EFFECTS OF TESTING AND ENACTMENT ON MEMORY

Veit Kubik

Doctoral Thesis in Psychology at Stockholm University, Sweden 2014
Effects of Testing and Enactment on Memory

Veit Kubik
“We shall not cease from exploration
   And the end of all our exploring
Will be to arrive where we started
   And know the place for the first time.”

T. S. Eliot
Learning occurs not only when we encode information but also when we test our memory for this information at a later time. In three empirical studies, I investigated the individual and combined effects of interleaved testing (via repeated rounds of study and test practice) and encoding (via motor enactment) during learning on later cued-recall performance for action phrases. Such materials (e.g., “water the flowers”) contain a verb and a noun and approximate everyday memory that typically revolves around past and future actions. Study I demonstrated that both interleaved testing (vs. study only) and enactment (vs. verbal encoding) individually reduced the forgetting rate over a period of 1 week, but these effects were nonadditive. That is, the *direct testing effect* on the forgetting rate occurred for verbal, but not for enactive encoding; enactment reduced the forgetting rate for the study-only condition, but not for the study–test condition. A possible explanation of these findings is that both study techniques sufficiently elicit verb–noun relational processing that cannot be increased further by combining them. In Studies II and III, I replicated these testing-effect results and investigated whether they varied as a function of recall type (i.e., noun-cued recall of verbs and verb-cued recall of nouns). For verbal encoding (Study II), the direct testing effect was of similar size for both noun- and verb-cued recall. For enactive encoding, the direct testing effect was lacking irrespective of recall type. In addition, interleaved tests enhanced subsequent re-encoding of action phrases, leading to an accelerated learning. This *indirect testing effect* was increased for the noun-cued recall of verbs—for both verbal and enactive encoding. A possible explanation is that because nouns are semantically more stable, in that the meaning of nouns changes less over time and across different contexts, they are more recognizable. Hence, associated information (e.g., about the recall status) may be more available to the learner during restudy that, in turn, can initiate more effective re-encoding. The two different testing benefits (i.e., direct and indirect) may, partly, engage different mechanisms, as they were influenced differentially by the manipulations of encoding type and recall type. The findings presented in the thesis provide new knowledge regarding the combined effects of strategies and materials that influence memory.

*Keywords:* testing effect, test-potentiated learning, enactment effect, cued recall, forgetting, episodic memory
Vi lär oss inte bara genom instuderening, utan även när vi efter instudereringen testar hur väl vi lärt oss något, vilket ofta benämns som testeffekten. Om det är handlingar vi ska komma ihåg (t.ex. ”vattna blommorna”), så lär vi oss också bättre genom att utföra handlingarna motoriskt jämfört med att bara läsa en beskrivning av dem. I tre empiriska studier har jag undersökt hur upprepade studie- och testställen (jämfört med endast instuderning utan minnestestning), samt hur motorisk inkodning (jämfört med endast verbal inkodning) påverkar lärandet och hur väl man minns informationen efter en vecka. Instuderingsmaterialet bestod av olika handlingsfraser, som exempelvis ”vattna blommorna”, då vardagligt minne ofta involverar förflutna och framtida handlingar. Studie I visade att både minnestestning, såväl som att motoriskt utföra en handling ledd till minskad glömska sett över en veckas tid, men att de två effekterna var icke-additiva (d.v.s. man kom inte ihåg materialet bättre genom att kombinera minnestestning med motorisk inkodning). Den positiva inverkan som minnestestning under inlärningsfasen hade på hur väl deltagarna mindes handlingsfraser efter en vecka uppstod dock enbart när fraserna lärt in verbalt, inte när de lärt in genom motoriskt utförande. En möjlig förklaring till varför de två studieteknikerna i kombination inte ledde till en additiv effekt kan vara att båda studieteknikerna involverar samma minnesmekanism, så kallat relationellt processande mellan verbet och substantivet i en fras. I studie II och III bekräftade jag resultaten från den första studien i termen av testeffekten (d.v.s. testeffekten uppstod endast för verbalt inkodat material, inte för motoriskt utförda handlingar), och undersökte om effekten berodde på vilken ledtråd som gavs vid minnestesten. Deltagarna fick antingen substantivet som ledtråd för att plocka fram verbet, eller så fick de verbet som ledtråd för att plocka fram substantivet. När fraserna lärt in verbalt (Studie II) var testeffekten lika stor oavsett om ledtråden var ett verb eller ett substantiv. Efter motorisk inkodning uteblev testeffekten oberoende av vilken ledtråd som gavs vid minnestesten (Studie III). Testning kan förutom att påverka hur väl man minns efter ett längre retentionsintervall (t.ex. en vecka) också förbättra inkodningen av information under senare inlärningstillfällen. I både Studie II och III så lärde sig deltagarna fler nya ord vid ett ytterligare inkodningsstillfälle efter att ha genomgått ett minnestest med substantivet som ledtråd (”?”–blommorna”), jämfört med när verbet var ledtråd (”vattna–?”). Sammanfattningsvis visar jag i avhandlingen att minnestestning leder till minskad glömska i samma utsträckning som
motorisk inkodning. Vidare leder minnestestning under inlärningsfasen till att man vid ett efterföljande studietillfälle lär sig orden bättre och då speciellt de man inte tidigare kom ihåg. De två effekterna som testning har kan delvis hänföras till olika mekanismer, eftersom de påverkas på olika sätt av inkodnings- och framplockningsformatet.
Acknowledgments

The time during my PhD at Stockholm University has been an exciting and humbling experience. It was with great enthusiasm that I started my little research projects, with the aim to learn as much as possible about memory (and beyond). Indeed, in the last few years, I learned a lot about memory, and managed to conduct quite a number of experiments. Though, rather than arriving at ultimate answers to my research questions, ideas of reformulating or addressing them differently entered my mind. Time has passed quickly (which is a good sign 😊) and commanded to the completion of my thesis. Time to move on.

Throughout this academic journey, I have had the privilege of pursuing, sharing and discussing my research interests with many bright and enthusiastic people. They have not only helped me sharpen and test my research ideas, but also to broaden my views and knowledge on many topics outside of my doctoral theme(s). While receiving such valuable guidance, I have had the pleasure of enjoying a great deal of freedom, as I was able to choose my own topic(s) of interest, the people to engage in research with and to travel the world hunting for the important conferences in my field. These experiences have really made my PhD experience and have enabled me to learn more than my thesis has revealed.

First and foremost, I would like to thank my supervisors. You, Lars-Göran Nilsson, are primarily responsible for me having pursued my PhD in Stockholm—“the city with the most beautiful university campus on the world” (these were your words when you welcomed me here for the first time :). Yes, indeed! I am truly grateful for your help, your patience (with the long and always returning manuscripts) and for your belief in me becoming a good scientist. You set a very high standard with your work ethic (being “an early bird”), your devotion to science and yet at the same time being so humble and gentle. You inspired me with your work, for instance on the enactment effect, which is one part of this thesis, and you encouraged me to get to know the researchers in my field and to collaborate with them. I did! I have gained and learned a lot from you throughout many discussions over the past few years! Thank you for your encouraging and sensible words in moments when I was doubtful and when my research became frustrating. You have contributed greatly in making me a more independent and better researcher.
Fredrik Jönsson, you have been a heavy influence and a great inspiration to me. I learned many aspects of research from you and you set many standards that made me a better researcher (yes, also to communicate my ideas a bit more reader-oriented :) I am grateful that you integrated me into your new research group and I have wonderful memories of our conference visits in San Sebastian and Budapest! It is fair to say that my (later) interest and enthusiasm for the testing effect and also for metacognition originated from you. In countless emails, conversations and meetings, we discussed the latest papers, our own research and of course better and new future experiments (which I always end up running regardless/because of the time limitations, as you may have noticed ;). These discussions were one of the most giving and fulfilling research times during my PhD and fulfilled the reason why I am doing science. These discussions were the starting point of a thrilling, obsessive research and travel “trip” in search for further enlightenment. We still have some scientific “leftovers” from this trip; let’s get them out ☺! Despite your increasing working load, you have always been available and a reliable source of support and advice (“What do you think?!?” ;)).

Jonas Olofsson, I vividly remember when you defended your PhD in Umeå and now it is my turn to finish mine—with you as one of my supervisors. How cool is that (!) and what moving journey this has been in between! I am grateful for your contributions on my manuscripts. You helped me very efficiently to further improve my writing in a short time, to set priorities and to focus on ONE good story. You are the master! Your positive and pragmatic view on science taught me for sure the one or the other lesson and I enjoyed many good discussions about important topics in science, life and MMA. Looking forward to more lunch sessions in Kräftriket, more UFC-evenings and continued collaboration!

Monika Knopf, I am deeply indebted to you for your support throughout these years. I do remember clearly when you asked me for a meeting in the second term in order to tell me, convincingly, that I should pursue a scientific career and that you wanted to put me on the “scientific track.” Since then, I have been following this track with pleasure and addiction. Although there is a great deal to say about your academic influence, I want to thank you here more for all the doors that you have pointed to and even opened for me. Based on your suggestion, I received the scholarship of the German National Academic Foundation as an undergraduate and PhD student, and therefore I had the opportunity to attend “doctoral forums,” “Sommerakademien,” and many more events, where I made good, inspiring friends. You also encouraged me to study abroad and to start research already as a student in Brighton and Stockholm, this is where I got to know Jenny Rusted, Reza Kormi-Nouri and Lars-Göran Nilsson—all fabulous people who became close to me. Beyond the specific support of projects or papers, you always emphasized the development of my own views and to grow as an independent, self-confident person in science. This is what I have appreciated most! Despite travelling a lot and
finally doing my PhD in Stockholm, I have ALWAYS been more than welcome in your lab in Frankfurt and, in fact, being part of it. A special thank and greeting to the whole group, which includes, Sven, Isy, Frauke, Jantina, Gabi, Thorsten, Sonja, Christian, Monika, Margit and Frau Weyershäuser. You all are important that I feel very happy being at my other academic home! Thanks!

Thank you for very valuable comments and suggestions on the kappa to Timo Mäntylä and Linus Holm, and during my half-time seminar again to Timo Mäntylä and also to Stefan Wiens! You all improved my work to the better! My gratitude also to Hedvig Söderlund and Jenny Rusted for stimulating discussion and for being such engaged and positive co-authors.

During my doctoral time, I had the opportunity to visit well-known experts in my field: Simon Farrell in Bristol and Robert Bjork (and Nick Soderstrom) in Los Angeles. They all welcomed me to their research laboratories and I sophisticated and broadened my views on learning and memory a great deal. These were one of the prime times of my PhD! Also a special thank you to Lars Nyberg, Bernt Jonsson and Linnea Karlsson in Umeå, to Henry L. Roediger and John Nestojko in St. Louis as well as Melanie Steffens and Janette Schult in Jena, whom I had the opportunity to present and discuss various research topics with.

Throughout my time in Stockholm, I have had the great privilege to discuss and/or collaborate with many other researchers who have directly or indirectly contributed to my thesis: Hedvig Söderlund, Jenny Rusted, Nick Soderstrom, Andreas Jemstedt, Sven Obermeyer, Stefan Wiens, Gesa van den Broek, Artin Arshamian, Reza Kormi-Nouri, Maria Sandgren, Sverker Sikström, Wolfgang Mack, Timo Mäntylä, Tina Faust, Tanja Schatz, Petter Marklund, Eva Neidhardt, John Nestojko, Alan Castel, Kou Murayama, Elizabeth L. Bjork, Ivo Todorov, Viktoriya Voloshyna, Marc Borner, Karin Schermelleh-Engel, Helfried Moosbrugger, Anders Sand, Constanze Eib, Max Larsson Sundqvist, Mats Nilsson, Stina Cornell Kärnekull, Ola Svensson, and Åke Wahlin.

Thank you to the German National Academic Foundation (Studienstiftung des deutschenVolkes) for granting me an “academic” stipend during my doctoral time. This allowed me to participate in many doctoral forums and to get to know many doctoral students and their multidisciplinary projects over a period of three years. I had amazing times with you guys from the “Werkstatt” (Bea, Stefan, Manuel and Susanne), with you Marc, Johannes, Michael, Christoph, Marion, Eric, Maria, Adrian, Thomas, Steffis, Tobias, and so many more. I am happy to have gained you as friends.

I am more than blessed to have good friends and colleagues in my life, who made this journey worthwhile and provided me with superb times and needed
distractions. It is impossible to acknowledge each of you. I am glad to have you—Thomas, Undra, Nico, Johannes, Nat, David, Simon, Bea, Kristina, Stina, Anne, Ivo, Anders, Tine, Andreas, Henni, Kiki, Shara, Levke, Henrik, Tina, Geoffrey, Jelena, Stephanie F., Daniel, Mattea, Nathalie, Maria L., Maria Ö., Johanna S., Stina, Tonya, Stina, Tonya, Ann-Sofie, Ann-Charlotte, Cilla, Gabriella, Max, Håkan A., Håkan F., Jelena, Jesper, Joakim, Jonas P., Kristina, Margaretha, Maria S., Martin, Nischel, Pehr, and Ola Sternäng. Special thanks to Artin and Consti for just being such stimulating, loyal, helpful and fun friends during my doctoral time. Aram, my friend, we had some good times together! I am glad to have had the best office mates I can imagine, Joao and Sven—We had great times in Frankfurt and Toronto and shared the highs and lows in very good company! Thanks Wolfgang and Eva for a superb time in Wien and many interesting and thought-provoking discussions throughout the years!

Thank you to Lotta Lindqvist, Camilla Nelson, Monika Karlsson, Tommy Olin, Henric Bergqvist, Henrik Dunér, Eva Sjöqvist, Eva Persdotter and Stefan Sanderhem for all your help regarding administrative and IT matters.

I am indebted to my stimulating second working and living room, the café Ritorno and its fun personnel! “A cappuccino and a cheesecake, please!”

And finally all my gratitude belongs to my family! I am truly blessed having you, Nikolaj, my smarter alter ego. Having you on this planet, gives me hope, strength and comfort! Nobody has the tools as you to make me a better person. Your honesty, insane energy, and endless willingness to help (if I need it) drives me (at times) nuts! Mama, you are simply the best. Thank you for making the journey into the world of science and research possible for me in many ways (!) and encouraging me to do my (doctoral) studies abroad and to go my way. Your support and curiosity to see life through my eyes and thoughts over all these years makes me feel loved. The just so marvelous and humorous mails helped me “to keep a stiff upper lip”, as you used to say, and kept me going, often with a smile even in difficult times. Papa, you have surely passed on some of your overwhelming curiosity, energy and your explorative mind to me: “Immer strebe nach dem Ganzem.” Carmilotta, my little niece with only 11 months old, you managed to make me laugh and wonder in moments when it seemed impossible. You are magic!
List of studies

The present doctoral thesis is based on the following studies:


## Contents

Introduction ................................................................................................. 1

Memory ........................................................................................................ 3
 Memory Classifications .............................................................................. 3
 What is Remembering? ............................................................................ 5
 Concept of Episodic Memory .................................................................... 5
 Three Stages of Remembering and Forgetting ....................................... 7
 Encoding and Retrieval Factors ............................................................... 8
 Remembering Action Events ................................................................. 10

How Does Testing Affect Memory? .......................................................... 11
 Direct and Indirect Testing Effects .......................................................... 12
 Direct Testing Effect ................................................................................ 12
 Indirect Testing Effect ............................................................................ 13
 Additional Testing Effects ....................................................................... 15
 Explanations of the Testing Effects ......................................................... 17
 Direct Testing Effect ................................................................................ 17
 Indirect Testing Effect ............................................................................ 21

How Does Motor Enactment Affect Memory? .......................................... 24
 Enactment Effect ..................................................................................... 24
 Explanations of the Enactment Effect ....................................................... 26
 Theoretical Questions .............................................................................. 28
 Item-Specific versus Relational Processing ............................................ 28
 Conceptual versus Motoric Elaboration ................................................. 30
 Automatic versus Strategic Processing .................................................. 31
 Methodological Considerations ............................................................. 32

Background, Research Objectives, and Methods of Studies .................. 34
 Background ............................................................................................. 34
 Research Objectives ............................................................................... 37
 Methods of Studies ................................................................................ 38

Overview of Studies .................................................................................. 40
 Study I ....................................................................................................... 40
 Objective .................................................................................................. 40
 Background ............................................................................................. 40
 Method ...................................................................................................... 41
 Results ...................................................................................................... 42
 Conclusions ............................................................................................. 43
 Study II .................................................................................................... 44
 Objective .................................................................................................. 44
 Background ............................................................................................. 44
 Method ...................................................................................................... 44
 Results ...................................................................................................... 45
Conclusions ................................................................................................................. 47
Study III ...................................................................................................................... 49
  Objective .................................................................................................................. 49
  Background .............................................................................................................. 49
  Method ...................................................................................................................... 49
  Results ...................................................................................................................... 50
  Conclusions .............................................................................................................. 52

General Discussion .................................................................................................. 53
  Testing Effects ......................................................................................................... 53
    Direct Testing Effect .............................................................................................. 53
    Indirect Testing Effect ........................................................................................... 55
  Enactment Effect ...................................................................................................... 57
  Alternative Memory-Strength Explanation .............................................................. 60

Challenges and Directions for Future Research ...................................................... 62
  Methodological Challenges ..................................................................................... 62
    Isolating Direct and Indirect Testing Effects ......................................................... 62
    Measuring Forgetting ............................................................................................ 65
    Action Phrases—Asymmetric Learning Material .................................................. 67
  Theoretical Challenges ............................................................................................ 68
    Multiple Implementations of Testing ................................................................... 68
    A Synthesis—An Item–Context Account ................................................................ 69
  Practical Challenges ................................................................................................ 72

Concluding Remarks ................................................................................................ 74
Introduction

We often view learning as the process of taking in new information. To improve our learning outcomes, we focus on optimizing the encoding of information. For example, we may develop visual images or read the to-be-learned information aloud. Both study techniques typically result in better memory performance than silent reading (MacLeod et al., 2010; Paivio, 1971). From this traditional perspective, testing memory serves mainly to measure the degree of learning associated with different encoding techniques. However, retrieving information from memory is not neutral for learning. Contrary to our everyday intuitions, retrieval can promote learning. Each successful retrieval changes the retrieved information in memory and enhances its recallability in the future. This phenomenon is supported by recent evidence on the so-called testing effect. Tests during learning facilitate subsequent re-encoding and reduce the rate at which people forget information, relative to simple restudy of the materials (Pyc, Agarwal, & Roediger, 2014; Roediger & Butler, 2011). Thus, testing does not only assess learning, it is an effective tool to enhance it.

In addition, we usually view memory as a place where the encoded information is stored—similar to a “storehouse” of input elements (Koriat & Goldsmith, 1996). In this sense, memory is often investigated in experimental situations, in which learners passively observe a list of words or word pairs that are presented to them. However, in everyday life people are actively involved in events and have to remember actions that they have recently performed or will need to perform. For example, before leaving your flat, you need to remember whether you turned off the oven and closed the windows (past actions) or that you need to buy medicine after lunch (future obligation). This primacy of memory for actions in daily life reveals, for example, in superior recall for enacted information. Research demonstrated that motorically performing action phrases (e.g., “water the flowers”) during encoding typically trumps any form of verbal encoding in enhancing later retention, in the so-called enactment effect (for reviews, see Zaromb & Roediger, 2010; Zimmer et al., 2001). It is possible that human memory has evolved over time in such a way that our ability to keep action-relevant information in mind has improved (Glenberg, 1997; Ratner & Foley, 2001). To understand the nature of human learning and memory, it is important to consider the active role of retrieval and the primacy of actions for everyday memory.
The overall aim of this thesis is to contribute to the understanding of how retrieval and encoding can support human learning and memory of actions. Both study techniques have the potential to be more systematically incorporated into educational practice and used as effective learning tools that lead to durable and flexible knowledge. The present thesis has two general objectives. The first objective is to evaluate the effects of interleaved testing (i.e., repeated study–test practice) with more ecologically valid learning material. In the studies included in the present thesis, I used action phrases (e.g., “water the flowers”) to investigate memory through realistic learning materials that mimic the basic actions required in everyday life. Testing effects have been observed across various learning materials (e.g., word pairs and test passages). However, they have not been investigated with action-related information, even though everyday memory is highly concerned with such information. The second objective is to investigate the relative and combined efficacy of interleaved testing and motor enactment on learning and retention. Both retrieval via testing and enactment facilitate later retention of information and have generated a great deal of attention (cf., Roediger & Butler, 2011; Roediger & Zaromb, 2010). However, little is known about their joint effects. Do both study techniques have additive effects when they are combined or do they interact to produce nonadditive effects? Additive effects would suggest that both study techniques rely on independent mechanisms, while nonadditive effects would suggest that they rely in parts on the same mechanism(s).

In the following sections, I first present an overview of the conception and classification of human memory with an emphasis on episodic memory, since this is the main memory system addressed in the thesis. Then, I introduce the facilitative effects of testing and enactment on episodic memory. Next, I turn to the empirical part of this thesis and provide a brief overview of its three empirical studies. Finally, I discuss the findings from these studies in the context of current theoretical accounts and suggest potential avenues for future research.
Memory

The modern investigation of human memory falls short of providing a precise, general definition of its object of investigation (Tulving, 1983), perhaps with good reason. Attempting to account for all memory-related phenomena leaves us with a very broad definition of memory, along the lines of “the representation of past experiences that exists over time” (Schacter, 2007, p. 25). Tulving (1983) emphasized that an effective definition of memory should be more comprehensive than simply the storage of information and should also pertain to encoding and retrieval processes. For example, in cases of retrieval failure, one would neither speak of memory nor conclude that there is no memory for a particular information. Instead, it may be that the encoded information is in storage, but it simply cannot be retrieved (i.e., it is not accessible at the present time), as the appropriate retrieval cues may not be present (Tulving & Pearlstone, 1966). Thus, encoding, storage, and retrieval together interactively contribute to what constitutes memory (Schacter, 1996). To separate the individual contributions of these memory constituents is a challenge in contemporary research (cf. Schacter, 2007).

