UNIVERSITÄT MANNHEIM

The Generation Effect and Source Memory

Inauguraldissertation zur Erlangung des akademischen Grades
einer Doktorin der Sozialwissenschaften
der Universität Mannheim

vorgelegt von

Dipl.-Psych. Sandra Daniela Mattern

Dekan der Fakultät für Sozialwissenschaften:

Prof. Dr. Berthold Rittberger

Gutachter: Prof. Dr. Edgar Erdfelder

Prof. Dr. Rüdiger Pohl

Tag der Disputation: 20. Dezember 2010

“A curious peculiarity of our memory is that things are impressed better
by *active* than by *passive* repetition.”
(James, 1890, p. 646, italics added)

“Most of us have probably encountered the informally expressed sentiment
that there is an especial advantage to learning by doing,
or that some kind of active or effortful involvement of the person in the learning process
is more beneficial than merely passive reception of the same information.
To what extent does this general notion have solid empirical support,
as opposed to a casual or anecdotal base, [...]?”
(Slamecka & Graf, 1978, p. 592)

Summary

Superior memory performance for self-generated (i.e., incomplete) items compared to read (i.e., complete) items is called the positive generation effect, whereas the reverse pattern is called the negative generation effect. For item memory tasks, a positive generation effect typically occurs (cf., Slamecka & Graf, 1978). In contrast to this, no clear picture exists as to whether a positive or a negative generation effect is bound to emerge for source memory of perceptual attributes, due to empirical evidence for both outcomes. Therefore, the two lines of research investigated in the course of the present dissertation deal with the generation effect and source memory. They aim to shed light on present inconsistencies and contradictions as well as to illuminate unanswered questions. All data were analysed using the multinomial processing tree model for crossed source dimensions (Meiser & Bröder, 2002).

In the first line of research, I addressed the role of the processing of perceptual attributes and of the processing of internal states in memory for the degree of completeness. Within a generation effect paradigm, memory for source attributes of an item can be studied in different ways. When considering the processing account by Mulligan (2004) and the dual-hypothesis by Riefer, Chien, and Reimer (2007), contradictory predictions exist for the source memory dimension degree of completeness – namely a negative generation effect in the former and a positive generation effect in the latter case. To overcome these contradictions, I hypothesised two processing modes: the processing of and memory for (a) perceptual attributes (PA) and (b) internal states (IS). PA and IS processing were implemented via weak instructions and strong instructions. Contrary to expectations, no effect of instruction type could be found. In contrast, across types of instructions a null effect emerged, when instructions were weak. When instructions were strong, however, a positive generation effect occurred for both types of instruction. This latter finding is in line with Riefer et al. (2007) and indicates the importance of self-reference (Rogers, Kuiper, & Kirker, 1977) in source memory and in generation effect studies.

In the second line of research, I investigated the role of increased conceptual processing for memory of presentation colour, for which a negative generation effect has been found

consistently (e.g., Mulligan, 2004; Mulligan, Lozito, and Rosner, 2006; Riefer et al., 2007). Some theories state that conceptual processing takes place to a higher degree for self-generated items than for complete items (Donaldson & Bass, 1980; Graf, 1980; Mulligan, 2004; Slamecka & Graf, 1978), while other theories additionally claim that perceptual processing takes place to a lesser degree for incomplete as compared to complete items (Mulligan, 2004). These points were called into question by Mulligan et al. (2006). They challenged the appropriateness of a processing account which emphasises perceptual and conceptual processing, and considered it too specific. Instead, Mulligan and colleagues suggested a more general processing account, which is concerned with visual and non-visual processing instead. Since self-generation of items is strongly indicated to lead to higher conceptual processing, I argued that whether or not an incomplete item has a schema-typical colour influences colour memory performance; moreover, I theorised that it does so to a higher degree for incomplete items than is true for simply read items. To test these predictions, experiments were conducted, in which the supposed strength of the effect of conceptual processing was varied. When less natural stimuli were employed, no effect of schema-typicality of presentation colour or of completeness of the items could be found on source memory for colour. In contrast, when more natural stimuli were employed, a negative generation effect occurred. However, no additional effect of incompleteness and thus no additional effect of the theorised increased conceptual processing of incomplete items emerged. The present findings can thus be interpreted as evidence in favour of Mulligan et al. (2006), who suggest a more general visual and non-visual processing account.

I conclude, (a) that self-reference plays a critical role for generation effect studies when attempting to investigate source memory, and (b) that conceptual and perceptual processing seem to play a less critical role than assumed – results could rather be interpreted as evidence in favour of a more general account involving visual and non-visual processing.

In addition, crucial design issues for any research agenda concerned with the generation effect were revised and evaluated for the current studies. Finally, I discussed remaining research questions and suggested further research opportunities as well as future directions.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

The present dissertation comprises four chapters and is concerned with specific research questions regarding source memory and effects of self-generation.

Chapter 1 provides an outline of the status quo in research on the *generation effect*. Background information is illustrated by empirical evidence, theories, and hypotheses concerning the effects of self-generation on *item memory* as well as on *source memory* tasks.

Chapter 2 explores the roles of processing of perceptual attributes and processing of internal states in source memory (i.e., for memory for the degree of completeness). First, an introduction is presented in which the most significant papers are described, leading to the research problem, which consists of a partial contradiction between two prominent theories – the dual processing account (Mulligan, 2004) and the dual hypothesis (Riefer, Chien, & Reimer, 2007). Then, the solution of introducing different experimental changes to differentially enhance perceptual and conceptual processing is presented. Next, Experiments 1 (implementing a weak manipulation) and 2 (implementing a strong manipulation) are described and results are discussed briefly, before giving a more comprehensive final discussion of the research presented in Chapter 2.

Chapter 3 deals with the role of increased conceptual processing for source memory (i.e., for memory for presentation colour). In this chapter, relevant research papers are introduced concerning the generation effect and the conditions under which increased conceptual processing is assumed to occur. In addition, the reader is informed of literature pertaining to colour processing, concept colour effects, and concept activation, before the research hypotheses are stated. Descriptions and short discussions of the methods and results of Experiments 3 (implementing drawings in an intentional encoding design), 4 (implementing photos in an intentional encoding design), and 5 (implementing photos in an incidental encoding design) follow. A broader final discussion of Chapter 3 is then provided.

Finally, Chapter 4 comprises a summary and a description of the comprehensive meaning of and conclusions from research presented in Chapters 2 and 3. Then, critical research issues and caveats for any research agenda concerned with the generation effect are revised and critically evaluated for the current studies. Finally, remaining research questions are discussed and further opportunities for research are suggested.

1.1 Item Memory and Effects of Self-Generation

The *generation effect* (or the *positive generation effect*) is a memory phenomenon that has been researched in cognitive psychology for over three decades. It asserts a memory advantage for self-generated stimuli in comparison to material that was simply perceived. In other words, it has been found that one can more easily remember information that one has helped to produce rather than information that one has just read.

Over 30 years ago, Slamecka and Graf (1978) introduced the term generation effect. In contrast to most research realised in experimental psychology, the impetus for their experiments were neither extant theoretical issues nor previously published findings. They rather intended to find solid empirical support for the general notion “that there is an especial advantage to learning by doing, or that some kind of active or effortful involvement of the person in the learning process is more beneficial than mere passive reception of the same information” (Slamecka & Graf, 1978, p. 592). Consequently, Slamecka and Graf designed five experiments to delineate the outlines of this phenomenon. Their approach involved presenting a cue word that was semantically related to the target along with the first few letters of the response to be generated (i.e., the incomplete *target*), and a rule describing the relationship between cue and target (i.e., the *generation rule*). Incomplete target words were mixed with complete ones, which could simply be read.

For a traditional research design, consider the example shown in Figure 1.1. Usually, participants were asked to remember a list of target words, which were shown sequentially. Some of the items were presented complete (“Shoe”, in the read condition), whereas other items were presented in an incomplete form (“Sh__”, in the generate condition). To help participants to identify an item correctly, a semantically associated cue word (“Foot”) was shown alongside the target. In a recognition memory test, participants were presented with a list of words; some of which matched the previously studied target words, whereas others were entirely new words within the context of the study. Participants then indicated which of the words they remembered as being old, i.e., previously studied. Typically, incomplete items, which had been studied in the generate condition, had a significantly higher probability of being remembered correctly, as compared to complete items. This pattern was called the (positive) generation effect in item memory.

