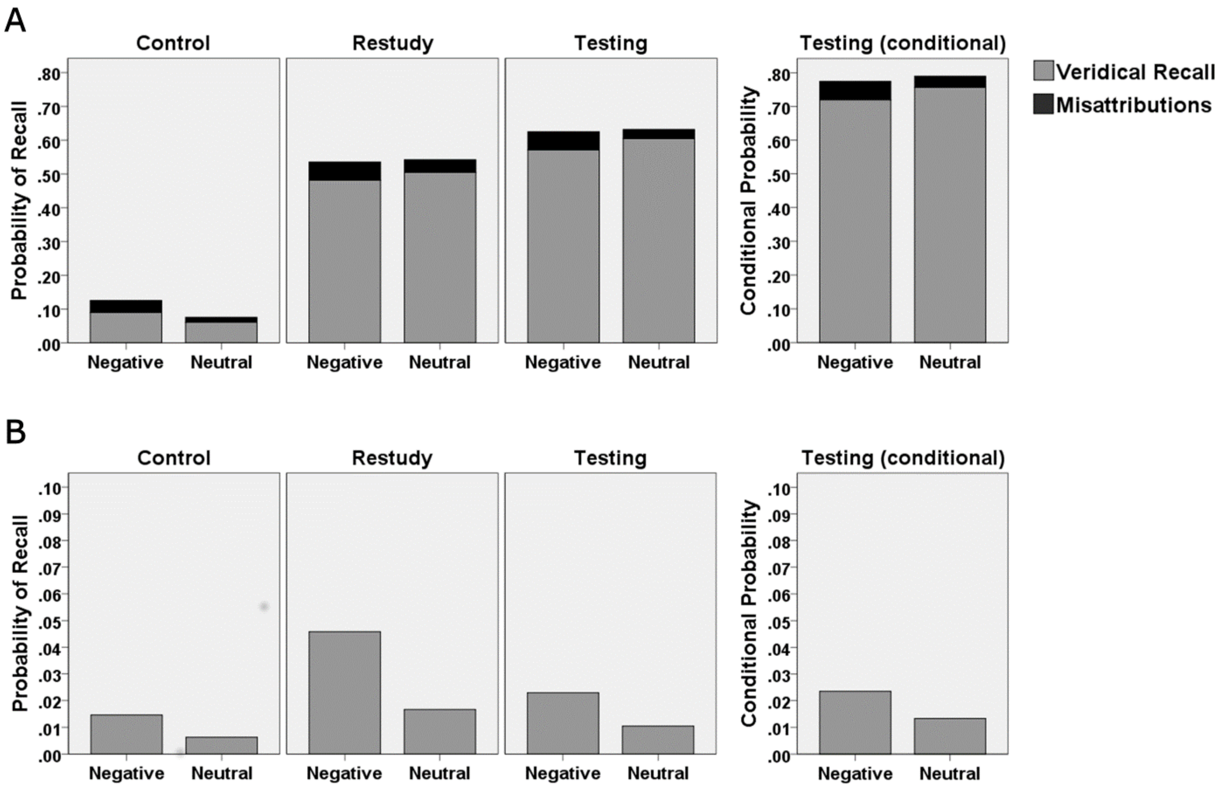

Figure 2. Memory performance after 1 week. **(A)** Performance in the cued recall test. Probability of recall of target pictures as a function of post-encoding condition (control, restudy, testing) and emotion (negative, neutral) is shown in the left panels. Conditional probabilities of recall given successful recall in the immediate test 1 week before are shown in the right panel. Grey bars show the proportion of pictures correctly recalled; black bars show the proportion of pictures misattributed to the wrong cue. **(B)** Performance in the free recall test. Probability of recall of target pictures as a function of post-encoding condition (control, restudy, testing) and emotion (negative, neutral) is shown in the left panels. Conditional probabilities of recall given successful recall in the immediate test 1 week before are shown in the right panel.

Cued recall test. First, the proportion of targets was determined that was correctly retrieved in the initial cued recall test via the provided retrieval cues (see Fig. 2A, grey bars). A 3 (Post-Encoding Condition: control, restudy, testing) × 2 (Emotion: negative, neutral) repeated-measure ANOVA showed a significant main effect of Post-Encoding Condition, $F(2, 118) = 219.06$, $p < .001$, $\eta_p^2 = .788$. The proportion of targets that could be accessed via the retrieval cue was higher for restudied targets than for control targets ($M_{\text{Restudy}} = .49$, $SD = .24$ vs. $M_{\text{Control}} = .08$, $SD = .10$), $t(59) = 14.91$ $p < .001$, $g_{av} = 2.23$, and higher for tested targets

($M_{\text{Testing}} = .59$, $SD = .22$) than for both control targets, $t(59) = 18.88$, $p < .001$, $g_{av} = 2.94$, and restudied targets, $t(59) = 4.18$, $p < .001$, $g_{av} = 0.41$, thus replicating the typical testing effect. However, no main effect was found for Emotion ($M_{\text{Neutral}} = .39$, $SD = .16$ vs. $M_{\text{Negative}} = .38$, $SD = .18$, collapsed over post-encoding conditions), $F(1, 59) = 0.40$, $p = .523$, $\eta_p^2 = .007$, neither was there a significant Post-Encoding Condition × Emotion interaction, Greenhouse-Geisser corrected $F(1.67, 98.68) = 1.62$, $p = .202$, $\eta_p^2 = .027$.

In addition to veridical recall in the cued recall test, we also examined misattributions (i.e., target pictures that were falsely retrieved in answer to a retrieval cue). The proportion of misattributions is shown by the black bars in Figure 2A. A 2 (Emotion: negative, neutral) × 3 (Post-Encoding Condition: control, restudy, testing) repeated-measure ANOVA revealed a significant main effect of Emotion, $F(1, 59) = 9.74$, $p = .003$, $\eta_p^2 = .142$. Collapsed across all post-encoding conditions, more negative than neutral target pictures were misattributed ($M_{\text{Negative}} = .05$, $SD = .05$ vs. $M_{\text{Neutral}} = .03$, $SD = .03$). There was no significant main effect of Post-Encoding Condition, $F(2, 118) = 2.72$, $p = .070$, $\eta_p^2 = .044$, and no significant Post-Encoding condition × Emotion interaction, $F(2, 118) = 0.18$, $p = .840$, $\eta_p^2 = .003$.

Free recall test. Memory performance in the subsequent free recall test for the target pictures is shown in Figure 2B. A repeated-measure ANOVA revealed a significant main effect of Emotion, $F(1, 59) = 6.41$, $p = .014$, $\eta_p^2 = .098$. Collapsed across all post-encoding conditions, more negative than neutral target pictures were additionally remembered independently of the cues ($M_{\text{Negative}} = .03$, $SD = .05$ vs. $M_{\text{Neutral}} = .01$, $SD = .03$). There was also a significant main effect of Post-Encoding Condition, Greenhouse-Geisser corrected $F(1.70, 100.07) = 5.20$, $p = .010$, $\eta_p^2 = .081$. Free recall rates were significantly higher for restudied targets ($M_{\text{Restudy}} = .03$, $SD = .05$) than for control targets ($M_{\text{Control}} = .01$, $SD = .03$), $t(59) = 2.77$, $p = .008$, $g_{av} = 0.48$, and tested targets ($M_{\text{Testing}} = .02$, $SD = .04$), $t(59) = 2.08$, $p = .042$, $g_{av} = 0.30$. Free recall rates did not significantly differ between tested targets and control targets, $t(59) = 1.23$, $p = .224$, $g_{av} = 0.19$. There was no Post-Encoding Condition × Emotion interaction, Greenhouse-Geisser corrected $F(1.60, 94.31) = 1.43$, $p = .243$, $\eta_p^2 = .024$.

Conditional memory performance. Finally, we restricted the analysis to target pictures that were successfully retrieved in the immediate test; that is, we determined the conditional probabilities of recall in the delayed memory test given successful recall in the immediate test (see Fig. 2, right bars). Conditional probabilities of veridical recall in the cued recall test did not significantly differ for neutral targets compared to negative targets ($M_{\text{Neutral}} = .76$, $SD =$

.20 vs. M_{Negative} = .72, *SD* = .23), *t*(59) = 1.46, *p* = .150, g_{av} = 0.17, neither did conditional probabilities of misattributions in the cued recall test (M_{Neutral} = .03, *SD* = .08 vs. M_{Negative} = .06, *SD* = .11), *t*(59) = −1.28, *p* = .207, g_{av} = 0.22, nor conditional probabilities of veridical recall in the subsequent free recall test (M_{Neutral} = .01 *SD* = .05 vs. M_{Negative} = .02, *SD* = .08), *t*(59) = −1.05, *p* = .300, g_{av} = 0.16.

Discussion

Basically, the results of the present study replicate two well-documented memory phenomena. For neutral target pictures, a typical testing effect pattern was observed: Long-term memory for the cue-picture pairs was higher for initially tested cue-picture pairs than for initially restudied cue-picture pairs (for a review, see Roediger & Butler, 2011). For negative target pictures, a typical emotional memory enhancement pattern was observed: The availability of the target pictures in memory independently of the cue was enhanced for negative compared to neutral targets, as evident both in the increased number of negative target pictures that were remembered in the cue-independent free recall and the increased number of negative target pictures that were falsely retrieved in answer to a retrieval cue in the cued recall test (misattributions); however, this boost in storage strength for negative target pictures did not transfer to the ability to retrieve stored memories via retrieval cues, as indicated by equal veridical recall rates in the cued recall test (for a review, see Mather, 2007).

Interestingly, although more negative than neutral targets were stored in memory, negative and neutral targets did not differ in the effects of testing and restudying. Testing and restudying increased long-term memory to exactly the same extent for both negative and neutral targets, leading to equal testing effects. At first glance, the absence of any effects of negative emotion on the testing effect seems to be simply attributable to the fact that emotional significance did not influence the ability to retrieve targets via retrieval cues so that equal numbers of negative and neutral memories benefited from successful retrieval. However, still, one may have expected that the benefit received from successful retrieval for later memory is higher for negative than for neutral targets, similar to the emotional effects observed during initial encoding. However, the results showed that even the benefit received from successful retrieval was equal for negative and neutral targets, as revealed by equal conditional probabilities of correct recall in the delayed recall test given successful recall in the immediate test.

The lack of influences of negative emotions on the effects of testing or restudying previously stored items is in line with previous findings, suggesting that the emotional memory enhancement effect is limited to initial encoding, while post-encoding factors seem to affect negative memories in no other ways than non-emotional memories (e.g. Kuhbandner et al., 2009; Sakaki et al., 2014; Wessel & Merckelbach, 2006). Going beyond previous studies, the present findings demonstrate that this holds not only for detrimental post-encoding factors that induce forgetting of memories, but also for post-encoding factors that enhance future remembering.

The present study also raises several questions that should be addressed in future research. First, both in our study on beneficial post-encoding factors and in previous studies on detrimental post-encoding factors, emotionally negative materials have been used (Kuhbandner et al., 2009; Sakaki et al., 2014; Wessel & Merckelbach, 2006). Thus, the question arises whether the observed lack of emotional effects generalizes to emotionally positive materials. Second, in the present study, non-emotional cues have been used to access memories that were unrelated to the depicted target picture. While this seems rather typical for real-life contexts such as witness interviewing where emotional memories are accessed via other stimuli that were present in the scene, the situation may be different when cues are used that are part of the emotional target itself. In such a case, emotions have been demonstrated to boost not only target memory, but also the associations to the cue (e.g., Mather, 2007). Thus, under such conditions, an increased testing effect for negative stimuli may occur.

Conclusion

The emotional significance of stimuli has been shown to impact all stages of information processing from perception to memory. However, such emotional effects seem to be restricted to the phase of initial encoding of stimuli in memory. As revealed by the present study, and consistent with previous studies on detrimental post-encoding factors, factors that are known to influence memories *after* the initial encoding took place seem not to be modulated by the emotional significance of stimuli. While replicating the testing effect as well as the emotional enhancement effect, we did not find any difference between negative and neutral memories in the benefit received from testing and restudying. Hence, post-encoding factors seem not to affect negative memories in any other way than neutral memories.

Funding

This work was supported by the German Research Foundation (DFG) [Deutsche Forschungsgemeinschaft] under grant number KU 3325/2-1 to Christof Kuhbandner.

STUDY 2

Testing Memories of Personally Experienced Events: The Testing Effect Seems Not to Persist in Autobiographical Memory

Kathrin J. Emmerdinger and Christof Kuhbandner

Abstract

Numerous studies have shown that retrieving contents from memory in a test improves long-term retention for those contents, even when compared to restudying (i.e., the testing effect). The beneficial effect of retrieval practice has been demonstrated for many different types of memory representations; however, one particularly important memory system has not been addressed in previous testing-effect research: autobiographical memory. The aim of the present study was to examine the effect of retrieving memories for personally experienced events on long-term memory for those events. In an initial elicitation session, participants described memories for personally experienced events in response to a variety of cue words. In a retrieval practice/restudy session the following day, they repeatedly practiced retrieval for half of their memories by recalling and writing down the previously described events; the other half of memories was restudied by rereading and copying the event descriptions. Long-term retention of all previously collected memories was assessed at two different retention intervals (2 weeks and 13 weeks). In the retrieval practice session, a hypermnesic effect emerged, with memory performance increasing across the practice cycles. Long-term memory performance significantly dropped from the 2-week to the 13-week retention interval, but no significant difference in memory performance was observed between previously repeatedly retrieved and previously repeatedly restudied memories. Thus, in autobiographical memory, retrieval practice seems to be no more beneficial for long-term retention than repeated re-exposure.

Numerous studies have shown that retrieving contents from memory in a test considerably improves long-term memory for those contents, even when compared to a condition where the contents are represented for restudying (i.e., the "testing effect," Carrier & Pashler, 1992; for recent meta-analyses see Adesope et al., 2017; Rowland, 2014). The beneficial effect of retrieval practice has been demonstrated for a wide range of test formats (e.g., cued recall, free recall, recognition memory), for a large variety of study materials (e.g., wordlists, vocabulary, prose texts), and even extends to procedural skills (Kromann et al., 2009) and emotional memories (Emmerdinger, Kuhbandner, & Berchtold, 2018). However, previous research in the field has neglected one particularly important memory system: autobiographical memory. While a few studies have examined the effect of retrieving autobiographical events from memory compared to non-retrieved autobiographical events (Barnier et al., 2004; Stone, Barnier, Sutton, & Hirst, 2013), to our knowledge, no study to date has examined the effect of testing autobiographical memories compared to repeated re-exposure to those memories.

Autobiographical memories, the recollections of personally experienced events, are a unique type of memory representations characterized by a high degree of complexity and a strong interconnectedness in an associative network. Autobiographical memories are thought to be hierarchically organized, such that detailed recollections of specific events are embedded in a rich context of more abstract knowledge about the personal past (Conway & Pleydell-Pearce, 2000). Importantly, autobiographical memories are also highly linked to the self, thus interrelated with personal motives and evaluations, and often emotionally significant (Conway, 2005).

It is an open question if the retrieval benefits reliably observed for non-personal learning materials persist for personal, self-related information. Generally, it has been shown that previously retrieved autobiographical events are better retained than non-retrieved autobiographical events (Barnier et al., 2004; Stone et al., 2013). However, less is known regarding the relative significance of retrieving personally experienced events compared to re-exposure to those events, as may occur for example when listening to rehearsals of socially shared personally experienced events in conversations. Interestingly, there is some evidence that retrieving socially shared personally experienced events in the role of the speaker benefits memory retention similarly to being re-exposed to those events in the role of the listener (Stone et al., 2013). This finding may hint at the possibility that retrieval practice may not necessarily benefit long-term memory for personally experienced events more than repeated re-exposure to those events.

Indeed, based on theoretical accounts of why retrieving information in a test benefits long-term memory, one may speculate that the testing effect may be less pronounced or even disappear for autobiographical memories of personally experienced events. Existing explanations for the testing effect rest upon the assumption that the successful retrieval of a memory initiates elaborative processes that update and strengthen the memory trace through the establishment of new relations (Carpenter, 2009, 2011; Kornell et al., 2015; Pyc & Rawson, 2010). However, this benefit received from testing may be reduced or even absent for autobiographical memories, which typically exhibit inherently strong links in an associative network. In fact, recent studies failed to replicate the testing effect for complex, highly associated materials (De Jonge et al., 2015; Van Gog et al., 2015; for a review, see Van Gog & Sweller, 2015). For example, in the study of De Jonge et al. (2015), while taking a test on a previously studied strongly interrelated text did not benefit memory any more than restudying the text, the typical retrieval benefit did emerge when the same text was scrambled and presented for studying and retrieval practice as single, nonrelated facts. On the other hand, a simple re-exposure to the material, as typically realized in the restudy condition, may benefit memory for personally relevant information more than memory for non-personal information. Numerous studies have shown increased memory performance for information that is relevant to the self (for a review, see Symons & Johnson, 1997). Due to the highly associatively organized and self-related nature of autobiographical memory representations, repeated re-exposure alone may lead to similar memory benefits than actively retrieving the information, and in consequence, attenuate the typical pattern of the testing effect. Following a distribution-based perspective (Kornell et al., 2011), the testing effect may thus emerge, if at all, only after comparably long retention intervals, because not only previously successfully retrieved, but also restudied memory representations would stay rather long well above the recall threshold.

In the current study, we extended the testing-effect paradigm to autobiographical memories. For this purpose, we adapted an experimental procedure developed by Barnier et al. (2004) using autobiographical memories for personally experienced events elicited in response to cue words. The present experiment consisted of four sessions: An initial collection session, a retrieval practice/restudy session the following day, and two delayed memory test sessions at different retention intervals (2 weeks and 13 weeks). In the collection session, participants were asked to describe events they had personally experienced in the last 6 months in response to a variety of cue words. To control for possible effects of emotional significance, one third of the cue words was emotionally negative, one third neutral, and the remaining third positive. The retrieval practice/restudy session took place 1 day after the collection session. Each participant was provided with the descriptions of half of his or her personal events of each emotional quality together with the corresponding cue for restudy. Participants were asked to carefully read the cue word and the event description and copy the event description by hand. For the other half of the events, only the cues were presented, and participants were asked to retrieve and write down the corresponding memories. After 2 weeks, we assessed delayed memory performance for all originally collected autobiographical events; participants were asked to recall and write down all originally described events in response to the corresponding cue words. To control for the possibility that for autobiographical material the benefit of retrieval practice might only emerge after a comparably long retention interval, the delayed memory test was repeated 13 weeks after the retrieval practice/restudy session.

Method

Participants

A power analysis was performed for sample size estimation and revealed that to achieve a power of .80 for detecting small to medium sized effects ($d = 0.40$, $\alpha = .05$; G*Power 3.1.7; Faul et al., 2007), a sample size of at least 41 would be required. Thus, we decided to recruit 48 undergraduate students (45 females, $M_{\text{Age}} = 21.3$, $SD = 5.5$) who participated for course credit. The study was approved by the Ethics Committee of the University of Regensburg. All participants provided written informed consent in accordance with the Declaration of Helsinki.

Material

In an initial collection session, participants were asked to describe personally experienced events from the last 6 months in response to cue words. For this purpose, three cue word lists were constructed that contained either 36 neutral (e.g., shoe, table, buy), 36 emotionally positive (e.g., entertaining, friendship, happiness), or 36 emotionally negative (e.g., sick, quarrel, lonely) German words (for the complete cue word lists in German and their English translations, see Supplementary Material, Table 4). Cue words were selected from previous studies using the cue word method for the collection of autobiographical memories (Barnier et al., 2004, 2007; Kuyken & Dalgleish, 2011; Maccallum, McConkey, Bryant, & Barnier, 2000; Robinson, 1976), and translated into German, and from the Berlin Affective Words List Reloaded (Võ et al., 2009), a list of affective German words.

We slightly modified the collection procedure applied by Barnier et al. (2004), who had instructed participants to describe events for each of the presented cue words, and allowed participants to choose individually from a broader selection of cue words in order to assure that participants were able to describe a sufficient sample of authentic autobiographical events. Participants received the neutral, positive, and negative cue word lists and a booklet to write down their autobiographical events and to rate them according to several characteristics (see below). For each cue word list, participants were asked to describe a specific event they had personally experienced within the last 6 months in response to 16 cue words, and to write down a description of the event in one or two sentences. Participants were told not to describe routines (e.g., “Thursdays I usually go swimming”) but specific events. They were instructed to think of non-emotionally significant events in response to the neutral cue words, and of emotionally positive or emotionally negative events in response to the emotionally positive or negative cue words. To prevent ceiling effects due to divergent cognitive processing of extremely emotionally charged events (e.g., Brown & Kulik, 1977; Wagenaar & Groeneweg, 1990; for a review, see Christianson & Safer, 1996), for emotionally positive and negative cue words, participants were asked to think of medium intense emotionally positive and negative events in their everyday life, rather than of profound, drastic events.

Participants worked successively on all three cue word lists; the order in which the neutral, positive, and negative cue word lists were provided was counterbalanced across participants. Immediately after having written down the description of an event, participants rated the event on 7-point-scales according to its clarity (“How clear is your memory of the event?”; 1 = not clear at all, 7 = very clear), personal relevance (“How personally relevant/significant is this memory for you?”; 1 = not relevant at all, 7 = very relevant), emotional valence (“How positive or negative is this memory for you?”; 1 = very negative, 7 = very positive), emotional arousal (“How emotionally arousing is this memory for you?”; 1 = not at all emotionally arousing, 7 = very emotionally arousing), and the frequency with which they had previously thought of this event or told others about it (“How often did you think about this memory or told others about it?”; 1 = not at all, 7 = very often). Table 1 shows the characteristics of the collected sample of autobiographical events (for a supplemental analysis of ratings of memory characteristics as a function of emotion condition and assigned practice condition, see Supplementary Material).

Table 1

Mean Ratings for the Characteristics of the Sample of Autobiographical Memories Retrieved by Participants in Response to Neutral, Positive, and Negative Cue Words.

	Emotional quality of cue word					
	Neutral		Positive		Negative	
Ratings of memory characteristics	*M*	*SD*	*M*	*SD*	*M*	*SD*
Clarity	5.07	0.98	5.98	0.58	5.76	0.48
Personal relevance	2.51	0.91	4.83	0.88	4.43	0.92
Emotional valence	4.07	0.25	6.01	0.47	2.25	0.40
Emotional arousal	2.23	1.02	4.71	0.83	4.80	0.83
Frequency of previous retrieval	1.99	0.66	3.99	0.91	3.86	0.90

Note. Participants rated their memories in terms of clarity (1 = not clear at all, 7 = very clear), personal relevance (1 = not relevant at all, 7 = very relevant), emotional valence (1 = very negative, 7 = very positive), emotional arousal (1 = not at all emotionally arousing, 7 = very emotionally arousing), and the frequency with which they had previously thought of this event or told others about it (1 = not at all, 7 = very often).