Memory Classifications

There are two major cognitive approaches to understand human memory—system-oriented and process-oriented approaches. The system-oriented approach views human memory not as a single entity but as being composed of multiple functionally and biologically distinguishable subsystems (Squire, 2004). Depending on the researchers’ focus, different classification systems have been developed, varying along the time dimension (sensory memory, short-term memory, and long-term memory), content and state of awareness (procedural vs. declarative memory and episodic vs. semantic memory), or other criteria. Although the exact number of systems is a matter of ongoing debate and their nature is rather hypothetical, the system-oriented approach provides a useful framework to integrate and discuss findings on human memory.
Many contemporary researchers agree that memory can be divided into five interrelated memory systems, that is, primary memory, episodic memory, semantic memory, the perceptual representation system, and procedural memory (Nyberg & Tulving, 1996; Tulving, 1983, 1985, 1993). Primary memory (also referred to as working memory) enables us to retain incoming information for relatively short time periods and to process it. Episodic memory refers to the conscious recollection of personally experienced information or events that occurred at a specific time and place. It captures aspects related to the “what,” “when,” and “where” of information (Nyberg et al., 1996). Semantic memory permits the implicit retrieval of general, decontextualized knowledge of concepts (e.g., the meaning of words) and facts about the world (e.g., the capital of Sweden). The perceptual representation system mainly permits rapid identification of formerly encountered perceptual stimuli, and procedural memory is concerned with the acquisition and retention of behavioral skills or procedures. These memory types are presumed to have evolved sequentially—both across the phylo- and ontogenesis—, with procedural memory developing earliest and episodic memory developing last. Furthermore, these memory systems are associated with different states of awareness (Tulving, 1993, 2002). Episodic memory allows the conscious recollection of previously occurred events from a specific episode (i.e., autonoetic awareness). Retrieval from semantic memory, however, is implicit and is accompanied by feelings of familiarity or knowing (i.e., noetic awareness). Retrieval from primary memory is associated with a fleeting awareness of thoughts and perceptions, as long as we consciously attend to them. In contrast to these later developing memory systems, retrieval from the perceptual representation system and procedural memory lack any form of conscious awareness.

As episodic and semantic memory are most directly relevant to the present thesis, it should also be noted that they have been classified as the two declarative forms of long-term memory (in contrast to the nondeclarative forms of procedural memory and priming). Both facts and events stored over long time spans (i.e., years or even decades) can be deliberately recalled and verbally expressed (Squire, Knowlton, & Musen, 1993). Though, based on several findings, episodic and semantic memory should be viewed as separate systems. For example, some brain-damaged patients suffering from brain damage exhibit specific deficits in episodic memory, but not in semantic memory (Vargha-Khadem et al., 1997); the two memory systems involve in parts different brain areas (Cabeza & Nyberg, 2000), and episodic memory has been

---

1 Retrieval from semantic memory (e.g., “Is Stockholm the capital of Sweden?”) can be viewed as implicit in the specific sense that learners do not deliberately attempt to retrieve information from a specific episode and that they do not consciously recollect the specific episode. However, semantic retrieval can be also viewed as “explicit” in the sense that learners can voluntarily retrieve previously acquired knowledge and be aware of it (Tulving, 2000).
shown to decline earlier in the life span (Rönnlund, Nyberg, Bäckman, & Nilsson, 2005).

The process-oriented approach views memory as one aspect of the general cognitive system (Craik, 2007) that is produced by perceiving, understanding, and acting upon the world (Craik, 2002; Craik & Lockhart, 1972; Craik & Tulving, 1975). Memory is regarded as a mental activity that is best analyzed in terms of the encoding and retrieval processes involved in specific memory tasks. Encoding processes, on one hand, are presumed to vary in depth of processing, with deeper, semantic processing leading to better memory performance (on explicit memory tests) than shallower, phonological processing. This is referred to as the levels-of-processing effect (Craik & Tulving, 1975). Retrieval processes, on the other hand, increase memory performance to the extent that they recapitulate the initial encoding processes and cues; this is called the encoding-specificity principle (Tulving & Thomson, 1973). For example, the semantic processing of information increases memory performance on those tests that tap on semantic information (e.g., recall and recognition tests), and thus encoding processes benefit later memory performance only if they can be transferred appropriately to the test at retrieval (this is referred to as the transfer-appropriate processing principle; see Morris, Bransford, & Franks, 1977). From this process-oriented approach, it is not necessary to assume specific memory stores.

The two approaches—one delineating the structure of memory, the other modeling the underlying processes invoked in the various memory tasks—complement each other and contribute to the understanding of human memory. One contemporary trend is to integrate both approaches in so-called process-oriented systems.

The present thesis does not strictly adopt one or the other of these approaches but, rather, relates to both of them. On one hand, I refer to various processing accounts, such as item-specific and relational processing, to explain the results pattern of the testing and enactment effects on cued recall. On the other hand, I assume that the studies of list learning included in the present thesis tap episodic memory (Kahana, Howard, & Polyn, 2008; Tulving, 1972), although other memory systems, most notably semantic memory, are also likely to be involved.

What is Remembering?

Concept of Episodic Memory

Episodic memory refers to the capability to remember information or events that occurred at a specific place and time. Critically, not only the information per se is remembered but also information associated with it, such as the surrounding temporal, spatial, or social context (i.e., when, where, and with
whom; see Yonelinas, 2007). The query “What did you do last weekend?” or “Can you recall the words that I presented to you a week ago and write them down?” requires episodic memory. It enables the learner to travel back in time (Tulving, 1983, 2002) and to re-experience events of one’s personal past. Episodic memory can be directed not only back toward the past but also toward the future (Schacter & Addis, 2007), as it enables the learner to travel forward and “pre-experience” himself or herself in a novel situation (i.e., episodic future thinking; for a review, see Szpunar, 2010). A critical feature of episodic memory is that it is associated with autonoetic consciousness (Tulving, 1985, 2001), that is, the self-awareness that it is me who is remembering my own past or thinking about my own future. Another less discussed feature is that remembering involves a judgment that the retrieved information is a memory of a specific event (Yonelinas, 2007).

Episodic memory is related to autobiographical memory, though the two concepts can be distinguished. Autobiographical memory comprises self-knowledge (e.g., knowing that one can work well under pressure), episodic memories related to self-knowledge (e.g., remembering the sunrise that occurred when one was preparing for the philosophy exam early in the morning), and also facts from semantic memory (e.g., schematic knowledge about an exam). Critically, all these different types of information (self-knowledge, episodes, facts, and procedures) are related to the self and to one’s personal history. As autobiographical memory is a more comprehensive concept than episodic memory (Conway, 2001; cf. Roediger & Marsh, 2003), it engages additional brain regions beyond those typically engaged in episodic memory (Cabeza et al., 2004) and is studied using separate methods (e.g., diary studies and the event-cuing technique; see Roediger & Marsh, 2003).

In the included studies of this thesis, participants learned a list of action phrases (verb–noun phrases, such as “water the flowers”) and memory was assessed by a cued-recall test (i.e., one item was given and the participant was asked to recall the other). This form of testing required them to retrieve parts of the previously encoded item pair and the formed associations between them, which are typical features of episodic memory. However, other features of episodic memory, such as the location or the recollective experience, were not assessed. This list-learning paradigm taps specific aspects of episodic memory but less specific aspects of autobiographical memory, as the learned action phrases are likely not relevant in terms of self-knowledge and personal history, although action events have been argued to be more self-involving (Kormi-Nouri & Nilsson, 2001). It should be acknowledged that semantic memory may contribute to successful recall in such situations; encoding requires the participants to understand the meaning of the action-phrase material and recall performance is also influenced by pre-experimental, semantic associations between items—that is, by prior knowledge (Tulving, 2002). Thus, it is possible that episodic and semantic memory interact in typical list-learning experiments, as they can be regarded as parallel and partially overlapping information processing systems within long-term memory (Tulving, 1972).
Three Stages of Remembering and Forgetting

Most studies assess episodic memory by means of so-called list-learning experiments, in which subjects learn a series of independent items. The procedure of these experiments can be divided into three stages (McDermott, 2007): encoding (i.e., acquisition), storage (i.e., retention), and retrieval (i.e., access to or reconstruction) of information. For example, you may be asked to learn 20 action phrases such as “lift the glass” (i.e., the encoding stage). After a given time period, perhaps one week (the retention or storage stage), the verb from each phrase (“lift”) is presented as a cue and you are asked to recall the noun (“the glass”); this cued recall test constitutes the retrieval stage. Alternatively, more demanding recall tasks could be presented, such as requiring you to retrieve all the action phrases freely in any order (i.e., free recall) or in the original order of study (i.e., serial recall). Or you may be presented with a list of action phrases and asked to indicate which ones were previously presented during encoding and which were not (i.e., a yes–no recognition test; for a methodological and theoretical description of these tests, see Kahana, 2012).

Although these tests do not solely tap episodic memory, they assess at least one critical aspect of it, as the subject needs to deliberatively retrieve the information provided at a specific time (and also at a specific location if the to-be-remembered events happen outside the laboratory context). To further distinguish episodic memory’s contributions from those of semantic memory, you may be asked to describe his or her subjective status of awareness—for example, by additionally indicating whether the recognition decisions originated from remembering the item’s presentation or simply from knowing that it was presented (Gardiner, 1988; Tulving, 1985).

All these types of episodic memory tests demonstrate that successful remembering requires passing through the stages of encoding, storage, and retrieval. Let us assume that you have successfully recalled 12 out of 20 nouns. To remember each of these 12 nouns, you needed to encode, store, and consolidate them in long-term retention, as well as to successfully retrieve them after 1 week when prompted with the verb cues. However, what can be said about the eight forgotten nouns? One possibility is that the memory trace had been completely erased from your storage. Irrespective of the retrieval conditions, you cannot retrieve these nouns any more. However, this scenario is a rather implausible one, except perhaps in extreme forms of dementia (Davis, 2007). More likely, your inability to recall these forgotten items is evidence of retrieval failure, although the underlying cause may be related to any of the three stages. Perhaps these action phrases, after being briefly held in working memory, were simply not stored in long-term memory (i.e., an encoding failure). Alternatively, a memory trace of an event may have been stored in long-term memory, but its strength has decayed over time (McGeoch, 1932), so that the trace is still available but no longer accessible (Tulving & Pearlstone,
A third possibility is that the memory trace is still available and accessible but that the retrieval cues have temporarily failed to elicit retrieval; in this instance, the memory may be retrievable at a later time or under stronger cuing conditions (i.e., cue-dependent forgetting; see Tulving, 1974). As an illustration of this possibility, perhaps you may not be able to recall the nouns, but when you are given a recognition test of previously studied and new target items, you may still recognize the “forgotten” nouns as items that you learned previously (Tulving & Pearlstone, 1966).

This example suggests that forgetting and remembering may be seen as two sides of the same coin: successful remembering reveals when forgetting does not occur, and forgetting reveals when retrieval fails. However, such a definition turns out to be inaccurate when we consider the process of forgetting over time. Long-term forgetting of information can be accurately assessed only if the memory has been successfully encoded into long-term memory, and the same retrieval cue is given repeatedly across retention intervals (RIs). For this reason, in the present studies, memory was also assessed immediately after the learning phase to determine the baseline level of retrieval success and thereby to investigate the degree to which previously retrieved information subsequently becomes unretrievable (i.e., the retention curves). In this way, forgetting can be more precisely defined as the “inability to retrieve information that was once consciously accessible under the cuing conditions that are again in effect now” (Wixted, 2007, p. 330). This modified definition turns forgetting and remembering into dependent rather than opposing concepts. Either way, the study of forgetting is an important approach in investigating episodic memory.

Many mechanisms have been proposed to explain the empirical phenomenon of retrieval failure. One explanation revolves around the notion that the strength of the memory trace decays over time (i.e., decay theory; for a classic evaluation, see McGeoch, 1932). Another explanation is that forgetting results from interference by other events, which can stem from previous learning (proactive interference) or encountering new information (retroactive interference). For example, if you want to remember your current bank account number, you may remember instead the one that you had earlier (proactive interference). A final explanation suggests that, over time, additional information becomes associated with a retrieval cue (known as cue overload theory; see Watkins & Watkins, 1975. The more the retrieval cue is “overloaded,” the less effective it becomes in eliciting the retrieval of the specific target information (this is). As the studies included in this thesis did not attempt to adjudicate between these theoretical options, I will not evaluate them further.

Encoding and Retrieval Factors
As noted earlier, episodic memory involves encoding, storage, and retrieval of information. Importantly, these components interact with each other. For example, encoding of information involves the retrieval of previous events or
general knowledge, and the act of retrieval itself can be regarded as an encoding event that changes and probably elaborates the memory of the stored event.

However, research and educational practices have generally focused on the encoding side of learning. For example, it has been reliably observed that generating a word (e.g., “rapid”) from a cue (e.g., “fast–r…”) increases memory performance compared to just reading that word (the generation effect, Slamecka & Graf, 1978). Similarly, memory advantages have been demonstrated when words were pronounced rather than not pronounced (the production effect, MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) or semantically processed rather than phonemically processed (the levels-of-processing effect; Craik & Lockhart, 1972). However, these encoding effects also depend on the retrieval test given. For example, the generation effect (Bertsch et al., 2007) and the production effect (Fawcett, 2013) are more pronounced in recognition than in free-recall. This can be explained by the so-called encoding specificity principle, in that study techniques enhance mainly item-specific information (i.e., the individual features of items) relative to silent reading, but less so relational information between items (e.g., temporal or semantic associations; Jones & Pyc, 2013; Mulligan & Lozito, 2004). As the recognition test mainly taps item-specific information, both encoding effects are increased with this memory test as compared to their size in free recall, which additionally requires inter-item relational information. Thus, the success of recall depends on the interaction of encoding and retrieval processes.

To complicate things further, the success of learning is not determined only by the way in which information is encoded into memory and assessed by the memory tests. The act of retrieval and reconstruction of knowledge from memory in itself can alter the accessibility of information in memory (Bjork, 1975; Karpicke, 2012). For instance, taking a test on the material to be learned leads to better long-term retention than simply rereading it, probably because it reduces the rate of forgetting. The resulting memory advantage for tested items over (in this example) restudied items is known as the direct testing effect. However, testing can also have detrimental effects on memory for information, which is in competition with to-be-recalled information. For example, selective retrieval of a subset of studied items can inhibit or interfere with subsequent retrieval of the other, nonretrieved items. This result memory disadvantage is referred to as retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; Storm, Bjork, & Bjork, 2012). Thus, retrieval can induce both learning and forgetting of information, critically depending on the specific situation (Bäuml & Schlichting, 2014; Bjork, 1975). Retrieval is a key process to understand learning (Karpicke, 2012; Tulving, 1991). The present thesis aims to investigate the impact on memory of action phrases of both a retrieval manipulation (interleaved testing vs. study-only practice) and an encoding manipulation (enactive vs. verbal encoding). The focus is set on the positive testing effects on learning of action-relevant information.
Remembering Action Events

Episodic memory is an integral part of our everyday activities. The most basic actions that we perform in the present are determined by our ability to remember what we have done in the past and what we intend to do in the future. However, research on episodic memory often investigates the encoding, retention, and retrieval of verbal material (mostly lists of words or word pairs) in the laboratory. Participants are thereby often assigned the role of passive observers who must take in and memorize the input presented to them. Thus, memory is studied in an experimental situation in which it is supposed to function like a “storehouse” (Koriat & Goldsmith, 1996) of input elements. In contrast, the use of memory in everyday life typically occurs in situations where the learner is actively involved in events and social interactions. It is not confined to the intake of information, but is also concerned with the “overt (and covert) (re-) actions as a response to the environmental stimulus” (Zimmer & Cohen, 2001, p. 4). People have to remember the actions that they have taken or that they intend to perform at some future point in time. Thus, it can be argued that the brain’s primary function is to enable actions that promote survival and reproduction in a changing environment. Human episodic memory may have been designed partially to keep action-relevant information in mind.

From this perspective, memory research in the early 1980s started to use action phrases (often verb–noun phrases) as stimulus material and let participants motorically perform them as a form of enacted learning. These self-performed, simple actions mimic the everyday affordances of memory in a more realistic way compared with word lists. These simple action phrases (“lift the glass”) were used as the learning material in the empirical studies included in this thesis, and they were either motorically performed or read aloud as part of a single study trial. However, these self-performed actions still do not capture other aspects of their real-world counterparts, which are more complex, goal-directed, and of higher personal relevance (Foley & Ratner, 2001). It may be important to note that retrieving the verbal action phrases primarily measures episodic memory. Though, procedural memory may also be involved to some extent when one performs or even reads the action phrases (Cohen, 1989), although it deals mainly with the acquisition and reproduction of motor skills through intensive practice. Procedural memory is, for example, assessed with the rotary pursuit task (i.e., tracking a revolving target with some indicator device).

In the present thesis, I will use action phrases to study the testing effects on memory for action phrases, both for verbal encoding (Studies I and II) and for enactive encoding (Studies I and III). In the next section, I will turn to the effects of retrieval on learning and forgetting, the so-called testing effects (for reviews, see Roediger & Karpicke, 2006b; Karpicke, Lehman, & Aue, 2014). Subsequently, I focus on the pervasive encoding phenomenon that has attracted much research, the enactment effect (for reviews, see Nilsson, 2000; Zaromb & Roediger, 2010; Zimmer et al., 2001).
How Does Testing Affect Memory?

“Retrieval from memory is often assumed, implicitly or explicitly, as a process analogous to the way in which the contents of a memory location in a computer are read out, that is, a process that does not, by itself, modify the state of the retrieved item in memory. In my opinion, however, there is ample of evidence for a kind of Heisenberg principle with respect to retrieval processes: an item can seldom, if ever, be retrieved from memory without modifying the representation of that item in memory in significant ways.” (Bjork, 1975, p. 123)

In our current educational system, tests and examinations are still primarily regarded as assessment tools used to measure student learning. Since testing is seen as fulfilling only this role, it is limited to the minimum necessary in school classrooms and even at the university level, and it often occurs cumulatively at the end of the term or the curricular period (cf. Roediger & Karpicke, 2006b). However, testing is an active learning event that alters the state of memory (Bjork, 1975). Taking a memory test can enhance learning and long-term retention more than restudying, and these test-related benefits can increase further when testing is repeated (Karpicke & Roediger, 2010) or spaced out temporally during learning (Landauer & Bjork, 1978; but see Karpicke & Roediger, 2010). Nevertheless, the mnemonic influence of tests on learning curves has been rather neglected, even in action-memory research (Koriat & Pearlman-Avnion, 2003; Kubik, Obermeyer, Meier, & Knopf, 2014a, 2014b), despite the fact that test-related benefits on memory are one of the more reliable phenomena in learning and memory research (Karpicke et al., 2014; Roediger & Butler, 2011). In the empirical studies of this thesis, I investigate the role of testing in the frequently used multitrial learning paradigm by adding a study-only condition, and I extend the positive testing effect to another, ecological meaningful, action-relevant learning material.
Direct and Indirect Testing Effects

Recently, direct and indirect effects of testing have been distinguished and have received some empirical scrutiny (Arnold & McDermott, 2013b; Karpicke et al., 2014; Roediger & Karpicke, 2006b). Although it is difficult to disambiguate these two types of effects in many experimental situations, it is important to consider them separately when possible.

Direct Testing Effect

Most research in the last decade has been devoted to the direct effects of testing (or retrieval practice). Direct effects refer to the benefits in (long-term) learning that accrue from the very act of retrieving information from memory itself (reviewed in Roediger & Butler, 2011; Roediger & Karpicke, 2006b; for a meta-analytic review, see Rowland, 2014). These benefits are usually investigated in an experimental paradigm in which, after an initial study trial, participants either attempt to retrieve the learning material in a memory test, restudy the material, or receive no study activity at all. Typically, retrieval practice enables learned items to be recalled better and over longer time periods than if one does not engage in retrieval practice or receives complementary study time (Halamish & Bjork, 2011; Kornell, Bjork, & Garcia, 2011). However, after shorter RIs, additional restudy trumps retrieval practice (Roediger & Karpicke, 2006b; Wheeler, Ewers, & Buonnano, 2003). In short, the type of learning activity utilized by participants in the second trial (retrieval practice, restudy practice, or no practice) and the RI (short vs. long) interact (Roediger & Karpicke, 2006b; van den Broek, Takashima, Segers, & Verhoeven, 2013).

To illustrate this classical testing effect, I provide an example from a study by Roediger and Karpicke (2006a). They provided participants prose passages on different topics as learning material and asked the participants to free recall as many of the idea units as possible after five minutes and again after one week. The subjects had four experiences each of learning the passages, but in three different combinations of study (S) and testing (T), i.e., SSSS, SSST, and STTT. The results demonstrated that final recall performance after the short RI increased with the number of study trials, whereas long-term retention increased with the number of test trials. Thus, successfully retrieving items from memory during learning functions as a “memory enhancer” (Bjork, 1975), in that the future retrievability of these items was enhanced relative to if one had simply restudied them. This result led to the conclusion that testing slows the rate of forgetting (Carpenter, Pashler, Wixted, & Vul, 2008; Wheeler et al., 2003; but for alternative viewpoints, see Halamish & Bjork, 2011; Kornell et al., 2011; Storm, Friedman, Murayama, & Bjork, 2013).