Note that I henceforth speak of a positive generation effect when a memory advantage exists for self-generated (i.e., incomplete) items as compared to simply perceived (i.e., complete) items.

Generation rule: cue and target are semantic associates

Study pair Cue word	Target word	*Encoding condition*	*Type of target word*
Foot	Shoe	*Read condition*	*Complete item*
Foot	Sh__	*Generate condition*	*Incomplete item*

Figure 1.1: Illustration of a traditional design used in the investigation of the generation effect

After the publication of Slamecka and Graf (1978), years of scientific controversy started over the generation effect, resulting in an abundance of experiments. Psychologists have employed a wide range of tasks to probe the generation of a target, such as generation from antonyms (e.g., Mulligan, 2004), rhymes (e.g., Burns, 1990), semantic associates (e.g., Begg, Snider, Foley, & Goddard, 1989), anagrams (e.g., Gardiner, Dawson, & Sutton, 1989), phrases (e.g., Engelkamp & Dehn, 2000), definitions (Horton, 1987), multiplications (Gardiner & Rowley, 1984; McNamara & Healy, 2000; Pesta, Sanders, & Nemec, 1996), numbers (Gardiner & Hampton, 1985), and nonwords (e.g., Johns & Swanson, 1988; Nairne & Widner, 1987), as well as different kinds of test paradigms to measure memory, such as free recall, cued recall, and recognition tasks. Refer to Greene (1992), Mulligan (2001), and Bertsch, Pesta, Wiscott, and McDaniel (2007) for further reviews. In other experiments, positive generation effects emerged, even when participants had not successfully produced the to be self-generated target during study. This finding was labelled *the try-to-generate effect* (Kane & Anderson, 1978; Slamecka & Fevreiski, 1983). Practical applications of the generation effect can be found within the pedagogical field (e.g., deWinstanley & Bjork, 2004; Foss, Mora, & Tkacz, 1994; Metcalfe & Kornell, 2007; Metcalfe, Kornell, & Son, 2007) as well as within the field of neuropsychological diagnostics and rehabilitation (e.g., Barrett, Crucian, Schwartz, & Heilman, 2000; Lubinsky, Rich, & Anderson, 2009; O'Brien, Chiaravalloti, Arango-Lasprilla, Lengenfelder, & DeLuca, 2007; Troyer, Haflinger, Cadieux, & Craik, 2006).

Although the positive generation effect persisted in most studies, it reversed in others (e.g., Grosofsky, Payne, & Campbell, 1994; Hirshman & Bjork, 1988; Schmidt & Cherry, 1989; Slamecka & Katsaiti, 1987). This outcome was labelled the negative generation effect.
Note that I henceforth speak of a negative generation effect, when a memory advantage exists for simply perceived (i.e., complete) items as compared to self-generated (i.e., incomplete) items.

Taking previous research into consideration, I conclude that self-generation does not result in better item memory per se. Instead, certain factors concerning design and underlying processes taking place during self-generation play an important role and influence memory outcome (cf., Steffens & Erdfelder, 1998). One example is presentation of stimuli in mixed list designs (complete and incomplete items are studied together) or in pure list designs (complete and incomplete items are studied in separate lists). It could be shown that a positive generation effect tends to occur in mixed lists, whereas no effect tends to emerge in pure list designs. Moreover, Steffens and Erdfelder (1998) found a negative generation effect for free recall tasks as compared to the more commonly employed recognition tasks. Therefore, when designing a new experiment investigating the generation effect, one has to consider and heed these influencing factors.

Theories and Accounts

Slamecka and Graf pointed out that the generation effect should ideally "be explained in the sense that it is seen to be simply another manifestation of some more general overarching law of behavior. The next question is whether any well-founded principle is already available to do the job." (Slamecka & Graf, 1978, p. 601). Chechile and Soraci (1999) stated that possible explanations of generative processing have involved general notions such as elaboration, distinctiveness (Begg et al., 1989; Jacoby & Craik, 1979), and cognitive operations. Finding such a more general overarching law of behaviour has been attempted repeatedly and has lead to a variety of theories and accounts. In this section, only the most prominent ones are described briefly.

In 1972, Craik and Lockhart published a theoretical construct, called *levels of processing* (*LOP*) framework, which raised much interest among cognitive psychologists. The LOP framework focuses on how stimuli and events are encoded. The main rationale is that memory strength of an item increases as depth and elaboration of item processing increase. The deeper and the more elaborate an item has been encoded, the better it is retained in memory. Craik and Lockhart (1972) had participants study three lists of words. For items in the first list, participants were to judge whether words were presented in upper or in lower case (i.e., physical judgement). For items in the second list, participants were to state whether words rhymed with other words in a pair (i.e., phonetic judgement). For items in the final list, participants were to judge whether words fit into sentences presented with other words (i.e., semantic judgement). As a result, best memory performances were obtained for items judged semantically, followed by memory for those judged phonetically. Worst performances were recorded for words judged on a physical dimension. This article was one of the most influential cognitive psychology papers since Miller's (1956) paper on the magical number seven. However, the LOP framework was criticised for being untestable und thus unfalsifiable (Baddeley, 1978), and for being circular (Lockhart, 2002). However, a systematic critical assessment of the LOP framework is beyond the scope of the present work. Most important for the current description is the fact that the LOP idea was applied to generation effect paradigms (Nairne, Pusen, & Widner, 1985; Slamecka & Graf, 1978). It was argued that self-generating a target by following a certain generation rule entails deeper processing and thus better memory compared to processing occurring for simply reading items. Hence, a positive generation effect was theorised to emerge.

The *lexical activation hypothesis* appeals to the special involvement of lexical memory processes while generating a target (e.g., Donaldson & Bass, 1980; Graf, 1980; Slamecka & Graf, 1978). The term lexical or semantic memory represents a person's knowledge system, including knowledge about verbal information, such as knowledge about particular words and their according definitions. This knowledge is acquired through interacting with the environment (cf., prior knowledge). Thus, some researchers started from the question of whether or not the generation advantage resulted from searching and then accessing an entry in the mental lexicon. Nairne et al. (1985) presumed that in the read condition, participants should be able to access their mental lexicons easily, whereas in the generate condition, words would have to be searched for *actively*. For clarification purposes, an image was used of mentally walking through a maze when searching through the lexicon and of coming to dead ends, before finally arriving at the correct target – somehow by some type of trial and error learning. Therefore it was assumed that this search demanded more processing, which would thus translate into a stronger or more accessible memory trace for self-generated items (Tyler, Hertel, McCallum, & Ellis, 1979).

The *Effort hypothesis*, on the other hand, is an account stressing the effortful nature of the self-generation process, without a need for lexical memory involvement (e.g., McFarland, Frey, Rhodes, 1980). According to this idea, the memory advantage for self-generated material is due only to the increased effort of correctly identifying the target in the generate condition. However, a specification of which types of cognitive operations are assumed to be necessary for the generation effect to emerge is mostly missing. Crutcher and Healy (1989) believed lexical activation hypothesis and effort hypothesis to be two sides of the same coin. Mental effort and semantic activation were regarded as components of the same theory that cannot be disentangled from one another. Thus, semantic activation and effort were seen as complementary elements of the mental processes carried out by participants while self-generating targets.

Finally, Einstein and Hunt (1980) highlighted the distinction present in the study phase between relational information on the one hand and item-specific information on the other hand. In their view, three types of processing could be enhanced in the generate condition: first, processing of *item-specific features*, i.e., better processing of features that make an item unique in comparison with other items of the list; second, processing of *cue-target relations* referring to the relation between the cues presented to facilitate generation of incomplete items and reading of complete items; and third, processing of *inter-target relations* also called whole-list processing, which is especially relevant for achieving high scores in a free recall task. The *two factor theory* (Hirshman & Bjork, 1988) was based on the first two factors described above and stated that both were enhanced in the generate condition. In contrast, processing of inter-target relations may be hindered by self-generating targets. McDaniel, Wadill, and Einstein (1988) supported the *three factor theory* and had a more general idea of the processes going on while generating an item. They promoted the idea of regarding item generation as a problem solving task and suggested that *all* information available could be used to accomplish this goal. According to their view, all three factors described can be enhanced by generating an item. This final idea is also referred to as the *multifactor account* (Hunt & McDaniel, 1993; Steffens & Erdfelder, 1998).