Design and Procedure

Figure 3 depicts the procedure of the experiment. The experiment consisted of four sessions. An initial collection session, a retrieval practice/restudy session the following day, and two delayed memory test sessions that took place at different retention intervals (2 weeks and 13 weeks after the retrieval practice/restudy session). The first three sessions took place in the laboratory and the second delayed memory test after 13 weeks was conducted via an online survey tool (SoSci Survey; Leiner, 2014) where participants accessed their individualized tests with a code.

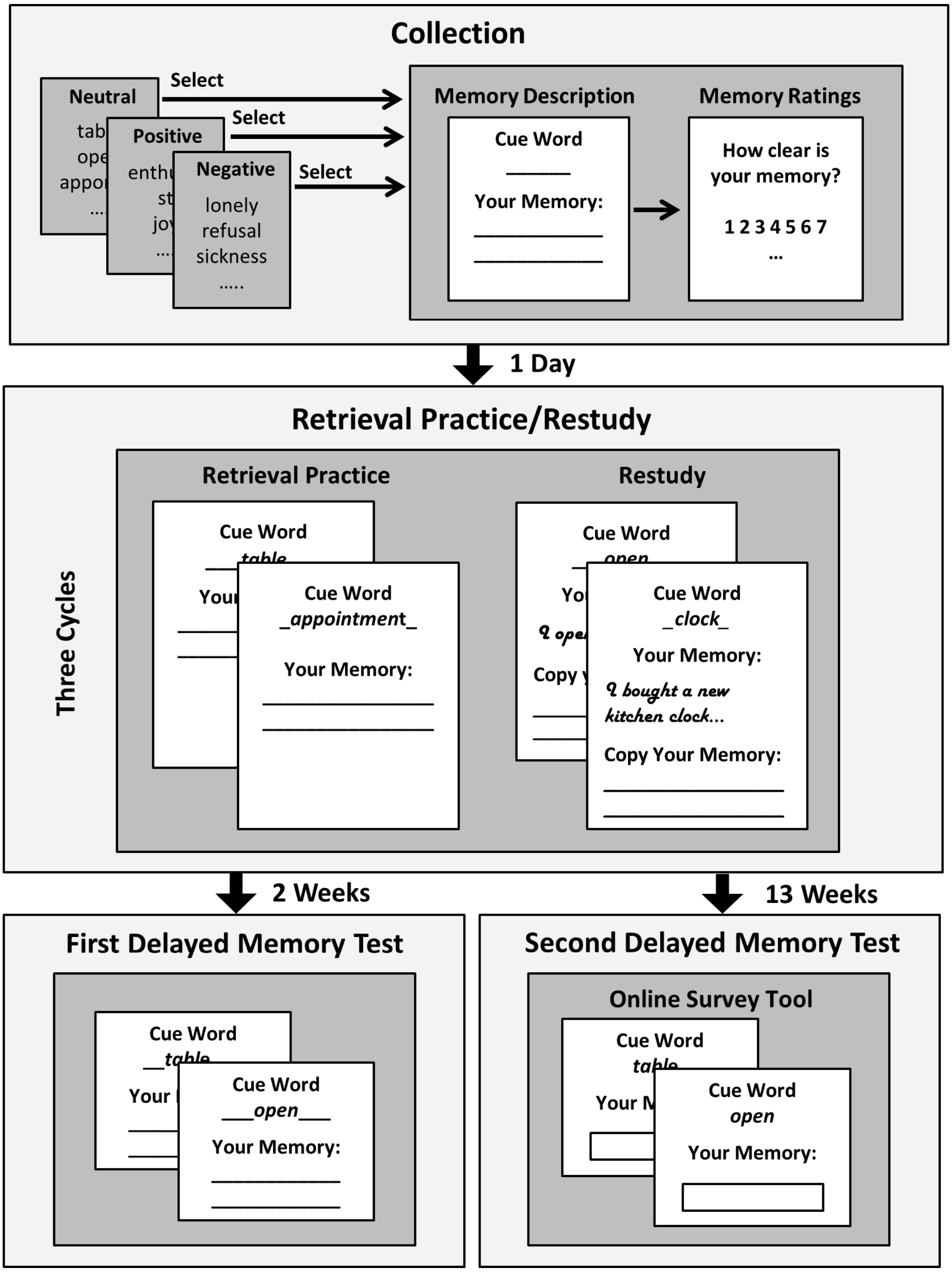

Figure 3. Procedure of the experiment. In an initial collection session, participants described 48 autobiographical events (16 neutral, 16 positive and 16 negative) in response to cue words. In a retrieval practice/restudy session the following day, for half of the collected events, participants received only the cue words, and they were asked to remember and write down the corresponding event (retrieval practice); for the other half of the events, the cue words were presented together with the collected descriptions, and they were asked to restudy and copy the descriptions (restudy). Overall, participants completed three retrieval practice/restudy cycles. In two delayed memory test sessions (2 weeks and 13 weeks after the retrieval practice/restudy session), memory for all 48 originally collected autobiographical events were tested (for details, see Method section).

Collection session. At the beginning of the initial collection session, participants were told that they were taking part in an experiment investigating how people remember and cope with autobiographical memories of emotional and non-emotional events, and that they would be asked to describe personally experienced events in the present session, and work in different ways on their memories about these events in the following sessions. They were informed that the experimenter would have no access to the booklets containing their event descriptions, but that independent research assistants would collect and code their anonymized booklets, thus ensuring that none of the collected events would be assignable to any individual person. Participants were instructed that they would subsequently work on three cue word lists and that, for each list, their task consisted in sequentially selecting a total of 16 cue words and describe personally experienced events from the last 6 months in response to them. They were told that there was a list with neutral, one with emotionally positive, and one with emotionally negative cue words, and that each participant would start with another list.

Participants then received a booklet on which they noted a personal code. Each page of the booklet contained a space designated to write down the cue word and a short description of the corresponding event in one or two sentences, and a series of 7-point scales designated to rate the event according to its clarity, personal significance, valence, emotional arousal, and the frequency with which participants had previously thought about the event or told it to others. Participants were asked to think of events that they had personally experienced during the last 6 months, and that had lasted seconds, minutes or even hours, but not several days. They were instructed to describe unique, specific events, and not regularly recurring events or personal routines (for example, a description of a specific event would be "Last Monday when I went to the gym, I met Thomas and had a short conversation about his new apartment with him," but not "I usually go to the gym on Monday"). Finally, they were instructed to think of a new event for each selected cue word, thus collecting a total of 16 different autobiographical events for each list. For the neutral cue words, participants were asked to think of events that did not have any emotional significance to them. For the emotionally positive and negative cue words, they were instructed to think of emotionally positive or negative events of medium intensity experienced in their daily life (thus, their daily hassles and daily uplifts), rather than of profound, intensely emotional events (as for example marriage or death of a significant other). To ensure that event descriptions would not largely vary in detail and length across events and participants, following previous research using the cue word method for the collection of autobiographical memories (Barnier et al., 2004, 2007), participants were instructed to write down a relatively brief description of one to two sentences for each event.

Depending on counterbalancing group, the experimenter first handed the list containing the neutral, emotionally positive, or emotionally negative cue words to the participants. There were no time restrictions, and the session ended when participants had recalled 16 memories for each of the three lists. It took participants between 90 and 120 min to complete the collection session. Before leaving the laboratory, participants inserted their completed booklets through a slot into a locked box. Afterwards, the booklets were collected by two independent research assistants who prepared the individualized booklets for the retrieval practice/restudy session and the first delayed memory test session.

Retrieval practice/restudy session. The retrieval practice/restudy session took place 1 day after the collection session. At the beginning of the session, participants received via their personal code individualized booklets that were prepared before the session. For each participant, the completed booklet of the collection session was scanned, and the cue words and corresponding event descriptions of each page were cut out and stored as image files. The events were assigned alternating one by one to the retrieval practice condition or to the restudy condition, following the order in which they had been described in the collection session; the assignment of the first event was counterbalanced across participants. In this way, for each participant, half of the events of each emotional quality were retrieval practiced, and the other half restudied. For retrieval practiced events, only the image of the cue word was pasted on a single page; for restudied memories, the image of the cue word and the image of the corresponding event description were pasted on a single page. The order of retrieval practice/restudy was blocked by emotional quality, following the same order as in the collection phase. Within each emotional quality block, events were presented blocked for retrieval practice or restudy; the order of retrieval practice/restudy was counterbalanced across participants. Overall, participants completed three retrieval practice/restudy cycles, with a short break of 40 s between cycles.

Participants were instructed that for half of the events they had collected the previous day, only the cue word would be presented in the booklet, and that in this case their task would be to remember the corresponding event and describe it in the designated space beneath the cue word. For the other half of the memories, the cue word would be presented together with the corresponding event description, and in this case, their task would be to carefully read the cue word and the event description, and to copy the event description in the designated space beneath. A time slot of 40 s was allotted for restudying or retrieving one event. When time was up, participants were notified by an acoustic signal to turn the page in their booklet and continue working on the next event. The total duration of the retrieval practice/restudy session was around 100 min; at the end of the session, participants again inserted their completed booklets into the locked box.

First delayed memory test session. Two weeks after the retrieval practice/restudy session, participants returned to the laboratory for the first delayed memory test session. At the beginning of the session, they received their individualized test booklet via their personal code. The test booklets contained, for all 48 originally collected autobiographical events, a page showing only the corresponding cue word; the order followed the order in which the events had been described in the collection session. Participants were instructed to remember the autobiographical event they had described in response to the cue word in the first session of the experiment, and to write a description of the events in the designated space beneath the cue word. For each event, a time slot of 40 s was allotted, after which a signal tone notified participants to turn the page of the booklet and try to remember the next event. After finishing the test, participants answered a questionnaire unrelated to the study question. Then, they inserted their completed test booklets into the locked box.

Second delayed memory test session. Thirteen weeks after the retrieval practice/restudy session, participants received an email with a link to the second delayed memory test that was performed within an online survey environment (SoSci Survey; Leiner, 2014). Participants could access their individualized memory test via their personal code. The cue words participants had originally selected in the collection phase were presented one by one on a single page each, following the order in which the events had been described in the collection session. Participants were prompted to remember the autobiographical event they had described in response to the cue word in the first session of the experiment and type it in the designated space beneath the cue word; they were asked to write "no memory" if they could not recall the corresponding event. There was no time restriction; however, to ensure that participants would really try to recover their memories rather than only click through the test, they had to stay at the same page for at least 15 s until a button appeared through which they could progress to the next cue word. The whole test had to be completed within one session that took participants on average 27 min. After the memory test, participants answered a questionnaire unrelated to the study question.

Twenty-nine participants undertook the memory test on the same day that they had received the invitation email or on the following day, 18 participants undertook the memory test within 1 week after receiving the invitation email, and one participant did not take part in the delayed memory test. Thus, the participants' individual retention intervals between retrieval practice/restudy and the second delayed memory test lay between 13 and maximal 14 weeks.

Scoring

Memories were scored as correctly remembered if the event description at recall corresponded to the event description given for the corresponding cue word in the collection session. Following the scoring procedure applied by Barnier et al. (2004, 2007) exact correspondences were not required, but there had to be a clear relation between the descriptions of the events described for the same cue word at recall and in the collection session, that is, they had to contain at least some of the same information and unambiguously refer to the same event. For instance, if the original description was “my boyfriend and I were drinking wine in front of the Eiffel tower,” then “wine under Eiffel tower” was scored as correctly, but “trip to Paris” not, because the latter description does not unambiguously refer to the same event. Similarly, if the original description was “a few weeks ago I went running and discovered a nice little park,” then “I discovered a nice park” was scored as correctly, but “a sunny day and going for a walk” not, because the latter description does not clearly relate to the original description.

Two raters independently scored all recall protocols of the retrieval practice and delayed memory test sessions. Cohen’s Kappa (κ) was performed to determine consistency among raters and indicated high inter-rater reliability, $\kappa = .94$ ($p < .001$), 95% CI [0.93, 0.95]. Finally, any discrepancies between raters were solved by a third rater.

Results

Memory performance in the retrieval practice/restudy session (cycle 1, cycle 2, cycle 3) and in the delayed memory tests (2 weeks, 13 weeks) as a function of the type of previous practice (retrieval practice, restudy) is shown in Figure 4 (for mean recall rates among emotion conditions, see Supplementary Material, Table 3).

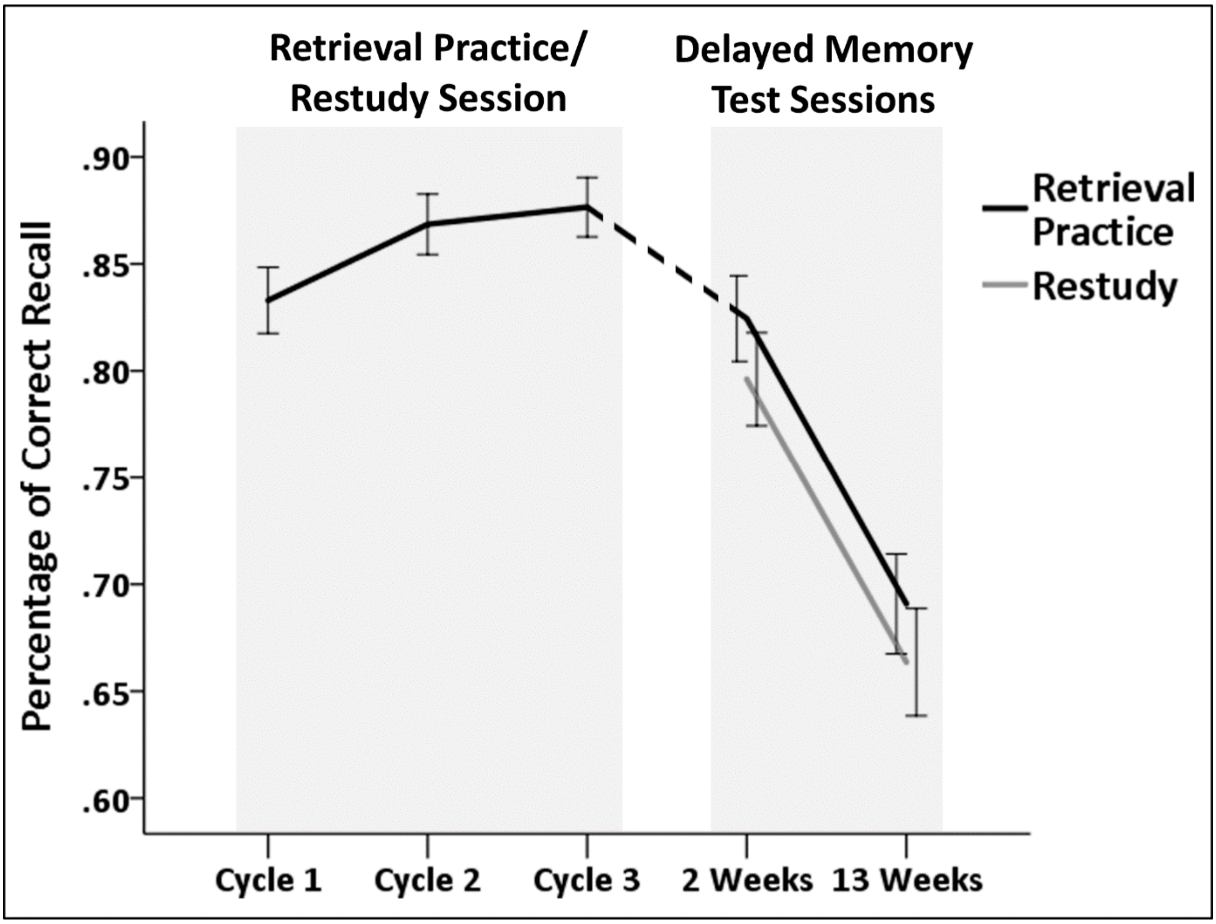

Figure 4. Percentage of correct recall of autobiographical events in the retrieval practice/ restudy session (cycle 1, cycle 2, and cycle 3) and in the delayed memory test sessions (2 weeks, 13 weeks) as a function of the type of previous practice (retrieval practice, restudy). Error bars represent standard errors of the means.

Retrieval practice. For the retrieval practice/restudy session, a 3 (Retrieval Practice Cycle: cycle 1, cycle 2, cycle 3) × 3 (Emotion: neutral, positive, negative) analysis of variance (ANOVA) was performed on the recall rates obtained in the three consecutive retrieval practice cycles. The analysis showed a significant main effect of Retrieval Practice Cycle, Greenhouse-Geisser corrected $F(1.39, 65.51) = 28.25$, $p < .001$, $\eta_p^2 = .375$, but no significant Retrieval Practice Cycle by Emotion interaction, Greenhouse-Geisser corrected $F(3.05, 143.36) = 0.96$, $p = .413$, $\eta_p^2 = .020$, indicating that correct recall rates, independently of emotion condition, increased throughout the three successive cycles. Pairwise comparisons, collapsed over emotion conditions, showed a significant increase in memory performance from the first ($M_{\text{Cycle 1}} = .83$, $SD = .11$) to the second retrieval practice cycle ($M_{\text{Cycle 2}} = .87$, $SD = .10$), $F(1,47) = 29.89$, $p < .001$, $\eta_p^2 = .389$, and a less substantial, but significant increase from the second to the third retrieval practice cycle ($M_{\text{Cycle 3}} = .88$, $SD = .10$), $F(1,47) = 4.74$, $p = .035$, $\eta_p^2 = .092$. There was also a significant main effect of Emotion, $F(2, 94) = 3.80$, $p = .026$, $\eta_p^2 = .075$. Pairwise comparisons, collapsed over retrieval practice cycles, revealed that this main effect was driven by a significant difference between recall rates for neutral ($M_{\text{Neutral}} = .89$, $SD = .13$) and positive memories ($M_{\text{Positive}} = .83$, $SD = .15$), $F(1,47) = 7.13$, $p = .010$, $\eta_p^2 = .132$. There were no significant differences between recall rates for negative ($M_{\text{Negative}} = .86$, $SD = .13$) and neutral memories, $F(1,47) = 2.60$, $p = .114$, $\eta_p^2 = .052$, nor between recall rates for negative and positive memories, $F(1,47) = 1.45$, $p = .235$, $\eta_p^2 = .030$.

Delayed memory tests. One participant dropped out before the second delayed memory test session. Therefore, all analyses of correct recall rates in the delayed memory tests are based on the remaining sample of 47 participants (45 females, $M_{\text{Age}} = 21.3$, $SD = 5.6$). A 2 (Retention Interval: 2 weeks, 13 weeks) × 2 (Type of Practice: retrieval practice, restudy) × 3 (Emotion: neutral, positive, negative) repeated measure ANOVA revealed a significant main effect of Retention Interval, $F(1, 46) = 118.53$, $p < .001$, $\eta_p^2 = .720$, indicating a significant decrease in correct recall rates after a retention interval of 13 weeks compared to a retention interval of 2 weeks ($M_{\text{2 weeks}} = .82$, $SD = .12$ vs. $M_{\text{13 weeks}} = .68$, $SD = .15$, collapsed over emotion and practice conditions). There was no main effect of Emotion, $F(2, 92) = 2.60$, $p = .080$, $\eta_p^2 = .054$, nor a significant Emotion by Retention Interval interaction, $F(2, 92) = 2.41$, $p = .095$, $\eta_p^2 = .050$, indicating that correct recall rates did not significantly differ for neutral, positive and negative memories. The analysis showed no significant main effect for Type of Practice, $F(1, 46) = 2.53$, $p = .118$, $\eta_p^2 = .052$, nor a significant Type of Practice by Retention Interval interaction, $F(1, 46) = 0.02$, $p = .904$, $\eta_p^2 < .001$. Hence, recall rates for repeatedly retrieved memories did not significantly differ from recall rates for repeatedly restudied memories, neither after a retention interval of 2 weeks ($M_{\text{Retrieval Practice}} = .82$, $SD = .14$ vs. $M_{\text{Restudy}} = .80$, $SD = .15$, collapsed over emotion conditions), nor after a retention interval of 13 weeks ($M_{\text{Retrieval Practice}} = .69$, $SD = .16$ vs. $M_{\text{Restudy}} = .66$, $SD = .17$, collapsed over emotion conditions) (see Fig. 4). There was also no significant Type of Practice by Emotion interaction, $F(2, 92) = 0.85$, $p = .431$, $\eta_p^2 = .018$, nor a significant three-way interaction between Retention Interval, Type of Practice and Emotion, $F(2, 92) = 2.49$, $p = .089$, $\eta_p^2 = .051$.

Discussion

The present research addressed the question whether the testing effect, that is, the finding that active retrieval from memory benefits long-term retention more than repeatedly restudying, extends to autobiographical memory. For this purpose, participants repeatedly retrieved or repeatedly restudied personally experienced autobiographical events that they had experienced within the last 6 months. In a delayed memory test 2 weeks after the retrieval practice/restudy session, no significant difference emerged between recall rates for previously repeatedly retrieved and previously repeatedly restudied autobiographical events. Importantly, this pattern persisted even after a long retention interval of about three months; while overall recall performance significantly decreased from the medium to the long retention interval, this decline was not any different for previously retrieved than for previously restudied events, thus ruling out the possibility that the benefit of retrieval practice in autobiographical memory might only emerge after comparably long retention intervals (Kornell et al., 2011). Thus, it seems that the typical testing effect commonly found for non-personal information does not extend to autobiographical memory representations of personally experienced events.

There are several possible explanations for this finding. Autobiographical memory is typically conceived as highly associatively organized (Conway, 2005; Conway & Pleydell-Pearce, 2000). This unique characteristics of autobiographical memory representations may at the one hand entail a comparably stronger benefit of simple re-exposure in the form of restudying, as it is well documented in previous research that memory performance is increased for highly associatively organized information (for a review, see Bower, 1970). On the other hand, the memory boost received from retrieval practice is often explained by the enhancement and formation of new associative links between memory traces following successful retrieval (e.g., Carpenter, 2011), and may thus be less pronounced for autobiographical information which is inherently organized in a highly associative network. Both mechanisms, either on their own or in conjunction, can account for the present finding that repeatedly retrieving memories for autobiographical events does not benefit long-term retention for those memories any more than repeatedly restudying them. Indeed, similar theoretical assertions have been made for the finding that the testing effect is less pronounced or disappears when inherently relationally organized learning material is employed (Van Gog & Sweller, 2015). Furthermore, it may also be that the decreased effectiveness of testing compared to restudying in autobiographical memory is attributable to other characteristics of autobiographical memories such as, for instance, high self-relevance (Conway, 2005; Conway & Pleydell-Pearce, 2000). Memory performance for self-relevant information is typically increased (for a meta-analysis, see Symons & Johnson, 1997), and high self-relevance may lead to an enhanced associative processing during restudying due to increased interest in the restudied contents (for a review, see, e.g., Schiefele, 2001). Indeed, an interesting avenue for future research would be to determine the specific mechanisms underlying the decreased testing effect in autobiographical memory by comparing, for instance, the effect of testing on memories for autobiographical and matched non-autobiographical events.