Several factors make reliable contributions to the efficacy of retrieval practice (Roediger & Butler, 2011). One factor is the number of test events given.
The repetition of tests increases the facilitative effects on long-term retention (Roediger & Karpicke, 2006a). Another factor is the spacing of testing events. The more widely they are distributed across time—rather than occurring close to each other, almost as a single event—the more mnemonically efficient they are (Jacoby, 1978; Landauer & Bjork, 1978; but see Karpicke & Roediger, 2010). Third, the difficulty of successfully retrieving the items to be remembered determines the efficacy of retrieval practice (see Bjork, 1975, who refers to a level of “desirable difficulty”). Through manipulating the type of memory assessment used during intermediate test trials, retrieval practice was shown to have stronger benefits in free-recall tests compared with cognitively less demanding cued-recall tests or recognition tests (Halamish & Bjork, 2011). Thus, with increasing difficulty, retrieval becomes mnemonically more effective. Finally, feedback following each test event provides restudy opportunities for items that are not successfully recalled and thereby leads to superior retention after shorter RIs (Kang et al., 2007). However, providing the learner with such restudy opportunities adds another benefit to learning, the indirect effect of testing.

Indirect Testing Effect

There has also been a recent surge of interest in the indirect effect of testing, which is alternatively called test-potentiated learning (Izawa, 1966; Karpicke & Roediger, 2007). This term refers to the finding that taking tests enhances subsequent re-encoding of information, specifically when the attempted retrieval of information from memory has failed. When items are not recalled on a test, they are likely to receive more or enhanced encoding during an ensuing restudy and/or feedback event when the information is presented again. This, in turn, enhances subsequent memory performance even after relatively short delays, matching restudy practice in some cases (Nelson, Arnold, Gilmore, & McDermott, 2013; Vestergren & Nyberg, 2014) and outperforming it in others (Karpicke, 2009; Pyc & Rawson, 2012). Typically, this indirect testing effect has been demonstrated by applying a multitrial learning paradigm in which trials of study trials also follow test trials (e.g., STSTST) as well as precede them (as is the case when examining the direct effects of testing, such as by using a schedule of SSTT). That is, conditions involving at least one test–restudy sequence (i.e., initial study, then testing, and then restudy) have been compared to conditions where only restudy occurs after the initial study trial.

Early studies (Izawa, 1966, 1970, 1971; Lachman & Laughery, 1968) showed reliably that increasing the number of test trials preceding a restudy trial leads to a faster learning rate (i.e., steeper learning curves). Izawa (1966, 1970) assumed that no learning (or forgetting) occurs during test trials; she attributed the accelerated learning rate to test-potentiated learning. That is, testing facilitates subsequent encoding even when no immediate feedback is
provided and the restudy opportunity occurs in a separate trial, with many other items occurring between them.

However, Arnold and McDermott (2013b) questioned the assumption that learning does not occur during testing, as recent research has shown reliable direct effects of testing (see previous section). Consequently, the finding that learning rates increase with the number of interpolated tests could possibly originate from both indirect (potentiated) and direct effects of testing. To isolate these effects more clearly from each other, Arnold and McDermott (2013b) differentiated between items that are retained from before (pre-test) to after (post-test) a restudy trial and items that are newly retrieved. The retained items may reflect direct and indirect effects of testing, whereas the newly retrieved items possibly reflect the potentiating effects of testing. Based on this reasoning, they applied a 20-trial cued-recall paradigm to the task of learning foreign-language vocabulary pairs. Following an initial study trial, participants received either five tests or one test before restudying the material. Results revealed that the learning rate was greater for the five-test condition compared with the one-test condition, thus replicating previous research (Izawa, 1966, 1971). Importantly, the proportion of newly retrieved items following the first and second restudy trials was greater for the five-test condition, suggesting that test-potentiated learning increased with the number of interleaved test trials. As the proportion of retained items after the first restudy trial was also heightened for the five-test condition, an increased direct testing effect via successful recall attempts may also have contributed to the accelerated increase in the learning rate with more preceding test trials. Recent evidence indicated that test-potentiated learning might apply not only to unsuccessfully recalled but also to successfully recalled items (Vestergren & Nyberg, 2014). Irrespective of the previous recall status, tested items (when compared with ones that were additionally studied) may receive more focused attention during restudy according to the authors, as indicated by increased activations in the anterior insula (AI). However, in a behavioral paradigm, the indirect testing effect can be examined only in terms of new learning of items not successfully recalled on the pre-restudy test.

Another paradigm was reintroduced (Kornell, Hays, & Bjork, 2009) to examine the effects of attempted retrieval on subsequent encoding. In this “pre-test” paradigm (Slamecka & Fevreiski, 1983), participants learned word pairs in a study trial. Some of the word pairs were “pre-tested” immediately beforehand by presenting the cue word and letting participants attempt to retrieve the target word from semantic memory (without successfully generating the target information). On a subsequent recall test, pre-tested items were better recalled than ones that were only studied. This result pattern suggests that just attempting retrieval by itself can facilitate subsequent encoding of the pre-tested items (Grimaldi & Karpicke, 2012; Hays, Kornell, & Bjork, 2013; Kornell, 2014; Kornell et al., 2009; Wissman, Rawson, & Pyc, 2011). However, this phenomenon should perhaps be called generation-potentiated learning (cf. Arnold & McDermott, 2013b), as pre-testing of targets involves generated
target guesses retrieved from semantic memory rather than retrieval attempts from episodic memory. Several findings support this differentiation. For example, increasing the delay between the pre-test and study decreases or even obviates the potentiating effects of pre-tests (Grimaldi & Karpicke, 2012; Hays et al., 2013, Experiments 1 and 2, but see Experiment 3 for a different result; Huels & Metcalfe, 2012; Kornell, 2014), and the pre-test effect was not observed for weakly related word pairs (Grimaldi & Karpicke, 2012; Hays et al., 2013; Kornell et al., 2009). However, test-potentiated learning in multitrial learning paradigms facilitates the re-encoding of unrelated items (e.g., weakly associated words in Nelson et al., 2013; foreign-language word pairs in Arnold & McDermott, 2013b; Vestergren & Nyberg, 2014) and pertains to restudy opportunities that are provided in a separate trial after many other test events occurring in between (Arnold & McDermott, 2013a, 2013b; Vestergren & Nyberg, 2014).

In sum, testing is a powerful study technique that can enhance the future retainability of information through successful retrieval attempts from memory. Furthermore, it can enhance the impact of subsequent restudy in terms of learning unsuccessfully retrieved items (Arnold & McDermott, 2013b; Izawa, 1969, 1970) and perhaps also successfully retrieved items (Vestergren & Nyberg, 2014).

Additional Testing Effects
In addition to the direct and indirect benefits of testing, recent studies have indicated that testing provides the learner with a multitude of other benefits when compared with restudying. For example, testing bestows the learner with a metacognitive benefit. That is, retrieval practice enhances the learner’s ability to accurately predict his or her own future performance (Nelson & Dunlosky, 1991; for a review, see Rhodes & Tauber, 2011), presumably because retrieval practice exposes the gaps in one’s own knowledge. Taking a test enables learners to discover which items they can and cannot remember. Such a diagnostic retrieval experience guides the learner to select those items that need the greatest amount of restudy, choose effective learning strategies, and efficiently allocate the resources of study time and attention (Finn & Metcalfe, 2007; Koriat, Sheffer, & Ma’ayan, 2002; Nelson & Dunlosky, 1991). Testing apparently also has positive effects on the metacognitive control of learning. For example, Soderstrom and Bjork (2014) investigated how far restudy versus retrieval practice affect self-regulated study time. They showed that learners, after an intermediate test, deployed more study time to items that were previously missed than those that were successfully recalled and increasingly so the more difficult the items were. Learners, after restudy, however, did not spend more time on the items than learners, after intermediate testing, did on successfully recalled items (see also Son & Kornell, 2008) and the study-time allocation did not reflect differences in item difficulty. Without a testing experience, learners overestimate the amount to which they have mastered the
learning of the to-be-remembered materials (for a review, see Bjork, Dunlosky, & Kornell, 2013). In sum, testing facilitates metacognitive monitoring (Kornell & Rhodes, 2013; Kubik, Soderstrom, Jemstedt, & Jönsson, 2014a, 2014b) and metacognitive control of learning (Karpicke, 2009).

Testing facilitates not only long-term retention of previously studied information but also learning of subsequently presented new information (for a review, see Pastötter & Bäuml, 2014). Szpunar, McDermott, and Roediger (2008) demonstrated this so-called “forward effect” of testing in a multiple-list learning experiment. Participants learned five consecutively presented lists of words for a final cumulative recall test. Between studying Lists 1–4, they solved either mathematical tasks, restudied the information, or attempted to recall the most recently presented list. When recalling List 5, participants in the tested condition recalled significantly more items and with significantly fewer intrusions from previous lists than participants from the two nontested conditions. Thus, testing shows not only a “backward effect” on previously learned information but also a positive forward effect on the future learning of new information.

However, as noted previously, selective retrieval of some events (e.g., a birthday party) can be detrimental to the retrieval of other events. Similarly, the retrieval of specific details of an event can attenuate the retrieval of other details of the event. Such retrieval-induced forgetting (RIF) has been typically studied in the retrieval-practice paradigm (for a seminal study, see Anderson et al., 1994). Subjects study category–exemplar pairs (e.g., “fruit–apple,” “animal–elephant,” “fruit–banana,” “animal–giraffe”) before they are (repeatedly) prompted to retrieve half the items from half of the categories (e.g., “fruit–ap__”). The critical finding is that nonpracticed items from the practiced categories (e.g., “giraffe”) are remembered worse on a subsequent memory test than the control items from the nonpracticed categories (e.g., “apple,” “banana”). RIF is a reliable phenomenon (for narrative reviews, see Raaijmakers & Jakab, 2013; Storm & Levy, 2012; for a meta-analytic review, see Murayama, Miyatsu, Buchli, & Storm, 2014) and seems to be most pronounced when nonpracticed (competitor) items are moderately associated to the category and thus induce medium levels of interference during prior retrieval practice (Keresztes & Racsmány, 2013). It has been proposed that selective retrieval practice either inhibits or blocks access to nonpracticed information in memory at future retrieval occasions, depend on how retrieval-induced forgetting is measured (Storm & Levy, 2012; Murayama et al., 2014). Conversely, in some situations retrieval of some specific information can support and guide the retrieval of other information (Bäuml & Samenieh, 2010; Collins & Loftus, 1975). For example, positive effects of selective retrieval have been observed when the retention interval between study and retrieval is prolonged or when the internal, spatial or social or context is changed (e.g., Bäuml & Schlichting, 2014). In these cases, the encoding context is less available and may need to be reactivated (Bäuml & Samenieh, 2012). Such reinstatement of the context is enhanced through retrieval (Karpicke et al., 2014).
Thus, the memory dynamics after selective retrieval can be self-restricting or self-propagating, in part depending on the presence or absence of contextual change.

In conclusion, testing benefits learning directly via the completion of the successful retrieval act and indirectly by enhancing subsequent re-encoding. However, the retrieval of particular information from memory can also impair the retrieval of other information.

**Explanations of the Testing Effects**

Despite the extensive research on the multitude of test-related effects, the progress of theoretical understanding lags behind (cf. Roediger & Butler, 2011), and the underlying mechanisms remain unclear. Notably, no general theory of the various testing phenomena has been formulated yet; rather, we only have separate accounts of each type of testing effect. In the following discussion, I will outline the main accounts of the positive effects of testing, as I primarily focus on them in the present thesis.

**Direct Testing Effect**

Numerous accounts have been proposed to explain the direct effect of testing. I will focus here on the item-specific and relational processing account, the semantic-elaboration account, and the memory-strength account. These accounts have received much empirical support and are discussed in the studies conducted as part of this thesis. Two other theories, the amount-of-processing and transfer-appropriate processing accounts, will be mentioned first, as they have been influential but have rendered less empirical support in recent years.

**Early Accounts**

The *amount-of-processing account* (Dempster, 1996; Slamecka & Katsaiti, 1988; Thompson et al., 1978) states that long-term retention is not promoted by testing *per se*, but by overlearning of the subset of items that are re-exposed during successful retrieval. This early account has not received much support. For example, tested material (i.e., ST) is typically better retained than non-tested material that was not further processed (i.e., S–) or material that received additional restudy (i.e., SS). Strictly speaking, the restudy condition permits even more re-exposure to the material than the testing condition, as all materials are restudied, whereas often not all materials are successfully recalled in a test situation (Karpicke & Roediger, 2008). Thus, additional re-exposure to material during testing can be neglected as a necessary factor explaining the direct testing effect.

The *transfer-appropriate processing account* (Blaxton, 1989; Morris, Bransford, & Franks, 1977) states that the final test performance is a function
of the match between processes required at the final test and those initiated during encoding. The greater the encoding–retrieval match, the better is the memory performance (Kolers & Roediger, 1984; Morris et al., 1977). This account can explain the testing effect in that interpolating tests during the learning phase increases the similarity to the final test. However, the prediction that identical test formats between intermediate and final tests should result in the strongest testing effects has not always been supported. In fact, the testing effect increases as a function of retrieval difficulty during both initial and final tests (cf. Halamish & Bjork, 2011). The testing effect is more pronounced when more demanding retrieval measures are used, such as free recall (Carpenter & Delosh, 2006; Glover, 1989) or tests involving brief answers (Kang, McDermott, & Roediger, 2007; McDaniel, Anderson, Derbish, & Morisette, 2007), instead of less demanding measures such as cued-recall or recognition tests (Halamish & Bjork, 2011).

These two early accounts are descriptive and do not provide a satisfactory explanation for several findings, such as the impact of retrieval effort on the direct testing effect.

**Item-Specific—Relational Processing Account**

The item-specific and relational processing account has successfully explained the testing effect for word pairs (Karpicke & Zaromb, 2010; Mulligan & Peterson, 2014; Peterson & Mulligan, 2013; Zaromb & Roediger, 2010) and for visual–spatial information (Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). Recent research has shown that testing via cued recall leads to higher retention compared with reading or generating information, both in final recognition and in free-recall tests (when intermixing the presentation of both study types). Both findings indicate that an increased amount of item-specific processing is elicited through testing (Karpicke & Zaromb, 2010). Furthermore, testing also enhances cue–target relational processing—a type of processing that is especially required in cued-recall tests (Carpenter, 2011; Kubik, Söderlund, Nilsson, & Jönsson, 2014; Peterson & Mulligan, 2013). When free recall (and not cued-recall) is used in the intermediate tests and lists of categorized items are given, testing also enhances the semantic processing between items, as indicated by an increased number of recalled categories and a correlation with free-recall performance (Congleton & Rajaram, 2011; Zaromb & Roediger, 2010). However, testing seems to disrupt the retention of order information (Karpicke & Zaromb, 2010; Mulligan & Peterson, 2014). Thus, the way, in which testing is implemented, influences the nature of the elicited retrieval processes.

The most prominent contribution of this account is the theoretical prediction of a negative testing effect. Peterson and Mulligan (2013) assumed that testing via cued recall enhances item-specific processing and the encoding of cue–target relations at the expense of inter-item relational information (e.g., categorical or temporal relationships between items across trials). To investigate this notion, the authors applied a typical testing-effect paradigm, which
included three phases. In Phase 1, 36 rhyming cue–target word pairs (with targets stemming from six semantic categories) were studied and subsequently either restudied or tested by cued recall. In Phase 2, the items were presented in their category groups. After a 15-min distractor task, a final test was given (either free recall, cued recall, or recognition; Mulligan & Peterson, 2014; Peterson & Mulligan, 2013). As the testing was presumed to encode item-specific and cue–target relational information during Phase 2 at the expense of inter-item information (i.e., the items’ categories), a positive testing effect was predicted in recognition and cued-recall tests that drew on item-specific and cue–target relational information, respectively. In contrast, a negative testing effect was predicted for free recall of targets, as this memory test additionally relies on categorical information that should have been less encoded for tested items during Phase 2. These predictions were supported: testing produced worse recall performance and less categorical clustering than restudy in the free-recall test, whereas positive effects were found in cued-recall and recognition tests.

In sum, testing promoted item-specific and cue–relational processing at the expense of inter-item relational processing when retrieval was practiced with a cued-recall test. However, the processes underlying testing are adaptive to the specific task requirements (e.g., type of memory tests). Testing with free recall can facilitate both item-specific and inter-item relational processing.

Semantic-Elaboration Account
According to the semantic-elaboration account, memory retrieval elicits semantic elaboration in terms of adding extra information to the representation of the learned items (the elaborative retrieval hypothesis; Carpenter, 2009; Carpenter & DeLosh, 2006). When one retrieves the target word (e.g., “bread”) from a cue word (e.g., “basket”), other items are activated that are semantically related to the cue (e.g., “eggs,” “flour”). These semantic mediators (Carpenter, 2011) help to link cue and target words and provide additional retrieval routes to the target word. The more difficult the act of retrieval is, the more elaborate is the semantic search of memory of the target word, and the more semantic associates are activated to mediate the cue–target relationship. This benefits later retention of the target word. However, restudy involves little elaboration when only the entire cue–target pair is presented.

In support of this assumption, the testing effect was greater for weakly related compared with strongly related cue–target pairs (Carpenter, 2009), for more difficult tests (e.g., free recall compared with recognition; Kang, McDermott, & Roediger, 2007), and after longer RIs (Karpicke & Roediger, 2007; Pyc & Rawson, 2007). Furthermore, testing cue–target pairs increased the false alarm rate of target-associated distractor items (i.e., semantic mediators) compared to restudying, and it increased later target recall when the semantic mediator was used as the cue word (Carpenter, 2011). These findings suggest that testing activates additional, semantically related information.
Memory-Strength Account

An early and still influential idea used to explain the testing effect is related to the strength of memory traces or representations (Bjork, 1975, 1994). According to the memory-strength view, any form of studying strengthens the underlying representations of the learned information. Furthermore, it is assumed that memory retrieval leads to even greater strengthening than restudy, as it requires more effort and thereby produces deeper retrieval processing (retrieval-effort hypothesis, Bjork, 1975). To measure the retrieval effort, several factors have been varied, such as the cue support given for retrieval practice during learning (Carpenter & DeLosh, 2006; Finley, Benjamin, Hays, & Bjork, 2011) or the criterion to be achieved at the end of the learning phase (Pyc & Rawson, 2009).

Testing is thought to improve long-term memory by reducing forgetting, but more recent implementations of this idea have proposed that testing only appears to reduce forgetting relative to restudy and that the testing effect may instead be explained through different memory-strength distributions resulting from testing and restudy (Halamish & Bjork, 2011; Kornell et al., 2011). This bifurcated-distribution model—a simplified version of the classical New Theory of Disuse (Bjork & Bjork, 1992)—assumes that the to-be-learned items are normally distributed in memory strength (as a first approximation; cf. Halamish & Bjork, 2011), which changes as a function of study type. In the testing condition, correctly recalled items are presumed to gain dramatically in memory strength, whereas nonrecalled items are presumed to remain unchanged (Halamish & Bjork, 2011; Kornell et al., 2011) or even decline in memory strength (given the evidence for retrieval-induced forgetting, see Anderson et al., 1994). The result is a bifurcated distribution of memory strength for tested items. In contrast, in the restudy condition, all items are re-exposed and receive a similar boost in memory strength, such that the items remain normally distributed across items and participants. Even assuming equal rates of forgetting, these different item-strength distributions would give the memory advantage to restudy conditions after shorter RIs (i.e., the items studied more extensively would have a memory strength above the threshold) and to testing conditions after longer RIs; most successfully tested items would still reside above the threshold despite a decrease in memory strength over time.

Importantly, most of the central findings related to the testing effect are consistent with this account, including the moderating influence of retrieval difficulty of the intermediate and final test on the direct testing effect.

---

2 The new theory of disuse further distinguishes between retrieval strength (i.e., the current target access given the cue word) and storage strength (i.e., the degree of association strength between the target word and other item representations). Retrieval strength determines the current recall performance at any given moment (i.e., memory strength), while storage strength acts as a latent variable that moderates the gain and loss functions of retrieval strength across time and restudy events. However, it is beyond the scope of this review to further elaborate on this dual-strength version of the distribution model.
(Halamish & Bjork, 2011), spacing effect of testing (Landauer & Bjork, 1978), and role of feedback or restudy (Kornell et al., 2011). This account has led to new theoretical predictions (Storm, Friedman, Murayama, & Bjork, 2014; van den Broek et al., 2013). For example, as memory strength is higher on average for retrieved items than for restudied items, retrieval latencies are predicted to be shortened for tested items after a short RI. Such a finding of test-reduced retrieval latencies in combination with the test–delay interaction in retrieval performance has recently been provided (van den Broek et al., 2013).

Indirect Testing Effect

Compared with the direct testing effect, the indirect effect of testing has received much less empirical attention and the theoretical explanations of why tests potentiate subsequent encoding are underdeveloped. As reviewed earlier, they have been investigated based on several research designs: the inclusion of additional restudy following the test trial (Arnold & McDermott, 2013a, 2013b; Nelson et al., 2013; Vestergren & Nyberg, 2014); immediate restudy, given item by item during the test trial; providing corrective feedback; or pre-testing the items (Grimaldi & Karpicke, 2012; Hays et al., 2013; Kornell, 2014; Kornell et al., 2009). As the studies included in this thesis concerned repeated study–test practice in a multitrial paradigm, I will focus on the theoretical notions suggested to explain the indirect testing effect involving separate, subsequent restudy (Grimaldi & Karpicke, 2012; Kornell, 2014 for accounts related to “generation-potentiated learning” on immediate restudy).