Evidence, however, exists both for and against any of the above described attempts of explaining generation effect findings. Thus, in the introduction part of their meta-analytic review, Bertsch et al. (2007) state:

> Over the last 20-plus years, a substantial body of research has evolved around this seemingly simple cognitive task. Nevertheless, controversy still exists over many of the particulars of the generation effect, including its true magnitude [...], the underlying cognitive processes that are responsible for it [...], the exact experimental conditions that are required to produce it [...], the influences that moderate its size [...], and even the conjecture as to whether it is real or merely an experimental design artefact [...]. (p. 201)

Eventually, they conclude their paper by writing:

> Additional studies (and the cumulation of those studies) can only help to clarify existing theories or to assist in the creation of new ones. Regardless of what the underlying cognitive mechanisms may be, the generation effect appears to be a real phenomenon that deserves further empirical study. (p. 207)

Although important conditions and variables influencing the experimental memory outcome in generation effect studies could be identified so far, it seems that (a) no clear picture exists by now and that (b) a reasonable need for further research is obvious – research in which previous results and conclusions have to be heeded and appraised permanently.

1.2 Source Memory and Effects of Self-Generation

So far in the theoretical outline, only item memory discrimination has been discussed. The corresponding question at a test of memory refers to whether or not an item has been presented earlier. In the present section, source memory - another concept that can be and has been employed to a variety of memory phenomena, the generation effect being one of them - is reviewed. The corresponding question at a test of memory refers to the encoding context of a study item.

Simply put, when one learns something (e.g., a Spanish vocabulary written on a sheet of paper), many things are encoded along with the information one actually intends to memorise. This includes, for example, information about whether the word was written by hand or printed out in a certain font, the size of the word, the colour of the sheet of paper, what the room looked like (e.g., desk, walls, shelves, plants), who else was in the room, what it smelled like, whether it was warm or cold, what mood one was in, and many more details of the study context. In the present example, item memory reflects memory for the intended object (i.e., the Spanish vocabulary), whereas source memory reflects memory for any imaginable detail of the study context.

One can summarise by saying that source memory "means memory for contextual information that was acquired during the encoding episode of a given item or fact and that indicates the origin of the item or fact knowledge" (Meiser & Bröder, 2002, p. 116). In contrast to item memory, which depends more upon *semantic detail* and on cognitive operations that influence item strength itself, source memory depends more upon external and internal *contextual details* of a study episode.

In 1993, Johnson, Hashtroudi, and Lindsay published an influential paper titled simply "Source Memory". In this paper, Johnson and colleagues stated that, generally speaking, decisions that are associated with source memory "capitalize on average differences in characteristics of memories from various sources" (Johnson et al., 1993, p. 4) in combination with judgment processes. Among the most significant characteristics are "records of perceptual information (e.g., sound and colour), contextual information (spatial and temporal), semantic detail, affective information (e.g., emotional reactions), and cognitive operations (e.g., records of organizing, elaboration, retrieving, and identifying) that were established when the memory was formed" (Johnson et al., 1993, p. 4). According to their view, source memory can additionally be regarded as a type of source-discrimination problem focussing either (a) on the origin of information (e.g., Was the statement made by my brother or my sister? Did I tell my friend about this or did my husband tell her?) or (b) on the state of information being actual, i.e., public, or imaginative, i.e., private (e.g., Did I actually voice this idea or did I only think about it? Did I imagine putting the key on the table or did I really put it there?).

To summarise, Johnson et al. (1993) argued that source memory can either involve discriminating between two or more external sources (e.g., Person A vs. Person B) or distinguishing between two or more internal sources (e.g., said overtly vs. thought covertly). Moreover, the decision can also be a combination of external and internal sources, which is referred to as reality monitoring (e.g., Did Person A state this or did I?).

Meiser and Bröder (2002) emphasised that source information can essentially be regarded as being multidimensional. They highlighted that multiple features (e.g., temporal, spatial, auditive, or visual features) of the encoding episode can be stored in memory and can thus contribute to source attributions made at a later point.

Source Memory and Generation Effect Studies

This section presents a general overview of research on source memory within generation effect designs. Therefore, it cannot be comprehensive and fully explicated, but rather provides significant literature and explanations on source memory and effects of self-generation. More details on this topic are given in Chapters 2 and 3.

Many of the previously described types of source memory have been implemented in studies of the generation effect. Although to date source memory for generated versus perceived information has been examined frequently, results are mixed and thus coherent conclusions cannot be drawn easily. In some studies, better source memory in the read condition was reported (e.g., Jurica & Shimamura, 1999; Mulligan, 2004; Riefer, Chien, & Reimer, 2007), whereas in others, better source memory in the generate condition was reported (e.g., Gegham & Multhaup, 2004; Kinjo & Snodgrass, 2000; Marsh, 2006; Marsh, Edelman, & Bower, 2001; Riefer et al. 2007). Other researchers have found equivalent source memory performance in both conditions (Mulligan, 2004; Slamecka & Graf, 1978; Voss, Vesonder, Post, & Ney, 1987) and still others have suggested that the effect of generation on source memory depends on the population being studied, namely younger versus older adults (e.g., Rabinowitz, 1989; Taconnat, Baudouin, Fay, Clarys, Vanneste, Tournelle, & Isingrini, 2006; Taconnat, Froger, Sacher, & Isingrini, 2008; Taconnat & Isingrini, 2004), or on the type of generation task employed (Johnson, Raye, Foley, & Foley, 1981).

In some generation effect studies, source was implemented in a special and rather uncommon way. *Context* was viewed either as *spatio-temporal background* (a) or as *audience* (b). To illustrate this point, I present several experiments that studied the generation effect investigating context in either of these ways.

Regarding (a): Studies in which external study settings were manipulated and tested for, e.g., study rooms, can be considered as examples in which context was defined as spatio-temporal background. Koriat, Ben-Zur, and Druch (1991) had participants study high-frequency words first in an enclosed lab room and then in an office with a window. Hence, the variable “room” was manipulated within-subject. After the study phase, participants were tested in a third room on old-new recognition memory and on source memory (lab room vs. office). The authors found that memory for room was significantly better for generated than for read words. Thus, a positive generation effect emerged. Other means of varying the spatio-temporal background were applied by Johnson, Raye, Foley, and Kim (1982), who used different *positions of words on the computer screen* (left vs. right). However, no significant results were obtained, although a negative generation effect could be seen when inspecting the descriptive data. In a second experiment, Johnson and colleagues used the relative temporal order, namely *earlier or later in the presentation of study material*, and replicated their previous findings.

Regarding (b): Studies in which other people in addition to the learner are present and are involved in the encoding process, can be considered as examples of context being defined as audience. In several studies, participants had to decide which source (i.e., self or other) a piece of information originated from. Note that this is a commonplace decision, often required in everyday life, such as when asking oneself whether one had read about a fire or an accident in the newspaper or was told about it by a well-informed friend or whether one's colleagues had just come up with a clever argument or whether it was one-self. Studies emphasising context as audience were conducted by Jurica and Shimamura (1999), Koriat et al. (1991), and Brown, Jones, and Davis (1995). The study by Brown and colleagues was modelled to follow the example of *group conversations*. A group conversation Here, a group situation was created in which categories (e.g., favourite recent movie or hardest school subject) were offered as probes and category members were offered as responses. The conditions that are of most relevance for the present purpose follow: Participants alternated in simply watching a question-response-interaction (i.e., bystander role) and generating examples for a category themselves (i.e., responder role). Participants were asked later to name the questioner when they had been a mere bystander (i.e., read condition) or when they had been the responder (i.e., generate condition). Results showed no significant effect of encoding condition.

Note that the distinction between (a) context as spatio-temporal background and (b) context as audience does not necessarily have anything to do with the distinction between (c) memory for external source and (d) reality monitoring, as described previously.

Theories and Accounts

In the current section, theories and hypotheses on the role of memory for source in generation effect paradigms are presented briefly along with more recent empirical evidence for them. The most influential ideas are demonstrated: *item-source enhancement* hypothesis, *item-source trade-off* hypothesis, *processing accounts*, and *dual-hypothesis*.