The present findings are also relevant for the literature on autobiographic memory rehearsal. In previous literature, rehearsal of personal events in social communication and various types of private autobiographical memory rehearsal have been distinguished (e.g., Walker et al., 2009). However, since typically rehearsal types have been examined that involve retrieving personally experienced events from memory, less is known about the relative significance of retrieving autobiographical events compared to simple re-exposure to those events. From an applied perspective, the latter may occur when listening to rehearsals of socially shared personally experienced events in conversations, but also individually, for example through re-exposure to autobiographical events stored via different media (e.g., diaries, videos, social media postings; see Wang, Lee, & Hou, 2017). In line with previous results demonstrating that retrieval by talking about socially shared personally experienced events benefits long-term retention similarly to re-exposure by listening to the rehearsal of those events (Stone et al., 2013), the present results show that repeatedly retrieving autobiographical events does not benefit long-term retention significantly more than repeated re-exposure to those events. This indicates that the pattern reported for memory rehearsal in social contexts may also extend to forms of individual rehearsal of and re-exposure to autobiographical events.

It is important to note that the autobiographical memory paradigm (Barnier et al., 2004) that was adapted in the present study to examine post-encoding effects in autobiographical memory involves retrieval components in the initial collection phase: Individual autobiographical memories are collected in response to cue words. In fact, alternative ways of collecting autobiographical memories that do not involve initial retrieval by participants (e.g., recording the daily life of participants) seem hardly feasible for both economical and ethical reasons. Thus, the autobiographical memories in the restudy condition had been at least one time retrieved before memory was measured in the 2-week delayed memory test. However, this does not necessarily reflect a problem since it seems to be a natural characteristic of autobiographical memories that they are retrieved from time to time, as also shown in the present study by the ratings of frequency of previous retrieval of the collected autobiographical memories (mean of 3.3 on a scale ranging from 1 = not at all to 7 = very often). In fact, it may be that previous retrieval is one of the reasons why testing may be less effective than restudying in autobiographical memory compared to other types of memories.

One potential caveat regarding the interpretation of the memory results in the 13-week delayed memory test may be that memory performance may have been influenced by the fact that both tested and restudied memories had already been tested once in the 2-week delayed memory test. More precisely, according to distribution-based explanations of the testing effect (Halamish & Bjork, 2011; Kornell et al., 2011), it may be that actually existing differences in memory strength between tested and restudied autobiographical memories are not detectable at a retention interval of 2 weeks because the memory strengths of both types of memories are still above retrieval thresholds, but only after longer retention intervals. As both tested and restudied memories were tested in the 2-week delayed memory test, both may have received a (comparable) additional boost in memory strength so that actually existing differences were still not detectable at the time of the 13-week delayed memory test. However, such a possibility would presuppose relatively slow forgetting rates due to the relatively long delay of the final test (13 weeks), which seems unlikely given that a comparatively strong memory decay was observed in the present study between the 2-week delayed and the 13-week delayed memory tests. Still, it is important to note that future research should examine potential effects of moderating variables such as the number or the spacing of retrieval practice in order to investigate the generalizability of the present findings. Thus, from a broader perspective, the present research represents an important starting point for future research on the effects of testing on long-term retention of autobiographical memories.

The fact that we did not observe any difference between a testing and a restudy condition indicates that the testing effect (in the sense of a comparison between testing and restudy) seems not to occur in autobiographical memory. Still, it may be an interesting question for future research to compare the effects of re-exposure (testing or restudy) to a no-treatment control condition in autobiographical memory. Furthermore, future research is also necessary to evaluate whether the present findings generalize beyond specific circumstances of the present experiment. First, as the present sample consisted mainly of female undergraduate students, future research should investigate the generalizability of the findings across gender and age groups. Second, as the second delayed memory test took place in an online setting, future research should examine, whether similar long-term effects are found when testing in a laboratory setting.

The present study also contributes to another line of research addressing the effects of repeated testing. In the retrieval practice/restudy session, the amount of correctly recalled autobiographical events significantly increased throughout the three retrieval cycles, a phenomenon commonly referred to as “hypermnesia,” that is, the improvement of memory performance across varying retention intervals (for a review, see Erdelyi, 2010; Payne, 1987). Concerning autobiographical memories, previous research has demonstrated hypermnesia to occur across various retellings of one autobiographical event (Bluck, Levine, & Laulhere, 1999). The recall pattern observed across retrieval cycles in the retrieval practice/restudy session of the current study indicates that the hypermnesic effect also extends to autobiographical memories of multiple events retrieved in response to external cues. This observation is in line with previous interpretations of hypermnesia emphasizing the role of imagery encoded study material (e.g., Erdelyi, 2010), as autobiographical memories have been demonstrated to be highly imagery (Brewer, 1992).

Interestingly, in the present study, emotional memories were not better remembered than neutral memories. By contrast, in the retrieval practice phase, neutral memories were even significantly better recalled than positive memories. At first glance, this seems to be at odds with the common finding that memory is enhanced for emotionally significant compared to neutral events (for a review, see Reisberg & Heuer, 2004). However, previous studies have shown that the emotional enhancement effect disappears when controlling for memory-enhancing cognitive factors such as the distinctiveness or relatedness of the stimuli (Schmidt & Saari, 2007; Talmi, Luk, McGarry, & Moscovitch, 2007). Talmi et al. (2007) offer two possible explanations for these findings. First, because emotional stimuli are usually inherently more distinct, related, or attention grabbing than neutral stimuli, the typically observed emotional enhancement effect might at least be partially mediated by these cognitive factors. Alternatively, the manipulation of these factors may have simply raised memory performance for neutral stimuli to the same level as for emotional stimuli, regardless of possibly different underlying processes for emotional and neutral stimuli. Thus, in the current study, the collected neutral autobiographical events may have exhibited features that made them somehow cognitively distinct, especially as most autobiographical memories are typically emotionally significant (Brewer, 1988). More precisely, as participants were specifically instructed to think of neutral autobiographical events, they may have selected events that stood out compared to other neutral events (e.g., events that involved the breaking of a routine or low-frequent events; Brewer, 1988), which in turn may have led to equal memory performance for neutral and emotionally significant memories. Additionally, in the present study, we focused on moderately emotionally intense autobiographical events. As many explanations of the emotional enhancement effect emphasize the central role of emotional arousal (for a review, see McGaugh, 2004), it may also be that the events collected in the present study were not emotionally arousing enough for the emotional enhancement effect to emerge, especially if neutral events also exhibited distinct, memory-supporting features. Thus, concerning the effect of retrieval practice on autobiographical memories, it may well be that the pattern of results would differ for intensely emotionally arousing autobiographical events, an interesting question that should be addressed in future research.

Conclusion

Retrieving events from memory in a test has been demonstrated to provide a powerful boost for long-term retention of those events, even compared to repeatedly restudying the event. While the benefits received from testing have been shown across a variety of stimuli material and recall tasks, one important memory system has not been addressed in previous research: autobiographical memory. In the present study, the testing-effect paradigm was extended to autobiographical memories of personally experienced events. Across two delayed memory tests with retention intervals of up to three months, no significant memory benefit emerged for previously repeatedly retrieved compared to previously repeatedly restudied autobiographical events. Thus, it seems that the testing effect does not persist in autobiographical memory.

Funding

This work was supported by the German Research Foundation (DFG) [Deutsche Forschungsgemeinschaft] under grant number KU 3325/2-1 to Christof Kuhbandner.

Supplementary Material

The Supplementary Material for this article is presented fully below. It can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00810/full#supplementary-material

Supplementary Data: Further Analysis of Memory Characteristic Ratings

Results. Table 2 shows participants' mean ratings for the characteristics (clarity, personal relevance, emotional valence, emotional arousal and frequency of previous retrieval) of the autobiographical memories they described in response to neutral, negative or positive cue words as a function of the assigned practice condition (retrieval practice, restudy).

Table 2

Participants' Mean Ratings for the Characteristics of the Autobiographical Memories Described in Response to Neutral, Positive, and Negative Cue Words as a Function of Assigned Practice Condition (Retrieval Practice, Restudy).

		Emotional quality of cue word					
		Neutral		Positive		Negative	
Ratings of memory characteristics	Practice condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Clarity	Retrieval practice	5.05	1.09	5.98	0.60	5.72	0.55
	Restudy	5.08	0.98	5.97	0.69	5.79	0.54
Personal relevance	Retrieval practice	2.50	1.00	4.82	0.97	4.46	0.87
	Restudy	2.51	0.90	4.85	0.91	4.40	1.10
Emotional valence	Retrieval practice	4.07	0.30	6.00	0.53	2.25	0.43
	Restudy	4.08	0.28	6.02	0.48	2.24	0.50
Emotional arousal	Retrieval practice	2.24	1.08	4.71	0.91	4.80	0.81
	Restudy	2.21	1.01	4.71	0.86	4.81	0.96
Frequency of previous retrieval	Retrieval practice	1.98	0.73	3.96	1.04	3.80	0.99
	Restudy	2.01	0.67	4.02	0.90	3.93	0.97

Note. Participants rated their memories in terms of clarity (1 = not clear at all, 7 = very clear), personal relevance (1 = not relevant at all, 7 = very relevant), emotional valence (1 = very negative, 7 = very positive), emotional arousal (1 = not at all emotionally arousing, 7 = very emotionally arousing), and the frequency with which they had previously thought of this event or told others about it (1 = not at all, 7 = very often).

For each memory characteristic, a 3 (Emotion: neutral, positive, negative) × 2 (Type of Practice: retrieval practice, restudy) ANOVA was computed on participants' ratings.

Clarity. The analysis revealed a significant main effect of Emotion, Greenhouse-Geisser corrected $F(1.48, 69.68) = 44.99$, $p < .001$, $\eta_p^2 = .489$, but no significant main effect for Type of Practice, $F(1, 47) = 0.36$, $p = .553$, $\eta_p^2 = .008$, nor a significant Emotion by Type of Practice interaction, $F(2, 94) = 0.21$, $p = .814$, $\eta_p^2 = .004$, indicating that ratings of memory clarity were higher for emotional than for neutral memories, and that memories assigned to the retrieval practice condition and memories assigned to the restudy condition did not differ in terms of memory clarity.

Personal relevance. The analysis revealed a significant main effect of Emotion, Greenhouse-Geisser corrected $F(1.75, 82.39) = 180.79$, $p < .001$, $\eta_p^2 = .794$, but no significant main effect for Type of Practice, $F(1, 47) = 0.02$, $p = .879$, $\eta_p^2 < .001$, nor a significant Emotion by Type of Practice interaction, $F(2, 94) = 0.28$, $p = .755$, $\eta_p^2 = .006$, indicating that ratings of personal relevance were higher for emotional than for neutral memories, and that memories assigned to the retrieval practice condition and memories assigned to the restudy condition did not differ in terms of personal relevance.

Emotional valence. The analysis revealed a significant main effect of Emotion, Greenhouse-Geisser corrected $F(1.46, 68.50) = 1192.207$, $p < .001$, $\eta_p^2 = .962$, but no significant main effect for Type of Practice, $F(1, 47) = 0.01$, $p = .916$, $\eta_p^2 < .001$, nor a significant Emotion by Type of Practice interaction, $F(2, 94) = 0.05$, $p = .950$, $\eta_p^2 = .001$. Thus, the analysis indicated that memories described in response to negative cue words were indeed rated as more negative, $F(1, 47) = 935.02$, $p < .001$, $\eta_p^2 = .952$, and memories described in response to positive cue words as more positive, $F(1, 47) = 788.30$, $p < .001$, $\eta_p^2 = .944$, than memories described in response to neutral cue words, and that memories assigned to the retrieval practice condition and memories assigned to the restudy condition did not differ in terms of emotional valence.

Emotional arousal. The analysis revealed a significant main effect of Emotion, Greenhouse-Geisser corrected $F(1.66, 78.09) = 206.52$, $p < .001$, $\eta_p^2 = .815$, but no significant main effect for Type of Practice, $F(1, 47) = 0.01$, $p = .907$, $\eta_p^2 < .001$, nor a significant Emotion by Type of Practice interaction, Greenhouse-Geisser corrected $F(1.77, 83.14) = 0.06$, $p = .925$, $\eta_p^2 = .001$, indicating that ratings of emotional arousal were higher for emotional than for neutral memories, and that memories assigned to the retrieval practice condition and memories assigned to the restudy condition did not differ in terms of emotional arousal.

Frequency of previous retrieval. The analysis revealed a significant main effect of Emotion, $F(2, 94) = 144.72$, $p < .001$, $\eta_p^2 = .755$, but no significant main effect for Type of Practice, $F(1, 47) = 1.71$, $p = .198$, $\eta_p^2 < .035$, nor a significant Emotion by Type of Practice interaction, $F(2, 94) = 0.32$, $p = .730$, $\eta_p^2 = .007$, indicating that participants rated the frequency with which they had previously thought of this event or told others about it as higher for emotional than for neutral memories, and that memories assigned to the retrieval practice condition and memories assigned to the restudy condition did not differ in terms of the frequency of previous retrieval.

Discussion. For all rated memory characteristics, significant differences emerged between ratings for the characteristics of memories described in response to emotional cue words and ratings for memories described in response to neutral cue words. The ratings of emotional valence indicated that emotional valence differed between emotional conditions as intended, with memories described in response to negative cue words rated as more negative, and memories described in response to positive cue words rated as more positive than memories described in response to neutral cue words. Additionally, memories described in response to both positive and negative cue words were rated as more emotionally arousing than memories described in response to neutral cue words. Regarding further memory characteristics, memories collected in the negative and positive conditions were also rated as slightly clearer, as more personally significant and more frequently previously retrieved than memories collected in the neutral condition, thus replicating the pattern for emotional and non-emotional memories found in previous studies (Barnier et al., 2004, 2007; Walker et al., 2009). Importantly however, memories assigned to the retrieval practice condition and memories assigned to the restudy condition did not differ in any of the rated memory characteristics.

Supplementary Tables

Table 3

Mean Recall Rates for Autobiographical Events During Retrieval Practice (Cycle 1, Cycle 2, Cycle 3) as a Function of Emotion Condition (Neutral, Positive, Negative) and Mean Recall Rates for Autobiographical Events in the Delayed Memory Tests (2 Weeks, 13 Weeks) as a Function of Emotion Condition (Neutral, Positive, Negative) and the Type of Previous Practice (Retrieval Practice, Restudy).

			Emotion condition					
			Neutral		Positive		Negative	
			M	*SD*	*M*	*SD*	*M*	*SD*
Retrieval practice	Cycle 1		.87	.15	.79	.17	.83	.14
	Cycle 2		.90	.13	.84	.15	.86	.14
	Cycle 3		.90	.13	.85	.14	.88	.12
Delayed memory tests	2 Weeks	Retrieval practice	.86	.18	.81	.17	.82	.17
		Restudy	.80	.18	.80	.19	.81	.17
	13 Weeks	Retrieval practice	.73	.20	.67	.23	.68	.22
		Restudy	.70	.21	.68	.21	.61	.22

Table 4

Neutral, Positive and Negative Cue Word Lists Used in the Experiment (Original German Cue Words, English Translations in Brackets).

Cue word lists					
Neutral		Positive		Negative	
ausleihen *(to borrow/to lend)*	ordnen *(to arrange/to organize)*	Erfolg *(success)*	herzlich *(cordial)*	Streit *(quarrel)*	peinlich *(embarrassing)*
Mensa *(canteen)*	Frage *(question)*	beschwingt *(elated)*	Sport *(sports)*	Abschied *(farewell)*	krank *(sick)*
Uhr *(clock/watch)*	Jacke *(jacket)*	sonnig *(sunny)*	Freundschaft *(friendship)*	Zeitnot *(shortage of time)*	fürchten *(fear)*
Fenster *(window)*	Termin *(appointment)*	helfen *(to help)*	lecker *(delicious)*	frieren *(to be cold)*	tragisch *(tragic)*
anmelden *(to register)*	Boden *(floor)*	Feier *(celebration/ party)*	Natur *(nature)*	nervös *(nervous)*	versagen *(to fail)*
Haushalt *(housekeeping)*	drucken *(to print)*	Glück *(happiness/ luck)*	tanzen *(to dance)*	einsam *(lonely)*	Lüge *(lie)*
Schere *(scissors)*	Frisur *(hairstyle)*	Ausflug *(excursion)*	gesellig *(convivial)*	unfair *(unfair)*	Pech *(misfortune)*
aufladen *(to charge)*	Tisch *(table)*	spielen *(to play)*	Tier *(animal)*	Ärger *(anger)*	Übelkeit *(nausea)*
Stift *(pen)*	gehen *(to walk)*	entspannt *(relaxed)*	lachen *(to laugh)*	verirren *(to get lost)*	gemein *(mean)*
Schuh *(shoe)*	egal *(indifferent)*	Kompliment *(compliment)*	kreativ *(creative)*	Fehler *(mistake)*	verpassen *(to miss [e.g. the train])*
öffnen *(to open)*	Werkzeug *(tool)*	unterhaltsam *(entertaining)*	Belohnung *(reward)*	verlieren *(to loose)*	anekeln *(to disgust)*

Table 4 (continued)

einkaufen *(to buy/to shop)*	aufheben *(to pick up)*	Geschenk *(present)*	energiegeladen *(energetic)*	traurig *(sad)*	Panne *(mishap/ breakdown)*
Lesesaal *(reading room)*	Brief *(letter)*	Familie *(family)*	umarmen *(to embrace)*	Absage *(rejection)*	entsetzt *(appalled)*
E-Mail *(e-mail)*	mitgeben *(to give to take with them)*	freuen *(to rejoice)*	Reiseerlebnis *(travel experience)*	Gestank *(stench)*	allergisch *(allergic)*
Lampe *(lamp)*	Wohnung *(apartment)*	begeistert *(enthusiastic)*	verliebt *(enamored)*	Problem *(problem)*	Vorwurf *(reproach)*
klingeln *(to ring/to ring the bell)*	fahren *(to drive/to go [by vehicle])*	Musik *(music)*	Gespräch *(conversation)*	verletzen *(to injure)*	machtlos *(powerless)*
Tasche *(bag)*	Eimer *(bucket)*	Kind *(child)*	warm *(warm)*	Sorge *(worry)*	schlaflos *(sleepless)*

STUDY 3

Tests Improve Memory – No Matter if You Feel Good or Bad While Taking Them

Kathrin J. Emmerdinger and Christof Kuhbandner

Abstract

Based on studies demonstrating that testing promotes better long-term retention than restudying (i.e., the testing effect), testing has been recommended as a powerful tool to boost knowledge acquisition in educational settings. However, a factor ubiquitous in real-life learning contexts has been ignored to date: the learner's affective state. To examine whether the learner's affective state influences the testing effect, we conducted two experiments. We employed a standard testing-effect paradigm consisting of an initial study phase and a subsequent restudy/ testing phase, and induced negative, neutral, or positive affective states either before participants initially studied short expository texts (Experiment 1) or before they restudied or were tested on them (Experiment 2). After 1 week, memory for the texts was tested. In both experiments, previously tested material was better remembered than previously restudied material. However, in none of the experiments did the memory advantage of testing over restudying vary as a function of affect condition. Hence, the present results suggest that testing seems to benefit long-term retention independently of the learner's affective state.

Based on numerous studies demonstrating that being tested on studied material promotes better long-term retention than restudying the material (i.e., the testing effect, Carrier & Pashler, 1992), more frequent testing has been recommended as a powerful technique to boost long-term knowledge acquisition in educational settings (e.g., Roediger & Karpicke, 2006a; Roediger & Pyc, 2012). Indeed, the benefit of testing over restudying has been replicated for a large variety of materials, across various test formats (see Rowland, 2014, for a meta-analysis), and has been extended from laboratory studies to field studies in the classroom (e.g., McDaniel et al., 2007).

However, the potential influence of various psychological factors that are ubiquitous in real-life educational settings is largely unknown. For instance, recent studies have shown that factors such as trait test anxiety (Tse & Pu, 2012), performance pressure (Hinze & Rapp, 2014) and extrinsic motivation (Kuhbandner et al., 2016) can reduce the benefit received from testing. Another factor that seems to be especially relevant in real-life educational settings is the affective state learners experience when taking a test (e.g., Pekrun et al., 2002).

There is a large body of research showing that positive and negative affect promote different cognitive processing styles (for a review, see Clore & Huntsinger, 2007, 2009; Fiedler & Hütter, 2014), which may have consequences for the effects of testing as well. Negative affect has been demonstrated to promote a predominantly item-specific processing style and positive affect a predominantly relational processing style. Item-specific processing refers to the processing of an event's distinctive attributes and relational processing to the processing of associations between events in relation to other concepts in memory (e.g., Hunt & McDaniel, 1993). For instance, positive affect has been shown to increase and negative affect to decrease the breadth of attention (e.g., Gasper & Clore, 2002; Rowe et al., 2007; Vanlessen et al., 2016), the storage of visual information as integrated object versus independent feature representations (Spachtholz et al., 2016; Spachtholz & Kuhbandner, 2017), and the activation of associative networks in memory (e.g., Bäuml & Kuhbandner, 2009; Huntsinger et al., 2009; Storbeck, 2013).

By promoting different processing styles, affect may influence the benefit received from testing as well. In search of a given target information during testing, related concepts in memory are activated and thus may serve to update and strengthen the reactivated memory trace by establishing additional relations (e.g., Carpenter, 2009, 2011; Kornell et al., 2015; Pyc & Rawson, 2010). Negative affect, by promoting an item-specific processing style, may attenuate the spreading of activation in the associative network, which in turn may reduce the benefit received from successful retrieval. Positive affect, by promoting a relational processing style, may enhance the activation of associations and thus increase the benefit received from successful retrieval.

The aim of the present study was to investigate whether the effect of testing on long-term memory may differ depending on the learner's affective state during testing. For this purpose, we planned to conduct two experiments. As in real-life learning settings affective states are typically already present during studying (Pekrun et al., 2002), and theories on mood-dependent memory suggest that humans encode their affective experience together with the learned material and reactivate it when retrieving the material (e.g., Bower, 1981), in Experiment 1, affective states were induced at the beginning of the whole learning process. In Experiment 2, affective states were induced specifically during the restudy/testing phase.