Researchers have commonly invoked the idea of elaborative retrieval involving the use and shift of semantic mediators. This notion was most explicitly formulated in the semantic-mediator account by Pyc and Rawson (2012). This account broadly consists of two hypotheses. First, it is posited that learners produce higher recall rates under repeated test–restudy practice as testing generates more effective mediators that link the cue and target words in paired associates (Pyc & Rawson, 2010). Importantly, this increased mediator effectiveness is assumed to arise partially from a mediator shift that is elicited by interleaved testing during restudy. More specifically, by attempting to recall the target word during the test trial, learners have the experience of retrieval success and retrieval failure for each item. In cases of retrieval failure, learners are prompted to replace the previously used mediator(s) with more effective ones during restudy; in cases of retrieval success, they likely continue to generate the previous mediators. On the other hand, as restudying the material (without testing) does not provide systematic retrieval experience, only a slight mediator shift is prompted, even in cases where it would be useful.

Support for this account was given with a specific keyword-generation procedure (Pyc & Rawson, 2012). During initial study, participants learned Swahili–English word pairs (e.g., “wingu–cloud”) and subsequently received either repeated test–restudy or restudy-only practice. Critically, after each (re-
studied item, a keyword mediator (“wing”) that is phonologically similar to the Swahili cue word (“wingu”) and semantically connected to the English target word (“cloud”) should be generated and reported. After two days, a final cued-recall test was given, followed by a memory test for the keyword mediators. The results supported the theoretical predictions. Beyond the increased recall performance for the test–restudy compared with the restudy-only condition, keyword mediators were more likely to be shifted in the test–restudy condition and shifts occurred more often following retrieval failure compared with retrieval success. In addition, after interleaved testing, more keyword mediators were recalled.

Several points remain unclear, however. For example, participants do not normally receive instruction to generate semantic mediators (e.g., keywords) during encoding, and they might not spontaneously generate mediators at all if not prompted to do so. It is possible, though, that nonrecalled or incorrectly recalled items may at least receive additional attention and thus may be associated with more effective encoding strategies during restudy (the strategy-shift hypothesis, Bahrick & Hall, 2005). Furthermore, it is still unknown how the learner is informed about the items’ retrieval status (recalled vs. nonrecalled) during restudy. This diagnostic information needs to be conveyed somehow to the learner at the time of restudy, as it is separated from prior testing through many items presented in between.

Two recent theoretical suggestions shed light on this issue. First, Nelson et al. (2013) suggested that retrieval processes occur involuntarily during restudy (Greene, 1989) and that these retrieval processes are increased through test-potentiated learning. They reported as evidence that not only retrieval practice per se (compared with restudy or no study at all) but specifically the amount of new learning was associated with increased activations in the left lateral and medial parietal cortex (i.e., the left posterior inferior parietal lobule [pIPL]/dorsal angular gyrus [AG]). Participants who learned a higher proportion of the previously nonrecalled items during restudy exhibited greater activation in this brain region, though this correlation was not present with regard to recall performance for previously recalled, restudied, or nonstudied items. As the pIPL/AG region features often successful recognition performance (Nelson et al., 2010) or involuntary attention to an unexpectedly familiar stimulus (Jaeger, Konkel, & Dobbins, 2013), these activations may reflect reminders of the previous, unsuccessful recall attempts (Hintzman, 2004) or at least of the tested cues (as the targets were not recalled).

Second, Vestergren and Nyberg (2014) suggested that tested items may be differentiated from other items during restudy by means of an automatic tagging process; they proposed that the cues of tested items are tagged during the recall attempt with regard to their testing status. The cue word and its associated tag may then be retrieved during restudy, making the tested items more salient. As evidence, they reported that the AI—a region prominently involved in attention and specifically in salience tagging of items (Menon & Uddin, 2010)—showed more activation during the restudy of tested items that were
then successfully recalled in the post-restudy test compared with those items that were not recalled. In turn, tested items may receive more and deeper re-encoding, as the authors demonstrated in terms of increased activations in the left hippocampus and the ventrolateral prefrontal cortex (left orbital part of the inferior frontal gyrus [LOIFG]). Both these brain regions are involved in the encoding processes of memory; the left hippocampus has been linked to the binding of items within an event (Sullivan Giovanello, Schnyrer, & Verfaellie, 2004) and the LOIFG to the controlled access to concepts (Badre & Wagner, 2007). However, the activations in both regions did not differ as a function of the recall status of tested items. Thus, the indirect testing effect may occur for not only unsuccessful but also successful retrieval attempts—that is, for all tested items in general.

In sum, testing enhances the re-encoding of information during restudy. However, the underlying mechanism for such test-potentiated learning remains unspecified, partly because behavioral paradigms provide only indirect evidence about the mechanisms employed during restudy by analyzing pre- and post-restudy recall. Recent findings suggest that in addition to enhanced re-encoding, study retrieval (in the form of remindings) and an attention-modulating tagging mechanism help to identify and boost the encoding of tested items when restudy is given at a separate time after a delay.
How Does Motor Enactment Affect Memory?

Enactment Effect

Several research groups (Cohen, 1981; Engelkamp & Krumnacker, 1980; Saltz & Donnenwerth-Nolan, 1981) independently developed the experimental paradigm of subject-performed tasks (SPTs; for reviews, see Engelkamp, 1998, 2001; Nilsson, 2000) with two different goals. Cohen (1981) investigated memory for actions as a means for testing the generalizability of principles that have been reliably established in research on verbal learning and memory with word materials. Engelkamp and Krumnacker (1980) sought to explore ways to enhance memory by optimizing encoding strategies.

In SPTs, the participants are requested to encode a list of action phrases (mainly in the form of verb–noun phrases, e.g., “open the window”) and perform the specified actions. SPTs have traditionally been compared with verbal tasks (VTs), in which the same items are verbally encoded under standard learning conditions (visual or acoustical presentation) without the need to additionally enact them. The by now well-replicated finding is that SPTs are better remembered than VTs. This enactment effect improves memory performance by approximately 20%–30% (Nyberg & Nilsson, 1995) and has been reliably demonstrated across various recall tests, such as free recall (Peterson & Mulligan, 2010; Schatz, Spranger, Kubik, & Knopf, 2011; see also Kubik & Knopf, 2014), cued-recall (Kormi-Nouri, 1995; Steffens, Jelenec, & Mecklenbräuker, 2009), and recognition tests (Engelkamp, Zimmer, Mohr, & Sel- len, 1994; Russ, Mack, Grama, Lanfermann, & Knopf, 2003; Steffens, Jele- nec, Mecklenbräuker, & Thompson, 2006). However, in addition to verbal encoding, other control conditions have been employed, such as motor imagery (Ecker & Engelkamp, 1995) and observing the experimenter performing the actions (experimenter-performed tasks (EPTs), Cohen, 1981; Koriat, Ben- Zur, & Druch, 1991; Schult, Stülpnagel, & Steffens, 2014). Typically, SPTs are also better retrieved than these control tasks, but this depends on the experimental design and the chosen memory test (for a review, see Zaromb & Roediger, 2010).

The enactment effect is also reliable across various encoding variables (intentional and incidental learning: Zimmer & Engelkamp, 1984; blindfolded participants: Engelkamp, Zimmer, & Biegelmann, 1991; Kormi-Nouri, 2000) and types of memory tests, as reported earlier. Analyses from the population-based Betula study, which included 1000 participants in the age range of 35–

Although this encoding effect seems to be robust in nature, some studies have failed to replicate it. For example, people diagnosed with schizophrenia (Daprati, Nico, Saimpont, Franck, & Sirigu, 2005) did not exhibit any enactment effect nor did people diagnosed with frontal lobe syndrome, suggesting that processes related to action planning may contribute to the enactment effect (Knopf et al., 2005). Furthermore, Engelkamp (1986) showed that for paired-associate learning, enactment disrupts rather than enhances later memory performance. More specifically, during the learning of verb–verb pairs under verbal or enactive encoding instructions, enactment leads to worse cued recall (i.e., recall of one the verbs prompted by the other) compared with verbal encoding. However, when the participants were explicitly instructed to integrate both verbs into an action, the negative enactment effect disappeared (Engelkamp, Mohr, & Zimmer, 1991; Helstrup, 1991). Additionally, enactment does not spontaneously facilitate the retention of temporal information. That is, in serial recall tests, when the participants tried to recall the learned items in the original order of encoding or reconstruct the temporal order when the items were presented to them, they did not perform better after enactment compared with verbal encoding (Olofsson, 1996). Finally, after enactment of the action phrases, the participants could not effectively predict their future recall performance with respect to which item would later be remembered and which would not be remembered. Memory predictions for word material were reliably higher (Cohen, 1983, 1988, 1989; Cohen & Bryant, 1991), suggesting that enactment seems to hamper metacognitive monitoring.

Strikingly, several reliable phenomena established in verbal learning research have failed to be replicated in SPTs. For example, deep (i.e., semantic) relative to shallow (i.e., phonological) processing typically boosts memory performance (for a seminal study, see Craik & Lockhart, 1972). This levels-of-processing effect was nonexistent or reduced in SPTs (Cohen, 1981; Nilsson & Craik, 1990; Zimmer & Engelkamp, 1999). Similarly, the mnemonic effects of generating relative to reading or hearing information were not observed for SPTs (i.e., generation effect; Nilsson & Cohen, 1988). Furthermore, it was shown that enactment changes the typical U-shaped serial position
curve by shortening or eliminating the primacy effect and extending the recency effect (Cohen, 1981; Schatz et al., 2010; Seiler & Engelkamp, 2003; Zimmer et al., 2000).

In sum, although the enactment effect is a pervasive encoding phenomenon, there are experimental circumstances and conditions in which this effect does not emerge. Furthermore, a peculiarity of active learning and memory for actions (i.e., SPTs) is that several robust memory phenomena from the verbal domain are not replicated in an enactive encoding context or are at least diminished.

Explanations of the Enactment Effect

Several robust phenomena of verbal learning research are decreased or lacking in SPTs. These findings have raised the possibility that other laws may be at work in the memory for actions. However, memory research has not supported this claim, as these findings are not universal but rather dependent on the specific experimental conditions (e.g., design and retrieval tests). Thus, it might be claimed that for memory in general (Roediger, 2008) and action memory in particular (Zaromb & Roediger, 2010), no general law-like statements can be made. Instead, four main explanatory accounts have been proposed to account for the existence and variable size of the enactment effect (for more comprehensive reviews, see Engelkamp, 1998; Nilsson, 2000; Zaromb & Roediger, 2010; Zimmer et al., 2001).

Cohen (1981, 1983) invoked the notion of enactment as an optimal form of encoding that does not require any study strategies—such as rehearsal or creating associations among to-be-learned items—to enhance later memory performance. Given the “automatic” nature of the enactment effect, it should increase in size when strategy use is minimal during study or retrieval. Later, Cohen (1989) added the assumption that motoric execution is a critical aspect in producing the enactment effect; however, he provided no specific mechanistic explanation.

Bäckman and colleagues (Bäckman & Nilsson, 1984, 1985; Bäckman, Nilsson, & Chalom, 1986) proposed that enactment relies on not only nonstrategic but also strategic processes (e.g., rehearsal). Reading the action phrases employs strategic processes for both VTs and SPTs; however, SPTs involve additional physical components (e.g., movement, involvement of real objects) that are encoded nonstrategically (Bäckman et al., 1991, 1993). Furthermore, they assumed that the retrieval of verbal components is explicit, whereas the retrieval of physical information is implicit (Nilsson & Bäckman, 1989). On the basis of this dual-component notion, the enactment effect should increase in size when physical aspects are added and, in contrast, should decrease when the amount of strategy use and thereby the verbal aspects are experimentally reduced (e.g., by dividing the attention through a distractor task).
Engelkamp and colleagues (Engelkamp & Zimmer, 1983, 1984, 1985) proposed a third account. They claimed that the additional encoding of motor information leads to the enactment effect. In later years, they developed a system-oriented approach, according to which episodic memory can be divided into independent, modality-specific systems. Expanding upon the dual-code theory of Paivio (1971), they assumed separate systems for verbal and visual input modality and for motoric output modality, with each of them comprising specific codes and representational formats. According to this view, performing actions activates an additional motor system and enriches the representations, which in turn results in the enactment effect. Engelkamp and colleagues (Engelkamp, 1990) added the notion of item-specific and relational processing to action memory—a notion that has become popular in verbal learning research (for a seminal paper, see Hunt & Einstein, 1981; Hunt, 2006). Enactment was supposed to foster the individual features of single items (i.e., verb and noun). Relational processing between the single items and the semantic categories to which the noun components of the phrases belong were supposed to be independent of the enactment effect (Engelkamp, Seiler, & Zimmer, 2005; Engelkamp & Zimmer, 1994). Later, they included the assumption that enactment depends primarily on nonstrategic processes and, more specifically, that it may trigger an automatic “pop-out” retrieval mechanism (Zimmer et al., 2000).

Finally, another account was proposed by Kormi-Nouri and colleagues (Kormi-Nouri, 1995; Kormi-Nouri & Nilsson, 1998, 1999). In contrast to the previous accounts, they assumed that enactive encoding is strategic and that enactive and verbal encoding are two different variants of conceptual processing, differing only in the degree to which the learner becomes involved during encoding. The increased self-involvement following enactment brings a more fine-grained registration of the situation and leads to the following “integration” effects. First, action phrases become more strongly related to each other, for example, by virtue of the shared semantic category (“between-item integration”). Second, the verb and noun components within action phrases become more related to each other (“within-item integration”). Third, the participants become more integrated with the environment (“subject–environment integration,” Kormi-Nouri & Nilsson, 2001). As all three integrative effects are critical for the enactment effect and occur simultaneously during an encoding episode, they called it the “episodic integration account” (Kormi-Nouri & Nilsson, 1998). Translated to the item-specific–relational account, these authors assumed that enactment fosters not only item-specific processing but also category-relational and verb–noun relational processing. Furthermore, they viewed episodic memory as a unitary memory system (Tulving, 1972, 1983) that permits one to recollect past experience, irrespective of whether information is encoded verbally, visually, or motorically. Thus, in contrast to the account by Engelkamp and colleagues, no separate modality-specific systems are required to explain the enactment effect.
Theoretical Questions

In general, the accounts reviewed above present three theoretical dimensions that can be viewed as three theoretical questions. Which among the following does enactment rely on: item-specific versus relational processing, conceptual versus motoric elaboration, or automatic versus strategic processing? Contemporary research attempts to answer these more specific research questions rather than the previously outlined accounts. Much attention has been focused on the notion of item-specific versus item-relational processing, which is also central to the studies comprising this thesis.

Item-Specific versus Relational Processing

Theoretical discourse has been extensively evaluated in terms of the assumptions regarding item-specific and relational processing, which is a central account in the verbal learning tradition (Einstein & Hunt, 1981; Hunt, 2006; Hunt & McDaniel, 1993; Mulligan & Lozito, 2004). According to this account, enactment can potentially facilitate two processes: item-specific and item-relational processing. Item-specific processing refers to the encoding of the items’ individual features that makes them distinct from each other and thereby better recallable at retrieval. Cue–target relational processing refers to encoding of the association between the cue and target word and, in the case of action phrases, between the verb and noun (i.e., verb–noun relational processing). As the whole phrase is considered as the unit of analysis, this binding mechanism is often referred to as (within-) item integration and regarded as a type of item-specific processing (Kormi-Nouri, 1995; Kormi-Nouri & Nilsson, 1998, 2001; Kubik, Söderlund et al., 2014; Steffens, Buchner, Wender, & Decker, 2007; Steffens et al., 2009). Inter-item relational processing refers to the encoding of information that is shared between items, such as temporal, semantic, and/or idiosyncratic associations among them, and is supposed to constrain the search for the to-be-remembered items (Hunt, 2006; Hunt & McDaniel, 1993). For example, “apple,” “banana,” and “cherry” all share the same category-relational information “fruit.”

The prevalent theoretical discourse generally agrees that enactment promotes item-specific and cue–target relational processing of the action phrase (Koriat, 1995; Seiler & Engelkamp, 2003; Steffens et al., 2006, 2009), which makes action phrases more distinctive (Spranger et al., 2008; Zimmer, Helstrup, & Engelkamp, 2000). For example, evidence has been provided that the enactment effect is pronounced in cued-recall (Steffens et al., 2006; Wippich & Mecklenbräuker, 1995) and recognition tests (Perrig & Hofer, 1989; Steffens et al., 2009) that mainly tap item-specific information. However, in free-recall tests, the enactment effect is often diminished (Earles & Kersten, 2000) and occasionally even nonexistent (Brooks & Gardiner, 1994; Knopf, 1995; Steffens, 1999; cf. Steffens et al., 2009). These findings have been explained by the fact that successful free recall relies on additional inter-item relational
information, which is not enhanced in SPTs, at least not in cases of symbolic enactment, i.e., where no action-related objects are used (Engelkamp & Zimmer, 1996; Koriat, Pearlman-Avnion, & Ben-Zur, 1998; Steffens et al., 2009; Zimmer, 1991).

For example, Steffens and colleagues (2006, 2009) confirmed these findings by studying lists of action phrases as a function of the encoding type and applying free- and cued-recall tests as well as verb- and noun-recognition tests. For greater theoretical validity, they applied a multinomial modeling approach in which memory performance was decomposed into parameters related to item-specific processing for verbs and nouns, verb–noun relational processing, and inter-item relational processing (which they termed “retrieval”; see Steffens et al., 2006, 2009, for more details). In addition to a substantial enactment effect, they demonstrated that enactment mainly increases the parameters measuring item-specific and verb–noun relational processing. These results confirm that enactment facilitates verb- and noun-specific processing—as earlier shown in increased recognition and free recall performance of verbs (Earles & Kersten, 2000; Engelkamp, Zimmer, & Kurbyuweit, 1995; Engelkamp, Zimmer, & Mohr, 1990; Freeman & Ellis, 2003) and nouns (Earles & Kersten, 2000). Furthermore, cue–target relational processing is enhanced, which supports the previous findings of enhanced cued recall following enactment. However, this seems to be the case only for verb–noun phrases (Helstrup, 1993; Kormi-Nouri, 1995). Enactment is shown to hamper the cued recall of verb–verb and noun–noun phrases (Engelkamp, 1986, 1995), or at least enactment does not enhance the cued recall when no specific instruction is given to integrate the phrases (Engelkamp, Mohr, & Zimmer, 1991; Helstrup, 1989, 1991).

Regarding inter-item relational processing, there exists substantial evidence supporting the notion that enactment does not enhance this mechanism relative to verbal encoding. For example, using lists of independent items, Steffens et al. (2009) showed that the parameter measuring inter-item relational processing was not enhanced (Steffens et al., 2006) and was even hampered in certain cases (Steffens et al., 2009) after enactive encoding. This is in line with research demonstrating that idiosyncratic relations between items are facilitated to a similar degree between encoding types (Koriat et al., 1998). The temporal order among items (Olofsson, 1996) and the relation between item and context (Koriat et al., 1991; Zimmer, 1994, 1996) is even less encoded during enactment. Furthermore, when providing items that are related to a common taxonomic category, verbal and enactive encoding lead to similar semantic clustering scores of items during free recall (Engelkamp et al., 2003; Engelkamp & Zimmer, 2002), suggesting similar degrees of elicited category relational processing. However, when including physical objects, SPTs were more clustered than VTs in relation to the categories (Bäckman & Nilsson, 1984, 1985; Bäckman et al., 1986; Kormi-Nouri & Nilsson, 1999), even though these clustering scores were often not positively correlated to memory performance (Engelkamp & Zimmer, 1996; Zimmer & Engelkamp, 1989).
Thus, the involvement of real objects may moderate the influence of enactment on category-relational processing (see also Nilsson, 2000). However, Koriat and Pearlman-Avnion (2003) showed that encoding types facilitate different forms of organization; verbal encoding increased clustering along episodic–semantic concepts (e.g., gardening: “water a plant,” “snip a rose”), whereas enactive encoding clustered action phrases with regard to the similarity of the bodily movements (“twist a toothpaste cap,” “screw a wire in place”). In sum, enactment seems to enhance item-specific information in terms of the individual features of the verb and noun as well as their relation within action phrases. Enactment, relative to verbal encoding of action phrases, may also enhance organizational processing, probably along a similarity dimension of body movements (Koriat & Pearlman-Avnion, 2003). However, further research is needed to substantiate this hypothesis.

Conceptual versus Motoric Elaboration

It is widely agreed that enactive relative to verbal encoding provides an elaboration of the memory trace in one way or the other. However, it is debated whether motor information is necessarily included in these elaborated memory representations and how critical it is (cf. Engelkamp, 2001; Kormi-Nouri, & Nilsson, 2001; Zimmer, 2001). The accounts from Cohen, Engelkamp/Zimmer, and Bäckman/Nyberg/Nilsson agree in more recent formulations on the notion that the motor component is critical in the enactment effect. In addition, Engelkamp and Zimmer (see Engelkamp, 2001) added the assumption that episodic memory needs to be differentiated in independent information-processing systems that are related to visual, verbal, and motor modalities (Zimmer et al., 2001). In contrast, other researchers explained the enactment effect within one unitary system of episodic memory and assumed that the underlying representations of enacted action phrases are elaborated using the same set of conceptual processes that are also involved in other encoding manipulations (e.g., differentiation vs. integration, Helstrup, 1987; episodic integration, Kormi-Nouri & Nilsson, 2001; see also Knopf, 1992).