Item-Source Enhancement Hypothesis and Item-Source Trade-off Hypothesis

One prominent attempt of explaining item memory and source memory patterns in generation effect paradigms has been termed *item-source enhancement hypothesis*. It states that an enhancement in item memory concomitantly leads to an enhancement in source memory. Hence, a variable influencing memory encoding should generally affect item memory and source memory in the same direction and to the same extent. This hypothesis is primarily based on assuming the existence of one single type of memory processing. Evidence favouring the item-source enhancement hypothesis is, for example, provided by

Marsh et al. (2001), Marsh (2006), and Gegham and Multhaup (2004), who re-examined a study conducted by Jurica and Shimamura (1999). Marsh et al. (2001) present three studies in which they asked participants to remember in which of two contexts read or self-generated items had been encoded. Contexts were either two separate rooms (Experiment 1), or two separate computer screens (Experiment 3), or were operationalised as separate perceptual attributes of target words (Experiment 4). In these experiments, memory for contexts and items was superior for self-generated words. Since Mulligan (2004) found no effect of self-generation for memory of location of the target on a computer screen, Marsh (2006) conducted a new study, in which stimuli were presented either on the left or on the right side of a computer screen. Participants in Condition A were asked to generate and read items silently, whereas in Condition B, participants were asked to write down their responses on a sheet of paper. Marsh found a memory advantage for location for self-generated items in both conditions and a positive generation effect in item memory. This difference however was much more pronounced for silent participants. In comparison, the effect was reduced when participants wrote down the study words. Marsh concluded that self-generation can effectively enhance memory for location in addition to item memory, but only if experimental parameters (i.e., overt responding) do not interfere with the processing benefit of generation.

In contrast to the item-source enhancement hypothesis, the *item-source trade-off hypothesis* assumes other types of memory processing. This account is based on the supposition that item memory and source memory are different in the modes of processing required for good performance. In addition, a possible limit on the cognitive resources available is emphasised. A trade-off between “the amount of processing allocated to item memory and the amount of processing allocated to source memory” (Jurica & Shimamura, 1999, p. 649) is argued to result. It was therefore stated that better item memory implied worse source memory, and vice versa. According to Johnson et al. (1993), a trade-off between perceptual or contextual information on the one hand and semantic detail or item strength on the other hand, is possible under some circumstances. This idea is consistent with the item-source trade-off hypothesis. Johnson et al. (1993) further argued that experiencing external events would lead to greater encoding of perceptual and spatio-temporal information than internally generating information would. Consequently, their framework and the item-context trade-off account concur in predicting that self-generation (i.e., internally generating information) leads to worse context memory. Some studies conducted in the field of neuropsychology suggest that item memory and source memory are dissociable from one another to a certain degree; however, most cases result in a simple dissociation instead of a double dissociation. For example, Schacter, Harbluk, and McLachlan (1984) and Shimamura and Squire (1987, 1991) showed that patients that were amnestic as a result of brain injury exhibited an impairment in their source memory performances that was out of proportion to the impairment shown in item memory. The same result was reported by Multhaup and Balota (1997), who studied healthy older adults, adults diagnosed with very mild dementia of the Alzheimer type (DAT), and adults diagnosed with mild DAT. A positive generation effect emerged for the DAT groups in item memory, whereas source discrimination was disproportionately impaired.

Evidence in favour of the item-source trade-off hypothesis within generation effect tasks was, among others, provided by Jurica and Shimamura (1999). They implemented a task that simulated a social conversation, in which participants either read statements presented by faces on the computer screen (i.e., read condition) or generated answers to questions (i.e., generate condition). Participants could more easily remember topics they gave answers to, than topics they read statements about. Additional to this positive generation effect in item memory, a negative generation effect in source memory occurred; participants' ability to recall the source of a topic was disrupted when they had been asked to self-generate answers.

Processing Accounts and Dual Hypothesis

Both the processing account (Jacoby, 1983; Mulligan, 2004; Mulligan, Lozito, and Rosner, 2006) and the dual hypothesis (Riefer et al., 2007) are discussed in greater detail in Chapters 2 and 3. For that reason, it is sufficient to explain the ideas only briefly at the present point.

According to the *processing account*, two types of processing arise when studying an item: perceptual processing and conceptual processing. While reading and while self-generating an item, both types of processing occur to a certain degree. Different amounts of perceptual and conceptual processing lead to superior memory for self-generated items in item memory tasks, while at the same time leading to inferior memory for self-generated items in source memory tasks. In spite of agreeing on some predictions, the processing account differs from item-source trade-off hypothesis on others. Evidence for the processing account was provided by Mulligan (2004). Mulligan et al. (2006) in principle followed the same logic as Mulligan (2004), nevertheless they questioned the roles of conceptual and perceptual processing as the basic underlying mechanisms. Instead, Mulligan and collaborators argued in favour of a more general version involving visual and non-visual processing.

Riefer et al. (2007) highlighted that two source memory paradigms have been implemented in generation effect studies to date: namely external source monitoring on the one hand and reality monitoring on the other hand. According to their view, empirical evidence had been regarded as diverse due to ignorance of these two types of paradigms. Therefore, they stated a new hypothesis in which both were included – the dual-hypothesis. Riefer and colleagues predicted and found the emergence of a positive generation effect in a reality monitoring paradigm and the emergence of a negative generation effect in a source-monitoring paradigm. These paradigms concern design and procedure of an experiment and were thus found to serve as a powerful moderator of the generation effect in source memory.

1.3 Interim Summary and Outlook

The occurrence of a positive generation effect in item memory proved to be quite robust and easily replicable (under certain conditions). For source memory, that is for the recall of the attributes of an item and its encoding conditions (such as colour, font, or degree of completeness), further papers exist investigating the effect of self-generation of items. However, compared with empirical findings on item memory and self-generation, less consensus concerning data and theory has been reached until now in studies of source memory and self-generated material.

Therefore, the two lines of research investigated in the course of this dissertation deal with the generation effect and source memory and aim to shed light on present inconsistencies, contradictions or aim to illuminate unanswered questions. Chapter 2 addresses the role of the processing of perceptual attributes and the processing of internal states in memory for source (i.e., in memory for the degree of completeness). Chapter 3 is concerned with the role of increased conceptual processing for source (i.e., for memory for presentation colour).

Chapter 2: The Role of the Processing of Perceptual Attributes and Internal States

The present chapter addresses the role of the processing of perceptual attributes and the processing of internal states in memory for source (i.e., in memory for the degree of completeness).

2.1 Introduction to Chapter 2

The purpose of the research introduced in Chapter 2 is twofold: First, it is aimed at solving the contradictions between the theories of Mulligan (2004) and Riefer et al. (2007), especially concerning the dimension degree of completeness. A solution was aimed for by implementing the processing of perceptual attributes and the processing of internal states for the specific case of degree of completeness. Second, the present research is aimed at advancing current theories in the field, by adding a new and more abstract layer concerning types of processing. These two goals assent and harmonise, because on the one hand the specific case of degree of completeness exemplifies the problem. On the other hand, in order to test the truth and existence of this abstract concept, one has to find a useful application; degree of completeness is perfect as an application. Thus, degree of completeness can be regarded on the one hand as the impetus for creating a new layer in the first place and on the other hand as the suitable means and method for testing it.

In the following part of the introduction, the basic research idea is outlined in greater detail. First, the paper by Mulligan (2004) is discussed thoroughly. Next, the hypothesis by Riefer et al. (2007) is summarised and critically evaluated. The research problem is then presented in detail with a suggested new solution to the problem following. Finally, applications of the suggested solution are given in Experiments 1 and 2.

The Studies by Mulligan (2004)

In his 2004 paper, Neil W. Mulligan studied the generation effect for colour memory of presented words. Due to previous findings that lead to an inconclusive picture (Jurica & Shimamura, 1999; Marsh et al., 2001; Rabinowitz, 1989), Mulligan attempted the matter once more. He conducted as many as 12 separate laboratory studies varying type of stimulus presentation (between-subjects lists vs. within-subject lists), target colour, and background colour, among other things. Material was held constant across studies: Antonym pairs, such as "superior – inferior", were taken from Masson and MacLeod (1992, Appendix B). Mulligan attempted to find a coherent picture of how the generation effect paradigm influences memory for item and, more importantly in the present research, memory for source. More specifically, he used the wording *memory for contextual details* and described this rather as one sub-part within the "broader notion of source memory" (Mulligan, 2004, p. 840) as presented in Johnson's source framework (Johnson, et al., 1993; Mitchell & Johnson, 2000). This distinction from the more general concept of s*ource memory* as well as the rather subtle but nevertheless crucial difference between *memory for contextual detail* and *source monitoring* are highly important and are addressed in greater detail later in this work.