We employed the testing-effect paradigm introduced by Roediger and Karpicke (2006b) where participants initially study short scientific texts, followed by a restudy/testing phase where participants either restudy the texts or are tested on them. Affective states were manipulated either before participants started to study the texts (Experiment 1), or before they started to restudy the texts or were tested on them (Experiment 2). In both experiments, after 1 week, memory for all texts was tested (for an illustration, see Figure 5).

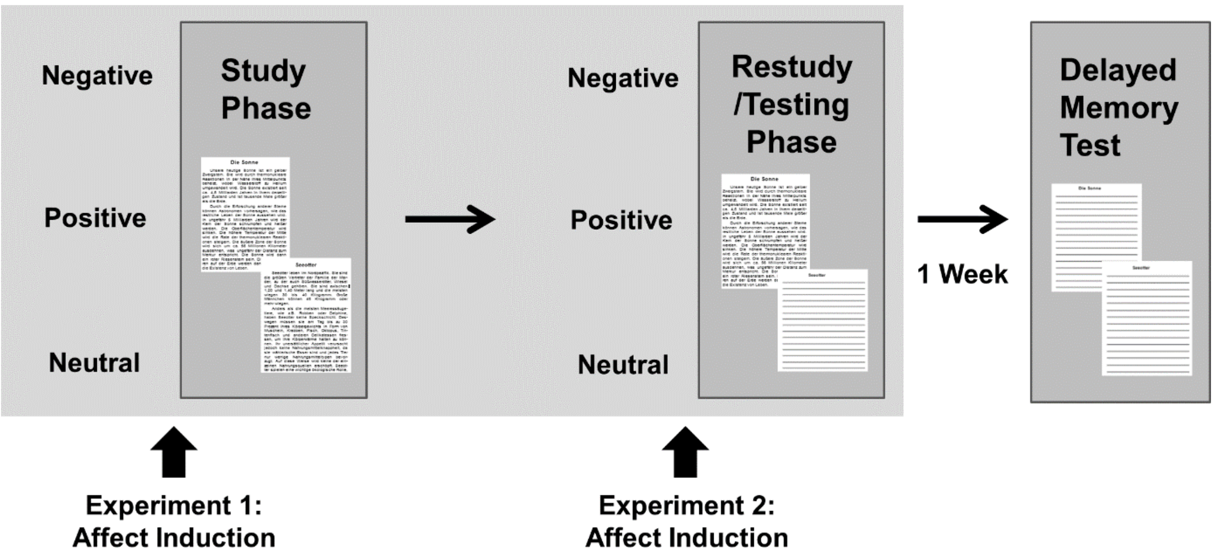

Figure 5. Procedure of the experiments. Both experiments consisted of an initial study phase, a subsequent restudy/testing phase where participants restudied one of the previously studied texts and were tested on the other text, and a delayed memory test on both texts after 1 week. Depending on affect condition, negative, positive, or neutral affect was induced before participants started to initially study the texts (Experiment 1) or specifically before they started to restudy or were tested on them (Experiment 2).

We expected to replicate the typical pattern of the testing effect for the neutral affect conditions, that is, superior recall in the delayed memory test for initially tested compared to initially restudied texts. However, our main interest was to investigate whether the pattern of the testing effect would change in the negative or positive affect conditions, relative to the neutral affect conditions. We assumed that negative affect may decrease, and positive affect may increase the benefit received from testing, resulting in the advantage for tested compared to restudied material increasing when experiencing positive affect, and decreasing or even disappearing when experiencing negative affect.

Experiment 1

Method

Participants. A power analysis revealed that to achieve a power of .90 for detecting small to medium sized effects (repeated measures ANOVA, within-between interaction, $f = 0.20$, $\alpha = .05$, $r = .30$ between repeated measures[1]; G*Power 3.1.7; Faul et al., 2007), a sample size of at least 114 would be required. Thus, we decided to recruit 120 undergraduate students (110 females, $M_{\text{Age}} = 21.50$, $SD = 3.05$)[2] who participated for course credit or a small monetary reward. The study was approved by the Ethics Committee of the University of Regensburg, and all participants provided informed written consent.

Material. Two different texts ("the sun" and "sea otters") were taken from Roediger and Karpicke (2006b) and translated into German. The translated texts were 249 and 267 words in length. To induce affective states, neutral, positive or negative film clips were presented. Two neutral ("Weather Forecasts"), two positive ("Harry and Sally" and "All about Mary") and two negative ("The Champ" and "Schindler's List") film clips were selected from data bases of standardized film stimuli serving the induction of affective states (Gross & Levenson, 1995; Schaefer, Nils, Sanchez, & Philippot, 2010). To measure the success of the affect manipulation, the affect grid (Russell et al., 1989) was used, which permits participants to rate their experienced affective state on the dimensions of valence (1 = extremely negative, 9 = extremely positive) and arousal (1 = low arousal, 9 = high arousal).

Design and Procedure. We used a 2 × 3 mixed-factorial design with the within-participants factor of learning condition (restudy, testing) and the between-participants factor of affect

---

1 Due to the application of different texts, small to medium sized correlations between repeated measures were expected (e.g., Kuhbandner & Pekrun, 2013).

2 One participant did not indicate age and gender.

condition (neutral, positive, negative). We applied the testing-effect paradigm introduced by Roediger and Karpicke (2006b), involving an initial study phase, a subsequent restudy/testing phase, and a delayed memory test phase taking place after 1 week, and induced neutral, positive or negative affective states during the initial study phase. Participants were randomly assigned to the affect conditions (n = 40 per condition).

All participants were tested individually in separate cubicles. They received a test booklet containing the texts and tests, and were informed that they were going to study texts for an upcoming memory test in the following week. Instructions, film clips, and the affect grid were presented on a computer screen using E-Prime 2.0 (Psychology Software Tools, Sharpsburg, PA). In the initial study phase, participants studied the texts for 7 min each. Study order of texts was counterbalanced across participants. To induce affective states, depending on condition, two neutral, two positive, or two negative film clips were presented; participants watched the first film clip and then studied the first text; subsequently they watched the second film clip and then studied the second text; presentation order of film clips was held constant for all participants. To measure the success of the affect manipulation, participants rated the affective state they had experienced during study immediately after studying each text on the affect grid. After a 5-min break during which participants solved simple arithmetic tasks, the restudy/testing phase followed. Participants were tested on one of the texts and restudied the other text for 7 min each. In case of testing, the title of the respective text was printed at the top of a page of the test booklet, and participants were instructed to write down as many text contents as they could recall without regard for order or verbal phrasing. In case of restudying, the respective text was presented again on a page of the test booklet, and participants were asked to reread the entire text as often as possible within the given time slot and memorize the contents as well as possible. Following Roediger and Karpicke's (2006b) procedure, they were asked to record the number of times they read the entire text by making marks on a separate sheet. Presentation order of learning conditions and assignment of texts to learning conditions were counterbalanced across participants. Immediately after working on the respective texts, participants rated the affective states they had experienced during restudying and testing using the affect grid. Between testing and restudying, participants solved simple arithmetic tasks for 2 min.

At the end of the session, participants indicated in their test booklets if they had read the same texts before the experiment. They were then thanked and asked to return to the laboratory 1 week later. In the delayed memory test phase after 1 week, participants' memory for both texts was tested. Participants received test booklets which contained separate pages with the titles of the texts, and they were asked to write down as many contents of the corresponding text as they could recall, without regard for order or verbal phrasing. Each memory test lasted 7 min. The order of the initially restudied and tested texts was counterbalanced across participants.

Results

Three participants indicated that they had already known the studied texts before the experiment and were excluded from data analysis (all from the neutral affect condition). All analyses were performed based on the remaining sample of 117 undergraduates (107 females, $M_{\text{Age}} = 21.53$, $SD = 3.08$). Including these participants did not change the significance of any of the reported results.

Affect manipulation check. The effect of the affect manipulation was determined for the initial study phase and the subsequent restudy/testing phase. For the initial study phase, participants' ratings of their affect experienced during initial studying were averaged across studied texts; for the restudy/testing phase, participants' ratings of their affect experienced during restudying and testing were averaged.

Regarding valence, in the initial study phase, the ratings were more negative in the negative condition ($M = 4.15$, $SD = 1.28$) compared to the neutral ($M = 5.11$, $SD = 1.44$) and positive conditions ($M = 5.70$, $SD = 1.33$), $t(75) = -3.10$, $p = .003$, $d = 0.72$, and $t(78) = -5.32$, $p < .001$, $d = 1.20$; the difference between the neutral and positive conditions was non-significant, $t(75) = -1.88$, $p = .065$, $d = 0.43$. However, a 3 (Affect Condition: neutral, positive, negative) × 2 (Time of Affect Rating: initial study phase, restudy/testing phase) ANOVA revealed a significant Affect Condition × Time of Affect Rating interaction, $F(2, 114) = 24.59$, $p < .001$, $\eta_p^2 = .301$, indicating that the affective state induced at the beginning of the learning process was not similarly stable across affect conditions. Whereas valence ratings did stay relatively stable in the neutral ($M = 4.81$, $SD = 1.27$; $t(36) = 1.65$, $p = .108$, $g_{av} = 0.21$) and positive ($M = 5.38$, $SD = 1.15$; $t(39) = 2.13$, $p = .039$, $g_{av} = 0.26$) conditions, valence ratings did strongly increase in the negative condition ($M = 5.39$, $SD = 1.03$), $t(39) = -5.97$, $p < .001$, $g_{av} = 1.04$). As a result, in the restudy/testing phase, valence ratings were more positive in the positive compared to the neutral conditions, $t(75) = 2.04$, $p = .045$, $d = 0.47$, and also more positive in the negative compared to the neutral conditions, $t(75) = 2.19$, $p = .032$, $d = 0.51$, with valence ratings not differing between the positive and negative conditions, $t(78) = 0.51$, $p = .959$, $d = 0.01$.

Regarding arousal, in the initial study phase, the ratings were higher in the positive ($M = 6.03$, $SD = 1.66$) compared to the neutral ($M = 4.92$, $SD = 1.83$) and negative ($M = 5.09$, $SD =$

1.64) affect conditions, $t(75) = 2.78$, $p = .007$, $d = 0.64$, and $t(78) = 2.54$, $p = .013$, $d = 0.58$; the difference between the neutral and negative conditions was not significant, $t(75) = -0.43$, $p = .671$, $d = 0.10$. A 3 (Affect Condition: neutral, positive, negative) × 2 (Time of Affect Rating: initial study phase, restudy/testing phase) ANOVA revealed a significant main effect of Affect Condition, $F(2, 114) = 3.38$, $p = .037$, $\eta_p^2 = .056$, but neither a main effect of Time of Affect Rating, $F(1, 114) = 0.52$, $p = .433$, $\eta_p^2 = .005$, nor a significant Affect Condition × Time of Affect Rating interaction, $F(2, 114) = 1.81$, $p = .169$, $\eta_p^2 = .031$, indicating that arousal did not change from the initial study phase to the subsequent restudy/testing phase ($M_{\text{Neutral}} = 4.99$, $SD = 1.97$; $M_{\text{Positive}} = 5.58$, $SD = 1.41$; $M_{\text{Negative}} = 5.16$, $SD = 1.55$).

Scoring. To assure consistent scoring of recall responses, each single piece of information was considered as one idea unit, resulting in detailed scoring schemes of 42 idea units (“the sun”) and 51 idea units (“sea otters”).[3] Recall responses were scored by giving one point for each correctly recalled idea unit. Initially, one third of the recall protocols were scored separately by two independent raters. Cohen's Kappa (κ) was performed to determine consistency among raters, and given the high inter-rater reliability, $\kappa = .81$ ($p < .001$), 95% CI [0.795, 0.827], the remaining protocols were scored by only one rater.

Main analysis. To explore the effects of affective states and learning conditions on participants’ recall rates in the memory tests, analyses of variance were performed. In addition, the recall data were examined by estimating Bayes factors (e.g., Wagenmakers et al., 2018), indicating the likelihood of the observed data under the alternative hypothesis, compared to the likelihood of the observed data under the null hypothesis (BF_{10}). For this purpose, the data were analyzed with the JASP software (JASP Team, 2018) using the Bayesian ANOVA methodology proposed by Rouder, Morey, Speckman, and Province (2012). According to current rules of thumb (e.g., Jarosz & Wiley, 2014; Wetzels et al., 2011), a BF of 1–3 is considered as anecdotal or weak evidence, a BF of 3–30 as substantial to strong evidence, a BF of 30–150 as very strong evidence and a BF of > 150 as decisive evidence.

Immediate memory test. Participants recalled on average 52.7% of the texts. Recall rates did not significantly differ between affect conditions ($M_{\text{Neutral}} = .55$, $SD = .14$ vs. $M_{\text{Positive}} =$

[3] Roediger and Karpicke (2006b) divided each text in 30 idea units. However, in their scoring scheme, single idea units often include several pieces of information, and they do not report which cut-off criterion was used to determine if a given idea unit was scored as correct. Hence, we decided to use a more fine-grained scoring scheme. In both experiments, all data were analysed on the basis of the scoring scheme provided by Roediger and Karpicke as well, and as the pattern of results did not differ, the results of the present study are only reported based on the fine-grained coding scheme.

.53, SD = .18 vs. $M_{Negative}$ = .50, SD = .15), $F(2, 114) = 0.69$, $p = .506$, $\eta_p^2 = .012$. The estimated BF_{10} was 0.14, corresponding to a BF_{01} of 7.03, indicating that the observed data are about 7 times more likely under the null hypothesis than under the alternative hypothesis.

Delayed memory test. Figure 6A shows recall performance in the 1-week delayed memory test as a function of learning condition (restudy, testing) and affect condition (negative, positive, neutral).

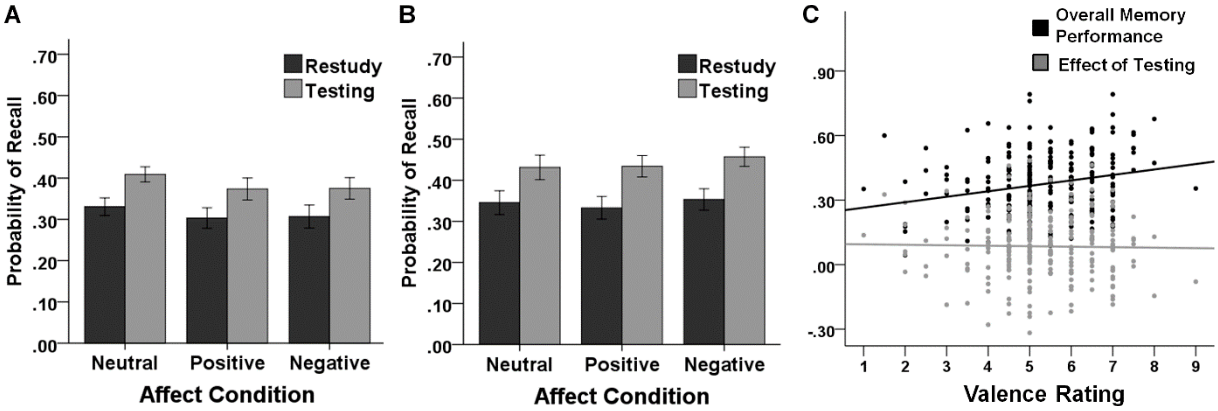

Figure 6. Results of the experiments. Memory performance in the delayed test after 1 week in Experiment 1 **(A)** and Experiment 2 **(B)** is shown as a function of learning condition (restudy, testing) and affect condition (neutral, positive, negative). Error bars represent standard errors. **(C)** shows the observed relationship in the combined data set (N = 237) between the emotional valence experienced during the restudy/testing phase and overall memory performance in the memory test after 1 week, and the individual effect of testing (i.e., difference between recall rates for tested and restudied material).

A 2 (Learning Condition: restudy, testing) × 3 (Affect Condition: neutral, positive, negative) ANOVA was conducted on the recall rates in the delayed memory test. The analysis revealed a significant main effect of Learning Condition ($M_{Testing}$ = .39, SD = .15 vs. $M_{Restudy}$ = .31, SD = .16, collapsed over affect conditions), $F(1, 114) = 34.51$, $p < .001$, $\eta_p^2 = .232$, but no significant main effect of Affect Condition, ($M_{Neutral}$ = .37, SD = .10 vs. $M_{Positive}$ = .34, SD = .15 vs. $M_{Negative}$ = .34, SD = .15, collapsed over learning conditions), $F(2, 114) = 0.60$, $p = .549$, $\eta_p^2 = .010$, nor a significant Learning Condition by Affect Condition interaction, $F(2, 114) = 0.06$, $p = .944$, $\eta_p^2 = .001$. The Bayesian analysis provides, compared to the null model, decisive evidence supporting the model containing the effect of Learning Condition (BF_{10} = 292,775), but substantial evidence against the model containing the effect of Affect Condition (BF_{10} = 0.19, corresponding to BF_{01} = 5.28 in support of the null model). Furthermore, a comparison of the model containing solely the two main effects with the model adding the Learning Condition by Affective Condition interaction (see Wagenmakers et al., 2018) indicates that the observed data are about 12 times less likely under the model including the interaction (BF = 64,974/5,428 = 11.97).

Conditional probabilities. Finally, we examined the impact of affective states on memories that were initially successfully retrieved; that is, we determined the conditional probabilities of recall in the delayed memory test given successful recall in the immediate test. Conditional probabilities did not significantly differ between affect conditions (M_{Neutral} = .65*, SD* = .13 vs. M_{Positive} = .59*, SD* = .21 vs. M_{Negative} = .61*, SD* = .21), $F(2, 114) = 1.20$, $p = .305$, η_p^2 = .021. The estimated BF_{10} was 0.22, corresponding to a BF_{01} of 4.63, indicating that the observed data are about 5 times more likely under the null hypothesis than under the alternative hypothesis.

Discussion

Basically, the results of Experiment 1 replicate the typical testing effect: a free recall test on initially studied text material led to superior recall performance in a 1-week delayed memory test in comparison to restudying the same material. More importantly, the effects of testing did not vary as a function of affect condition. Independent of affect condition, the same advantage of testing over restudying was found, and the same benefit was observed for initially successfully retrieved material. Additionally, Bayesian analyses provide substantial to strong evidence in support of the null hypothesis assuming no effect of affective states on the advantage of testing over restudying, and on the benefit received from initial successful retrieval. However, the results of the affect manipulation check revealed that participants in the positive affect condition rated their affect experienced during initial studying only marginally, statistically non-significant, more positive than participants in the neutral affect condition, and that the affect induced at the beginning of the learning process was not similarly stable across affect conditions. Consistent with theories on mood-dependent memory (e.g., Bower, 1981), participants in the positive affect condition rated their affective state still more positive in the restudy/testing phase compared to participants in the neutral condition. However, in the negative affect condition, the participants' affective state substantially changed. Although participants in the negative affect condition rated their affective state more negative than participants in the neutral and positive affect conditions in the initial study phase, in the subsequent restudy/ testing phase, they rated their affective state more positive than participants in the neutral affect condition, and equally positive than participants in the positive affect condition. Indeed, such a finding is well in line with mood repair theories, which suggest that the experience of negative affect activates mechanisms by which the unpleasant feelings are overcome (e.g., Morris & Reilly, 1987). However, as the affective state of participants in the negative affect condition was no longer negative in the restudy/testing phase, it remains to be shown whether the testing effect is indeed robust against the experience of negative affect during testing.

Experiment 2

Method

Participants. The required sample size was determined by the same power analysis as in Experiment 1. We recruited 120 undergraduate students (105 females, $M_{\text{Age}} = 21.97$, $SD = 3.52$) who participated for course credit or a small monetary reward. The study was approved by the Ethics Committee of the University of Regensburg, and all participants provided informed written consent.

Material. The same texts and film clips as in Experiment 1 were used again.

Design and procedure. The design and procedure were identical to Experiment 1, with the only exception that neutral, positive and negative affective states were induced specifically in the restudy/testing phase. Depending on affect condition, two neutral, two positive or two negative film clips were presented. Participants watched the first film clip and then restudied or took a test on the first text; subsequently, they watched the second film clip and then restudied or took a test on the second text; the order of the restudy and testing conditions was counterbalanced across participants. To measure the success of the affect manipulation, participants rated the affective state they had experienced during restudying and testing immediately after working on the respective texts.

Results

Affect manipulation check. To determine the success of the affect manipulation, participants' ratings of their affect experienced during restudying and testing were averaged. Valence ratings were more negative in the negative condition ($M = 4.71$, $SD = 1.38$) compared to the neutral ($M = 5.44$, $SD = 1.38$) and positive ($M = 5.93$, $SD = 1.07$) conditions, $t(78) = -2.35$, $p = .022$, $d = 0.53$, and $t(78) = -4.39$, $p < .001$, $d = 0.99$; the difference between the neutral and positive conditions was non-significant, $t(78) = -1.76$, $p = .082$, $d = 0.40$. Arousal ratings did not significantly differ between affect conditions ($M_{\text{Neutral}} = 5.31$, $SD = 1.53$; $M_{\text{Positive}} = 5.75$, $SD = 1.35$; $M_{\text{Negative}} = 5.46$, $SD = 1.59$), $F(2, 117) = 0.89$, $p = .414$, $\eta_p^2 = .015$.

Scoring. The same scoring procedure as in Experiment 1 was applied. Again, two independent raters scored one third of the recall protocols. Cohen's Kappa (κ) was performed and given the high inter-rater reliability, $\kappa = .85$ ($p < .001$), 95% CI [0.832, 0.859], the remaining tests were scored by only one rater.

Main analysis. The same analyses of variance and Bayesian analyses as in Experiment 1 were performed.

Immediate memory test. Participants recalled on average 57.7% of the texts. Recall rates did not significantly differ between affect conditions ($M_{\text{Neutral}} = .59$, $SD = .17$ vs. $M_{\text{Positive}} = .57$, $SD = .16$ vs. $M_{\text{Negative}} = .57$, $SD = .13$), $F(2, 117) = 0.14$, $p = .866$, $\eta_p^2 = .002$. The BF_{10} was 0.09, corresponding to a BF_{01} of 11.21, indicating that the observed data are about 11 times more likely under the null hypothesis than under the alternative hypothesis.

Delayed memory test. Figure 6B shows recall performance in the 1-week delayed memory test as a function of learning condition (restudy, testing) and affect condition (negative, positive, neutral).