Research has not provided conclusive evidence to solve this issue. One line of argument refers to the idea that if the motor component is critical, self-performed actions (SPTs) should be better remembered than observed actions (EPTs). Several studies confirmed this prediction (Engelkamp & Zimmer, 1997); however, others did not, and they showed equivalent memory performance for both encoding tasks (Cohen, 1981; Steffens et al., 2007; Schult, Stülpnagel, & Steffens, 2014). List length and experimental design seem to moderate this pattern of results (Engelkamp & Zimmer, 1997): when EPTs and SPTs were provided across subjects and in short lists, no enactment effect emerged. However, superior SPT recall was observed when the list length was increased (from 24 to 48 items) and both tasks were randomly mixed within the lists during encoding. Thus, the results obtained were inconclusive. A sec-
ond line of evidence refers to brain-imaging studies that showed that the enactment effect is related to the motor areas in the brain. For example, when measuring the differences in the regional blood flow via positron emission tomography, brain regions—including the premotor cortex, motor cortex, and sensosomatory cortex—showed increased activity both during encoding and cued recall of action phrases (Nilsson et al., 2000; Nyberg et al., 2001). These results suggest that encoded motor information is reactivated during retrieval. However, these regions were similarly involved when no physical movement was requested in the motor-imagery condition and likely also in purely “verbal” processing of action phrases, though to a lesser degree. Thus, there seems to be no qualitative difference in the underlying mechanism of reading and acting out a phrase; rather, a quantitative difference is observed.

Automatic versus Strategic Processing
The notion that enactive encoding is nonstrategic has also been discussed. As reviewed earlier, several robust memory phenomena that are related to the employment of strategic processing have not been replicated in SPTs, or at least these effects were substantially reduced, such as in the levels-of-processing effect (Cohen, 1981; Cohen & Bryant, 1991; Nilsson & Craik, 1990), generation effect (Nilsson & Cohen, 1988; see, however, Österreich & Köddig, 1995), and primacy effect (Bäckman & Nilsson, 1984; Cohen, 1981; Seiler & Engelkamp, 2003; Zimmer et al., 2000). All these effects are presumed to reflect the usage of strategies in one way or the other. The levels-of-processing effect refers to the finding that words profit mnemonically from deeper (i.e., semantic) relative to superficial (i.e., phonological) study strategies. The primacy effect is assumed to emerge as a result of increased rehearsal of the first items in working memory. With increasing serial position, the items in the study list are presumed to receive a decreasing amount of rehearsal. Another line of evidence in favor of a nonstrategic account of SPT enhancement on memory was that memory performance after enactment did not vary with age (in children: Cohen & Stewart, 1982; in older adults: Bäckman & Nilsson, 1984, 1985) or intelligence (group with mental retardation vs. control group, Cohen & Bean, 1983). Given that age and intelligence influence verbal memory performance and the ability to employ study strategies, these early findings were considered as evidence for enactment as an optimal, nonstrategic form of encoding that does not require any study strategies.

However, several findings dispute the nonstrategic, optimal nature of enactive encoding. First, enactive encoding can be improved by several other study strategies (e.g., repetition: Schatz et al., 2010, 2011; semantic clustering: Bäckman et al., 1986). Additionally, dividing the attention during encoding through an interference task can deteriorate the recall performance of SPTs, suggesting that a certain degree of strategic processing is involved during enactment. Second, more recent studies showed developmental memory differences for SPTs (Knopf, 1991; Ratner & Hill, 1991), and a longitudinal study
on memory and aging showed similar trajectories of age-related decline in VTs and SPTs (Rönnlund et al., 2003), suggesting that enactment does not make people less resistant to strategy-related memory differences. Instead of defining enactive encoding as intrinsically strategic or nonstrategic encoding, Nilsson (2000) argued that the self-initiated use of strategies may rather depend on task demands such as task difficulty or cue support during retrieval (see also Knopf, 1992). In this regard, SPTs may require lesser strategy use in many experimental situations than VTs or words, but the difference may be quantitative rather than qualitative.

**Methodological Considerations**

Different research traditions have yielded different methodological approaches in action memory research. These differences have likely contributed to the empirical and theoretical inconsistencies.

The first discrepancy pertains to the provision of physical objects that are denoted by the noun in action phrases (e.g., a glass: “lift the glass”). Some research groups typically provide the physical objects for SPTs, but not for VTs (Rönnlund, Nyberg, Bäckman, & Nilsson, 2003; Nilsson, Adolfsson, Bäckman, de Frias, Molander, & Nyberg, 2004). One reason to do so is to instantiate enactment under more ecologically valid and realistic conditions. Another reason is theoretical in nature; some researchers view the enactment effect as related to the multimodal experience of performing an action with a physical object (weight, texture, color, smell, etc.; Bäckman, Nilsson, & Chalom, 1986; Kormi-Nouri & Nilsson, 1999). However, others (Engelkamp & Zimmer, 1996; Schatz et al., 2010; Steffens et al., 2009) view the pure motor execution of an action as critical. Typically, they let the participants symbolically enact the phrases’ content in SPTs in order not to confound the facilitative effects of object presence and motoric execution (SPTs-plus-objects vs. VTs; see Steffens et al., 2007, 2009). However, as it turns out, the additional provision of objects in SPTs is not crucial for the enactment effect (Engelkamp & Zimmer, 1997; Kormi-Nouri, 2000; Zimmer & Engelkamp, 2003), though it can moderate its magnitude under specific conditions (Steffens et al., 2007) and influence the extent of semantic clustering in free-recall performance (cf. Engelkamp et al., 2005).

Another less-discussed research practice refers to the “proper control condition” for SPTs. VTs usually imply only silent reading of the action phrases, whereas SPTs require reading and enacting the phrases. This permits researchers to evaluate the mnemonic value of enactment in addition to simply reading (Engelkamp & Jahn, 2003; Schatz et al., 2011; Steffens, Buchner, & Wender, 2003; Steffens et al., 2009). However, in other studies, participants attentively read and say the action phrases aloud (Spranger et al., 2008). Such vocally-produced action phrases provide a stricter control condition for SPTs, as vocal
production is presumed to enhance distinctiveness and, consequently, the retention of information (production effect; MacLeod et al., 2010). The relative advantage of motoric versus vocal production of action phrases can then be evaluated. Furthermore, some research used EPTs, and not VTs, as control conditions. As reviewed earlier, SPTs outperform EPTs under experimental conditions that facilitate item-specific processing (e.g., using long lists or giving recognition tests; Engelkamp, Jahn, & Seiler, 2003).

The last methodological discrepancy refers to the action-phrase material used. Generally, rather simple, discrete action phrases were used (e.g., “sharpen the pencil”), and these were selected with regard to specific criteria, such as excluding action phrases with body parts as nouns (e.g., “fold your arms”) as they would be available cues during retrieval (see Schatz et al., 2010; Spranger et al., 2008). To this end, the action phrase material was also rated in order to standardize and control the various dimensions in which the action phrases may differ (e.g., regarding familiarity and emotionality, see Molander & Arar, 1998). On the other hand, others (Cohen, 1981) used action phrases that differed largely and even included mental activities, such as “add 2 + 3,” whereas others applied complex, goal-oriented activities with the aim of approaching everyday actions (Foley & Ratner, 2001; Schult et al., 2014).

Given these research traditions that vary across research groups, the enactment effect has been demonstrated even under methodologically stricter conditions, that is, if no action-related objects were given for SPTs (and VTs), if the participants were asked in VTs to read and say the action phrases aloud (Spranger et al., 2008), or even when the participants watched the experimenter performing the actions (Engelkamp & Zimmer, 1983, 1997; for more complex, goal-directed action phrase materials, see Schult et al., 2014; Stefens, 2007). The empirical studies contained in this thesis were conducted under more methodologically “conservative” conditions, that is, use of simple action phrases (without body parts as noun objects), VTs were to be read aloud, and action-related objects were not provided for SPTs. EPTs were not applied as an additional reference condition to SPTs.
Background, Research Objectives, and Methods of Studies

Background

As noted in the preceding sections, retrieving information from memory not only assesses but also produces learning. The act of retrieval alters the future recallability of the retrieved information and enhances the subsequent re-encoding, specifically of nonrecalled information. These direct and indirect testing effects have been demonstrated for various learning materials, though they have not yet been investigated with action-relevant information. This is surprising as everyday memory is typically concerned with past and future actions and may have developed, largely, to keep the same in mind. Thus, a general aim of this thesis is to investigate the effects of testing with action (i.e., verb–noun) phrases as the material to be learned.

The testing effect is often demonstrated by comparing test conditions to nonstudying or additional, though ineffective, restudy conditions (cf. Kornell et al., 2012; see recent exceptions in Karpicke & Blunt, 2011; Karpicke & Smith, 2012). To evaluate the relative effectiveness of testing, active study techniques should be considered as the comparison conditions. To this end, I employed motor enactment as an active study technique. First, using action phrases as the learning material, motor enactment is the major and most effective encoding form, as it produces substantive memory benefits relative to verbal encoding (the enactment effect; for a review, see Roediger & Zaromb, 2010). Second, enacted phrases, in particular, mimic important aspects of real-life actions and could explain how testing effects translate to more ecologically valid objects of everyday memory. Finally, enactment can be viewed as a form of motoric production or generation of actions based on the verbal descriptions of verb–noun phrases. From this perspective, motor enactment instantiates a highly effective case of the more general, well-established encoding phenomena of the generation effect (Bertsch et al., 2007) and the production effect (MacLeod et al., 2010). Therefore, results with enacted phrases may be of more general importance. Thus, an additional general aim of this thesis is to compare the relative and combined effects of testing (vs. study-only) and motor enactment (vs. verbal encoding) on the learning of and memory for action phrases.

Prior studies on action memory primarily investigated the mnemonic effect of enactment in a single-trial learning paradigm, containing one study trial and
one test trial (for a review, see Nilsson, 2000). Only a few studies examined this encoding effect over the course of several study–test cycles (multitrial learning) and employed free-recall tests (Koriat et al., 2002; Koriat & Pearlman-Avnion, 2003; Kubik, Obermeyer et al., 2014a, 2014b; Nilsson et al., 1989). These studies revealed that the robust memory advantage of enactment decreases (Koriat & Pearlman-Avnion, 2003) or even vanishes in certain cases (Kubik, Obermeyer et al., 2014a, 2014b). To this end, the present thesis examines whether the enactment effect persists over a multitrial learning phase when it is assessed with a final cued recall. This was hypothesized to be the case as cued recall is more sensitive to the enhanced cue–target relational processing through enactment. Typically, tests in these multitrial learning paradigms were simply given to measure the degree of learning across learning trials (Koriat & Pearlman-Avnion, 2003; Kubik, Soderstrom et al., 2014b) or the rate of forgetting (Nilsson et al., 1989), though their impact on memory has not been controlled for. It should be noted that only one study (Nilsson et al., 1989) investigated the rate of forgetting for longer retention intervals (RIs) as a function of encoding type. During the learning phase, participants learned several lists of action phrases, either verbally or enactively, with immediate tests following each list. After several RIs (0, 1 days, or 1 week; Exp. 1), they were asked to retrieve the items from all lists on a final free-recall test (i.e., multiple-list learning paradigm: Nilsson et al., 1989; Nyberg, Nilsson, & Bäckman, 1992). Importantly, it was shown that the forgetting rate is similar for verbal and enactive encoding. However, this finding may be affected by the intermediate tests in that they have mitigated forgetting for the verbal more than the enactive encoding condition. To evaluate this possibility, the included studies considered the effects of interleaved testing in a multitrial learning paradigm (i.e., STSTST) by adding a study-only control condition (SSSSSS) in which the action phrases were exclusively restudied during the learning phase. An additional testing was implemented after 1 week in order to assess the rate of forgetting over time as a function of both study type (study–test vs. study-only) and encoding type (verbal vs. enactive encoding).

The item-specific–relational processing framework provides an account for several encoding phenomena in memory (e.g., production effect, generation effect, enactment effect) and also the direct effects of testing (Mulligan & Peterson, 2014; Peterson & Mulligan, 2013). From this theoretical standpoint, I attempt to explain the individual and combined effects of testing and enactment on the memory for action phrases. In the included studies, cued-recall tests were employed, which primarily measure the strength with which the cue item can elicit the target item. As the two study techniques are assumed to commonly rely on cue–target relational processing (Mulligan & Peterson, 2014; Steffens et al., 2009), both testing (relative to study-only) and enactment (relative to verbal encoding) should reduce the rate of forgetting with action phrase material, and they should combine less than additively. However, if both effects turn out to be additive when combined, they may rely on different mechanisms. For example, it has been proposed that testing does not rely on
cue–target or other elaborative processes, as typically assumed for encoding phenomena such as the enactment effect. Instead, direct testing effects are presumed to rely on retrieval-specific mechanisms that restrict the search set of potential target candidates to the cue word (Karpicke & Smith, 2012; Karpicke & Zaromb, 2010), for instance, by reinstating the previous items’ context (Karpicke et al., 2014).

In everyday life, we typically perceive objects that remind us about the performed and to-be-performed actions. It is still unclear whether the recall type (noun-cued recall of verbs vs. verb-cued recall of nouns) moderates the effects of testing on memory. Prior research on the learning of verb–noun pairs (Earles & Kersten, 2000; Kormi-Nouri, 1995; Steffens et al., 2009) and paired noun associates (for a review, see Ekstrand, 1966; Kahana, 2002) shows approximate symmetry with respect to the recall type, though no study has yet investigated whether the effects of repeated study–test practice are also symmetrical across recall types with respect to action phrase material. In Studies II and III, memory was tested through either verb-cued recall of nouns or noun-cued recall of verbs in order to examine whether the type of recall moderates the testing effect on the rate of forgetting. If testing with the two recall types strengthens the verb–noun association to a similar degree, the rate of forgetting should be similar for noun- and verb-cued recall.

So far, no study has addressed the indirect testing effects with action phrase material. In Studies I–III, testing was instantiated through repeated study–test practice. For this reason, an indirect testing effect can occur because testing enhances the re-encoding of unsuccessfully recalled action phrases during subsequent restudy. To evaluate this possibility, I calculated the number of newly retrieved items from the pre- to the post-restudy tests (Arnold & McDermott, 2013b) and examined whether this index of new learning differed between recall types. Any difference in the amount of new learning would reveal that the indirect testing effect occurs and differs in size for noun- versus verb-cued recall of action phrases. In contrast to the expectations for the direct testing effect, I hypothesized that the recall type may moderate the indirect effects of testing. Noun-cued recall of verbs may produce more new learning during subsequent restudy. Nouns—being semantically more stable and concrete than verbs (Earles & Kersten, 2000; Kersten & Earles, 2004)—should be better recognized as prior retrieval cues and thereby may provide more reliable information regarding the recall status during restudy as compared with verbs. This, in turn, should enhance the chances for previously nonrecalled items to receive effective re-encoding following noun-cued recall as compared with verb-cued recall. Although it is difficult to distinguish between both types of testing effects in multitrial learning paradigms, it is important to consider them to the extent possible.
Research Objectives

This thesis deals with the individual and combined effects of interleaved testing (i.e., repeated study–test practice) and motor enactment on the memory for action phrases. The studies included in this thesis are the first to aim at such an integrative approach of these retrieval and encoding effects in action memory. The main research objectives of the three empirical studies contained in the thesis were to investigate

1. the *direct testing effect* with action phrase material. I examined the relative effects of interleaved testing and study-only conditions on the rate of forgetting over an interval of 1 week for both verbal encoding (Studies I and II) and enactive encoding (Studies I and III).

2. the *indirect testing effect* with action phrase material. I examined the test-potentiating effect on the subsequent re-encoding of previously nonrecalled items for both verbal encoding (Study II) and enactive encoding (Study III).

3. *verb–noun asymmetry of testing effects.* Are the direct and indirect testing effects symmetrical with regard to recall type (noun-cued recall of verbs vs. verb-cued recall of nouns)? I investigated with regard to this question for both verbal encoding (Study II) and enactive encoding (Study III).

4. the *enactment effect.* I examined the relative effects of enactive and verbal encoding on short- and long-term retention as well as on the forgetting rate (Study I; Studies II vs. III).

5. the combined benefit of the *direct testing effect* and the *enactment effect* on the rate of forgetting. Do they interact to produce nonadditive effects, or are they additive? (Study I; Studies II vs. III).
Methods of Studies

Participants
In Studies I–III, the participants included undergraduate students from Stockholm University, Uppsala University, and Södertörn University. Participation was compensated with course credits or movie vouchers. The assignment of participants to the experimental groups (study type and recall type) was random. However, in Study I, the study-only groups were recollected because of a programming error. In general, similar participant characteristics were retained across groups, resulting in an equivalent gender ratio, mean age, and working memory capacity for Studies I–III. Thus, potential selection bias effects seemed to be marginal between groups in each study, if at all existent. However, in Study 2, however, there was a consistent difference observed in the age between groups. When the two oldest participants were excluded in preliminary analyses, the age difference was eliminated, but the result pattern remained the same. Thus, age did not influence the results.

Materials
Action phrases (e.g., “squeeze the lemon”) were used as stimuli and were drawn from normative item pools (Study I: 36 items from Molander & Arar, 1998; Studies II and III: 40 items from Kormi-Nouri, 1995). To control for item characteristics, action phrases were selected that were two to four words long, comprised a verb and a noun, and did not include nouns that described body parts (e.g., “lift the arm”) as these nouns would provide extra body-related cues during retrieval. Additionally, in Study I, action phrases were not selected if they were bizarre (i.e., stemming from the lower tertile of the familiarity-rating distribution, such as “share a kiwi fruit”) or highly emotional (i.e., stemming from the upper tertile of the emotionality-rating distribution).

Procedure
A multitrial cued-recall learning paradigm was employed as the experimental procedure, comprising a series of study and/or test trials. During the study trials, the participants were instructed to intentionally learn the action (i.e., verb–noun) phrases (e.g., “lift the glass”) displayed on the computer screen, either by reading the action phrases aloud (i.e., VTs) or by enacting them symbolically (i.e., SPTs). SPTs did not contain any physical objects corresponding to the noun, as memory performance had to be investigated purely as a function of enactment (and not the additional or interactive involvement of objects). SPTs (i.e., motorically produced actions) were compared against a strict control condition including another active study strategy of vocally producing the action phrases (MacLeod et al., 2010). The participants were instructed to

3 In the term “multitrial learning paradigm,” trial refers to the whole study or test phase in which a list of items or cues are presented. An event refers to the single presentation of each item or cue.
attentively read the action phrases and say them aloud. During test trials, the verb or the noun was given as a retrieval cue for the participant to recall the other item of the action phrase. In all three studies, final recall tests were given after a short and long RI.

Scoring and Analyses
Participants’ responses were scored as correct if the original noun or verb was entered on the keyboard. The data were analyzed with conventional analyses of variance (ANOVA) and t-tests, as these are suitable statistical techniques to investigate the statistical relationships between one or several independent variables (dichotomic) and one continuous dependent variable (Tabachnik & Fidell, 2005). An alpha level of .05 was used. Adjusted effect sizes are reported for ANOVAs and t-tests. For ANOVAs, generalized omega squared $[\hat{\omega}^2_G]$ (for a statistical discussion, see Olejnik & Algina, 2003) are used as effect sizes, as the same is less biased for small samples (Carol & Nordholm, 1975) and estimates the effect sizes for the population. The generalized form of omega squared is used, as it permits a comparison of the effect sizes across studies and study designs (within- vs. between-subjects), though it can be substantially smaller. Partial forms of eta (or omega) squared, which are more often used, are sensitive to design differences, as they partially reflect the error variance that is reduced in within-subject designs. Adjusted forms of Cohen’s $d$ are used for t-tests (for independent groups, Cohen’s $d$ unbiased $[d_{unb}]$; for dependent groups, Cohen’s $d_{av}$; Cumming, 2014). Furthermore, 95% confidence intervals for $d_{unb/av}$, based on the noncentral $t$-distribution, are reported. In cases when the assumption of sphericity was violated, the numbers using a Huynh–Feldt correction are reported.
Overview of Studies

Study I

Objective
The objective of Study I was to investigate the individual and combined effects of interleaved testing and enactment on the rate of forgetting and long-term retention of action phrases over a period of 1 week. I hypothesized that both the study techniques effectively engage in verb–noun relational processing that can only marginally, if at all, increase further when combining them. Based on this notion, I predicted that both the study techniques reduce the rate of forgetting, though their combined effects should be less than additive.

Background
As compared to restudying, interleaved testing reduces the rate of forgetting over longer time periods such that long-term retention is enhanced for tested information. This testing effect has been most frequently investigated in cued recall (Carpenter, 2009, 2011) and is due to the successful act of memory retrieval, which reduces the rate of forgetting. Several authors have argued that testing may specifically promote cue–target relational processing (Mulligan & Peterson, 2014; Peterson & Mulligan, 2013) and item-specific processing (Karpicke & Zaromb, 2010).

The enactment effect is a robust encoding effect that is specifically pronounced in cued recall and recognition, due to promoting cue–target relational and item-specific processing. However, little is known about how enactment influences the retention of action phrases over longer time periods. There is scarce evidence that enactment facilitates retention across RIs that are shorter than one day (Knopf, 1991; Nyberg, Nilsson, & Bäckman, 1992), though forgetting did not appear to be reduced over a 1-week interval, as Nilsson et al. (1989) demonstrated.
In Study I, the testing effect was investigated with action-phrase material. Its relative and combined effects with motor enactment on subsequent memory performance were evaluated. More specifically, verb-cued recall was given during intermediate and final tests, which primarily taps verb–noun relational processing in the case of action phrases. As the mnemonic effects of testing and enactment were commonly explained by this binding mechanism, both the study techniques were predicted to exhibit similar effects on forgetting and to have nonadditive effects when they are combined. The additive effects of testing and enactment suggest that they partly rely on different mechanisms. In contrast to enactment, testing may reduce the search set during cued recall instead of directly strengthening the specific verb–noun association.