Mulligan based his hypotheses and expectations on an account voiced previously by Larry L. Jacoby (1983): *the dual processing account*. This idea is applicable in principle to any type of material, but may lead to different qualitative memory outputs depending on the experimental conditions and paradigms under which material was studied (e.g., read vs. generate condition). According to the dual processing account, two types of processing occur when perceiving and elaborating on a stimulus: *perceptual processing* and *conceptual processing*.

Perceptual or data-driven processing refers to the processing (i.e., to the perception and the encoding) of the *perceptual attributes* of an item – that is, to its appearance. Examples of perceptual attributes are colour, font, or location of a target. An object's *perceivable* characteristics are processed. Note that this description holds across all sensory modalities. For auditory processing, for example, perceptual processing might include processing of voice characteristics, such as pitch or timbre. In contrast to this, conceptual or conceptually driven processing refers to the processing of the *concept* of an item itself – that is, to its meaning. Conceptual processing is also closely related to the concepts of priming, semantic activation, spreading activation, or conceptual activation (e.g., Anderson, 1976, 1983; Collins & Loftus, 1975; McClelland & Rumelhart, 1986; McNamara, 1992, 1994; Meyer & Schvaneveldt, 1971). The types of processing are theorised to take place simultaneously and to draw on the same executive resources or memory resources. Consequently, an *in*crease in the amount of *perceptual* processing (necessitated for instance by study condition) would therefore naturally lead to a *de*crease in the amount of *conceptual* processing, and vice versa.

Concomitantly, the type of processing at study determines the memory trace. When *perceptual processing has taken place predominantly at study*, the memory trace that was created is rich in perceptual and visual detail. What the item presents or what it exactly is or what it is semantically related to, is not encoded well. When *conceptual processing has taken place predominantly at study*, the memory trace is rich in meaning; which object is presented and its semantic network are encoded well. What this specific object looked like, on the other hand, is not.

Finally, when memory is tested, the quality of the memory output depends on the memory trace. For cases in which a great overlap exists between trace and test demands, memory should be good. This is the case (a) when a perceptually rich memory trace was formed and a test of perceptual attributes of the item follows, or (b) when a conceptually rich memory trace was formed and a test of conceptual meaning or conceptual features of the item follows. For cases in which this match is not perfect, memory performance should be diminished.

In short, Mulligan stated that the processing account "stems from a broader transfer-appropriate processing account of the generation effect (Jacoby, 1983) that argues that generation does not uniformly improve memory but, rather, produces a difference in what is encoded" (Mulligan, 2004, p. 851).

Jacoby (1983) had devised the processing account to be applicable to any type of task or material; Mulligan, on the other hand, applied the processing account to the specific case of generation effect tasks. Thus, Mulligan stated that for complete items contrasted with incomplete items, there is not a difference *as such* in what takes place when remembering colour. There is not *one thing* occurring for complete items (i.e., in the read condition) and a *different thing* occurring for incomplete items (i.e., in the generate condition). Instead, for both types of stimuli, perceptual and conceptual processing result. However, processing occurs to different degrees, which is induced and enforced by the diverse task demands brought forward by the highlighted differences in item presentation.

When *complete* words (that are coloured) are to be studied, their complete form is presented (e.g., "foot-shoe"). All coloured letters are displayed fully. Consequently, words can be identified easily and most processing can take place for perceptual attributes of the items. This leads primarily to perceptual processing. Thus, a memory trace results that is perceptually rich. When *incomplete* words (that are coloured) are to be studied, only parts of the target words are presented. The words can still be identified easily due to the co-occurrence of the cue word and knowledge of the generation rule, but the identification arises from a conceptual deduction from the cue word on the one hand and the remainders of the target word on the other hand ("foot-sh__"). Hence, conceptual processing occurs primarily and a memory trace results that is conceptually rich.

For item memory tests, a conceptually rich memory trace is advantageous, because memory for the concept of an item is tested. Consequently, incomplete items, which are processed in a predominantly conceptual fashion, are remembered better than complete items: A *positive generation effect* emerges. For source memory tests, a perceptually rich memory trace is beneficial, because a perceptual attribute (e.g., word colour) has to be remembered and retrieved. Hence, for complete items, which are processed in a predominantly perceptual fashion, colour is remembered better than is true for incomplete items. A *negative generation effect* emerges. An illustration of this explanation can be seen in Figure 2.1.

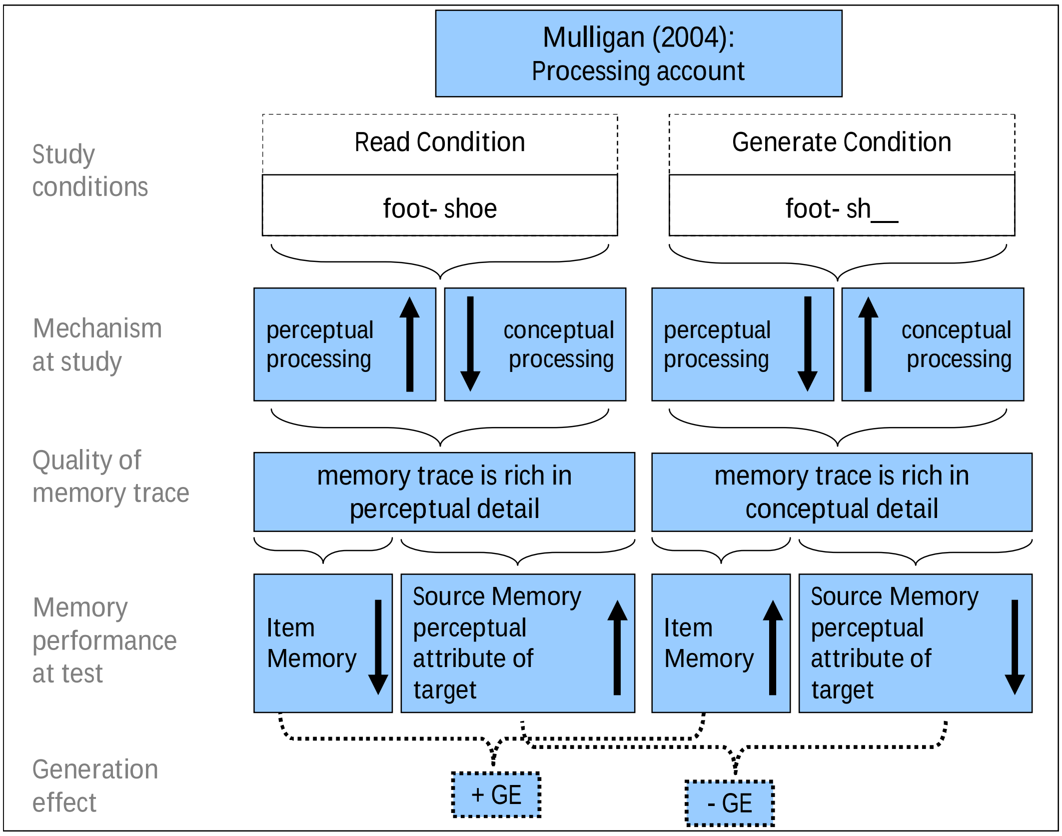

Figure 2.1: Illustration of the *cognitive mechanisms*, the created *memory trace*, and the resulting *test performance*, regarding a generation effect paradigm as theorised in the processing account by Mulligan (2004)

So far, predictions of the processing account and of the item-source trade-off hypothesis are congruent. In spite of being based on diverse theoretical assumptions, both predict a positive generation effect on item memory tasks and a negative generation effect on source memory tasks. However, Mulligan stressed that the item source trade-off hypothesis is rather general in its applicability: It does not discriminate between what type of perceptual attributes are relevant. In contrast to this, the dual processing account is rather specific. Mulligan thought the type of perceptual attribute to be crucial. Instead of separating perceptual attributes by task dimensions (colour vs. position vs. font, etc.), he differentiated between (a) *attributes pertaining to the target itself* and (b) *attributes independent of the target*. This means that the same task dimension, e.g. colour, can either pertain to the object - e.g., colour of the word - or be independent of the object - e.g., background colour of the screen -. Another source attribute, which is independent of the target and is also studied in Mulligan (2004), is the location of the target word on the computer screen. Mulligan highlighted that "the encoding of extratarget features is at neither an advantage nor a disadvantage in the generate condition" (Mulligan, 2004, p. 837).

Hence, when a perceptual attribute is part of the object, a negative generation effect should occur, whereas when a perceptual attribute is independent of the object, memory for this attribute should be equal for complete and incomplete items.