A 2 (Learning Condition: restudy, testing) × 3 (Affect Condition: negative, positive, neutral) ANOVA on the recall rates in the delayed memory test revealed a significant main effect of Learning Condition, $F(1, 117) = 46.02$, $p < .001$, $\eta_p^2 = .282$. Collapsed over affect conditions, initially tested material was remembered better than initially restudied material ($M_{\text{Testing}} = .44$, $SD = .17$ vs. $M_{\text{Restudy}} = .34$, $SD = .17$), thus replicating the typical pattern of the testing effect. However, there was no significant main effect of Affect Condition, ($M_{\text{Neutral}} = .39$, $SD = .17$ vs. $M_{\text{Positive}} = .39$, $SD = .15$ vs. $M_{\text{Negative}} = .41$, $SD = .13$, collapsed over learning conditions), $F(2, 117) = 0.22$, $p = .800$, $\eta_p^2 = .004$, and no significant Learning Condition by Affect Condition interaction, $F(2, 117) = 0.16$, $p = .855$, $\eta_p^2 = .003$. The Bayesian analysis provides, compared to the null model, decisive evidence supporting the model containing the effect of Learning Condition ($BF_{10} = 20{,}830{,}000$), but substantial evidence against the model containing the effect of Affect Condition ($BF_{10} = 0.12$, corresponding to a $BF_{01} = 8.25$). Furthermore, a comparison of the model containing solely the two main effects with the model adding the Learning Condition by Affective Condition interaction (see Wagenmakers et al., 2018) indicates that the observed data are about 13 times less likely under the model including the interaction (BF = 3,243,000/253,242 = 12.81).

Conditional probabilities. Finally, we determined the conditional probabilities of recall in the delayed memory test given successful recall in the immediate test. Conditional probabilities did not significantly differ between affect conditions ($M_{\text{Neutral}} = .63$, $SD = .21$ vs. $M_{\text{Positive}} =$

.65, SD = .18 vs. M_{Negative} = .70, SD = .16), $F(2, 117) = 1.46$, $p = .236$, $\eta_p^2 = .024$. The BF_{10} was 0.26, corresponding to a BF_{01} of 3.79, indicating that the observed data are about 4 times more likely under the null hypothesis than under the alternative hypothesis.

Discussion

The results of Experiment 2 again replicate the typical testing effect: a free recall test on initially studied text material led to better recall performance in a 1-week delayed memory test than restudying the material. Crucially, the effects of testing did not vary as a function of affect condition. Independent of affect condition, the same advantage of testing over restudying was found, and the same benefit was observed for initially successfully retrieved materials. Additionally, Bayesian analyses provide substantial to strong evidence in support of the null hypothesis assuming no effect of affective states on the advantage of testing over restudying, and on the benefits received from initial successful retrieval. However, it is important to note that only the negative affect induction, but not the positive affect induction, led to reliable differences in participants′ ratings of their affective state experienced during restudying/testing compared to the neutral affect condition.

Combined Data Set

One caveat for the interpretation of the experiments′ results is that the affect induction did sometimes yield only weak to moderate effects on participants′ ratings of their affective states experienced during testing and restudying. In Experiment 1, participants to whom negative affective states were induced during the initial study phase rated their affect experienced in the restudy/testing phase in fact as more positive than participants in the neutral affect condition. In Experiment 2, participants to whom positive affective states were induced directly in the restudy/testing phase rated their affect only marginally more positive than participants in the neutral affect condition.

As participants varied in the effects of the affect induction, we supplemented the analyses based on the experimental conditions with additional analyses based on participants′ individual ratings of their affect experienced during restudying/testing. In this way, we aimed to gain further insight into the influence the affect actually experienced during testing may have on the benefit received from testing.

For this purpose, we combined the data set from both experiments (N = 237 participants) and conducted simultaneous regression analyses using participants′ affect ratings as predictor variables and participants′ immediate and delayed memory performances, participants′ conditional probabilities of recall in the delayed memory test given successful recall in the immediate test and participants′ individual effects of testing as dependent variables. The individual effect of testing was defined as the difference between the recall rate for restudied and the recall rate for tested material, determined individually for each participant.[4]

Results

Simultaneous regression analyses were conducted using affect ratings as predictor variables. To analyze possible interactions between valence and arousal ratings, both variables were centered on their respective means, and mean-centered valence ratings, mean-centered arousal ratings, and their cross product were entered in the basic regression model used in all analyses. Regressing immediate memory test performance on this set of predictors revealed a significant effect for valence, $\beta = .21$, $t = 3.00$, $p = .003$, but no significant effect for arousal, $\beta = -.01$, $t = -0.16$, $p = .875$, and no significant interaction, $\beta = -.01$, $t = -0.17$, $p = .869$. Regressing memory performance in the delayed memory test on the same set of predictors also showed a significant effect for valence, $\beta = .21$, $t = 3.07$, $p = .002$, but no significant effect for arousal, $\beta = .13$, $t = 1.91$, $p = .058$, and no significant interaction, $\beta = .10$, $t = 1.44$, $p = .152$. Hence, the analyses demonstrate that valence is a significant independent predictor of memory performance in the immediate and delayed memory tests, with recall rates decreasing with higher negative affect, and increasing with higher positive affect (see Fig. 6C).

Next, the individual effect of testing was regressed on the same predictors, revealing that none of the variables was a significant predictor (all βs < .05 and > −0.3, all *ps* > .518), which indicates that participants' affective state during the restudy/testing phase does not influence the advantage of testing over restudying (see Fig. 6C). Finally, conditional probabilities of recall in the delayed memory test given successful recall in the immediate test were regressed on the same set of predictors. The analyses revealed significant effects for arousal, $\beta = .20$, $t = 2.91$, $p = .004$, but no significant effect for valence, $\beta = .08$, $t = 1.08$, $p = .282$, and no significant interaction, $\beta = .08$, $t = 1.22$, $p = .223$. Thus, while emotional valence did not influence the benefits received from testing for successfully retrieved memories, such benefits did increase with increased arousal.

---

[4] In order to control for possible effects of the assignment of texts to learning conditions, all reported regression analyses were performed including type of text assignment as an additional predictor, and as the results did not differ, only the results based on the original set of predictors are reported.

In addition, Bayesian linear regression analyses were conducted using the JASP software (JASP Team, 2018). We report the Bayes factor which indicates how many times more likely the data occurs under each model compared to the null model containing no predictor. For immediate test performance, the model containing valence as predictor outperformed the null model ($BF_{10} = 22.79$), but the model containing arousal as predictor ($BF_{10} = 0.21$) and the model containing the valence*arousal interaction ($BF_{10} = 0.28$) did not. For delayed test performance, the model containing valence as predictor outperformed the null model ($BF_{10} = 41.27$), as did the model containing arousal as predictor ($BF_{10} = 4.39$), while the model containing the valence*arousal interaction did not ($BF_{10} = 0.14$). For the individual effect of testing, the null model outperformed all other models (all $BF_{10} < 0.16$, all $BF_{01} > 6.22$), providing substantial evidence in favor of the null model. For conditional probabilities of recall in the delayed memory test given successful recall in the immediate test, the model containing arousal as predictor outperformed the null model ($BF_{10} = 15.62$), but the model containing valence as predictor ($BF_{10} = 0.58$) and the model containing the valence*arousal interaction ($BF_{10} = 0.14$) did not. Thus, the Bayesian analysis also provides evidence that emotional valence predicts immediate and delayed memory performance, but does not influence the individual effect of testing, nor the benefits received from testing for successfully retrieved memories.

Discussion

Participants′ ratings of their valence experienced during testing did neither predict the individual effect of testing nor the probability of recall in the delayed memory test given successful recall in the immediate test. Thus, both the results of the experiments and the regression analyses of the combined data set indicate that the benefits received from testing seem to be independent of the learner′s affective state. However, the regression analyses revealed that participants′ individual ratings of their valence experienced during restudying and testing significantly predicted their memory performances in the immediate and in the delayed tests ($\beta = .21$ in both analyses), indicating that memory performance increased with higher positive affect and decreased with higher negative affect. This pattern of results is well in line with previous meta-analyses demonstrating that overall the effects of negative affect on performance are detrimental but only small in magnitude (e.g., Hembree, 1988; Seipp, 1991). In addition, Bayesian regression analyses also provide strong to very strong evidence in support of the models containing emotional valence as predictor of immediate and delayed memory performance, but substantial evidence in favor of the null model containing no predictor of the individual effect of testing. Thus, the results suggest that while affect seems to influence overall memory performance, the benefits received from testing may be immune to affective influences, at least in the range of the affective intensities examined in the present research.

Interestingly, the regression analyses, both frequentist and Bayesian, suggest that there might be an exception to the latter rule. While emotional valence did neither influence the advantage of testing over restudying nor the benefits received from successful retrieval, arousal predicted the latter one, with benefits increasing with increased arousal.

General Discussion

The present research aimed to clarify the potential influence of the learner′s affective state on the effects of testing. For this purpose, we induced negative, positive, or neutral affective states before participants started to study short scientific texts (Experiment 1) or before they started to restudy or were tested on them (Experiment 2). In both experiments, performance in a delayed memory test after 1 week was better for previously tested than for previously restudied material, thus replicating the typical testing effect. However, the memory advantage of testing over restudying did not vary as a function of affect conditions in any of the experiments.

Interestingly, a recent study shows similar results than the present study with regard to the effects of emotional significance of learning materials on the testing effect (Emmerdinger et al., 2018). In that study, participants studied cue-picture pairs consisting of a neutral cue and an either emotionally negative or neutral picture, and subsequently restudied the cue-picture pairs or retrieved pictures via the associated cue. In a delayed memory test after 1 week, recall performance was superior for formerly tested compared to formerly restudied pictures, again replicating the testing effect. However, although replicating the typical effects of emotional significance on overall memory performance, emotionally negative and neutral pictures did not differ in the effects of testing and restudying. Hence, the testing effect seems to be unaffected both by the emotional significance of the tested memories themselves, as indicated in the study of Emmerdinger et al. (2018), and by the affective state of the learner, as indicated in the present research. Taken together, these results suggest that testing effects might be generally relatively immune to emotional influences, at least in the range of the examined affective intensities.

The fact that we did not find any differences in the benefit received from testing between affect conditions may seem surprising in view of previous findings showing a wide range of effects of affective states on cognitive processing (for a review, see Clore & Huntsinger, 2007, 2009). However, in most of these studies, effects of affect have been demonstrated for very specific cognitive tasks. By contrast, a complex task such as text learning reflects combined effects across all processing stages. Indeed, findings on the impact of affective states on more complex cognitive tasks remain discordant about whether affect improves or impedes performance (e.g., Jung, Wranke, Hamburger, & Knauff, 2014; Knörzer, Brünken, & Park, 2016). Thus, the effect of affective states on the testing effect may be more complex than initially assumed. For instance, the predominant processing styles generally found in negative and in positive affective states have also been described in terms of two adaptive functions, accommodation and assimilation (e.g., Fiedler, 2001). Accommodation refers to the updating of internal knowledge structures based on incoming information, while assimilation refers to the interpretation of the external world on the basis of internal knowledge structures. Thus, detrimental effects of negative and beneficial effects of positive affective states on elaborative processing may have been counteracted by reverse effects regarding the impact of the reactivated memory representation on internal knowledge structures. More precisely, successfully reactivated memory representations may have promoted updates of internal schemes more likely when experiencing negative affect, but less likely when experiencing positive affect. Consequently, from a broader perspective, the present research represents an important starting point for future research on the effects of affective states on the testing effect and its potentially underlying mechanisms.

The present research also raises several more specific questions that should be addressed in future research. First, it is important to note that in the current experiments, only moderate intense affective states have been examined. In particular, valence ratings indicated that affect experienced during initial studying in Experiment 1, and during restudying/testing in Experiment 2, did not significantly differ between the positive and the neutral affect conditions. Thus, a potential caveat when interpreting the present results is that effects may emerge when stronger affect intensities are elicited, especially concerning the positive affect inductions. This concern is partly alleviated by the fact that regression analyses, both frequentist and Bayesian, at the level of participants' individually experienced affect did not provide any evidence for effects of affective intensity: participants' individual ratings of their valence experienced during restudying and testing, while predicting overall memory performance, did neither predict the individual effect of testing nor the benefit observed for initially successfully retrieved materials. Furthermore, the fact that affect ratings did not always substantially differ between affect conditions may also be partly due to their measurement after the respective study tasks, as reliable affective changes when using the same affect induction procedures have been reported in studies where affect was measured directly after the induction (e.g., Kuhbandner & Pekrun, 2013; Kuhbandner, Pekrun, & Maier, 2010). We decided to apply the study and restudy/testing phases directly following affect inductions in order to prevent fading of affective intensity before engaging in the study or testing activities. However, this procedure may also have resulted in the affect measurement capturing already faded affect intensities. Still, it cannot be ruled out that effects on the advantage of testing compared to restudying might emerge if stronger affect intensities are elicited, a question that should be examined in future research.

Second, while the focus of the present research was on affective states in terms of valence, the results of the regression analyses suggest that, in contrast to valence, arousal may indeed influence the benefit received from successful retrieval. In line with the present result that arousal predicted the long-term benefit received from successful retrieval, previous research has shown that arousal induced immediately following initial recall enhances later retrieval for successfully retrieved memories (Finn et al., 2012; Finn & Roediger, 2011). On the one hand, it has been suggested that arousal can modulate memory consolidation (e.g., Cahill & McGaugh, 1998; McGaugh, 2000), corroborated by findings that memory enhancements for emotionally arousing stimuli often emerge only after comparably long retention intervals (e.g., Hamann, Ely, Grafton, & Kilts, 1999; Sharot & Yonelinas, 2008). On the other hand, the benefits of testing may in part rely on retrieval-induced reconsolidation of the memory trace (Finn et al., 2012; Finn & Roediger, 2011). Thus, moderate arousal may enhance the benefit received from successful retrieval by fostering retrieval-induced reconsolidation processes. Furthermore, regarding the effects of arousal, recent studies show that the testing effect is preserved even if very strong arousal is experienced during memory retrieval (Smith et al., 2016; Szőllősi et al., 2017), and that previous retrieval practice may even immunize memories against detrimental effects of very high arousal on memory retrieval (Smith et al., 2016). However, these studies differ from the present experiments in that the effects of testing in neutral conditions on the effects of arousal present at the final test were examined, while in the present study arousal during restudy/testing predicted benefits in a later test conducted under neutral conditions. In summary, while not the focus of the present research, a further examination of the role of arousal for the benefits of testing represents an interesting avenue for future research.

Conclusion

Based on the finding that testing improves long-term memory retention compared to restudying, more frequent testing in educational settings has been recommended (e.g., Roediger

& Pyc, 2012). However, the affective state of the learner, an important factor in real-life educational settings, has to date been ignored in testing-effect research. In the present study, in two experiments, we replicated the typically observed pattern of the testing effect, thus better memory performance in a 1-week delayed memory test for previously tested than for previously restudied materials. This pattern of results did not change across a neutral, negative and positive affect condition. Thus, extending previous research, the present experiments suggest that testing seems to benefit long-term retention independently of the learner's affective state. Future research is needed to corroborate these findings and clarify if effects may be different when stronger affect intensities are experienced.

Funding

This work was supported by the German Research Foundation (DFG) [Deutsche Forschungsgemeinschaft] under grant number KU 3325/2-1 to Christof Kuhbandner.

PART III

CONCLUDING DISCUSSION

Summary of Findings

As reviewed in the first part of this thesis, previous research has shown that testing memories for previously studied contents enhances long-term retention more than providing additional restudy opportunities, and numerous studies have identified influencing factors and boundary conditions for the testing effect (for a meta-analysis, see Rowland, 2014). However, despite its relevance for educational implications, one potentially moderating variable has received only marginal attention: the influence that emotions may exert on the testing effect. The present research aimed to provide a first step towards closing this research gap.

From a functional perspective, emotions may influence cognitive processing on two general levels (Kuhbandner & Pekrun, 2010, p. 37). First, emotionally significant memories may be processed qualitatively different than neutral memories. Thus, the first of the three peer-reviewed studies presented in the second part of this thesis examined whether the effects of testing and restudying may differ for emotionally significant learning events compared to neutral ones. In this context, memories of personally experienced events are a special type of emotionally significant memory representations. Autobiographical memory representations exhibit unique characteristics (Conway, 2005) that may potentially influence the testing effect. Hence, the second study examined the testing effect for emotional and unemotional memories of personally experienced events. Finally, the affective state of the learner may influence information processing, independent of whether the processed information is emotionally significant or neutral. Thus, the third study, across two experiments, pursued the question whether the learners' affective state during studying and testing may modulate the typically observed pattern of the testing effect.

STUDY 1 investigated whether the effects of testing differ for emotionally negative stimuli compared to neutral stimuli. Participants studied cue-target pairs consisting of neutral retrieval cues and either emotionally negative or neutral target pictures, and were subsequently tested on, or restudied the cue-target pairs. For both emotionally negative and neutral target pictures, the typical pattern of the testing effect emerged in a 1-week delayed cued recall test. Memory performance for previously restudied target pictures exceeded memory performance for control target pictures that had not been presented again, but memory performance for previously tested target pictures was highest. Additionally, equal conditional probabilities of correct recall in the delayed cued recall test given successful recall in the initial test suggest that even the benefit received from successful retrieval did not differ for emotionally negative and neutral targets. Importantly, this pattern emerged albeit a typical emotional enhancement effect was also observed. While veridical recall rates in the 1-week delayed cued recall test were about the same for emotionally negative and neutral target pictures, more negative than neutral target pictures were additionally recalled in a subsequent cue-independent free-recall test, or misattributed to another retrieval cue. Thus, although the accessibility via the retrieval cues was not enhanced for emotionally negative compared to neutral targets, the availability of the targets independently of the cues was enhanced. Taken together, the results of Study 1 indicate that the testing effect does not differ for emotionally negative information, compared to neutral information.

STUDY 2 extended the testing-effect paradigm to emotional and unemotional autobiographical memories of personally experienced events. Participants initially wrote down brief descriptions of personal memories in response to emotionally negative, emotionally positive, and neutral cue words. On the following day, participants were tested on half of their previously described memories, while the other half of their memory descriptions was presented for restudy. In a 2-week delayed cued recall test, recall rates for previously tested autobiographical memories did not significantly differ from recall rates for previously restudied autobiographical memories, and this pattern of results persisted after a long retention interval of 13 weeks. Importantly, while recall rates significantly dropped from the medium to the long retention interval, the rate of the decline did not differ for previously tested memories compared to previously restudied memories, ruling out the possibility that the testing effect may be indirectly manifested as comparably slower time-dependent forgetting (e.g., Abel et al., 2019; Kornell et al., 2011). Thus, the results indicate that the testing effect may not pertain to autobiographical memories of personally experienced events.

Finally, **STUDY 3** examined whether learners' affective states during studying and testing may modulate the testing effect. To address this issue, across two experiments, a testing-effect paradigm using educationally relevant text materials was employed (Roediger & Karpicke, 2006b), and positive, negative, or neutral affect was induced before initially presenting the texts (Experiment 1), or specifically before testing and restudying them (Experiment 2). In both experiments, in a 1-week delayed free recall memory test, the typical pattern of the testing effect emerged, with superior performance for previously tested contents compared to previously restudied contents. However, contrary to what was expected, affect did not influence the occurrence or magnitude of the testing effect in either experiment. As participants varied in the effects of the affect induction, a supplemental analysis based on participants' individual ratings of their affect experienced during restudying and testing was performed on the combined data set of both experiments. The analysis revealed a significant effect of participants' emotional valence ratings on memory performance that, albeit small ($\beta = .21$), was within the expected range (for a meta-analysis, see Seipp, 1991). However, participants' valence ratings did neither predict the advantage of testing over restudying, nor the benefit received from successful retrieval. Taken together, the results of Study 3 suggest that the testing effect is unaffected by mild to medium intense affective states experienced during studying or test taking.

Emotions as Potentially Influencing Factors on the Testing Effect

An important endeavor of testing-effect research is to identify influencing factors and boundary conditions for the benefits of testing. Emotions have been shown to exert effects on all levels of cognitive processing (for reviews, see Clore & Huntsinger, 2009; Fiedler & Hütter, 2014; Holland & Kensinger, 2010; Kensinger, 2009). Thus, a crucial, but hitherto largely ignored, question is whether emotions may also act as a moderating factor for the effects of testing. Regarding this question, two of the three peer-reviewed studies presented in the main part of this thesis suggest that the testing effect may be relatively immune to emotional influences, concerning both the emotional significance of tested materials and the affective states present during testing of neutral materials. The testing effect emerged when comparing the effects of testing for emotionally negative and neutral information (Study 1), and across positive, negative, and neutral affective states induced before initial studying or before restudying and testing (Study 3).

Study 1 replicated typical emotional memory effects by showing an enhanced availability of negative targets in memory, but no enhanced accessibility through retrieval cues, compared to neutral targets. This finding is well in line with previous research demonstrating that emotional memory enhancement often comes at the expense of weakened associations to other stimuli present in the same scene (for a review, see Kensinger, 2009; Mather, 2007). However, testing emotionally negative information did not influence later memory in any other way than it did for neutral information: the typical advantage of testing over restudying emerged for both emotional and unemotional target pictures. The persistence of the testing effect for emotionally negative information may in part be attributable to the fact that an equal proportion of emotionally negative and neutral targets were initially successfully retrieved and thus benefited from testing (Kornell et al., 2011). Additionally, equal conditional probabilities of correct final test recall given successful initial recall indicate that, in contrast to the emotional enhancement effect usually observable at the level of initial encoding, successful retrieval seems not to benefit memories for emotional information more than memories for neutral information.

A potential explanation for this pattern of results may be that, in line with the often-observed emotional memory trade-off, successful retrieval may have boosted memory strengths for negative targets more than for neutral targets, but associations to the retrieval cue less. As both components are decisive for successful retrieval, equal net memory performances may have resulted. In accordance with this interpretation, two recent studies found similar results when applying initial and delayed memory tests relying on associations (Jia, Gao, Cui, & Guo, 2018), but modulations of the testing effect for emotionally significant verbal material when applying memory tests that predominantly rely on the availability of a given target in memory (Jia, Gao, Cui, & Guo, 2019). Furthermore, the present results confirm and extend previous findings demonstrating that other factors that modulate memories after initial encoding also seem to influence emotionally significant memories in similar ways than neutral memories (e.g., Kuhbandner et al., 2009; Sakaki et al., 2014; Wessel & Merckelbach, 2006). In showing that the same seems to apply concerning the effects of testing previously encoded memories, the present results extend this research beyond the previously examined memory-impairing factors to a memory-enhancing factor.