Method
A total of 112 undergraduate students (i.e., four experimental groups of \( n = 28 \)) engaged in a learning phase and two retrieval phases (for a specific description, see Figure 1).


The learning phase comprised four trials, each separated by a 30-s arithmetic filler task. In the study-only groups, the participants learned 36 action phrases repeatedly during four study trials, either verbally (VT) or enactively.

---

4 In action-memory research, this mechanism is regarded as an item-specific process, as the entire action phrase is regarded as the “item unit” and becomes more distinctive. Thus, in Study I, I used the term “item-specific processing.” In the thesis, however, I will use the more precise term “verb–noun relational processing.”
(SPT). In the study–test groups, the participants alternately studied the learning material and cued-recalled it in intermediate tests. In the retrieval phases, a final verb-cued recall test was given after short (18-min) and long (1-week) RIs. A 2 (study type: study-only vs. study–test) × 2 (encoding type: SPT vs. VT) × 2 (RI: short vs. long) mixed factorial design was used, with the first two factors manipulated between subjects and the last factor manipulated within subjects.

Results

*Figure 2.* Final cued-recall performance as a function of study type (study-only/study–test) and retention interval (short/long), shown separately for verbal encoding (left panel) and enactive encoding (right panel). Error bars represent 95% confidence intervals (CIs) for the mean.


As can be seen in Figure 2, interleaved testing relative to restudy practice reduced the rate of forgetting for action phrases. Consequently, a reliable memory advantage emerged in favor of study–test items after the long RI, whereas memory performance after the short RI was equal for both the study types. However, this testing effect was only demonstrated for verbal and not for enactive encoding. Furthermore, a robust enactment effect was demonstrated after both the short and long RIs. The size of the encoding effect increased from short- to long-term retention only in the study-only condition due to the decreased rate of forgetting for enactive encoding. Thus, both testing and enactment reduced the rate of forgetting, though this effect was less than additive when the two were combined (study-only\textsubscript{verbal} > study-only\textsubscript{enactive} = study–test\textsubscript{verbal} = study–test\textsubscript{enactive}) as demonstrated by a reliable interaction effect between the study type and encoding type. As for long-term retention,
the combination of testing and enactment did not exceed enactment alone; however, it exceeded the long-term retention of testing alone.

Furthermore, memory performance was equivalent for both study types after the short RI, despite the fact that the action phrases in the study–test conditions were less often restudied than in the study-only conditions. This finding may stem in part from the indirect testing effect, that is, the facilitating effect of testing on the subsequent restudy trials, after both verbal and enactive encoding.

Conclusions
This study provides evidence that the testing effect generalizes to action-phrase material, similarly as for various other learning materials (English–Swahili word pairs, Pyc & Rawson, 2012; text materials, Karpicke & Blunt, 2011). Importantly, the testing effect exclusively occurred for verbal encoding but not for enactive encoding. One explanation for this resistance of enacted materials to the testing effect is that testing—similar to motor enactment—effectively establishes a close connection between the cue and target words (Peterson & Mulligan, 2013), and further verb–noun relational processing becomes less efficient. This is in accordance with previous findings that several otherwise robust memory phenomena (e.g., generation effect) have failed to be replicated with enacted materials (Nilsson & R. L. Cohen, 1988).

In addition to the general enactment-related memory benefit, enactment reduced the rate of forgetting; however, this was observed exclusively for the study-only condition. This finding complements previous research. Nilsson and colleagues (1989) demonstrated equal rates of forgetting for verbal and enactive encoding; however, they employed a multiple-list learning paradigm with intermediate tests given immediately after each encoded list of action phrases. These tests may have reduced the rate of forgetting to a higher degree after verbal rather than enactive encoding and equalized potential differences in the rates of forgetting.

However, the effects of enactment exceed those of testing after both the short and long RIs. Combining both study techniques led to superior cued-recall performance compared with testing, but not when compared with only enactment. Thus, enactment can contribute to long-term learning beyond the effects of testing.

Future studies should demonstrate the reliability of the presented findings across the design status of study type (within- vs. between-subjects) and recall type (verb- vs. noun-cued). Importantly, the indirect testing effect needs to be explicitly considered, which also contributes to the final recall performance in a repeated study–test versus study-only cued-recall paradigm.
Study II

Objective
The objective of Study II was to examine both the direct and indirect effects of testing on cued recall for *verbally encoded* action phrases and, importantly, whether recall type (verb- vs. noun-cued) moderates these testing effects.

Background
Testing during learning reduces the rate of forgetting as compared with an equal amount of restudy; I demonstrated this *direct testing effect* in Study I with verb-cued recall of action phrases that were verbally encoded. As outlined earlier, memory performance might also benefit from testing at shorter delays provided that the test trial(s) are followed by a restudy trial (Pyc & Rawson, 2012). In this case, testing may enhance the subsequent re-encoding of information (*indirect testing effect*; Izawa, 1966). However, these direct and indirect testing effects often co-occur and can be difficult to separate in a multitrial learning paradigm.

The main aim of this study was to substantiate the direct testing effect for verbally encoded action phrases in Study I with the study type manipulated within subjects and to examine whether the recall type (noun-cued recall of verbs vs. verb-cued recall of nouns) affects the size of the testing effect. I hypothesized that testing via cued recall strengthens the verb–noun relation more than study-only and predicted the occurrence of a testing effect. If testing with both the recall types engenders an equal amount of verb–noun relational processing, the direct testing effects should have a similar size.

A secondary, more exploratory aim was to examine the indirect testing effect for both the recall types. To isolate the indirect from the direct effects of testing, I conducted conditional analyses. I decomposed the recall performance of the post-restudy tests into items that were newly retrieved and those that were retained from the pre-restudy tests (Arnold & McDermott, 2013b). A differential amount of new learning between the recall types would indicate that the indirect testing effect occurs and its size differs for noun- versus verb-cued recall.

Method
The final convenience sample consisted of 31 Swedish participants (*age M [SD] = 25.65 [5.96]; range = 18–42 years; 22 females and 9 males). Four additional participants were recruited but were excluded from the analyses as incomplete data were obtained. The experimental procedure was similar to
Study I. The participants learned 40 action phrases (i.e., verb–noun) by reading them aloud. Half the action phrases were repeatedly studied throughout the six learning trials (study-only condition), whereas the other half were learned and cued-recalled in an alternating order (study–test condition). Each learning trial was separated by a 30-s arithmetic filler task. This learning phase was followed by two final cued-recall tests, one given after 2 min (short RI) and the other after 1 week (long RI). During the intermediate and final recall tests, I implemented either noun-cued recall (i.e., the noun was provided to retrieve the verb) or verb-cued recall (i.e., the verb was provided to retrieve the noun).

A 2 × 2 × 2 mixed-factorial design was applied. Recall type was manipulated between subjects, with 15 participants (11 of whom were women) receiving verb-cued recall and 16 participants (11 of whom were women) receiving noun-cued recall. The participants were randomly assigned to both recall-type conditions with the restriction that gender was counterbalanced. Retention interval, or RI, (short vs. long) and study type (study-only vs. study–test) were manipulated within subjects to increase power.

Results

A main finding was the reliable testing effect for verbally encoded action phrases (see Figure 3). That is, repeated study–test practice reduced the rate of forgetting relative to study-only practice, as indicated by a significant interaction effect between the study type and RI.

Figure 3. Final recall performance (mean proportion correct) for the 40 verbally encoded action phrases as a function of study type (study-only/study–test) and retention interval (short/long), separately shown for verb-cued recall (Panel A) and noun-cued recall (Panel B). Error bars represent standard errors of the mean (SEs).
This testing effect had a similar size for both recall types in terms of the rate of forgetting. Furthermore, study–test items were better recalled than study-only items across RIs, as indicated by a main effect of the study type. This testing effect on overall memory performance was increased for noun-cued recall—a finding that may stem in part from the indirect testing effect during the learning phase.

The other main finding relates to the indirect testing effect with action phrases, which was moderated in size by recall type. First, I observed that the learning rate increased for noun- as compared with verb-cued recall; however, this was observed only for the post-tests of the Initial Study and Restudy 1, as can be seen in Figure 4.

Figure 4. Learning curves. Recall performance (mean proportion correct) on the post-tests of Initial Study, Restudy 1, and Restudy 2 as a function of recall type (verb-cued/noun-cued).

Second, this difference in the learning rate early in the learning phase was related to an increased number of items that were newly retrieved from the pre- to post-restudy tests of Restudy 1. However, the number of retained items was equivalent across the recall types. The amount of newly retrieved and retained items for Restudy 2 were comparable across the recall types. I observed these findings even when adjusting for the available number of items that were newly learned or retained (see Figure 5).
Figure 5. Panel A: Newly retrieved targets. Number of targets that were not retrieved on the pre-test but were retrieved on the post-test of Restudies 1 and 2 as a function of recall type (noun-cued/verb-cued). For Restudies 1 and 2, means are adjusted separately for the number of targets available to be newly retrieved during restudy. Panel B: Retained targets. Number of targets that were retrieved on pre- and post-tests of Restudies 1 and 2 as a function of the recall type (noun-cued/verb-cued). For Restudies 1 and 2, means are adjusted separately for the number of targets available to be retained during restudy. Error bars represent standard errors of the mean (SEs).

Conclusions
Repeated study–test practice reduced the rate of forgetting and enhanced long-term retention, relative to study-only practice. This finding confirms the findings of Study I and supports the notion that testing directly strengthened the verb–noun relation within action phrases (Kubik, Söderlund et al., 2014; Peterson & Mulligan, 2013). However, as the type of recall did not moderate the size of this testing effect, the degree of verb–noun relational processing seems to be similar across the recall types.

Testing also enhanced the subsequent recall performance indirectly by enhancing the subsequent restudy of verbally encoded action phrases. This indirect testing effect was asymmetrically increased in favor of noun-cued recall of verbs. These findings were explained by the notion that during recall, the cues are tagged with regard to their recall status. During restudy, these tags may be retrieved (Vestergren & Nyberg, 2014), and the learners are reminded about the prior retrieval episodes (Nelson et al., 2013). I reasoned that as nouns are more stable and concrete, they may be better recognized as prior retrieval cues during restudy and may provide more reliable information about the recall status of items. This, in turn, may increase the chance of the action phrases (related to the noun cue) to receive enhanced re-encoding. Additionally, unsuccessful recall of verbs—the central components of the action phrase—may be more salient than not recalling the noun; therefore, the respective action phrases receive more effective re-encoding.
Together, the direct and indirect testing effects have contributed to the testing benefit on overall memory performance, which was increased for noun-cued recall. As a future step, it is important to investigate both the testing effects with enacted action phrases as a function of the recall type. It was expected that the direct testing effect would not occur after enactive encoding irrespective of whether the study type is manipulated within subjects and testing is implemented with either verb- or noun-cued recall. In contrast, the indirect testing effect on subsequent re-encoding should also be existent for enacted phrases and increased for noun-cued recall.
Study III


Objective
The objective of Study III was to examine the direct and indirect effects of testing on cued recall for enactively encoded action phrases (i.e., SPTs) and how they are affected by the type of recall.

Background
Study II demonstrated that the indirect testing effect occurred for verbally encoded action phrases and, importantly, that it was asymmetrically enhanced in favor of noun-cued recall of verbs. Although we usually remember the actions in our daily lives, no study has yet investigated test-potentiated learning for performed actions. Thus, Study III aimed to examine the indirect testing effects on the subsequent learning of performed actions and investigate whether they are asymmetrical with respect to recall type. As reasoned in Study II, if nouns provide more reliable information about the recall status of actions during restudy, I predicted an increased indirect testing effect for noun-cued recall of actions.

Although the direct testing effect has been robustly demonstrated with various learning materials, I did not observe a test-related decrease in the rate of forgetting with cued recall of enacted action phrases (Study I; Kubik, Söderlund et al., 2014). A main aim of the present study was to substantiate this finding by manipulating study type within subjects and examine the testing effect with two recall types: noun-cued recall of verbs and verb-cued recall of nouns. If testing with both recall types elicits a similar amount of verb–noun relational processing, I predicted no testing effect irrespective of the recall type.

Method
The final sample consisted of 32 Swedish-speaking participants (age $M$ [$SD$] = 24.75 [6.22]; range = 16–42 years; 21 women and 11 men). Two additional participants were recruited but were excluded from the analyses as only incomplete data were obtained. The design, materials, and procedure were identical to Study II, with the exception of encoding type. In Study III, participants enacted the action phrases instead of saying them aloud. Specifically, they
motorically performed the actions described in the verb–noun phrases displayed on the screen without any action-related object at hand. A $2 \times 2 \times 2$ mixed-factorial design was applied, with RI (short vs. long) and study type (study-only vs. study–test) as within-subjects factors. Recall type was manipulated between subjects, with 16 participants (11 of whom were women) receiving verb-cued recall of nouns and 16 participants (10 of whom were women) receiving noun-cued recall of verbs. Participants were randomly assigned to both the recall-type conditions with the restriction that a similar gender distribution should be achieved.

Results
A main finding of this study pertains to the indirect testing effect. The results suggested that the indirect testing effect occurred for enactive encoding and, importantly, that it was moderated by recall type. First, the learning rate was accelerated with noun-cued recall of verbs early in the learning phase. Though the recall performance was similar on the first intermediate test, the recall performance across intermediate tests increased more for noun-cued recall in the succeeding tests (Figure 6).

![Figure 6](image)

*Figure 6.* Learning curves (mean proportions correct and standard errors of the mean, SEs) are displayed as a function of post-tests (of Initial Study/Restudy-1/Restudy-2) and recall type (noun-cued/verb-cued).

Second, the number of newly learned items during Restudy 1 (though not Restudy 2) was increased for noun-cued recall, while the number of retained items between pre- and post-tests of Restudies 1 and 2 was equivalent across recall types. These results were robustly obtained even when adjusting for the available amount of newly retrievable/retainable items (Figure 7). Together, this result pattern suggests that noun-cued recall of verbs engenders a larger indirect testing effect compared with verb-cued recall of nouns.
The other main finding pertains to the direct effect on final recall performance. As can be seen in Figure 8, repeated study–test practice did not reduce the rate of forgetting relative to study-only practice.
Thus, the direct testing effect was not observed in terms of reducing the rate of forgetting, neither with noun-cued recall of verbs nor with verb-cued recall of nouns. However, a testing effect occurred in terms of a general memory advantage for study–test items on final recall performance. This overall testing effect was moderated by the recall type and was only reliable for noun-cued recall of verbs.

Conclusions
Testing enhanced subsequent re-encoding of enacted action phrases during the first restudy trial, and this indirect testing effect increased when the tests included noun-cued recall of verbs. These results support the notion that nouns may provide more reliable information regarding the recall status during restudy or alternatively that verbs produce a stronger tagging of the associated nouns, which then become more salient during restudy. Either way, the chances for action phrases to receive enhanced re-encoding increase to the extent that one of these mechanisms is activated (Vestergren & Nyberg, 2014), specifically when the phrases were previously not recalled (Nelson et al., 2013).

Repeated study–test practice did not reduce the rate of forgetting of performed actions, irrespective of recall type. These findings substantiate the results of Study I and extend them to another study design as well as to another recall type (i.e., noun-cued recall). They are congruent with the view that interleaved testing increased the verb–noun relation to a similar degree for both recall types.

Furthermore, there was an overall testing benefit for the final recall performance, which was reliable for noun-cued recall but not verb-cued recall. This finding stems largely from the indirect testing effect, because the direct testing effect was not reliable in terms of reducing the rate of forgetting over a period of 1 week, and retention across learning trials was equivalent between the recall types.
General Discussion

The aim of this doctoral thesis was to study the effects of testing (via study–test practice) and encoding (via enactment) on learning and memory with action phrases, an ecologically valid material. I used action-phrase materials (i.e., verb–noun phrases such as “water the flowers”) because they mimic the basic aspects of everyday actions. Of particular interest was to investigate the effects of interleaved testing on the rate of forgetting (i.e., the direct testing effect) and on subsequent restudy (i.e., the indirect testing effect) as well as how the type of recall (noun- vs. verb-cued) moderates these testing effects. In combination, the facilitative effects of encoding through motor enactment were investigated on memory relative to verbal encoding (i.e., the enactment effect).

In the following sections, I will discuss the main results from the three empirical studies along the themes of the testing effects and enactment effect. Subsequently, I will outline several challenges and directions for future research related to these themes before concluding with the final remarks.

Testing Effects

Although in our daily lives, we are often concerned with remembering performed and still to-be-performed actions, the facilitative effects of testing have not previously been investigated with action-relevant information. To this end, in Studies I–III, I investigated the effects of testing with action phrases, after both verbal and enactive encoding.

I will first report and discuss the direct testing effect on the rate of forgetting, following which I will discuss the indirect testing effect on the subsequent restudy of action phrases.

Direct Testing Effect

Studies I–III demonstrated the existence of a direct testing effect for action phrases. That is, interleaved testing reduced the rate of forgetting over a 1-week interval, compared to study only, and thereby enhanced long-term retention. However, the testing effect was only observed for verbal encoding (Studies I and II) and not for enactive encoding (Studies I and III). This results pattern was found when both study types (i.e., study-only vs. study–test) were manipulated between subjects (Study I) and within subjects (Studies II and
III). In addition, the type of recall (noun-cued recall of verbs vs. verb-cued recall of nouns) did not affect the size of the testing effect for either verbal encoding (Study II) or enactive encoding (Study III).

The present thesis extended the testing effect to action phrases, supporting its robustness across learning materials (Roediger & Butler, 2010). I hypothesized that testing by means of cued recall enhances verb–noun relational processing if the retrieval attempt was successful. Recent research supports this notion. Retrieval practice led to a testing effect when the final test was cued recall (Peterson & Mulligan, 2013)—a memory test mainly relying on the cue–target association strength (Peterson & Mulligan, 2013; Steffens et al., 2009). It has also been suggested that successfully recalling items activates more semantic mediators in the cue-related network compared to restudy practice and that these mediators may link cue and target more strongly together and, in turn, support later retrieval of the target (Carpenter, 2009, 2011). Although such semantic mediation of the cue–target relation may be a potential mechanism underlying memory for action phrases, it may be a more realistic scenario in other types of learning materials, such as noun pairs. In simple action phrases (e.g., “lift the glass”), the verb and noun are, rather, directly related: the verb specifies a concrete action (“water”) that can be directly performed with the imagined object denoted by the noun (“the flowers”).

Given that motor enactment of action phrases should already promote efficient verb–noun relational processing (Kormi-Nouri, 1995; Steffens et al., 2009), I hypothesized that this binding mechanism can only be marginally, if at all, enhanced through testing. In line with this notion, a direct testing effect, in terms of the rate of forgetting, was observed only for verbal encoding (Studies I and II) but not for enactive encoding (Studies I and III), irrespective of the design status of study type (within- vs. between-subjects). Thus, enactment and testing engendered nonadditive effects on the rate of forgetting (For potential statistical and design-specific qualifications, see section Methodological Challenges.). This results pattern is congruent with the observation that several other well-known memory-enhancing phenomena have not been observed for enactive encoding. Of particular relevance is the fact that generating additional information (e.g., producing a verb for the available noun) also does not improve retention beyond the effect of enactment (Nilsson & Cohen, 1988). This generation effect shares procedural similarities with the testing effect as the generation of associates to a cue can be viewed as retrieval practice from semantic memory. Generation and testing (by cued recall) show similar facilitative effects on memory (Burns, 1990, 1992; Carrier & Pashler, 1992; Mulligan & Peterson, 2014; Peterson & Mulligan, 2013). Furthermore, they are both assumed to enhance cue–target relational and item-specific processing, though not inter-item relational processing (generation effect, Bertsch et al., 2007; McDaniel & Bugg, 2008; testing effect, Karpicke & Zaromb, 2010; Peterson & Mulligan, 2013), similar to enactment (Steffens et al., 2009). However, repeated testing seems to facilitate item-specific processing to a larger extent than repeated study or item generation (Karpicke & Zaromb,
and can enhance inter-item relational processing when testing is implemented via free recall (Zaromb & Roediger, 2010; Congleton & Rajaram, 2012). A notable difference between generation and testing may be that the latter involves retrieving information from episodic memory through an intentional retrieval mode to consciously reconstruct the past (Karpicke & Zaromb, 2010).

The results of Studies II and III indicated that the type of recall did not moderate the size of the testing effect. For verbal encoding (Study II), testing reduced the rate of forgetting to a similar degree for both recall types, and for enactive encoding, no testing effect emerged for either recall type (Study III). Due to the relatively small sample sizes in Studies II and III, I could not exclude the possibility that the lack of influence of recall type was due to low power. However, additional analyses of the combined datasets of Studies II and III (N = 63) confirmed that the influence of recall type on the testing effect was negligible ($\hat{\omega}^2 < .01$). This pattern of results is congruent with the notion that testing via noun-cued recall of verbs engages verb–noun relational processing to a similar degree as verb-cued recall of nouns. The verb–noun association has been potentially strengthened in different directions or alternatively in both directions when testing with noun- versus verb-cued recall. The latter possibility accords with the finding that testing via cued recall enhanced the retention of word pairs (e.g., “train–plane”) more than restudy, irrespective of whether memory in the final test was assessed in the same recall direction (“train – ?”) or the opposite direction (“? – plane”; Carpenter, Pashler, & Vul, 2006). However, as the recall types were always congruent between the intermediate tests in the learning phase and the final tests in the retrieval phase of Studies I–III, the present thesis design could not answer this research question.