When reviewing the empirical findings by Mulligan (2004), four points can be emphasised in brief: First, no difference in result patterns emerged for studies conducted with within-subject designs and between-subjects designs. Note that in previous experiments, differences in the occurrence of the generation effect could be found depending on which of these designs was used (e.g., Grosofsky et al., 1994; Hirshman & Bjork, 1988; Schmidt & Cherry, 1989; Slamecka & Katsaiti, 1987). However, one has to emphasise that in Mulligan (2004) between-subjects variations and within-subject variations were conducted in *separate* studies. For a more precise test, it would be preferable to use both variations in only *one* study. Second, a positive generation effect in item memory was replicated consistently in Mulligan's investigation. Moreover, a negative generation effect in source memory emerged for the external perceptual attribute target colour. Hence, a negative generation effect occurred when the to-be-studied *perceptual attribute pertained to the actual object*. In contrast, in source memory for location and background colour a null effect resulted; for the latter, high statistical power could be assured. Overall, for studies in which a *perceptual attribute was independent of the actual object* a null effect appeared. Thus, it can be concluded that Mulligan found full support for his predictions deduced from the processing account. See table 2.1, for a summary of the respective studies.

Mulligan also studied a nuisance factor, which could itself be responsible for the negative generation effect found for the source attribute target colour. When considering the coloured incomplete words, one could argue that colour memory was worse for these items simply because *less colour* was displayed at presentation as compared to the amount of colour present for complete items (e.g., "future-p___" vs. "future-past"). In other words, superior colour memory for read items could simply be due to the presentation of a larger amount of colour. Mulligan (2004) acknowledged this circumstance as a potentially confounding factor. He attempted to rule it out (Exp. 5) by replacing the underscores with a continuous set of solid blocks presented in the word colour (e.g., "future-p████" vs. "future-past”). Thereby, Mulligan overcompensated the previous imbalance by implementing a lesser amount of colour in complete items. Nonetheless, the negative generation effect for source memory colour persisted, strongly supporting the interpretation that the negative generation effect is more likely to be attributable to the primary conceptual processing taking place for incomplete targets (see also Mulligan et al., 2006, Experiments 3A and 3B).

Table 2.1: Selected experiments from Mulligan (2004) that are relevant for the current presentation of evidence for the dual processing account; Result patterns include positive (“+”) generation effects, negative (“-“) generation effects and null effects

Selected experiments from Mulligan (2004)					
Experiment number	Type of perceptual attribute	Source memory dimension	Presentation of study material	Results item memory	Results source memory
1	Pertaining to object	Colour of target word	Within-subject manipulation	+ generation effect	- generation effect
2	Independent of object	Location of target word	Within-subject manipulation	+ generation effect	Null effect
3	Pertaining to object	Colour of target word	Between-subjects manipulation	+ generation effect	- generation effect
4	Independent of object	Location of target word	Between-subjects manipulation	+ generation effect	Null effect
6	Independent of object	Background colour	Within-subject manipulation	+ generation effect	Null effect (reported power = 0.95 for n = 40, α = 0.05, one-tailed)

Note that Mulligan discussed the possibility that generation may enhance memory for more global aspects of spatio-temporal context, even though it did not enhance memory for background colour or for location in his studies. Koriat et al. (1991) and Marsh et al. (2001) found a positive generation effect for memory of the study room. These results may indicate that some aspects of spatio-temporal context may be encoded better under generate than under read conditions. However, Mulligan pointed out, that in their papers it did not become clear what type of information was remembered better – either temporal information, memory for perceptual attributes of the rooms, or memory for associations between study words and details of the room.

In conclusion, Mulligan (2004) showed a concise theoretical foundation of and evidence for the dual processing account.

Interim Summary

To reiterate the basic points of Mulligan (2004) that are of greatest importance in the context of Chapter 2, one can state the following: Mulligan (2004) demonstrated that in the classical generation effect design (i.e., when there are a semantically related cue and a target word, the latter of which is either presented in a complete or in an incomplete form), (a) a positive generation effect tends to result in item memory and more importantly (b) a negative generation effect tends to emerge for the contextual detail target colour. See Figure 2.2, for a simplified presentation of this statement.

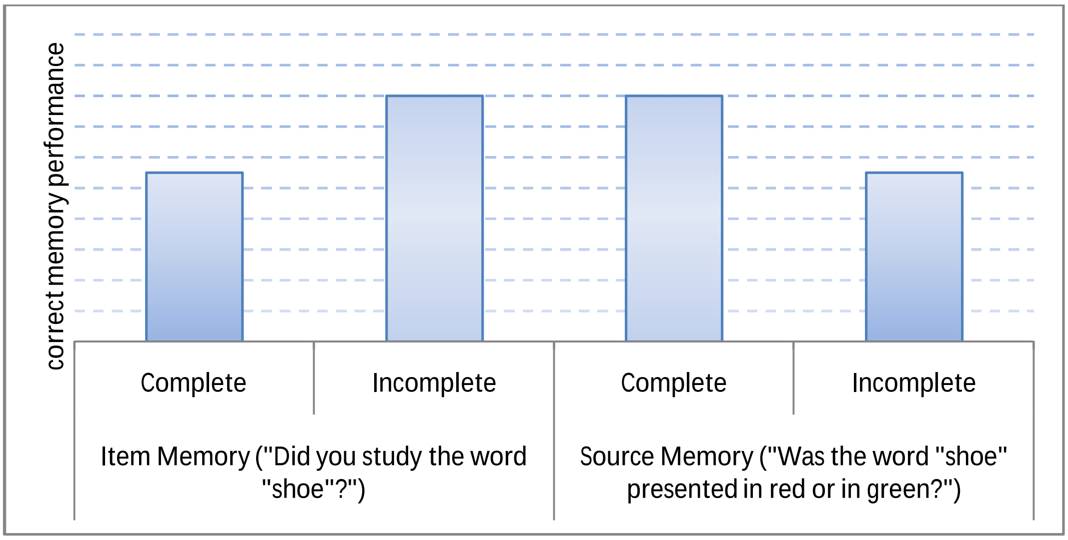

Figure 2.2: Illustration of the predictions drawn from Mulligan's (2004) processing account for item memory and source memory for colour

The Study by Riefer et al. (2007)

In the current section, the dual hypothesis by Riefer et al. (2007) is introduced. The authors distinguished between two types of experimental paradigms, namely between the *external source monitoring paradigm* and the *reality monitoring paradigm*. Riefer and colleagues state in their *dual hypothesis* that in external source monitoring paradigms, negative generation effects tend to emerge, whereas in reality monitoring paradigms, positive generation effects tend to emerge.

Mulligan (2004) argued that several researchers have examined the role of self-generation of material on memory for source features and he concluded that results were "quite mixed" (Mulligan, 2004, p. 840). He therefore settled for working on source dimensions that were usually employed in the literature saying that his "generation task and materials were chosen to be typical of research on the generation effect" (Mulligan, 2004, p. 841). However, Riefer et al. (2007) arrived at a *different conclusion*. Riefer and colleagues subdivided previous studies into two categories representing diverse types of experimental paradigms.
The first category comprised studies in which a reality-monitoring paradigm was employed. Here, participants were asked to encode and recall whether the target had been produced actively by themselves or whether it had been conjured by someone else (e.g., another person or the computer) and could therefore simply be read by the participant. In short, participants had to decide whether the source of production was themselves or someone else. The second category included studies in which an external source-monitoring paradigm was employed. Here, participants were asked to encode and recall an external attribute (i.e., a perceptual attribute referring to the appearance of the target) of the item, such as colour or font.
After having exemplified and identified this difference, Riefer and colleagues asserted that they were able to classify negative and positive generation effects found in the literature by employing these two categories. They formulated a dual-hypothesis, in which they stated that in a reality monitoring design, a positive generation effect should emerge and that in an external source monitoring paradigm, a negative generation effect should emerge. Riefer et al. (2007) gave two reasons why they considered the dual hypothesis theoretically tractable (see also Figure 2.3):

(a) For positive generation effects tending to occur in reality monitoring paradigms, Riefer et al. (2007) referred back to theories of why a positive generation effect robustly occurs in item memory, arguing that both follow the same logic. Riefer and colleagues highlighted that these theories assume that self-generating information results in "extra mental operations that improve various aspects of the encoding of that information. It is logical to theorise that this enhanced encoding not only benefits the overall recognition of this information but also memory for its source as well." (Riefer et al., 2007, p. 1403)

(b) For negative generation effects that tend to occur in external source-monitoring paradigms, Riefer and colleagues gave two examples of possible explanations, namely the item-source trade-off hypothesis on the one hand and Mulligan's (2004) processing account on the other hand.