Besides the emotional significance of the learning materials per se, the present research also indicates that positive and negative affect experienced during studying or during testing and restudying does not seem to modulate the testing effect. This finding is especially relevant considering potential educational applications, as learners experience a variety of emotions in real-life educational environments (Pekrun et al., 2002). In Study 3, across two experiments, positive or negative affect induced before initial studying or before testing and restudying did not change the pattern of the testing effect, compared to neutral affect conditions. Furthermore, supplemental analyses based on participants' individual affect ratings demonstrated that valence significantly predicted initial and delayed memory performance but did not predict the advantage of testing compared to restudying, nor the benefit received from successful retrieval. The fact that, contrary to what was expected, the testing effect remained unaffected by the learners' current affective states seems surprising given previous research demonstrating that affect modulates current modes of cognitive processing (for a review, see Clore & Huntsinger, 2007; 2009; Fiedler & Hütter, 2014). However, most of these studies focused on very specific cognitive tasks and subprocesses, such as, for instance, visual selective attention (e.g., Rowe et al., 2007; for a review see Vanlessen et al., 2016), working memory capacity (Spachtholz, Kuhbandner, & Pekrun, 2014), cognitive priming (Storbeck & Clore, 2008), implicitly measured stereotypes (Huntsinger et al., 2009), or intrusion of false memories in list-learning procedures (Storbeck, 2013). Indeed, findings concerning affective influences on more complex tasks, such as logical reasoning or learning from complex instructional materials are mixed, and depend in part on the specific outcome measures (e.g., Jung et al., 2014; Knörzer et al., 2016). Performance in the text-learning tasks applied in the present Study 3 reflects the interplay of various lower-level and higher-level cognitive subprocesses such as orienting and sustaining attention, word recognition, semantic proposition encoding, inferences based on available cognitive schemes, and updating of cognitive schemes (see Wylie et al., 2018). Thus, as discussed in more detail in the General Discussion section of Study 3, since the observations at the level of the testing effect reflect the net effect of these underlying cognitive subprocesses, the present pattern of results may also reflect a combination of divergent emotion effects on the involved cognitive processes.

To date, only a few studies have examined the testing effect in the context of emotions, showing, for instance, that retrieval practice under neutral conditions can protect against detrimental effects of stress in a delayed memory test (Smith et al., 2016, 2018, 2019) and decrease students' test anxiety in a final exam (Agarwal et al., 2014). Concerning educational implications of testing-effect research, the present studies complement these studies by showing that even if present during initial testing, emotions do not seem to impede the beneficial effects of testing. Furthermore, emotional arousal may even boost the benefits received from testing, as indicated in the results of the regression analysis in Study 3 (see also Finn & Roediger, 2011; Finn et al., 2012). Taken together, these findings seem to imply that testing may indeed be a promising technique to boost knowledge-acquisition in educational settings, as its effectiveness does not appear to be affected by the experience of emotions or can even be enhanced by it. However, on the other hand, the present Study 3 focused on medium intense affective states induced by external sources. Thus, the situation may be different for more extreme affect intensities, and when affect is induced by the test situation itself. Both aspects will be discussed in more detail further below.

While Study 1 and Study 3 showed robust testing effects across emotional conditions, Study 2 revealed a potential boundary condition: The testing effect did not emerge for autobiographical memories of personally experienced events. Previous studies demonstrated benefits of retrieval practice of autobiographical memories compared to non-retrieved autobiographical memories (e.g., Barnier et al., 2004; García-Bajos & Migueles, 2017; Stone et al., 2013). However, there is some evidence that re-exposure to autobiographical memories may benefit later memory similarly than actively retrieving them (Stone et al., 2013), which was also the case in the present Study 2. A potential explanation for the present results is that restudying may benefit autobiographical memories more than other learning events, as memory performance is typically increased for highly associatively organized (Bower, 1970) and self-relevant materials (for a review, see Symons & Johnson, 1997), which are both inherent characteristics of autobiographical memories (Conway, 2005). On the other hand, because they are already inherently organized in a highly associative network, autobiographical memories may profit less from the elaborative processes initiated through retrieval. Both mechanisms, alone or in conjunction, may explain the lack of an advantage of testing compared to restudying for autobiographical memories. In addition, the proposed mechanisms are consistent with findings implying that the testing effect may be attenuated for highly associatively organized materials (for a review, see Van Gog & Sweller, 2015).

Limitations and Future Directions

The present research contributes to the existing testing-effect literature by addressing a current research gap that, among others, has important implications for educational applications: the impact of emotions on the testing effect. The presented findings provide first indications that the testing effect may remain relatively stable across various emotional influences but not persist in autobiographical memory. However, there are some limitations and additional questions that should be addressed in future research. In summary, the present research represents an important starting point for research on the impact of emotions on the testing effect, which needs to be corroborated and expanded on in future research.

Emotional Manipulations

Regarding the kind of emotional manipulations applied in the present studies, it must be kept in mind that in the present research, the focus was on the impact of moderate emotional intensities. Thus, it may be that the effects would differ when very strong emotional intensities were examined. In particular, in Study 3, no significant differences were observed between the positive and neutral affect conditions concerning participants' ratings of emotional valence experienced during initial studying (Experiment 1) and during restudying and testing (Experiment 2). As discussed in more detail in the General Discussion section of Study 3, affect ratings may in part have reflected already faded affect intensities due to the measurement after the respective learning tasks, as reliable affective changes have been induced by the same induction procedures when affect was rated directly after the induction procedure (e.g., Kuhbandner & Pekrun, 2013; Kuhbandner et al., 2010), and supplemental regression analyses at the level of participants' individually experienced affect did not deliver any indications that affective intensity may modulate the testing effect either. Still, especially concerning the positive affect inductions, it cannot be excluded that the pattern of the testing effect may be influenced by stronger affect intensities – an important caveat to be kept in mind when considering educational applications.

Moreover, the regression analyses based on individual affect ratings in Study 3 indicated that, unlike valence, arousal may modulate the testing effect: Arousal was a significant predictor of the benefit received from successful retrieval. In particular, this finding fits with previous research showing that arousal induced immediately after successful retrieval enhances later memory (Finn & Roediger, 2011; Finn et al., 2012). These findings may relate to the central role of arousal in memory consolidation. Both animal studies and studies with human subjects show that arousal can modulate the consolidation of memories, presumably because regions within the affect processing system, most notably the amygdala, modulate processes in brain regions involved in memory consolidation, such as the hippocampus (e.g., Cahill & McGaugh, 1998; McGaugh, 2000). At the same time, when memories are retrieved, they enter into a labile state that makes them amenable to changes, initiating subsequent reconsolidation processes (for reviews, see Elsey, Van Ast, & Kindt, 2018; Sara, 2010). These retrieval-induced reconsolidation processes have been assumed to play an important role in explaining the benefits of testing (Finn et al., 2012; Finn & Roediger, 2011), and arousal may enhance those processes. Thus, one interesting avenue for further research may be to further disentangle the influences of (moderate) arousal on the effects of testing and their underlying processes.

However, effects may be less beneficial when strong arousal or strong affect intensities in general are concerned. Previous research on the effects of highly intense emotions on memory has revealed mixed results, showing both beneficial and detrimental effects, but in most cases, stress during retrieval seems to impair memory (for a meta-analysis, see Shields, Sazma, McCullough, & Yonelinas, 2017). Concerning the testing effect, a few previous studies found the typical testing-effect pattern even when the delayed memory test occurred under memory-impeding conditions of acute stress (Smith et al., 2016, 2018; Szőllősi et al., 2017), and even showed that previous retrieval practice can immunize memories against the detrimental effects of stress-induced strong emotional arousal (Smith et al., 2016, 2018, 2019). Still, it remains an open question how strongly the intensity of the experienced affect present during the initial acquisition phase – either during initial encoding or during restudying and testing –

influences the testing effect observed in a delayed memory test. If strong affect during retrieval impairs memory performance, the benefit received from testing may be attenuated because of reduced initial retrieval success.

Yet, when comparing the effects of retrieval practice and of restudying, different patterns of results are conceivable based on previous research. In educational settings, strong affect may be more likely to occur during testing (e.g., in an exam) than during restudying. When comparing the effect of retrieval practice under conditions of performance-related affect with the effect of restudying in neutral conditions, Hinze and Rapp (2014) found no testing effect and even a reversed testing effect for some question types. However, the situation may be more complex when emotions are present during both testing and restudying. On the one hand, restudying may be less affected by affect-induced memory impairments, because the whole material is presented again. On the other hand, affect-induced memory impairments may be due to task-irrelevant and distracting thoughts, which divide attention and deplete cognitive resources available for the task at hand (e.g., Ellis & Ashbrook, 1988; Eysenck & Calvo, 1992; Eysenck, Derakshan, Santos, & Calvo, 2007). These processes may have an equal or even larger effect on the reencoding of memory traces during restudying compared to during testing. Task-irrelevant thoughts induced by affect may lead to similar conditions than dividing attention between concurrent tasks. Divided attention, however, mainly disrupts the encoding process and has less impact on the retrieval of information (e.g., Naveh-Benjamin, Craik, Perretta, & Tonev, 2000; Naveh-Benjamin, Kilb, & Fisher, 2006), and testing effects have been found both under divided attention and under full attention conditions (Buchin & Mulligan, 2017, 2019; Zhu, Olechowski, & Habib, 2017). Still, findings on potential divergent effects of divided attention on testing and restudying are mixed, sometimes showing detrimental effects only for restudying (Buchin & Mulligan, 2017) and sometimes showing detrimental effects for both testing and restudying (Zhu et al., 2017).

Furthermore, it may make a difference whether the influence of externally induced emotions present during testing is investigated, as it was the case in the present Study 3, or the influence of emotions triggered by the test situation itself is examined (i.e., test anxiety). Accordingly, a few studies indicate that test anxiety may decrease the beneficial effects of testing on long-term memory (Hinze & Rapp, 2014; Khanna, 2015; Khanna & Cortese, 2016). Finally, although dimensional affect models (e.g., Russell, 1980; Russell & Barrett, 1999) posit that effects of emotions are driven by changes in the underlying dimensions of valence and arousal, categorical affect theories (e.g., Ekman, 1992, 1999; Izard, 2009) suggest that additional mechanisms may play an important role, such as different underlying appraisal dimensions or action tendencies associated with discrete emotions (e.g., anxiety, anger, sadness) (for a review, see Angie, Connelly, Waples, & Kligyte, 2011). In summary, future research is needed to resolve the question whether the effects observed in the present studies may or may not be different when more emotionally intense memories or more intense affective states are examined, and whether the source of the current emotion or other mechanisms specific to discrete emotions matter.

Surprisingly, contrary to previous findings (for a review, see Holland & Kensinger, 2010), in Study 2, no memory enhancement was encountered for autobiographical memories of emotionally significant events compared to unemotional events, and during retrieval practice neutral memories were even recalled slightly better than positive memories. Interestingly, slight advantages for unemotional compared to emotional autobiographical memories in cued recall tests were also found in previous studies using the cue-word methodology for initial memory collection (Barnier et al., 2004; 2007). Thus, because participants were explicitly asked to collect memories for neutral events, they may have selected memories that were somewhat cognitively distinct from other neutral events (e.g., low-frequent events or events that involved breaking a routine; Brewer, 1988), which may have made them equally memorable than emotional memories. As previous research has presumed that the emotional enhancement effect for non-arousing emotional information may rely on non-emotion-specific mechanisms such as distinctiveness and relatedness (e.g., Schmidt & Saari, 2007; Talmi et al., 2007) or self-referential processing (Holland & Kensinger, 2010), this may especially have been the case here, because participants were asked to think of medium intense events (thus, their daily hassles and daily uplifts) rather than of profoundly intense emotional experiences. However, as in Study 1 an emotional enhancement effect emerged albeit participants rated the employed target pictures only slightly more emotionally intense ($M = 6.60$ on a 9-point scale) than participants in Study 2 rated their autobiographical memories ($M = 4.71$ for positive and $M = 4.80$ for negative memories on a 7-point scale), it seems likely that the absence of the emotional enhancement effect is mainly attributable to characteristics of the neutral autobiographical memories, which made them more memorable.

Another issue concerning the effects of emotions that should be addressed in future research is whether the testing effect remains equally robust when studying and testing emotionally positive information. Both in the present Study 1 and in another study using verbal learning material (Jia et al., 2018), emotionally negative stimuli were employed. There is some reason to believe that the pattern of results may be different for emotionally positive information. Beyond effects of emotional arousal, the emotional valence of the information may also contribute to the effects of emotional significance on memory. As mentioned in Part I of this thesis, the formations of item memory and associations between items seem to rely on different mechanisms (e.g., Bisby & Burgess, 2014; Davachi, 2006; Giovanello et al., 2003, 2004; Hannula et al., 2015), and conditions that promote item-specific processing do not necessarily coincide with conditions promoting relational processing. In particular, postulated differential effects of positive and negative emotions on relational versus item-specific processing (for reviews, see Clore & Huntsinger, 2007, 2009; Fiedler & Hütter, 2014) would entail the typical emotional memory trade-off to be more marked for emotionally negative stimuli. Indeed, previous studies showed that focal memory enhancements for central features at the expense of more peripheral details seem to be more likely for emotionally negative memories (e.g., Berntsen, 2002; for reviews, see Holland & Kensinger, 2010; Kensinger, 2009; Mather, 2007). By contrast, the relational processing style predominant when experiencing positive affect may attenuate the emotional memory trade-off by fostering associative connections. In turn, this may lead to different effects when regarding the effects of testing for emotionally positive information – a question that should be explored in future research.

Characteristics of the Employed Tests and Materials

Future research should also address whether the present results can be generalized beyond the specific characteristics of the tests and learning materials employed in the present research. Regarding the effects of testing emotionally significant information, different results may emerge if integral features of the emotional target itself serve as retrieval cues, instead of neutral retrieval cues unrelated to the emotionally significant target as employed in the present Study 1. Previous research shows memory enhancements for both central and more peripheral details if emotions are tied to the complete episode instead of only to salient visual stimuli (Laney, Campbell, Heuer, & Reisberg, 2004), or if the peripheral details constitute an integral feature of the emotion-eliciting object, such as color, shape, size (D'Argembeau & Van der Linden, 2004; Kensinger, Garoff-Eaton, & Schacter, 2007b), or spatial location (D'Argembeau & Van der Linden, 2004; Mather & Nesmith, 2008). Thus, when integral features of the emotional target itself serve as retrieval cues, item memory and memory for associations may receive similar boosts in storage strength, which may lead to larger benefits received from testing compared to neutral targets.

Furthermore, results may differ when test formats are used that emphasize the availability of items in memory. In the present Study 1, a standard cued recall testing-effect paradigm and a two-stage delayed memory test were employed, aiming to capture both accessibility through the retrieval cues, and availability of the target memories independently of the retrieval cues, revealing that the testing effect did not differ for emotionally significant and neutral target memories. Similar results were found in another study when testing associations between emotional words and background scenes (Jia et al., 2018). However, stronger testing effects may emerge for emotionally significant information than for neutral information when employing test formats that aim to test the availability of a given target in memory, such as recognition tests. A recent study employed recognition tests and found that the pattern of the testing effect differed for emotionally negative and neutral items (Jia et al., 2019). However, in that study, contrary to what could be expected, emotionally negative items profited less from testing than neutral items. The authors explained this result with comparably less retrieval effort in the initial recognition test, thus less benefit from successful retrieval for the emotionally significant items than for the neutral items. In summary, further research needs to clarify the role of different test formats when examining the impact of testing for emotionally significant information.

Similarly, future research should clarify whether the robustness of the testing effect across various affective states, as demonstrated in Study 3, persists across different testing conditions and learning materials. Particularly, the testing effects observed in Study 3 were relatively modest, so further studies are needed to determine whether positive and negative affective states may influence the magnitude of the effect if a larger baseline testing effect is obtained. The relatively modest testing effects in the present experiments may be in part due to medium retrieval success in the initial tests. Thus, future research could extend the introduced paradigm for examining the effects of affective states on the testing effect in terms of test conditions or materials that are more likely to render high initial retrieval success (such as cued recall tests, learning to criterion, or simpler learning materials). Furthermore, as the results of the present research may have reflected a combination of divergent affective influences on different aspects of cognitive processing, future research needs to disentangle the effects of affective states on the testing effect and its potentially underlying cognitive mechanisms.

Another test-related factor that may have influenced the pattern of results in the present research is that in Study 2, autobiographical memories were retrieved at two different retention intervals. The idea of repeating the memory test after a comparably long retention interval of 13 weeks was that, following distribution-based models (Kornell et al., 2011), for memories with relatively strong initial memory strengths, the testing effect may only emerge after very long retention intervals. However, because memories had already been tested at the 2-week delay, one could argue that both previously tested and previously restudied memories have received comparable memory boosts so previously restudied memories may still have been above recall threshold even at the long retention interval of 13 weeks. Although the relatively strong memory decay between the 2-week delayed and the 13-week delayed memory test speaks against this argument, it is still important for further research to determine the effects of moderating variables such as the number or spacing of testing and restudying opportunities on the testing effect for autobiographical memories. For instance, test delay could be introduced as an additional experimental manipulation. However, it should be noted that in the present Study 2, a quite time-consuming first collection session (90 to 120 min) was necessary in order to collect enough memories to prevent ceiling effects. A within-manipulation of test delay would require collecting even more memories which might result in participants' loss of motivation or concentration, compromising the quality and reliability of the described events.

Another interesting avenue for future research is the reinvestigation of the effects of testing in autobiographical memory using another methodology for initial memory collection. As addressed in the Discussion section of Study 2, collecting autobiographical memories with the cue-word methodology is an established method in autobiographical memory research (e.g., Barnier et al., 2004, 2007; Kuyken & Dalgleish, 2011; Maccallum et al., 2000; Robinson, 1976), but it bears the problem that it involves initial retrieval of those memories. On the one hand, this may not be problematic as it may in any way be a natural characteristic of autobiographical memories to be retrieved from time to time, making a clear distinction between tested and exclusively restudied memories impossible. On the other hand, one could additionally speculate that the absent emotional enhancement effect for medium intense emotional autobiographical memories may in part be due to the fact that participants were explicitly asked to think of a neutral memory. As neutral memories typically come less easy to mind than emotional memories (for a review, see Holland & Kensinger, 2010), more retrieval effort may have been involved in collecting the neutral memories, boosting their memory strengths and thus making them more retrievable at later memory tests. As mentioned in the previous section, at least two other studies (Barnier et al., 2004, 2007) using the cue-word methodology also found a slight memory advantage for unemotional autobiographical memories in a delayed cued recall test compared to emotional autobiographical memories. As collection methodologies that would guarantee the exclusion of any retrieval components (e.g., the recording of participants' daily lives) are hardly feasible, one viable alternative might be to employ diary-methods where participants record their activities for a given time, although even this methodology involves some retrieval components.

The present research also has implications for recent debates on whether highly associatively structured learning material may or may not constitute a boundary condition of the testing effect (e.g., Rawson, 2015; Van Gog & Sweller, 2015). As described in the previous section and in the Discussion section of Study 2, the finding that the effects of testing and re-exposure do not differ for autobiographical memories may in part be due to the highly associatively organized nature of this kind of memory representations. However, it may also be explained by other typical characteristics of autobiographical memories, such as high self-relevance and personal involvement, as both factors are known to enhance later memory (e.g., Muscatell et al., 2010; Symons & Johnson, 1997). In order to disentangle the specific underlying mechanisms, future research could, for example, compare the effects of testing for autobiographical events and matched non-autobiographical events or examine the effects of testing on non-associatively structured materials that have been encoded with reference to oneself (for a review, see Symons & Johnson, 1997).

Conclusion

The present thesis investigated whether emotional factors may affect the mnemonic benefits of testing. Two of the three peer-reviewed studies presented in this thesis indicate that the testing effect indeed may be relatively immune to emotional influences, concerning both the emotional significance of studied information and the affective state of the learner. In these studies, equal advantages of testing compared to restudying, and equal benefits received from successful retrieval, emerged for emotionally negative and neutral information (Study 1), and across positive, negative, or neutral affective states experienced during studying and testing (Study 3). However, the present thesis also indicates a potential boundary condition for the testing effect. Independent of the emotional significance of the corresponding memories, across two delayed memory tests, no advantage for testing compared to re-exposure emerged for autobiographical memories of personally experienced events. As autobiographical memories stand out due to various characteristics such as self-relevance or highly associatively organized memory structure, future research needs to clarify potentially underlying mechanisms of this boundary condition.

The finding that testing effects may be relatively robust across various emotional influences seems promising in view of potential educational applications, as in real-life learning settings the experience of emotions is ubiquitous (e.g., Pekrun et al., 2002). However, further research should corroborate and extend the present findings before drawing more decisive conclusions. In particular, the findings should be replicated across various test and material characteristics. Additional emotional factors, such as the role of affect intensity, the role of arousal and valence, and the role of the specific emotional manipulations and sources of affect should be investigated. From that perspective, the present research provides a starting point to spark further research on emotional influences – an important, but hitherto largely overlooked topic in testing-effect research.