In sum, I argue that the successful cued recall of information enriches the memory trace of the action phrase in the form of cue–target relational processing (Hunt, 2006; Karpicke & Zaromb, 2010). The notion that successful recall constrains memory search to the correct targets (Karpicke et al., 2014; Thomas & McDaniel, 2013) cannot be excluded given the relatively high final recall performance and the experimental design of Studies I–III. They will be further discussed in the section “Theoretical Challenges.”

Indirect Testing Effect

In addition to the direct effect, testing also facilitates re-encoding of previously unrecollected items when a subsequent restudy opportunity is given. As testing was implemented via repeated study–test practice, I was also able to examine this indirect testing effect (Arnold & McDermott, 2013a, b; Izawa, 1966).

The results showed that memory performance was equivalent for the two study types (Studies I–III) or even increased in favor of study–test items (noun-cued recall condition in Study III) after the short retention interval. This
finding provides some evidence that the indirect testing effect occurred for action phrases, as study–test items—due to unsuccessful recall attempts at the intermediate tests—were less exposed than study-only items during the learning phase. Only if the re-encoding of nonrecalled items is enhanced through prior testing, study–test items can compensate for the fewer restudy occasions. This interpretation is supported by the fact that recall retention across learning trials was very high. Even though the direct testing effect may contribute to final recall performance by enhancing the retention of successfully recalled items, its mnemonic effect primarily unfolds across longer delays. This interpretation is congruent with previous research showing that after shorter delays, repeated study–test practice leads to equivalent or increased memory performance relative to restudy practice (Arnold & McDermott, 2013b; Karpicke, 2009; Pyc & Rawson, 2012; Vestergren & Nyberg, 2014) and clearly surpasses repeated testing without subsequent restudy, as the latter condition only relies on the direct testing effect (Karpicke & Roediger, 2007).

In addition, the learning rate was accelerated for noun-cued recall of verbs for verbal (Study II) and enactive encoding (Study III) due to more items being newly learned during Restudy 1. The number of retained items, however, was equally high for both recall types. This differential amount of new learning during Restudy 1 provides evidence for the indirect testing effect, at least following noun-cued recall, and its effectiveness to facilitate subsequent learning varies with recall type. I draw this conclusion because prior recall performance, during the first intermediate test, was similar for the two recall types. Furthermore, item retention between the intermediate tests did not vary across recall types—a finding that reflects, in part, the direct testing effect (Arnold & McDermott, 2013b). The increased indirect testing effect for noun-cued recall largely contributed to the test-related benefit in final memory performance, which was only reliable for noun- but not verb-cued recall in Studies I–III.

Together, the indirect testing effect was generalized to action-phrase material for both enactive and verbal encoding, but it was asymmetric with regard to recall type. These findings complement previous research that more directly manipulated the magnitude of the indirect testing effect. For example, when varying the number of testing rounds preceding the restudy trial, it was observed that the size of the indirect testing effect (i.e., the amount of new learning during testing ensuing restudy) increased the more pre-restudy testing rounds were given (Arnold & McDermott, 2013b). Convincing evidence for the indirect testing effect was also provided by Arnold and McDermott (2013a) who manipulated both the availability of tests before restudy and the availability of a restudy opportunity. They observed that the benefit of restudy increased with previous testing as compared to no pre-testing. The present thesis expanded upon these findings by demonstrating that even the type of recall on a single pre-restudy test can affect its potency to enhance subsequent encoding of action phrases.
Although several studies have addressed the indirect testing effect, its underlying mechanisms remain unclear. Vestergren and Nyberg (2014) suggested that the indirect testing effect may rely in part on automatically tagging the cue item of the tested pair associates during pre-restudy recall. Paired associates then become more salient during restudy when the cue items are recognized as previous retrieval probes and their associated tags are retrieved. In addition, Nelson et al. (2013) demonstrated that the amount of new learning was positively associated with retrieval processes during restudy, which may involuntarily remind the learner of unsuccessful retrieval attempts. Together, these mechanisms may then initiate an enhanced re-encoding of tested items.

The cue items during the restudy of the verb–noun phrase may play a significant role if the indirect testing effect primarily applies to items for which the target was not recalled. As nouns are semantically more stable and recognizable than verbs (Earles & Kersten, 2000; Kersten & Earles, 2004), I reasoned that nouns as retrieval cues should provide more reliable information about the associated (unsuccessful) recall status during restudy. This, in turn, may engender more effective re-encoding of the target items. The finding that the relative amount of new learning following noun-cued recall of verbs increased during Restudy 1 supports my hypothesis and indicates an increased indirect testing effect compared to the verb-cued recall of nouns. Furthermore, the unsuccessful retrieval of verb targets during noun-cued recall may produce a stronger tag for the associated cues than the unsuccessful retrieval of nouns because verbs constitute the main component in action phrases (Kersten & Earles, 2004). The more salient cue tags for nouns may function as stronger reminders of unsuccessful retrieval episodes and also initiate enhanced re-encoding during restudy (Nelson et al., 2013; Vestergren & Nyberg, 2014).

In sum, the indirect testing effect was generalized to action-phrase material for both encoding conditions and, critically, was moderated by recall type. One explanation pertains to the retrieval cue, which may function during restudy as a reminder of previous retrieval attempts and, in turn, influence the amount or nature of subsequent re-encoding. To further the understanding of the indirect testing effect in action-phrase material, future research should more directly manipulate the possibility and amount of test-potentiated learning and create a pool of action phrases for which the characteristics of their single component—noun, verb, and directive association strength—are better controlled, as discussed in the section “Methodological Challenges.”

Enactment Effect

The present results speak to the relative mnemonic efficacy of verbal and enactive encoding. As cued recall was used in the included studies, I predicted a reliable enactment effect based on the notion that enactment promotes verb–noun relational processing. The results supported this prediction. In Study I,
enactment facilitated final cued-recall more than verbal encoding after both retention intervals. This result was corroborated when Study II (verbal encoding) was compared to Study III (enactive encoding) regarding final recall performance; a robust enactment effect emerged across the studies, supporting previous findings (Kormi-Nouri, 1995; Steffens et al., 2009; Svensson & Nilsson, 1989). This cross-study comparison is internally valid, as the participants were randomly attributed to studies, and the participants’ characteristics were equivalent among them. The results extend those of prior studies, as the enactment effect is typically examined in single-trial paradigms including only one study trial and one ensuing test trial. However, the included studies demonstrated the enactment effect even for the final cued-recall tests following a multitrial learning phase. Only a few studies investigated the learning of action phrases across multiple trials, mainly including free-recall tests. These studies indicated that the enactment effect attenuated over the learning session with action-phrase material that can be clustered along various pre-experimental categories (Koriat & Pearlman-Avnion, 2003; Koriat, Pearlman-Avnion, & Ben-Zur, 1998) and was even absent for independent action phrases (Kubik et al., 2014). The discrepancy between the present and previous results might thus be explained by the additional reliance of free recall on inter-item relational processing—a mechanism that is not enhanced by enactment, specifically not for independent action phrases (Steffens et al., 2006, 2009). To examine the size and progression of the enactment effect in cued recall during the learning phase, I conducted additional analyses for Study I and the combined Studies II and III. A reliable enactment effect was revealed for the first intermediate test. Even though the enactment effect decreased in size over the learning phase, it remained reliable across all the tests that followed. Thus, this encoding effect seems to be more robust in cued than in free recall, conceivably because the memory test mainly taps the processing of the verb–noun relation, which is strongly enhanced by enactment (Kormi-Nouri, 1995; Steffens et al., 2006, 2009).

Critically, Study I also investigated how far enactment influenced the rate of forgetting over a 1-week retention interval. Only one study addressed this issue more systematically for longer retention intervals (2 days, 1 week; Nilsson et al., 1989) and observed that action phrases were forgotten at similar rates after verbal and enactive encoding. However, these findings may be affected by the tests, which were interleaved after each list to be learned during the learning phase. These additional tests may have reduced the rate of forgetting more for verbal than for enactive encoding and potentially obscured any differences in the rate of forgetting between encoding conditions. In the included studies, I tested this possibility by manipulating study type (repeated study-only vs. study–test practice). Confirming previous research, the results showed that while enactment did not reduce the rate of forgetting in the study–test conditions, it did so in the study-only conditions. Additional, cross-experimental analyses of Studies II and III revealed equivalent results and confirmed the results of Study I. Thus, interleaved testing obviated the finding
that enactment reduced the forgetting of action phrases relative to verbal encoding. This results pattern supported our hypothesis that motor enactment and testing via successful recall may both elicit verb–noun relational processing—a binding mechanism that is supposed to reduce the rate of forgetting in cued recall over time. This finding does not accommodate the observation that no encoding variable reduces the rate of forgetting (Slamecka & McElree, 1983).

Two caveats that may have influenced the results should be noted. First, the two encoding conditions produced different initial recall levels after the short retention interval, which may have influenced the rate of forgetting. However, enactment reduced the rate of forgetting exclusively in the study-only conditions even though initial recall levels were equivalent across study types. Second, the rate of forgetting was assessed by repeated tests, and therefore, the effect of forgetting over time may be confounded by testing effects and generally reduced. However, this circumstance should rather strengthen our conclusions, as the final test after the short delay may have likely reduced the rate of forgetting more for verbal than for enactive encoding. Given these limitations, the reported results suggest that enactment does reduce the rate of forgetting over time. However, further research should explore whether these findings generalize to experimental designs in which longer retention intervals are given, or initial recall levels are lower or equalized across encoding conditions.

Overall, in the present studies, the enactment effect surpassed the (direct and indirect) effects of testing in enhancing short- and long-term retention. Furthermore, combining the two study techniques resulted in superior cued-recall performance than testing alone but not than enactment alone. Thus, enactment maximizes long-term learning beyond the effects of testing, even though they have similar and nonadditive effects on the rate of forgetting. These findings, however, may rely on the fact that not all action phrases were successfully recalled and benefitted from retrieval processing. Other implementations of testing—for example, testing with feedback—may be equally or even more effective in enhancing long-term learning than enactment.

In the included studies, I did not provide any real objects (denoted by the nouns) during learning for either encoding condition. For this reason, the mnemonic benefits of enactment in Studies I–III can be merely attributed to motorically performing the item material and not to any effects related to objects. Even though providing objects would further enhance the ecological validity of the learning materials, it may confound the facilitating effects associated with enactment and object presence, specifically if objects are exclusively given for enactive encoding. Importantly, a verbal control condition was instantiated by asking participants to read the action phrases aloud. This vocal production of action phrases can be viewed as a conservative control condition because it facilitates later memory performance, particularly in cued recall and recognition, relative to the silent reading of items (MacLeod et al., 2010). This production effect has also been explained in terms of increased item-specific
and cue–target relational processing. Based on this assumption, the data in the present thesis suggest that enactment—viewed as a motoric production of actions—is even more effective than vocal production as it probably engages the same item-specific mechanism(s) to a higher degree than in vocal production. Thus, even increased mnemonic benefits of enactment are expected when silent reading is used as the control condition.

In sum, the research reported in the present thesis began from the currently dominant item-specific relational framework in action memory (Kubik, Söderlund et al., submitted; Peterson & Mulligan, 2010; Schult et al., 2014), which also successfully explains other encoding phenomena, such as the production effect, generation effect, enactment effect, and recently, the testing effect. The predictions of this framework for the current thesis were largely confirmed. An enactment effect was revealed in that intermediate and final memory performance were substantially increased and the rate of forgetting reduced (for study-only conditions), suggesting enhanced processing of the verb–noun relation by enactment (Kormi-Nouri, 1995; Steffens et al., 2009). Similarly, interleaved testing reduced the rate of forgetting for verbal but not enactive encoding, providing evidence that both study techniques exhibit non-additive effects on the rate of forgetting and may rely in part on the same mechanism.

Alternative Memory-Strength Explanation

The findings of Studies I–III can be alternatively interpreted in terms of the bifurcated distribution model (Halamish & Bjork, 2011; Kornell et al., 2011). According to this model, forgetting only appears to be mitigated by testing and can instead be explained by the different distribution or amount of memory strength for both study types. Successfully recalled items gain more in memory strength than studied items while nonrecalled items do not gain at all in memory strength (Halamish & Bjork, 2011; Kornell et al., 2011). Due to test-potentiated learning, specifically these nonrecalled items can receive more effective re-encoding and memory strength during subsequent restudy than study-only items, and additionally when being successfully recalled on the post-restudy test. A difference in the resultant item strength between the study types can then explain the observed testing effect, even when we assume equal rates of forgetting. The equivalent memory performance after shorter retention intervals reflects the same amount of study-only and study–test items above the critical memory threshold. The long-term advantage for study–test conditions results from the increased average memory strength for study–test items. They reside more likely above the threshold despite an eventual decrease in memory strength with time.

Similarly, the enactment effect can be explained by this account. Both enacted and verbally encoded items should produce normal item strength distributions whereby the distribution for enacted items would be shifted further
upwards. The increased number of enacted phrases above the threshold leads to a better recall performance while the increased memory strength following enactment leaves more items above the threshold as memory strength decreases with time. Thus, although the rates of forgetting were similar, the enactment effect increased over the retention interval. However, it should be noted that due to the increased recall level after enactive encoding, retrieval was perhaps “undesirably” easier than after verbal encoding, and this decreased retrieval difficulty may have reduced the relative gain in memory strength for testing enacted phrases. When combining testing and enactment, no additive effects on the rate of forgetting were observed. Due to the bifurcated-distribution model, combining testing and enactment may lead to a larger increase in memory strength than when considering each of the study techniques alone. However, this difference in memory strength may not be revealed as many items are too far away from the memory threshold. Based on the distribution model, I would expect that with longer retention intervals or a lower recall level, a testing effect would occur for enacted items and an enactment-related decrease would occur in the rate of forgetting for study–test items. Without any restudy trials following the interleaved tests, the distribution of memory strength should also be bifurcated for enacted phrases (i.e., recalled, but not unrecalled, items gain in memory strength) and thus lead to a typical cross-over interaction between study type and retention interval. Such an experimental design would provide more conclusive evidence about the direct testing effect with action-phrase material by preventing the possibility of test-potentiated learning. Ongoing research investigates these predictions of the bifurcated-distribution model.
Challenges and Directions for Future Research

The present thesis provides a starting point to the investigation of the effects of testing and motor enactment with action-phrase material. In the following sections, I will outline the methodological, theoretical, and practical challenges for future research on this topic.

Methodological Challenges

In this section, I will highlight the design features of Studies I–III that require a more thorough discussion. This discussion is aimed at providing suggestions for future research to overcome potential methodological shortcomings and to further the understanding of the mechanisms underlying the effects of testing and enactment on memory.

Isolating Direct and Indirect Testing Effects

All included studies applied a multitrial learning paradigm, involving both repeated restudy and repeated study–test practice. Since the test trials were followed by restudy opportunities, both direct and indirect testing effects could occur in combination. As recent research (Arnold & McDermott, 2013 a, b) highlights the distinction between these two types of testing effects, I attempted to disambiguate their specific contributions to learning and memory of action-phrase material in the empirical studies. Future research would benefit from isolating both testing effects—where relevant and to the extent possible—and studying them separately. This would allow us to uncover more clearly their common and distinctive underpinnings.

Evidence for the indirect testing effect, at least for noun-cued recall, was provided by the differential amount of new learning between recall types. In addition, the finding that final recall was equivalent across study types or even increased for study–test items—despite being less exposed during the learning phase than study-only items—indicated the existence of the indirect testing effect. However, this evidence may still be fairly indirect. An important next step would be to effectively separate the distinctive contribution of the indirect testing effect from the effect of successful retrieval and restudy. A plausible step would be to manipulate both the availability of a restudy opportunity following testing (e.g., STST vs. ST–T) and the number of tests preceding the
restudy trial (e.g., STTST vs. STST vs. S-ST). Such a design would potentially demonstrate the indirect testing effect beyond the facilitating effect of restudy practice per se (without previous testing). The additional manipulation of recall type would provide more direct evidence for the indirect testing effect following verb-cued recall as well as confirm the differential effectiveness between recall types to facilitate subsequent re-encoding. However, in behavioral paradigms, the evidence for test-potentiated learning remains rather indirect by means of pre- and post-restudy recall analysis. Future research examining neuroimaging may more directly investigate re-encoding during restudy and further elucidate the key mechanism(s) involved. An important area for future research is whether or not successfully recalled items also benefit from test-potentiated learning, relative to restudy per se. A behavioral paradigm cannot easily adjudicate between these possibilities, as the relative amount of new learning for nonrecalled items is the primary evidence for the indirect testing effect. Thus, in the present thesis, the indirect testing effect may have also occurred for the successfully tested items. Recent evidence (Vestergren & Nyberg, 2014)—using functional magnetic resonance imaging—suggests that irrespective of the previous recall status, tested items may receive more focused attention during restudy, consequently initiating deeper semantic processing and more associative binding between cue and target.

Another pertinent issue that can be explored in future research is to determine the relationship between the indirect testing effect and the “indirect generation effect.” That is, an unsuccessful attempt to generate the correct target (“tide–?”) in a pre-test trial facilitates the subsequent encoding of the complete word pair (“tide–beach”) during initial study (Grimaldi & Karpicke, 2012; Kornell, 2014; Kornell et al., 2009). Although both phenomena involve retrieval from memory (episodic vs. semantic), the indirect generation effect does not generalize to unrelated word pairs (“beach–spoon”) or to experimental designs in which initial study is delayed and given in a separate trial with many items interleaved. Future research, including brain imaging, may explore the common as well as distinct processes underlying these different types of the indirect “testing” effect.

The direct testing effect typically manifests in a cross-over interaction effect when no restudy opportunity follows testing. In other words, restudy practice leads to superior memory performance after the short retention interval, but testing surpasses the latter in long-term retention. This classical testing effect most likely results from the unbalanced re-exposure of items (i.e., retrieval success problem). The learner can process all items during restudy but only successfully recalled items during testing. In that regard, restudy—and every encoding manipulation—has an “unfair” advantage to begin with, and the testing effect on long-term retention occurs only due to mitigated forgetting (or the increased memory strength for successful recall; Kornell et al., 2011). In the present thesis, however, the testing effect did not emerge in the form of a crossover interaction for the verbal encoding condition (but see the verb-cued recall condition in Study II for an exception). In fact, the testing
effect is revealed as a pronounced advantage in long-term retention while the initial recall levels after the short delay were equivalent between study types, or the test advantage was smaller. These noncrossover interactions are likely due to the provision of restudy opportunities in the testing condition and the possibility of test-potentiated learning, which together enhance the recall performance for study–test items after the short retention interval. As the direct testing effect most evidently affects the future retainability of successfully recalled items, we analyzed this testing effect mainly in terms of the rate of forgetting. The latter refers to the recall drop of previously retrieved items with time and is arguably independent of the initial recall level (Slamecka & McElree, 1983), which is strongly influenced by the indirect testing effect. However, the possibility cannot be excluded that the rate of forgetting is also affected by the indirect testing effect. Nevertheless, there is no clear evidence yet for this possibility (Slamecka & Katsaiti, 1988). Previous research rather suggests that test-potentiated learning diminishes the interaction effect between study type and retention interval and elevates the overall recall level in keeping with the distribution model (Kornell et al., 2011). Future research in action memory should further investigate the testing effect by attempting to avoid the retrieval success problem. This would permit a fairer comparison between the testing manipulation and other effective encoding manipulations, such as motor enactment. One approach is to provide a re-encoding opportunity in a separate restudy trial, as was adopted in the included studies. Although I ensured high and comparable levels of retrieval success across the study types, it may still be argued that the learners re-experienced the study–test items less (approximately 80%–90%) than the study-only items (100%) in terms of processing time. Therefore, immediate restudy (i.e., feedback) may be provided after every item on the cued-recall test (100%). In this way, all items are processed to an equivalent amount even though some of them cannot be recalled. However, both variants make it difficult to disambiguate the direct and the indirect testing effects. An alternative approach is to ensure the successful retrieval of all items by means of learning to criterion (Karpicke & Roediger, 2007, 2008; Karpicke & Smith, 2012). That is, each tested item is recalled at least once (before dropping it from learning) or all items from each study type have been recalled at least once before implementing the restudy or testing manipulation. This may even lead to a direct testing effect after shorter delays (Karpicke et al., 2014). However, such a procedure—primarily orientated to achieve a behavioral criterion—neglects the spacing of the study and test events, which itself can foster voluntary retrieval attempts, (Greene, 1989). Furthermore, the underlying memory-strength distribution for both study types becomes less predictable.

All the aforementioned approaches to tackle the retrieval-success problem are associated with some methodological issues. To obtain greater clarity, I would suggest that future research explore the testing effects with more than one approach and examine whether or not the reported results pattern is robust. It should be noted, though, that manipulating the study type is never a
pure process. Retrieval can be initiated during restudy (specifically, when spaced apart and mixed with test opportunities), and tests can be ineffective in initiating retrieval practice, for example, when the answers to be given are readily available (open-book tests; Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008).

Measuring Forgetting
A thorough investigation of the temporal course of forgetting is a complex endeavor. Several factors may influence the course of forgetting.