How Riefer et al. (2007) arrived at their solution is described as a process, which commenced from scrutinising empirical findings. Riefer et al. (2007) stated that the separate methodologies inherent in reality monitoring and external source monitoring result in separate designs: "The distinction between reality monitoring and external source monitoring can account for many of the diverse results of prior research concerning the effects of generation on source memory" (Riefer et al., 2007, p. 1393). However, experimental details of supporting results as well as an allocation of papers or experiments to the two types of paradigms, which could substantiate their idea, are missing. Hence, an extended and more detailed list of empirical evidence differentiated by experimental paradigm can be found in Appendix A. When inspecting this list, it can be seen that the dual hypothesis by Riefer and colleagues accommodates only few of the results well. Several cases remain that disagree with their suggestion.

The dual hypothesis was tested in an intriguing experiment, which sustains the hypothesis' validity. Prior to Riefer et al. (2007), researchers studying the effect of self-generating material on memory for source attributes have implemented either the external source-monitoring paradigm or the reality monitoring paradigm exclusively. In these studies, the differences in paradigms were generally not explicitly theorised and discussed to influence experimental outcome. In contrast, Riefer and colleagues appreciated this difference as not only one in design but also one in item processing, which was theorised to concomitantly lead to differences in memory trace and in memory performance at test.

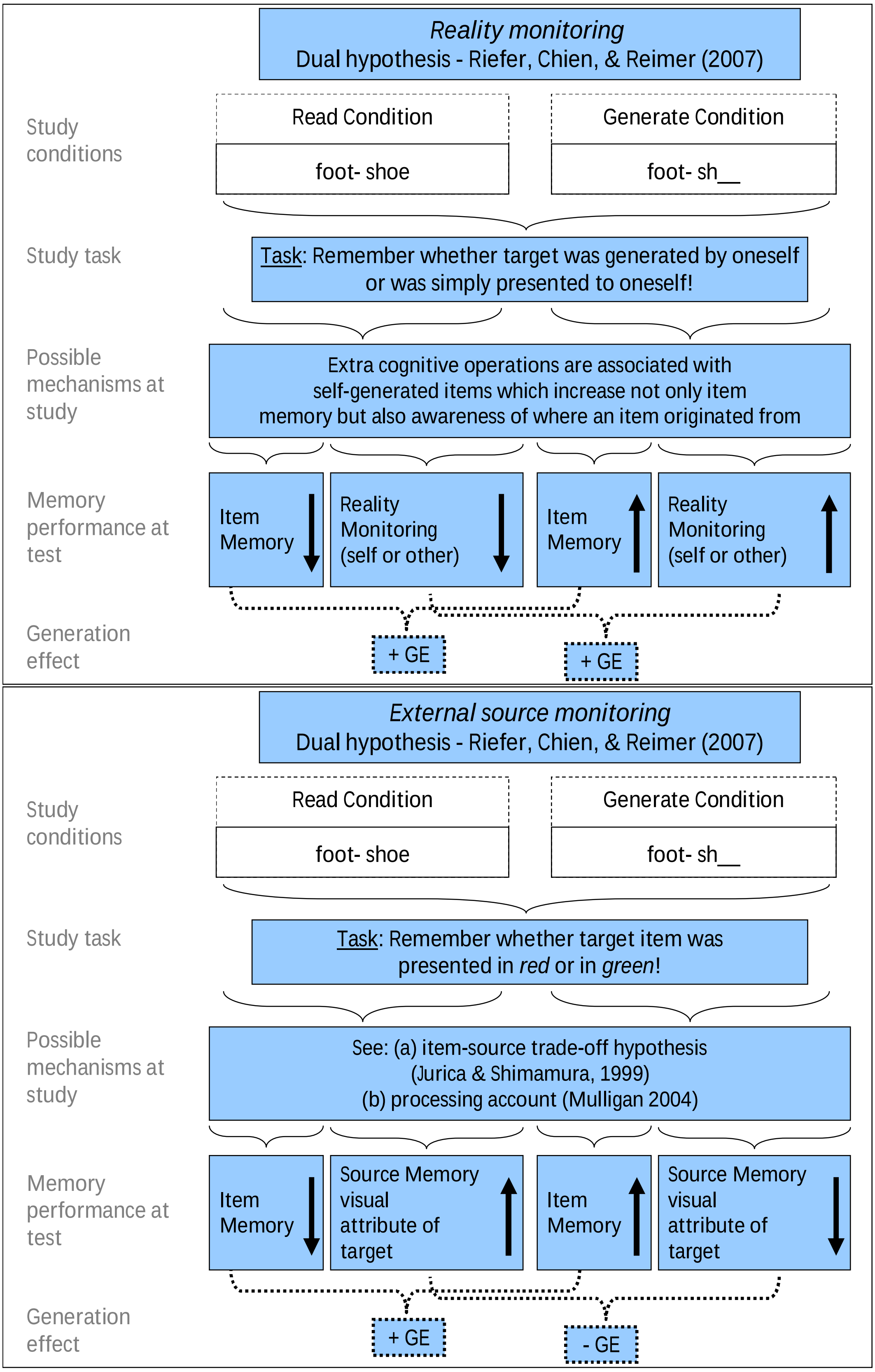

Figure 2.3: Illustration of the dual hypothesis theorised in Riefer et al. (2007) including suggested *mechanisms at study* and predicted *test performances* ("+ GE" = positive generation effect; "- GE" = negative generation effect)

Since Riefer et al. (2007) integrated both paradigms into only *one* theoretical construct – i.e., their dual hypothesis – it was rendered necessary to test the predicted outcomes in a more direct way. Therefore, they incorporated the two types of source-monitoring tasks into a single experiment. Concerning stimulus material and procedure, Riefer and colleagues replicated and extended Experiment 1 of Mulligan (2004). The extension consisted of adding a second source dimension by testing participants' memory also for the awareness of which items were generated and which were read. Thus, the experiment contained *multidimensional* source information: first, source information of target colour; second, source information of which targets were simply read and which targets were self-generated. In Riefer et al. (2007), memory for which targets were simply read (i.e., for complete items) and for which targets were self-generated (i.e., for incomplete items) was called *memory for generation*. However, for consistency reasons the expression *memory for degree of completeness* is instead used henceforth in the present dissertation.

Riefer et al. (2007) found a positive generation effect for item memory, accompanied by a negative generation effect for colour (i.e., in the external source monitoring paradigm) and a positive generation effect for memory for degree of completeness (i.e., in the reality monitoring paradigm). In other words, the standard finding of a positive generation effect in item memory could be replicated and – more interestingly – differences in experimental source memory paradigm influenced the experimental outcome significantly. Thereby an interaction between type of experimental paradigm and condition (read vs. generate condition) was exhibited. Differences in experimental paradigm moderated the generation effect in source memory.

It can be concluded that the empirical evidence gained by Riefer and colleagues supported the dual hypothesis, showing that type of source-monitoring task constitutes an important factor in determining whether positive or negative generation effects occur for source memory. Consequently, Riefer and colleagues inferred that "future research and theory development on this issue should take into consideration the type of source-monitoring paradigm being examined" (Riefer et al., 2007, p. 1403).

Note that the idea of regarding source memory tasks in this two-fold way - i.e., external source monitoring vs. reality monitoring - is not new (cf., Johnson et al., 1993). Nevertheless, the contributions of Riefer et al. (2007) can be seen (a) in the resurrection of the concept within the generation effect literature and (b) in the systematic description and illustration of the concept in their paper. Additionally, (c) Riefer et al. (2007) were the first to test these paradigms not just across experiments, but *within a single* experiment. In doing so, they showed that the type of source monitoring task is an important factor in determining whether positive or negative generation effects occur for source memory. Furthermore, Riefer and colleagues re-exemplified the influence of response biases in traditional source memory measures and gave a discussion of generation effect results gained through analysing data by employing multinomial processing tree models.

However, unlike Mulligan (2004), Riefer et al. (2007) failed to provide just one overarching theory. Instead they set up their dual hypothesis that helps to accommodate previous findings, but lacks just one sound meta-theoretical model. Although different theories are stated to account for different parts of the dual hypothesis, these accounts seem to be an application of diverse previous theories. Moreover, Riefer et al. (2007) did not take Mulligan's specificity assumption into consideration.