REFERENCES

Abel, M., Haller, V., Köck, H., Pötschke, S., Heib, D., Schabus, M., & Bäuml, K.-H. T. (2019). Sleep reduces the testing effect – but not after corrective feedback and prolonged retention interval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 272–287. https://doi.org/10.1037/xlm0000576

Abel, M., & Roediger, H. L., III (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition*, *45*, 81–92. https://doi.org/10.3758/s13421-016-0641-8

Abott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, *11*, 159–177. https://doi.org/10.1037/h0093018

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*, 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*, 189–209. https://doi.org/10.1037/edu0000282

Agarwal, P. K., D'Antonio, L., Roediger, H. L., III, McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, *3*, 131–139. https://doi.org/10.1016/j.jarmac.2014.07.002

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L., III (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, *25*, 764–771. https://doi.org/10.1080/09658211.2016.1220579

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22*, 861–876. https://doi.org/10.1002/acp.1391

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *8*, 463–470. https://doi.org/10.1016/S0022-5371(69)80090-3

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General*, *134*, 258–281. https://doi.org/10.1037/0096-3445.134.2.258

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445. https://doi.org/10.1016/j.jml.2003.08.006

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*, 366-369. https://doi.org/10.1038/35066572

Angie, A. D., Connelly, S., Waples, E. P., & Kligyte, V. (2011). The influence of discrete emotions on judgement and decision-making: A meta-analytic review. *Cognition and Emotion*, *25*, 1393–1422. https://doi.org/10.1080/02699931.2010.550751

Arnold, K. M., & McDermott, K. B. (2013). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review*, *20*, 507–513. https://doi.org/10.3758/s13423-012-0370-3

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-l. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, *85*, 89–99. https://doi.org/10.1080/00220671.1991.10702818

Barber, S. J., & Mather, M. (2012). Forgetting in context: The effects of age, emotion, and social factors on retrieval-induced forgetting. *Memory & Cognition*, *40*, 874–888. https://doi.org/10.3758/s13421-012-0202-8

Barnier, A. J., Conway, M. A., Mayoh, L., Speyer, J., Avizmil, O., & Harris, C. B. (2007). Directed forgetting of recently recalled autobiographical memories. *Journal of Experimental Psychology: General*, *136*, 301–322. https://doi.org/10.1037/0096-3445.136.2.301

Barnier, A. J., Hung, L., & Conway, M. (2004). Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition and Emotion*, *18*, 457–477. https://doi.org/10.1080/0269993034000392

Barrett, L. F. (2006a). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*, 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*, 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, *8*, 10–14. https://doi.org/10.1111/1467-8721.00003

Bartlett, J. C., & Tulving, E. (1974). Effects of temporal and semantic encoding in immediate recall upon subsequent retrieval. *Journal of Verbal Learning and Verbal Behavior*, *13*, 297–309. https://doi.org/10.1016/S0022-5371(74)80066-6

Bäuml, K.-H., & Kuhbandner, C. (2009). Positive moods can eliminate intentional forgetting. *Psychonomic Bulletin & Review*, *16*, 93–98. https://doi.org/10.3758/PBR.16.1.93

Berntsen, D. (2002). Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Memory & Cognition*, *30*, 1010–1020. https://doi.org/10.3758/BF03194319

Berntsen, D., & Rubin, D. C. (2002). Emotionally charged autobiographical memories across the life span: The recall of happy, sad, traumatic and involuntary memories. *Psychology and Aging*, *17*, 636–652. https://doi.org/10.1037//0882-7974.17.4.636

Bertilsson, F., Wiklund-Hörnqvist, C., Stenlund, T., & Jonsson, B. (2017). The testing effect and its relation to working memory capacity and personality characteristics. *Journal of Cognitive Education and Psychology*, *16*, 241–259. https://doi.org/10.1891/1945-8959.16.3.241

Bisby, J. A., & Burgess, N. (2014). Negative affect impairs associative memory but not item memory. *Learning & Memory*, *21*, 21–27. https://doi.org/10.1101/lm.032409.113

Bishara, A. J., & Jacoby, L. L. (2008). Aging, spaced retrieval, and inflexible memory performance. *Psychonomic Bulletin & Review*, *15*, 52–57. https://doi.org/10.3758/PBR.15.1.52

Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, *3*, 165–170. https://doi.org/10.1016/j.jarmac.2014.03.002

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bless, H., & Fiedler, K. (2006). Mood and the regulation of information processing and behavior. In J. P. Forgas (Ed.), *Affect in social thinking and behavior* (pp. 65–84). New York, NY: Psychology Press.

Bluck, S., Alea, N., Habermas, T., & Rubin, D. C. (2005). A TALE of three functions: The self–reported uses of autobiographical memory. *Social Cognition*, *23*, 91–117. https://doi.org/10.1521/soco.23.1.91.59198

Bluck, S., Levine, L. J., & Laulhere, T. M. (1999). Autobiographical remembering and hypermnesia: A comparison of older and younger adults. *Psychology and Aging*, *14*, 671–682. https://doi.org/10.1037/0882-7974.14.4.671

Bouwmeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, *65*, 32–41. https://doi.org/10.1016/j.jml.2011.02.005

Bower, G. H. (1970). Organizational factors in memory. *Cognitive Psychology*, *1*, 18–46. https://doi.org/10.1016/0010-0285(70)90003-4

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*, 129–148. https://doi.org/10.1037/0003-066X.36.2.129

Bower, G. H., & Forgas, J. P. (2000). Affect, memory, and social cognition. In E. Eich, J. F. Killstrom, G. H. Bower, J. P. Forgas, & P. M. Niedenthal (Eds.), *Cognition and emotion* (pp. 87–168). New York, NY: Oxford University Press.

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language*, *66*, 407–415. https://doi.org/10.1016/j.jml.2011.12.009

Brewer, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered. Ecological and traditional approaches to the study of memory* (pp. 21–90). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511664014.004

Brewer, W. F. (1992). Phenomenal experience in laboratory and autobiographical memory tasks. In M. A. Conway, D. C. Rubin, H. Spinnler, & W. A. Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory* (pp. 31–51). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-015-7967-4_3

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*, 73–99. https://doi.org/10.1016/0010-0277(77)90018-X

Brunyé, T. T., Smith, A. M., Hendel, D., Gardony, A. L., Martis, S. B., & Taylor, H. A. (2019). Retrieval practice enhances near but not far transfer of spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0000710

Buchanan, T., & Adolphs, R. (2002). The role of the human amygdala in emotional modulation of long-term declarative memory. In S. C. Moore & M. Oaksford (Eds.), *Emotional cognition. From brain to behaviour* (pp. 9–34). Amsterdam, Netherlands: Benjamins. https://doi.org/10.1075/aicr.44.02buc

Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1934–1947. https://doi.org/10.1037/xlm0000427

Buchin, Z. L., & Mulligan, N. W. (2019). Divided attention and the encoding effects of retrieval. *Quarterly Journal of Experimental Psychology*, *72*, 2474–2494. https://doi.org/10.1177/1747021819847141

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118–1133. https://doi.org/10.1037/a0019902

Butler, A. C., Black-Maier, A. C., Raley, N. D., & Marsh, E. J. (2017). Retrieving and applying knowledge to different examples promotes transfer of learning. *Journal of Experimental Psychology: Applied*, *23*, 433–446. https://doi.org/10.1037/xap0000142

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*, 273–281. https://doi.org/10.1037/1076-898X.13.4.273

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

Butler, A. C., & Roediger, H. L., III (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*, 604–616. https://doi.org/10.3758/MC.36.3.604

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*, 294–299. https://doi.org/10.1016/S0166-2236(97)01214-9

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*, 30–41. https://doi.org/10.1016/j.cedpsych.2008.07.001

Calvo, M. G., & Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: Motivationally biased attention. *Motivation and Emotion*, *28*, 221–243. https://doi.org/10.1023/B:MOEM.0000040153.26156.ed

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*, 1563–1569. https://doi.org/10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1547–1552. https://doi.org/10.1037/a0024140

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268–276. https://doi.org/10.3758/BF03193405

Carpenter, S. K., & Kelly, J. W. (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin & Review*, *19*, 443–448. https://doi.org/10.3758/s13423-012-0221-2

Carpenter, S. K., Lund, T. J. S., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. *Educational Psychology Review*, *28*, 353–375. https://doi.org/10.1007/s10648-015-9311-9

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, *14*, 474–478. https://doi.org/10.3758/BF03194092

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, *23*, 760–771. https://doi.org/10.1002/acp.1507

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, *13*, 826–830. https://doi.org/10.3758/BF03194004

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438–448. https://doi.org/10.3758/MC.36.2.438

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642. https://doi.org/10.3758/BF03202713

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*, 1111–1146. https://doi.org/10.1037/bul0000166

Cho, K. W., & Powers, A. (2019). Testing enhances both memorization and conceptual learning of categorical materials. *Journal of Applied Research in Memory and Cognition*, *8*, 166–177. https://doi.org/10.1016/j.jarmac.2019.01.003

Christianson, S.-Å., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition and Emotion*, *5*, 81–108. https://doi.org/10.1080/02699939108411027

Christianson, S.-Å., & Safer, M. A. (1996). Emotional events and emotions in autobiographical memories. In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 218–243). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511527913.009

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, *11*, 393–399. https://doi.org/10.1016/j.tics.2007.08.005

Clore, G. L., & Huntsinger, J. R. (2009). How the object of affect guides its impact. *Emotion Review*, *1*, 39–54. https://doi.org/10.1177/1754073908097185

Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Cognition and Education*, *2*, 95–100. https://doi.org/10.1016/j.jarmac.2013.04.001

Cogliano, M., Kardash, C. M., & Bernacki, M. L. (2019). The effects of retrieval practice and prior topic knowledge on test performance and confidence judgments. *Contemporary Educational Psychology*, *56*, 117–129. https://doi.org/10.1016/j.cedpsych.2018.12.001

Congleton, A., & Rajaram, S. (2012). The origin of the interaction between learning method and delay in the testing effect: The roles of processing and conceptual retrieval organization. *Memory & Cognition*, *40*, 528–539. https://doi.org/10.3758/s13421-011-0168-y

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*, 594–628. https://doi.org/10.1016/j.jml.2005.08.005

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*, 261–288. https://doi.org/10.1037//0033-295X.107.2.261

Conway, M. A., & Williams, H. L. (2008). Autobiographical memory. In J. H. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory* (1st ed., pp. 893–909). Oxford, England: Academic Press.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, *21*, 919–940. https://doi.org/10.1080/09541440802413505

D'Argembeau, A., & Van der Linden, M. (2004). Influence of affective meaning on memory for contextual information. *Emotion*, 173–188. https://doi.org/10.1037/1528-3542.4.2.173

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, *91*, 66–73. https://doi.org/10.1037/h0031836

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*, 693–700. https://doi.org/10.1016/j.conb.2006.10.012

De Jonge, M., Tabbers, H. K., & Rikers, R. M. J. P. (2015). The effect of testing on the retention of coherent and incoherent text material. *Educational Psychology Review*, *27*, 305–315. https://doi.org/10.1007/s10648-015-9300-z

Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2014). The testing effect for learning principles and procedures from texts. *The Journal of Educational Research*, *107*, 357–364. https://doi.org/10.1080/00220671.2013.823370

Dobson, J. L., & Linderholm, T. (2015). Self-testing promotes superior retention of anatomy and physiology information. *Advances in Health Sciences Education: Theory and Practice*, *20*, 149–161. https://doi.org/10.1007/s10459-014-9514-8

Eich, E., & Forgas, J. P. (2003). Mood, cognition, and memory. In I. B. Weiner (Series Ed.) & A. F. Healy & R. W. Proctor (Vol. Eds.), *Handbook of psychology: Vol. 4.*

Experimental psychology (pp. 61–83). Hoboken, NJ: Wiley. https://doi.org/10.1002/0471264385.wei0403

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*, 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Chichester, England: Wiley. https://doi.org/10.1002/0470013494.ch3

Ellis, H. C., & Ashbrook, P. W. (1988). Resource allocation model of the effects of depressed mood states on memory. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behaviour. New evidence and integrative attempts* (pp. 25–32). Toronto, Canada: Hogrefe.

Elsey, J. W. B., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin*, *144*, 797–848. https://doi.org/10.1037/bul0000152

Emmerdinger, K. J., Kuhbandner, C., & Berchtold, F. (2018). Testing emotional memories: Does negative emotional significance influence the benefit received from testing? *Cognition and Emotion*, *32*, 852–859. https://doi.org/10.1080/02699931.2017.1359496

E-Prime (Version 2.0) [Computer software]. Sharpsburg, PA: Psychology Software Tools.

Erdelyi, M. H. (2010). The ups and downs of memory. *American Psychologist*, *65*, 623–633. https://doi.org/10.1037/a0020440

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion*, *6*, 409–434. https://doi.org/10.1080/02699939208409696

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*, 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fazio, L. K., & Marsh, E. J. (2019). Retrieval-based learning in children. *Current Directions in Psychological Science*, *28*, 111–116. https://doi.org/10.1177/0963721418806673

Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 85–98). Mahwah, NJ: Erlbaum.

Fiedler, K., & Hütter, M. (2014). Memory and emotion. In T. J. Perfect & S. D. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 145–161). London, England: SAGE. https://doi.org/10.4135/9781446294703.n9

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 632–652. https://doi.org/10.1037/a0026215

Finn, B., & Roediger, H. L., III (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later recall. *Psychological Science*, *22*, 781–786. https://doi.org/10.1177/0956797611407932

Finn, B., Roediger, H. L., III, & Rosenzweig, E. (2012). Reconsolidation from negative emotional pictures: Is successful retrieval required? *Memory & Cognition*, *40*, 1031–1045. https://doi.org/10.3758/s13421-012-0203-7

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, *56*, 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*, 1367–1378. https://doi.org/10.1098/rstb.2004.1512

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion*, *19*, 313–332. https://doi.org/10.1080/02699930441000238

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*, 499–526. https://doi.org/10.1002/acp.1287

García-Bajos, E., & Migueles, M. (2017). Retrieval of past and future positive and negative autobiographical experiences. *Cognition and Emotion*, *31*, 1260–1267. https://doi.org/10.1080/02699931.2016.1204988

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, *13*, 34–40. https://doi.org/10.1111/1467-9280.00406

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *40*, 1–104.

Giovanello, K. S., Schnyer, D. M., & Verfaellie, M. (2004). A critical role for the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus*, *14*, 5–8. https://doi.org/10.1002/hipo.10182

Giovanello, K. S., Verfaellie, M., & Keane, M. M. (2003). Disproportionate deficit in associative recognition relative to item recognition in global amnesia. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 186–194. https://doi.org/10.3758/CABN.3.3.186

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399. https://doi.org/10.1037/0022-0663.81.3.392

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, *28*, 135–142. https://doi.org/10.1002/acp.2956

Gray, E. K., & Watson, D. (2007). Assessing positive and negative affect via self-report. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 171–183). New York, NY: Oxford University Press.

Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, *9*, 2412. https://doi.org/10.3389/fpsyg.2018.02412

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, *9*, 87–108. https://doi.org/10.1080/02699939508408966

Guillet, R., & Arndt, J. (2009). Taboo words: The effect of emotion on memory for peripheral information. *Memory & Cognition*, *37*, 866–879. https://doi.org/10.3758/MC.37.6.866

Haager, J. S., Kuhbandner, C., & Pekrun, R. (2014). Overcoming fixed mindsets: The role of affect. *Cognition and Emotion*, *28*, 756–767. https://doi.org/10.1080/02699931.2013.851645

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*, 801–812. https://doi.org/10.1037/a0023219

Hamann, S., Ely, T., Grafton, S. T., & Kilts, C. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, *2*, 289–293. https://doi.org/10.1038/6404

Hannula, D. E., Tranel, D., Allen, J. S., Kirchhoff, B. A., Nickel, A. E., & Cohen, N. J. (2015). Memory for items and relationships among items embedded in realistic scenes: Disproportionate relational memory impairments in amnesia. *Neuropsychology*, *29*, 126–138. https://doi.org/10.1037/neu0000119

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 290–296. https://doi.org/10.1037/a0028468

Heitmann, S., Grund, A., Berthold, K., Fries, S., & Roelle, J. (2018). Testing is more desirable when it is adaptive and still desirable when compared to note-taking. *Frontiers in Psychology*, *9*, 2596. https://doi.org/10.3389/fpsyg.2018.02596

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*, 47–77. https://doi.org/10.3102/00346543058001047

Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology*, *28*, 597–606. https://doi.org/10.1002/acp.3032

Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory*, *19*, 290–304. https://doi.org/10.1080/09658211.2011.560121

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562–567. https://doi.org/10.1016/S0022-5371(71)80029-4

Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of Life Reviews*, *7*, 88–131. https://doi.org/10.1016/j.plrev.2010.01.006

Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, *28*, 853–873. https://doi.org/10.1007/s10648-015-9349-8

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421–445. https://doi.org/10.1006/jmla.1993.1023

Huntsinger, J. R., Sinclair, S., & Clore, G. L. (2009). Affective regulation of implicitly measured stereotypes and attitudes: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *45*, 560–566. https://doi.org/10.1016/j.jesp.2009.01.007

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, *100*, 68–90. https://doi.org/10.1037/0033-295X.100.1.68

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology*, *60*, 1–25. https://doi.org/10.1146/annurev.psych.60.110707.163539

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*, 1441–1451. https://doi.org/10.1037/a0020636

Jarosz, A., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*, 2–9. https://doi.org/10.7771/1932-6246.1167

JASP Team (2018). JASP (Version 0.9.2) [Computer software]. Retrieved from https://jasp-stats.org/

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test…or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*, 307–329. https://doi.org/10.1007/s10648-013-9248-9

Jia, X., Gao, C., Cui, L., & Guo, C. (2018). Does emotion arousal influence the benefit received from testing: Insights from neural correlates of retrieval mode effect. *NeuroReport*, *29*, 1449–1455. https://doi.org/10.1097/WNR.0000000000001130

Jia, X., Gao, C., Cui, L., & Guo, C. (2019). The role of emotion arousal in the retrieval practice effect. *Experimental Brain Research*, *237*, 3241–3252. https://doi.org/10.1007/s00221-019-05658-0

Jones, A. C., Wardlow, L., Pan, S. C., Zepeda, C., Heyman, G. D., Dunlosky, J., & Rickard, T. C. (2016). Beyond the rainbow: Retrieval practice leads to better spelling than does rainbow writing. *Educational Psychology Review*, *28*, 385–400. https://doi.org/10.1007/s10648-015-9330-6

Jones, H. E. (1923-1924). The effects of examination on the performance of learning. *Archives of Psychology*, *10*, 1–70.

Jung, N., Wranke, C., Hamburger, K., & Knauff, M. (2014). How emotions affect logical reasoning: Evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. *Frontiers in Psychology*, *5*, 570. https://doi.org/10.3389/fpsyg.2014.00570

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*, *38*, 1009–1017. https://doi.org/10.3758/MC.38.8.1009

Kang, S. H. K., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin & Review*, *18*, 998–1005. https://doi.org/10.3758/s13423-011-0113-x

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558. https://doi.org/10.1080/09541440601056620

Kang, S. H.K., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? *Journal of Applied Research in Memory and Cognition, 3,* 183–188. https://doi.org/10.1016/j.jarmac.2014.05.006

Karpicke, J. D. (2016). A powerful way to improve learning and memory. *APA Psychological Science Agenda*, *30*. Retrieved from http://www.apa.org/science/about/psa/2016/06/learning-memory.aspx

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*, 317–326. https://doi.org/10.1007/s10648-015-9309-3

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1250–1257. https://doi.org/10.1037/a0023436

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772. https://doi.org/10.1126/science.1199327

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, *7*, 350. https://doi.org/10.3389/fpsyg.2016.00350

Karpicke, J. D., Blunt, J. R., Sumeracki, M., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary school children. *Journal of Applied Research in Memory and Cognition*, *3*, 198–206. https://doi.org/10.1016/j.jarmac.2014.07.008

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

Karpicke, J. D., & Roediger, H. L., III (2007a). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 704–719. https://doi.org/10.1037/0278-7393.33.4.704

Karpicke, J. D., & Roediger, H. L., III (2007b). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L., III (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, *38*, 116–124. https://doi.org/10.3758/MC.38.1.116

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*, 227–239. https://doi.org/10.1016/j.jml.2009.11.010

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, *14*, 785–794. https://doi.org/10.1162/08989290260138672

Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences*, *15*, 241–251. https://doi.org/10.1515/revneuro.2004.15.4.241

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*, 99–113. https://doi.org/10.1177/1754073908100432

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of* the United States of America, 101, 3310–3315. https://doi.org/10.1073/pnas.0306408101

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007a). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*, 575–591. https://doi.org/10.1016/j.jml.2006.05.004

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007b). How negative emotion enhances the visual specificity of a memory. *Journal of Cognitive Neuroscience*, *19*, 1872–1887. https://doi.org/10.1162/jocn.2007.19.11.1872

Khanna, M. M. (2015). Ungraded pop quizzes: Test-enhanced learning without all the anxiety. *Teaching of Psychology*, *42*, 174–178. https://doi.org/10.1177/0098628315573144

Khanna, M. M., & Cortese, M. J. (2016). The benefits of quizzing in content-focused versus skills-focused courses. *Scholarship of Teaching and Learning in Psychology*, *2*, 87–97. https://doi.org/10.1037/stl0000051

Knörzer, L., Brünken, R., & Park, B. (2016). Facilitators or suppressors: Effects of experimentally induced emotions on multimedia learning. *Learning and Instruction*, *44*, 97–107. https://doi.org/10.1016/j.learninstruc.2016.04.002

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*, 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 989–998. https://doi.org/10.1037/a0015729

Kornell, N., Klein, P. J., & Rawson, K. A. (2015). Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 283–294. https://doi.org/10.1037/a0037850

Kornell, N., Rabelo, V. C., & Klein, P. J. (2012). Tests enhance learning? Compared to what? *Journal of Applied Research in Memory and Cognition*, *1*, 257–259. https://doi.org/10.1016/j.jarmac.2012.10.002

Kromann, C. B., Jensen, M. L., & Ringsted, C. (2009). The effect of testing on skills learning. *Medical Education*, *43*, 21–27. https://doi.org/10.1111/j.1365-2923.2008.03245.x

Kuhbandner, C., Aslan, A., Emmerdinger, K., & Murayama, K. (2016). Providing extrinsic reward for test performance undermines long-term memory acquisition. *Frontiers in Psychology*, *7*, 79. https://doi.org/10.3389/fpsyg.2016.00079

Kuhbandner, C., Bäuml, K.-H., & Stiedl, F. C. (2009). Retrieval-induced forgetting of negative stimuli: The role of emotional intensity. *Cognition and Emotion*, *23*, 817–830. https://doi.org/10.1080/02699930802204768

Kuhbandner, C., Lichtenfeld, S., & Pekrun, R. (2011). Always look on the broad side of life: Happiness increases the breadth of sensory memory. *Emotion*, *11*, 958–964. https://doi.org/10.1037/a0024075

Kuhbandner, C., & Pekrun, R. (2010). Emotion und Gedächtnis [Emotion and memory]. In H.-P. Trolldenier, W. Lenhard, & P. Marx (Eds.), *Brennpunkte der Gedächtnisforschung. Entwicklungs- und pädagogisch-psychologische Perspektiven* (pp. 35–51). Göttingen: Hogrefe.