First, to obtain a pure measure of forgetting, memory should be tested for different groups of participants or sets of items at different points in time, as assessing memory increases the future retrievability (of successfully recalled items) at a subsequent point in time. In the present thesis, I measured forgetting by testing the memory for all participants and all items after the short and long retention intervals. This repeated testing of items and persons at both measurement points possibly reduced the effects of time on the later test (Deese, 1958; Hanawalt, 1937). Although testing and time effects may be confounded, the repeated-testing design used in the included studies still allows us to evaluate the relative amount of testing on the rate of forgetting, in other words, to compare a single testing round (i.e., “study-only”: SSSSSST) versus repeated testing rounds (i.e., study-test: STSTSTT). Further research may additionally manipulate the retention interval between subjects (i.e., by testing different groups at different points) and different sets of items to isolate the effects of testing and time on memory and thereby to obtain purer measures of forgetting.

Second, different measures of forgetting can lead to diverse results, as they reflect the different properties of the retention curves (Wixted, 1990). In the included studies, I analyzed the data in terms of both absolute forgetting scores (testing the vertical distance of forgetting curves; Slamecka, 1985; Slamecka & McElree 1983) and proportional forgetting scores (i.e., testing the horizontal distance between forgetting curves, Loftus, 1978; 1985). With both measures, equivalent results were obtained, supporting the reliability of our conclusions.

Third, the heightened initial recall level after enactment (relative to verbal encoding) may have influenced the rate of forgetting, rather than enactment per se, an aspect that needs to be considered, depending on the chosen measure of forgetting (Wixted, 1990). Based on the test of vertical parallelism, variations in the initial recall level do not affect the rate of forgetting (Slamecka & Katsaiti, 1988). In this case, it can be concluded that enactment per se reduced long-term forgetting. Based on the test of horizontal parallelism, however, this conclusion cannot be drawn, as higher initial recall levels already lead to reduced rates of forgetting (Loftus, 1985). As equivalent results were obtained with both measures, I interpreted the interaction effect as an indicator of forgetting. However, future research may also match the initial recall levels
between encoding types, for example, by increasing the number of item presentations for verbal relative to enactive encoding (Nilsson et al., 1989).

Fourth, relatively high recall levels after the short retention interval were achieved. This was true for the included studies, and may have two reasons: first, to diminish the unequal re-exposure of items in restudy and test trials and, second, to ensure strong direct testing effects that rely on successful recall. However, in particular, the recall levels after enactive encoding approximated the ceiling level, and therefore, the decrease in memory performance may underestimate the actual amount of forgetting, even though a significant enactment effect was demonstrated. It should be noted, though, that enactment did not reduce the rate of forgetting for study–test items, which had an equivalent initial recall level after the short retention interval as study-only items. Nonetheless, future research should attempt to confirm the finding that enactment reduces the rate of forgetting also on a lower performance level.

Fifth, observed differences in the rate of forgetting may be restricted to the level of observed data (Slamecka, 1988). To draw conclusions regarding an underlying psychological construct (e.g., strength of memory trace or forgetting), evidence of crossover interactions in the recall data are needed, as observed recall probabilities and memory strength are nonlinearly related (Loftus, 1985b; Wagenmakers, Krypotos, Criss, & Iversson; Wixted, 1990). Translating recall probabilities to the scale of memory strength can remove (i.e., transform away) noncrossover, but not crossover, interactions. Although most interactions in the included studies did not cross over, the testing-effect interactions nearly “touch” (for the short retention interval). Therefore, they can be viewed as “borderline nonremovable interactions” (Wagenmakers et al., 2012), and tentative conclusions regarding the psychological level can be arrived at. Nonetheless, conclusions regarding memory strength or forgetting can be drawn in the numerous cases for which the direct testing effect emerged as a crossover interaction effect when testing was not followed by subsequent restudy (Roediger & Karpicke, 2006b). Thus, as the ultimate aim is to understand the underlying psychological processes, it is important to design experimental studies in such a way that conclusions can be reached independent of the measurement scale. In addition, future research should aim to describe the mathematical form of forgetting curves following retrieval and restudy practice (Wixted, 1990). This more quantified approach provides parameters to quantify the rate of forgetting (and the degree of initial learning) across multiple measurement points. Using such a method, Carpenter et al. (2008), for example, assessed the rate of forgetting for testing (with immediate feedback) versus restudy after 5 min, or 1, 2, 7, 14, or 42 days. Even in the absence of an interaction effect between study type and retention interval, fitting a power function to the data revealed that testing reduced the rate of forgetting. It remains to be seen whether testing memory within subjects across multiple and longer retention intervals (e.g., 2 min, 1 day, 1 week, 2 weeks, and 4 weeks) resists direct testing effects for enacted action phrases.
Action Phrases—Asymmetric Learning Material

As action phrases consist of two different components, a verb (e.g., “wash”) and a noun (e.g., “car”), they are inherently asymmetric, unlike common learning material (e.g., noun–noun pairs). Consequently, across recall types, the cue and the target vary at the same time, and thus, it is not possible to separate the contributions of cue efficiency, target availability, and the strength of the verb–noun association. With regard to the indirect testing effect, though, cues may play a significant role during the pre-restudy test and the restudy trial. Only the cues (but not the targets) are available for unsuccessfully recalled items; during restudy, they may provide reliable information to the participant about the recall status and about the items that need effective re-encoding. However, it may be argued that failed retrieval of the central verb component results in a stronger tagging of the noun cue, which in turn helps trigger enhanced re-encoding of that particular action phrase. The combined effects of cue efficiency and target availability may enhance the indirect testing effect specifically for noun-cued recall of verbs. For this reason, I refer to this manipulation of noun-cued recall of verbs versus verb-cued recall of nouns as two recall types, which imply the simultaneous change in cue and target between them.

Future research may investigate the specific factors that determine how recall type affects testing. For example, in addition to action (i.e., verb–noun) phrases, the testing effect should be examined also with verb–verb and noun–noun phrases. This would allow to examine the influence of the cue (and the search process afforded by the cue) on the testing effects by holding the target constant, although this manipulation may affect verbal and enactive encoding differently (Engelkamp, 1991; Helstrup, 1991). Furthermore, it would merit research to provide normative data regarding the item characteristics of verbs, nouns, and their associations, that is, the individual components of action phrases. In addition to their grammatical classification, verbs and nouns differ in their degree of concreteness (Kersten & Earles, 2004). Nouns—used in the present thesis—are mainly basic level words, referring to single objects (e.g., “car”). Verbs, on the other hand, refer to broader categories of events and may rather resemble superordinate nouns (e.g., “vehicle”; Kersten & Billman, 1997). Furthermore, the verb/noun order within action phrases may be influential. In the current thesis, I applied Swedish action phrases in which the verb always precedes the noun. However, in other languages, the naturally given word order is reversed such that the verb follows the noun (e.g., in German). Finally, action phrases may differ in association strength from verb to noun and vice versa. Although verb–noun association strength was previously measured (item integration: Kormi-Nouri, 1995; Kormi-Nouri & Nilsson, 1998), the associative direction was neglected. Specifically, differences in “forward” and “backward” associative strength may affect the moderating role of recall type on the testing effects. In sum, further research should adju-
dicate the underlying factors pertaining to cue effectiveness, target availability, and directive association strength when investigating testing effects with the cued recall of action-phrase material.

Another venue of future research is to generalize testing effects further to goal-directed action sequences in an attempt to approach more realistic features of everyday activities than with simple action phrases. Goal-directed action sequences comprise several component actions that are organized and specified in terms of a goal and an outcome, for example, loading a dishwasher, cleaning a window, etc., in the context of cleaning (Schult et al., 2014).

Theoretical Challenges

To explain the effects of testing and enactment with one general account of memory presents many challenges, though it may be a promising step to effectively understand the similarities, differences, and dependencies between them. One of these theoretical challenges pertains to the various ways in which testing can be implemented (e.g., free recall, cued recall, or recognition). Implementations of testing affect the efficacy and underlying mechanisms of the testing effect. Another, theoretical challenge relates to the role of the episodic context for the testing and enactment effects—an aspect of episodic memory representations, which has been rather neglected (Karpicke et al., 2014). In this section, I will focus specifically on the direct testing effect and enactment effect.

Multiple Implementations of Testing

The present thesis is mainly concerned with testing by means of cued recall during the learning phase. This implementation of testing and motor enactment enhances cue–target relational and item-specific processing, as both these types of processing exhibit positive effects in final cued-recall tests (reliant on cue–target association) and final recognition tests (reliant on item-specific information; Karpicke & Zaromb, 2010; Mulligan & Peterson, 2014; Steffens et al., 2009). However, testing via cued recall and enactment do not facilitate inter-item relational processing. For example, negative testing occurred with final free recall, which is additionally reliant on inter-item relational information (Mulligan & Peterson, 2014), and the enactment effect, if not presented with real objects (Kormi-Nouri, 1995), is of smaller size in free recall than in other retrieval assessments (i.e., cued recall and recognition tests; Kubik et al., 2014; Steffens et al., 2009).

However, testing can be implemented in various ways, and thereby, its underlying mechanisms can vary. For example, if implemented by means of free recall during the learning phase, testing can enhance encoding on inter-item
relational information (semantic associations between items: Congleton & Rajaram, 2012; Zaromb & Roediger, 2010; temporal associations between items: Lehman, Smith, & Karpicke, 2014). In contrast, motor enactment (i.e., in the absence of real objects) generally does not enhance the encoding of inter-item relational information relative to verbal encoding (semantic associations: Engelkamp & Zimmer, 1996, 2002; temporal associations: Zimmer & Engelkamp, 1989; idiosyncratic associations: Koriat & Pearlman-Avnion, 2003; Koriat, Pearlman-Avnion, & Ben-Zur, 1998 for increased enactive clustering based on movement similarities). In sum, the degree to which enactment and testing produce similar effects and share, in part, the same mechanisms, depends on the specific ways in which testing is implemented. Thus, the conclusions of the present thesis are fairly restricted to testing via cued recall, which probably exhibits most similarities with the effects of enactment. Other, less experimentally controlled testing formats, such as using flash cards, quizzes, or short-answer tests, likely capitalize on multiple mechanisms.

A Synthesis—An Item–Context Account
The effects of testing and enactment can be successfully explained in terms of an item-specific—relational processing account. However, in my view, this account might be more intuitively framed in terms of an item-context model. In that way, both item events and the specific spatiotemporal context (e.g., “when,” “where,” learner’s mental state, inter-item relational information) related to each of them are encoded and stored in an episodic memory representation (Howard & Kahana, 2002; Lehman & Malmberg 2013).

As I hypothesized previously, testing and motor enactment likely share mechanism(s) that strengthen the item representation (or memory trace) of an action event. Both study techniques facilitate the encoding of item-specific information (i.e., verb or noun) or cue–target relational information (i.e., the verb–noun association), which makes the entire action phrase more discriminable from other action phrases. This enhanced strength and distinctiveness of the underlying memory traces may result from the fact that semantic, motoric, and/or multisensoric attributes are added to the item representations (enactment effect: Engelkamp, 2001; testing effect: Carpenter, 2009, 2011) and thereby enrich the memory representation.

The spatiotemporal context is the other ingredient in the episodic memory representation. The context changes (slightly) with time and events (Lehman & Malmberg, 2013) and can be later used to guide the search process of the target information, specifically when no other item cues are available during retrieval. With exception to inter-item relational information, little attention was devoted to the aspects of the episodic context (e.g., location and time of the item event) and how they contribute to the mnemonic effects of testing (Karpicke et al., 2014) and enactment (Koriat et al., 1991; Mecklenbräuker, Steffens, Jelenec, & Goergens, 2011). This is surprising because episodic
memory in daily life is heavily dependent on context information. For example, when we have performed a routinized activity, such as locking the apartment door in the morning, we may remember an occurrence of this action type, but we often are uncertain of whether we performed the action today. Thus, to remember whether an event occurred at a specific time or space, we need to encode and retrieve distinctive contextual features related to this episode.

I reason that both study techniques differ in the degree to which they can capitalize on the episodic context. *Enactment* typically focuses attention on performing the action, probably at the expense of encoding context information and creating item-context associations (Koriat et al., 1991). Prior research has shown that enactment does not enhance or even hamper the encoding of source information (how the item was encoded; Hornstein & Mulligan, 2004), spatial context (Helstrup, 1987, 1989), or item-order information (Ollofsson, 1996; Engelkamp & Dehn, 2000). However, the context is interactively encoded when it becomes part of the performed actions, for example, when real objects (denoted by the noun) are present (Mecklenbräuker et al., 2011; Steffens et al., 2007). In contrast, during *testing*, learners use available cues in the present context even when they are independent of the item event and create item-context associations. For example, it has been shown that *testing via free recall* enhances source memory during final recognition tests (discriminating the list from which the particular items stemmed; Brewer et al., 2010) and enhances retrieval when only the last out of several previously learned item lists should be recalled (Szpunar et al., 2008). Recently, Karpicke et al. (2014) suggested that when learners attempt to retrieve the target, they use the available cues in the present context and, if needed, reinstate the prior study context (i.e., context reinstatement and context encoding). Testing integrates features from study and recall contexts to an updated context representation. The latter can serve as an effective cue in future retrieval situations and can reduce the search set from retrieval candidates that are not associated with specific context features (“episodic context account”; Karpicke et al., 2014; Raajmakers & Shiffrin, 1981).

I reason that testing can both strengthen items’ memory trace (through cue–target and item-specific processing) and create item-context associations (through context encoding and reinstatement). Critically, though, the degree to which context encoding and reinstatement occurs depends on the nature of the retrieval situation (Karpicke et al., 2014). The more item cues are available during retrieval or the stronger they are, the lesser the learner needs to capitalize on current and previous contextual representation to probe the searched item information. Thus, testing via cued recall or recognition, similar to enactment, may strengthen items’ memory trace and capitalize less on contextual information, leading to the similar and nonadditive effects of both study techniques in the present thesis. However, testing via free recall should depend to a larger degree on context encoding and reinstatement, in addition to
strengthening the item representation. Therefore, the processing overlap between testing via free recall and motor enactment should be reduced. These assumptions are consistent with the finding that testing and enactment effects are both moderated by the type of memory test, but in opposite ways. The testing effect is more pronounced when less item cues are provided in the retrieval situation and more context information needs to be encoded and reinstated (free recall > cued recall > recognition; Karpicke et al., 2014; Kornell et al., 2011). In contrast, the enactment effect is largest when more item information is available and more retrieval relies on the strength or distinctiveness of the item representation (recognition tests > cued recall > free recall; Kubik, Obermeyer et al., 2014; Steffens et al., 2009). Future research should investigate how testing and enactment permit the encoding of contextual information and restrict the search for the target items to a specific temporal context. With the multiple-list learning paradigm, this possibility can be tested in a straightforward manner (Lehman et al., 2014; Szpunar et al., 2008). Participants study several lists of action phrases; however, later, they should exclusively retrieve action phrases from the last list, which can be viewed as a particular temporal context. Crucially, after each studied list, the participants should process all action phrases once more from the preceding list in one of the four conditions: (i) testing via free recall, (ii) testing via cued recall, (iii) enactive encoding, and (iv) receiving only a distractor task (control condition). Performance of correct recall and intrusions from previous lists should indicate how far these conditions can use context information to restrict the search set of target items to a specific list context. I assume that testing via free recall, relative to the control conditions, should increase correct recall and decrease prior list intrusions (Lehman et al., 2014). In contrast, enactment and testing via cued recall should decrease these performance measures of time-constrained retrieval if they simply elaborate the item representation but ignore or even hamper the encoding of context information. If testing via cued recall should additionally encode and reinstate the critical list-context information—a viable possibility also in the present studies—testing should also enhance the ability of time-constrained retrieval, though less so than testing via free recall.

In sum, future research should explore the role of episodic context information for the mnemonic effects of testing and enactment. Performing an action may enhance remembering its occurrence later on, but not its time and place. Testing may also enhance memory for the occurrence of events, though it may additionally encode the current and reinstate the previous episodic contexts if the retrieval situation affords it. Such an item-context model for encoding (via enactment) and testing effects remains tentative at this point. Future research should test specific predictions to elucidate shared and distinct processes of testing and enactment.
Practical Challenges

Beyond the academic endeavor to understand learning principles in experimentally controlled paradigms, translating them into effective educational practice is an important goal of memory research. Although the present thesis did not study the practical applications of the findings, they may have implications for learning in educational contexts and other real-world settings, as both enactment and testing are effective and readily available tools to enhance learning and memory.

Action memory research has reliably shown that performed actions are better remembered than actions we read about. However, can this memory benefit be generalized to other materials (beyond verb–noun phrases) and to more educational contexts? Research on memory for gestures may provide an answer. As gestures can be seen as symbolic actions accompanying speech (McNeill, 1992), they extend the mnemonic effects of actions to less literal forms of enactment. Similar to our findings regarding the enactment effect, learning with gestures enhances memory and delays forgetting over longer delays (Macedonia, 2003). Importantly, these mnemonic benefits generalized to various learning materials, such as word categories differing in abstractness (verbs, nouns, and adverbs), foreign language vocabulary, and new words from an artificial language corpus (for an overview, see Macedonia, 2003; Macedonia & von Kriegstein, 2012). Moreover, observing a speaker gesturing led to better memory performance than observing a speaker not gesturing or a gesturer not speaking (Thompson, 1995; Kelly, Barr, Church, & Lynch, 1999). These findings are in line with the embodied-cognition perspective that the body can be used as a learning tool in educational contexts to enhance learning and memory by producing actions that literally (i.e., enactment) or nonliterally (i.e., gesturing) reflect the content of the information.

Research on the testing effect has obvious implications for education practice. Often, tests and examinations are still regarded as assessment tools of learning but not as active tools to enhance learning and memory. Thus, they are kept to the necessary minimum and occur only cumulatively at the end of the term. Recent research has provided evidence that the positive testing effect from laboratory research translates to many situations in which real-world features of educational contexts are instantiated. For example, with respect to mnemonic benefits of testing were demonstrated for educationally relevant study materials (foreign language word pairs: Pyc & Rawson, 2007; idea units from prose passages: Roediger & Karpicke, 2006b) and educationally relevant retrieval formats (quizzes: McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; short-answer test: McDaniel et al., 2007) as well as both in the laboratory context (Butler & Roediger, 2007) and in less controlled classroom environments (McDaniel et al., 2007). Further research may evaluate how testing influences the acquisition of abstract, flexible abilities (such as reasoning or flexible and creative thinking), in addition to the learning of facts and text material, and determine the conditions that moderate the testing effects.
on learning and retention specifically in educational contexts. For example, should retrieval practice be spaced out over time, and should curricular contents be blocked within knowledge domains or rather interleaved across them? To answer these and related questions, retrieval practice and other learning principles may be “packed” into specific educational instructions and programs (Pashler et al., 2007). In this way, potential restrictions in implementing (and combining) effective learning principles into educational practice can be addressed (e.g., due to the curricular schedule).
Concluding Remarks

In the present thesis, I investigated the individual and combined effects of interleaved testing (via repeated study–test practice) and encoding (via motor enactment) during learning on later cued-recall performance of action phrases. An additional objective was to examine the moderating role of recall type (noun-cued recall of verbs vs. verb-cued recall of nouns) on the direct and indirect testing effects. The main findings are recapitulated below:

1. Testing directly enhanced long-term retention of action phrases by reducing the rate of forgetting over 1 week, relative to study-only.

2. Testing indirectly enhanced subsequent memory performance by facilitating subsequent re-encoding of previously nonrecalled action phrases.

3. Testing by means of noun-cued recall of verbs engendered a larger indirect testing effect than verb-cued recall of nouns. In contrast, the type of recall did not moderate the direct testing effect.

4. Motor enactment of action phrases enhanced subsequent memory performance and reduced the rate of forgetting, relative to verbal encoding of action phrases (i.e., the enactment effect).

5. When both study techniques were combined, the direct effect of testing (in terms of reduced forgetting) emerged only when the action phrases were verbally encoded, not when they were enacted. However, the indirect testing effect for action phrases occurred for both verbal and enactive encoding.

In conclusion, the present findings imply that testing memory benefits the later retention of action-phrase material in direct and indirect ways and thus clearly contributes to learning. In addition, the present findings extend the understanding of how testing might enhance learning in everyday situations, as remembering often involves action-relevant information. Testing should be more systematically implemented in educational practice as an effective learning tool to enhance the sustainability and transfer of learning. However, it is important to further examine the factors that might moderate the effectiveness
of testing in educational contexts. In this thesis, it was observed that the direct testing effect, in terms of the rate of forgetting, was restricted to verbal encoding while the indirect testing effect occurred for both encoding types, though it was increased for noun-cued recall. These findings highlight the need to implement testing with different retrieval formats, such as free recall and recognition, and in different schedules (with vs. without restudy opportunities). Future research may also separate more clearly the specific contributions of direct and indirect testing effects to better understand their common and distinct mechanisms.

In addition, the results of the thesis imply that not only testing but also encoding via enactment is an effective tool to enhance learning and memory. In fact, when learning was enhanced though enacting, the potential benefits of testing were limited to the indirect effects. Future research should generalize this study technique more systematically to other materials apart from action phrases and in general also to materials for which the contents to be learned may not necessarily map the performed actions. For example, enactment by means of gestures has been shown to enhance learning of foreign language vocabulary and of dialog materials in the context of professional acting (Noice & Noice, 2001; Noice, Noice, & Kennedy, 2000). In this way, it is possible to assess the value of enactment—both in literal and nonliteral forms—in learning materials and environments more relevant from an educational standpoint.

From being a neglected and almost forgotten phenomenon, testing has advanced to one of the most frequently investigated research topics on memory. Similarly, the enactment effect—being rather known in the circles of memory researchers—has lately received a resurgent uptake through the contemporary research trend on embodied cognition. Despite challenges in translating effective learning principles from controlled lab research into educational practice, I believe that both testing and enactment can greatly benefit educational instruction and programs in future to enhance learning and memory in various educational contexts.
References