Kuhbandner, C., & Pekrun, R. (2013). Affective state influences retrieval-induced forgetting for integrated knowledge. *PLoS ONE*, *8*, e56617. https://doi.org/10.1371/journal.pone.0056617

Kuhbandner, C., Pekrun, R., & Maier, M. A. (2010). The role of positive and negative affect in the “mirroring” of other persons' actions. *Cognition and Emotion*, *24*, 1182–1190. https://doi.org/10.1080/02699930903119196

Kuhbandner, C., Spitzer, B., & Pekrun, R. (2011). Read-out of emotional information from iconic memory: The longevity of threatening stimuli. *Psychological Science*, *22*, 695–700. https://doi.org/10.1177/0956797611406445

Kuiper, N. A., & Rogers, T. B. (1979). Encoding of personal information: Self–other differences. *Journal of Personality and Social Psychology*, *37*, 499–514. https://doi.org/10.1037/0022-3514.37.4.499

Kuyken, W., & Dalgleish, T. (2011). Overgeneral autobiographical memory in adolescents at risk for depression. *Memory*, *19*, 241–250. https://doi.org/10.1080/09658211.2011.554421

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*, 54–64. https://doi.org/10.1038/nrn1825

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, *9*, 490–493. https://doi.org/10.1111/1467-9280.00090

Landauer, T. K., & Eldridge, L. (1967). Effect of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology*, *75*, 290–298. https://doi.org/10.1037/h0025047

Laney, C., Campbell, H. V., Heuer, F., & Reisberg, D. (2004). Memory for thematically arousing events. *Memory & Cognition*, *32*, 1149–1159. https://doi.org/10.3758/BF03196888

Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, *27*, 291–304. https://doi.org/10.1007/s10648-015-9296-4

Leiner, D. J. (2014). SoSci Survey (Version 2.5.00-i) [Computer software]. Retrieved from https://www.soscisurvey.de

Levine, L. J., & Edelstein, R. S. (2009). Emotion and memory narrowing: A review and goal-relevance approach. *Cognition and Emotion*, *23*, 833–875. https://doi.org/10.1080/02699930902738863

Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology*, *50*, 994–1000. https://doi.org/10.1037/a0035202

Lundqvist, D., Bruce, N., & Öhman, A. (2015). Finding an emotional face in a crowd: Emotional and perceptual stimulus factors influence visual search efficiency. *Cognition and Emotion*, *29*, 621–633. https://doi.org/10.1080/02699931.2014.927352

Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. S. (2019). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review.* Advance online publication. https://doi.org/10.1007/s10648-019-09489-x

Maccallum, F., McConkey, K. M., Bryant, R. A., & Barnier, A. J. (2000). Specific autobiographical memory following hypnotically induced mood state. *The International Journal of Clinical and Experimental Hypnosis*, *48*, 361–373. https://doi.org/10.1080/00207140008410366

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.

Madan, C. R., Caplan, J. B., Lau, C. S. M., & Fujiwara, E. (2012). Emotional arousal does not enhance association-memory. *Journal of Memory and Language*, *66*, 695–716. https://doi.org/10.1016/j.jml.2012.04.001

Madigan, S. A., & McCabe, L. (1971). Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 101–106. https://doi.org/10.1016/S0022-5371(71)80101-9

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*, 33–52. https://doi.org/10.1111/j.1745-6916.2007.00028.x

Mather, M., & Nesmith, K. (2008). Arousal-enhanced location memory for pictures. *Journal of Memory and Language*, *58*, 449–464. https://doi.org/10.1016/j.jml.2007.01.004

McCabe, D., Langer, K. G., Borod, J. C., & Bender, H. A. (2011). Practice effects. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 1988–1989). New York, NY: Springer. https://doi.org/10.1007/978-0-387-79948-3_1139

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494–513. https://doi.org/10.1080/09541440701326154

McDaniel, M. A., Bugg, J. M., Liu, Y., & Brick, J. (2015). When does the test-study-test sequence optimize learning and retention? *Journal of Experimental Psychology: Applied*, *21*, 370–382. https://doi.org/10.1037/xap0000063

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

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L., III (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*, 360–372. https://doi.org/10.1002/acp.2914

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., III, & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*, 3–21. https://doi.org/10.1037/xap0000004

McGaugh, J. L. (2000). Memory – a century of consolidation. *Science*, *287*, 248–251. https://doi.org/10.1126/science.287.5451.248

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*, 1–28. https://doi.org/10.1146/annurev.neuro.27.070203.144157

Meyer, A. N. D., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*, 142–147. https://doi.org/10.1037/a0030890

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1474–1486. https://doi.org/10.1037/xlm0000486

Moors, A. (2013). On the causal role of appraisal in emotion. *Emotion Review*, *5*, 132–140. https://doi.org/10.1177/1754073912463601

Moors, A. (2017). Integration of two skeptical emotion theories: Dimensional appraisal theory and Russell's psychological construction theory. *Psychological Inquiry*, *28*, 1–19. https://doi.org/10.1080/1047840X.2017.1235900

Moors, A., & Scherer, K. R. (2013). The role of appraisal in emotion. In M. Robinson, E. Watkins, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 135–155). New York, NY: Guilford Press.

Moreira, B. F. T., Pinto, T. S. S., Justi, F. R. R., & Jaeger, A. (2019). Retrieval practice improves learning in children with diverse visual word recognition skills. *Memory*, *27*, 1423–1437. https://doi.org/10.1080/09658211.2019.1668017

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

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*, *19*, 779–798. https://doi.org/10.1002/acp.1115

Morris, W. N., & Reilly, N. P. (1987). Toward the self-regulation of mood: Theory and research. *Motivation and Emotion*, *11*, 215–249. https://doi.org/10.1007/BF01001412

Mulligan, N. W., & Peterson, D. J. (2015). The negative testing and negative generation effects are eliminated by delay. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1014–1025. https://doi.org/10.1037/xlm0000070

Murray, B. D., Muscatell, K. A., & Kensinger, E. A. (2011). Effects of emotion and age on performance during a think/no-think memory task. *Psychology and Aging*, *26*, 940–955. https://doi.org/10.1037/a0023214

Muscatell, K. A., Addis, D. R., & Kensinger, E. A. (2010). Self-involvement modulates the effective connectivity of the autobiographical memory network. *Social Cognitive and Affective Neuroscience*, *5*, 68–76. https://doi.org/10.1093/scan/nsp043

Nashiro, K., & Mather, M. (2011). How arousal affects younger and older adults' memory binding. *Experimental Aging Research*, *37*, 108–128. https://doi.org/10.1080/0361073X.2011.536746

Naveh-Benjamin, M., Craik, F. I., Perretta, J. G., & Tonev, S. T. (2000). The effects of divided attention on encoding and retrieval processes: The resiliency of retrieval processes. *The Quarterly Journal of Experimental Psychology Section A*, *53*, 609–625. https://doi.org/10.1080/713755914

Naveh-Benjamin, M., Kilb, A., & Fisher, T. (2006). Concurrent task effects on memory encoding and retrieval: Further support for an asymmetry. *Memory & Cognition*, *34*, 90–101. https://doi.org/10.3758/BF03193389

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*, 257–268. https://doi.org/10.1037/1528-3542.6.2.257

Nunes, L. D., & Karpicke, J. D. (2015). Retrieval-based learning: Research at the interface between cognitive science and education. In R. A. Scott, M. C. Buchmann, & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences. An interdisciplinary, searchable, and linkable resource.* Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118900772.etrds0289

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*, 18–22. https://doi.org/10.1037/0022-0663.74.1.18

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*, 466–478.

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*, 563–575. https://doi.org/10.1037/edu0000074

Pan, S. C., Pashler, H., Potter, Z. E., & Rickard, T. C. (2015). Testing enhances learning across a range of episodic memory abilities. *Journal of Memory and Language*, *83*, 53–61. https://doi.org/10.1016/j.jml.2015.04.001

Pan, S. C., & Rickard, T. C. (2017). Does retrieval practice enhance learning and transfer relative to restudy for term-definition facts? *Journal of Experimental Psychology: Applied*, *23*, 278–292. https://doi.org/10.1037/xap0000124

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*, 710–756. https://doi.org/10.1037/bul0000151

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, 286. https://doi.org/10.3389/fpsyg.2014.00286

Pastötter, B., & Bäuml, K.-H. T. (2019). Testing enhances subsequent learning in older adults. *Psychology and Aging*, *34*, 242–250. https://doi.org/10.1037/pag0000307

Payne, B. K., & Corrigan, E. (2007). Emotional constraints on intentional forgetting. *Journal of Experimental Social Psychology*, *43*, 780–786. https://doi.org/10.1016/j.jesp.2006.07.005

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, *101*, 5–27. https://doi.org/10.1037//0033-2909.101.1.5

Payne, J. D., Stickgold, R., Swanberg, K., & Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science*, *19*, 781–788. https://doi.org/10.1111/j.1467-9280.2008.02157.x

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105. https://doi.org/10.1207/S15326985EP3702_4

Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books.

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*, 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*, 335. https://doi.org/10.1126/science.1191465

Pyc, M. A., & Rawson, K. A. (2012). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 737–746. https://doi.org/10.1037/a0026166

Rawson, K. A. (2015). The status of the testing effect for complex materials: Still a winner. *Educational Psychology Review*, *27*, 327–331. https://doi.org/10.1007/s10648-015-9308-4

Reeder, G. D., McCormick, C. B., & Esselman, E. D. (1987). Self-referent processing and recall of prose. *Journal of Educational Psychology*, *79*, 243–248. https://doi.org/10.1037/0022-0663.79.3.243

Reisberg, D., & Heuer, F. (2004). Memory for emotional events. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 3–41). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195158564.003.0001

Riggs, L., McQuiggan, D. A., Farb, N., Anderson, A. K., & Ryan, J. D. (2011). The role of overt attention in emotion-modulated memory. *Emotion*, *11*, 776–785. https://doi.org/10.1037/a0022591

Robey, A. (2019). The benefits of testing: Individual differences based on student factors. *Journal of Memory and Language*, *108*, 104029. https://doi.org/10.1016/j.jml.2019.104029

Robinson, J. (1976). Sampling autobiographical memory. *Cognitive Psychology*, *8*, 578–595. https://doi.org/10.1016/0010-0285(76)90020-7

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382–395. https://doi.org/10.1037/a0026252

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., III, & Karpicke, J. D. (2018). Reflections on the resurgence of interest in the testing effect. *Perspectives on Psychological Science*, *13*, 236–241. https://doi.org/10.1177/1745691617718873

Roediger, H. L., III, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, *1*, 242–248. https://doi.org/10.1016/j.jarmac.2012.09.002

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677–688. https://doi.org/10.1037/0022-3514.35.9.677

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 233–239. https://doi.org/10.1037/a0017678

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236. https://doi.org/10.1068/p5117

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 383–388. https://doi.org/10.1073/pnas.0605198104

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432–1463. https://doi.org/10.1037/a0037559

Rowland, C. A., & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory*, *23*, 403–419. https://doi.org/10.1080/09658211.2014.889710

Rowland, C. A., Littrell-Baez, M. K., Sensenig, A. E., & DeLosh, E. L. (2014). Testing effects in mixed- versus pure-list designs. *Memory & Cognition*, *42*, 912–921. https://doi.org/10.3758/s13421-014-0404-3

Rubin, D. C., & Kozin, M. (1984). Vivid memories. *Cognition*, *16*, 81–95. https://doi.org/10.1016/0010-0277(84)90037-4

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion*, *23*, 1259–1283. https://doi.org/10.1080/02699930902809375

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805–819. https://doi.org/10.1037//0022-3514.76.5.805

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*, 493–502. https://doi.org/10.1037/0022-3514.57.3.493

Sakaki, M., Kuhbandner, C., Mather, M., & Pekrun, R. (2014). Memory suppression can help people "unlearn" behavioral responses – but only for nonemotional memories. *Psychonomic Bulletin & Review*, *21*, 136–141. https://doi.org/10.3758/s13423-013-0480-6

Sara, S. J. (2010). Reactivation, retrieval, replay and reconsolidation in and out of sleep: Connecting the dots. *Frontiers in Behavioral Neuroscience*, *4*, 185. https://doi.org/10.3389/fnbeh.2010.00185

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, *24*, 1153–1172. https://doi.org/10.1080/02699930903274322

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*, 695–729. https://doi.org/10.1177/0539018405058216

Schiefele, U. (2001). The role of interest in motivation and learning. In J. M. Collis & S. Messick (Eds.), *Intelligence and personality. Bridging the gap in theory and measurement* (pp. 163–194). Mahwah, NJ: Erlbaum.

Schmidt, S. R., & Saari, B. (2007). The emotional memory effect: Differential processing or item distinctiveness? *Memory & Cognition*, *35*, 1905–1916. https://doi.org/10.3758/BF03192924

Schwarz, N. (2012). Feelings-as-information theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 289–308). Thousand Oaks, CA: SAGE. https://doi.org/10.4135/9781446249215.n15

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, *4*, 27–41. https://doi.org/10.1080/08917779108248762

Sharot, T., & Yonelinas, A. P. (2008). Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*, *106*, 538–547. https://doi.org/10.1016/j.cognition.2007.03.002

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin*, *143*, 636–675. https://doi.org/10.1037/bul0000100

Smith, A. M., Davis, F. C., & Thomas, A. K. (2018). Criterial learning is not enough: Retrieval practice is necessary for improving post-stress memory accessibility. *Behavioral Neuroscience*, *132*, 161–170. https://doi.org/10.1037/bne0000240

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science*, *354*, 1046–1048. https://doi.org/10.1126/science.aah5067

Smith, A. M., Race, E., Davis, F. C., & Thomas, A. K. (2019). Retrieval practice improves item memory but not source memory in the context of stress. *Brain and Cognition*, *133*, 24–32. https://doi.org/10.1016/j.bandc.2018.12.005

Spachtholz, P., & Kuhbandner, C. (2017). Visual long-term memory is not unitary: Flexible storage of visual information as features or objects as a function of affect. *Cognitive, Affective, & Behavioral Neuroscience*, *17*, 1141–1150. https://doi.org/10.3758/s13415-017-0538-4

Spachtholz, P., Kuhbandner, C., & Pekrun, R. (2014). Negative affect improves the quality of memories: Trading capacity for precision in sensory and working memory. *Journal of Experimental Psychology: General*, *143*, 1450–1456. https://doi.org/10.1037/xge0000012

Spachtholz, P., Kuhbandner, C., & Pekrun, R. (2016). Affect influences feature binding in memory: Trading between richness and strength of memory representations. *Emotion*, *16*, 1067–1073. https://doi.org/10.1037/emo0000199

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656. https://doi.org/10.1037/h0063404

Stanny, C. J., & Johnson, T. C. (2000). Effects of stress induced by a simulated shooting on recall by police and citizen witnesses. *The American Journal of Psychology*, *113*, 359–386. https://doi.org/10.2307/1423364

Steinmetz, K. R. M., & Kensinger, E. A. (2013). The emotion-induced memory trade-off: More than an effect of overt attention? *Memory & Cognition*, *41*, 69–81. https://doi.org/10.3758/s13421-012-0247-8

Stone, C. B., Barnier, A. J., Sutton, J., & Hirst, W. (2013). Forgetting our personal past: Socially shared retrieval-induced forgetting of autobiographical memories. *Journal of Experimental Psychology: General*, *142*, 1084–1099. https://doi.org/10.1037/a0030739

Storbeck, J. (2013). Negative affect promotes encoding of and memory for details at the expense of the gist: Affect, encoding, and false memories. *Cognition and Emotion*, *27*, 800–819. https://doi.org/10.1080/02699931.2012.741060

Storbeck, J., & Clore, G. L. (2008). The affective regulation of cognitive priming. *Emotion*, *8*, 208–215. https://doi.org/10.1037/1528-3542.8.2.208

Sumowski, J. F., Wood, H. G., Chiaravalloti, N., Wylie, G. R., Lengenfelder, J., & DeLuca, J. (2010). Retrieval practice: A simple strategy for improving memory after traumatic brain injury. *Journal of the International Neuropsychological Society*, *16*, 1147–1150. https://doi.org/10.1017/S1355617710001128

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*, 371–394. https://doi.org/10.1037/0033-2909.121.3.371

Szőllősi, Á., Keresztes, A., Novák, B., Szászi, B., Kéri, S., & Racsmány, M. (2017). The testing effect is preserved in stressful final testing environment. *Applied Cognitive Psychology*, *31*, 615–622. https://doi.org/10.1002/acp.3363

Talmi, D., Luk, B. T. C., McGarry, L. M., & Moscovitch, M. (2007). The contribution of relatedness and distinctiveness to emotionally-enhanced memory. *Journal of Memory and Language*, *56*, 555–574. https://doi.org/10.1016/j.jml.2007.01.002

Talmi, D., & McGarry, L. M. (2012). Accounting for immediate emotional memory enhancement. *Journal of Memory and Language*, *66*, 93–108. https://doi.org/10.1016/j.jml.2011.07.009

Taylor, T. L., Quinlan, C. K., & Vullings, K. C. H. (2018). Decomposing item-method directed forgetting of emotional pictures: Equivalent costs and no benefits. *Memory & Cognition*, *46*, 132–147. https://doi.org/10.3758/s13421-017-0751-y

Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, *35*, 668–678. https://doi.org/10.3758/BF03193305

Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. L. (1996). *Autobiographical memory: Remembering what and remembering when*. Mahwah, NJ: Erlbaum.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*, 252–257. https://doi.org/10.1027/1618-3169.56.4.252

Tran, R., Rohrer, D., & Pashler, H. (2015). Retrieval practice: The lack of transfer to deductive inferences. *Psychonomic Bulletin & Review*, *22*, 135–140. https://doi.org/10.3758/s13423-014-0646-x

Trumbo, M. C., Leiting, K. A., McDaniel, M. A., & Hodge, G. K. (2016). Effects of reinforcement on test-enhanced learning in a large, diverse introductory college psychology course. *Journal of Experimental Psychology: Applied*, *22*, 148–160. https://doi.org/10.1037/xap0000082

Tse, C.-S., Balota, D. A., & Roediger, H. L., III (2010). The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. *Psychology and Aging*, *25*, 833–845. https://doi.org/10.1037/a0019933

Tse, C.-S., & Pu, X. (2012). The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. *Journal of Experimental Psychology: Applied*, *18*, 253–264. https://doi.org/10.1037/a0029190

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175–184. https://doi.org/10.1016/S0022-5371(67)80092-6

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

Van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science*, *36*, 1532–1541. https://doi.org/10.1111/cogs.12002

Van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, *27*, 265–289. https://doi.org/10.1007/s10648-015-9297-3

Van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*, 247–264. https://doi.org/10.1007/s10648-015-9310-x

Vanlessen, N., De Raedt, R., Koster, E. H. W., & Pourtois, G. (2016). Happy heart, smiling eyes: A systematic review of positive mood effects on broadening of visuospatial attention. *Neuroscience & Biobehavioral Reviews*, *68*, 816–837. https://doi.org/10.1016/j.neubiorev.2016.07.001

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*, 534–538. https://doi.org/10.3758/BRM.41.2.534

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Wagenaar, W. A., & Groeneweg, J. (1990). The memory of concentration camp survivors. *Applied Cognitive Psychology*, *4*, 77–87. https://doi.org/10.1002/acp.2350040202

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J.,. . . Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*, 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagner, U., Hallschmid, M., Rasch, B., & Born, J. (2006). Brief sleep after learning keeps emotional memories alive for years. *Biological Psychiatry*, *60*, 788–790. https://doi.org/10.1016/j.biopsych.2006.03.061

Walker, W. R., & Skowronski, J. J. (2009). The fading affect bias: But what the hell is it for? *Applied Cognitive Psychology*, *23*, 1122–1136. https://doi.org/10.1002/acp.1614

Walker, W. R., Skowronski, J. J., Gibbons, J. A., Vogl, R. J., & Ritchie, T. D. (2009). Why people rehearse their memories: Frequency of use and relations to the intensity of emotions associated with autobiographical memories. *Memory*, *17*, 760–773. https://doi.org/10.1080/09658210903107846

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant – and memory helps to keep it that way! *Review of General Psychology*, *7*, 203–210. https://doi.org/10.1037/1089-2680.7.2.203

Walker, W. R., Vogl, R. J., & Thompson, C. P. (1997). Autobiographical memory: Unpleasantness fades faster than pleasantness over time. *Applied Cognitive Psychology*, *11*, 399–413. https://doi.org/10.1002/(SICI)1099-0720(199710)11:5%3C399::AID-ACP462%3E3.0.CO;2-E

Wang, Q., Lee, D., & Hou, Y. (2017). Externalising the autobiographical self: Sharing personal memories online facilitated memory retention. *Memory*, *25*, 772–776. https://doi.org/10.1080/09658211.2016.1221115

Waring, J. D., & Kensinger, E. A. (2011). How emotion leads to selective memory: Neuroimaging evidence. *Neuropsychologia*, *49*, 1831–1842. https://doi.org/10.1016/j.neuropsychologia.2011.03.007

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*, 820–838. https://doi.org/10.1037/0022-3514.76.5.820

Weinstein, Y., McDermott, K. B., & Roediger, H. L., III (2010). A comparison of study strategies for passages: Rereading, answering questions, and generating questions. *Journal of Experimental Psychology: Applied*, *16*, 308–316. https://doi.org/10.1037/a0020992

Wessel, I., & Merckelbach, H. (2006). Forgetting “murder” is not harder than forgetting “circle”: Listwise-directed forgetting of emotional words. *Cognition and Emotion*, *20*, 129–137. https://doi.org/10.1080/02699930500260195

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*, 291–298. https://doi.org/10.1177/1745691611406923

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571–580. https://doi.org/10.1080/09658210244000414

White, R. (2002). Memory for events after twenty years. *Applied Cognitive Psychology*, *16*, 603–612. https://doi.org/10.1002/acp.819

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*, 10–16. https://doi.org/10.1111/sjop.12093

Wissman, K. T., & Rawson, K. A. (2018). Test-potentiated learning: Three independent replications, a disconfirmed hypothesis, and an unexpected boundary condition. *Memory*, *26*, 406–414. https://doi.org/10.1080/09658211.2017.1350717

Wooldridge, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Cognition and Education*, *3*, 214–221. https://doi.org/10.1016/j.jarmac.2014.07.001

Wylie, J., Thomson, J., Leppanen, P., Ackerman, R., Kanniainen, L., & Prieler, T. (2018). Cognitive processes and digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. W. van den Broek (Eds.), *Learning to read in a digital world* (pp. 57–90). Amsterdam: Benjamins. https://doi.org/10.1075/swll.17.03wyl

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*, 8. https://doi.org/10.1038/s41539-018-0024-y

Yeo, D. J., & Fazio, L. K. (2019). The optimal learning strategy depends on learning goals and processes: Retrieval practice versus worked examples. *Journal of Educational Psychology*, *111*, 73–90. https://doi.org/10.1037/edu0000268

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*, 3–47. https://doi.org/10.1080/02699930903205698

Zaromb, F. M., & Roediger, H. L., III (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*, 995–1008. https://doi.org/10.3758/MC.38.8.995

Zhu, J., Olechowski, A., & Habib, R. (2017). The re-encoding processes of restudy and testing are equally susceptible to the impairment of divided attention. *Cognition, Brain, Behavior. An Interdisciplinary Journal*, *21*, 65–84. https://doi.org/10.24193/cbb.2017.21.05