Interim Summary

To reiterate the basic points of Riefer et al. (2007) that are of greatest importance in the context of Chapter 2, one can state the following: Riefer and colleagues stated that two types of experimental paradigms exist in generation effect studies, which lead to different source memory outcomes. First, there is the external source monitoring paradigm, for which participants have to remember a contextual attribute of the target itself, such as its colour or font. Second, there is the reality monitoring paradigm, in which participants have to remember which source the item originated from, e.g., whether the item was presented by an external source or was produced actively by oneself. In external source monitoring paradigms, reading items instead of generating them leads to superior memory - i.e., a negative generation effect tends to occur. In contrast to this, in reality-monitoring paradigms, generating items instead of reading items leads to superior memory - i.e., a positive generation effect tends to occur. Thus, the type of experimental paradigm determines the experimental outcome for source memory. See Figure 2.4, for a simplified presentation of this statement.

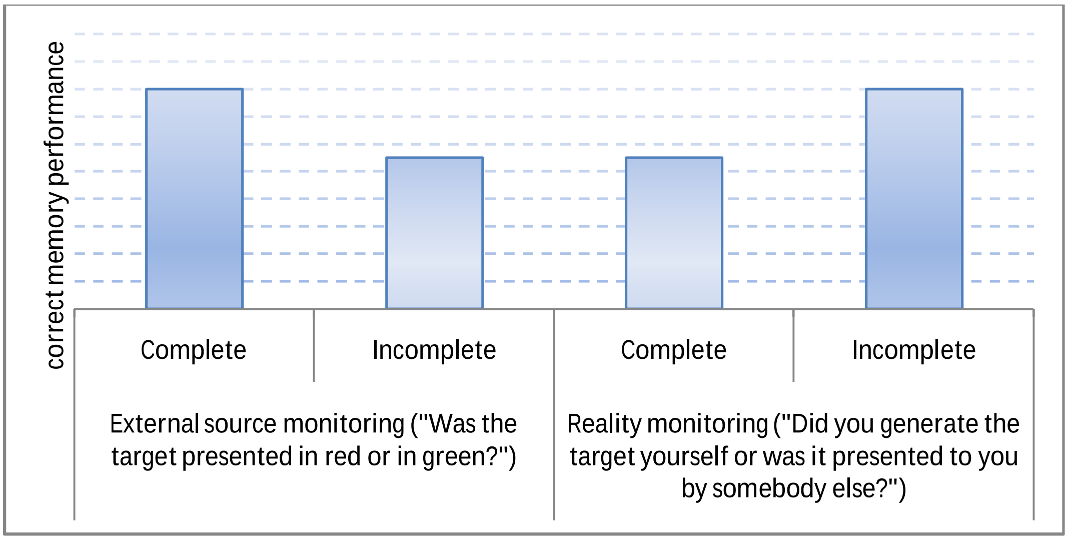

Figure 2.4: Illustration of the predictions for source memory drawn from the dual hypothesis (Riefer et al., 2007)

When considering the critical evaluations of Mulligan (2004) and Riefer et al. (2007) presented previously, one can see that both approaches show disadvantages and advantages. I conclude that the former approach seems to be more theoretically refined, whereas the latter approach seems to have evolved more from a practical and data-considering standpoint. However, both make robust and empirically well supported predictions and significantly advance knowledge in this field of research.

The Contradiction

When considering the theoretical accounts of Mulligan (2004) and Riefer et al. (2007), I conclude that they are, to a certain degree, incompatible. When attempting to integrate the accounts, the following points have to be considered: The approaches match in their predictions of a positive generation effect in item memory (a). Moreover, Mulligan (2004) as well as Riefer et al. (2007) both hypothesise the emergence of a negative generation effect for contextual details, when this information is about an external source attribute, such as colour (b). Mulligan further advanced this prediction by discriminating between contextual details that are part of the target, such as target colour, resulting in a negative generation effect, and contextual details that are independent of the target, such as background colour, entailing a null effect. This sub-clause to Mulligan's predictions is neither contradicted by

Riefer et al.'s (2007) hypothesis nor mentioned in it in any way. However, Riefer and colleagues added another source dimension, namely the distinction of whether an item was produced by oneself or by somebody else (i.e., reality monitoring) which is expected to result in a positive generation effect (c). In contrast, Mulligan did not mention reality monitoring in his account in any way.

In brief, concerning source memory, Mulligan's (2004) and Riefer et al.'s (2007) predictions match when external source attributes are concerned. Additionally, Riefer and his collaborators made an assumption about the emergence of a positive generation effect in reality monitoring paradigms, which Mulligan did not mention. See Figure 2.5, for an illustration of these points.

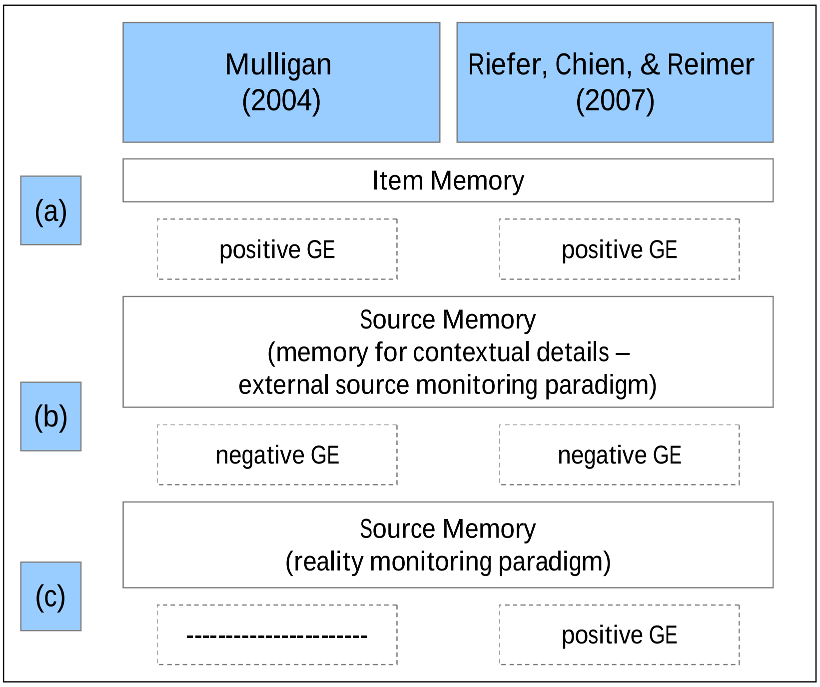

Figure 2.5: Illustration of the predictions of Mulligan (2004) and Riefer et al. (2007) concerning item memory and source memory ("GE" = generation effect)

Johnson et al. (1993) assumed that when confronted with a study item, many cues are encoded to a certain amount at once, e.g., the item itself, colour, sound, spatial location, emotional reactions, reflective processes, and elaborations. After evaluation of these encoded cues, one can arrive at a conclusion concerning memory and thereby regarding the prior study episode. Thus, I conclude that when considering an object, one can focus on different aspects.

For example, one can focus on and remember whether oneself or another person did something. Yet, when considering an object, one can also focus on and remember whether an object looked this or that way. In this abstract form, it does not seem to be implausible to make different judgements for *one and the same* stimulus. A painting, for instance, can be judged and evaluated concerning its artistic quality, its size, its colouring, or the emotional response it evokes in viewers, among other dimensions. Thus, diverse judgements can be made for the same stimulus and even the same judgment (e.g., "What can you see?") can lead to divergent results for the same stimulus ("I see an animal" vs. "I see a leopard"). Another example are optical illusions such as the young girl – old woman illusion in which a drawing can be interpreted as the image of a young girl and at the same time as the image of an old woman. In these examples, a certain object is considered under different foci. The same logic can be applied to the dimension degree of completeness, which is of relevance here.

When contemplating the paper by Riefer and colleagues, one could argue that the predictions and findings made by Mulligan (2004) simply fall under the category of the external source-monitoring paradigm. One could say that since Mulligan employs an external source-monitoring paradigm, he finds a negative generation effect in his studies. Likewise, one could continue to argue that there is no problem to this issue and one could suggest that Mulligan's account is simply a theory applicable to one part of the dual processing account by Riefer and colleagues.
However, I believe that when paying close attention to Mulligan (2004), this does not have to be the case. Mulligan does not mention different types of paradigms, likely in an attempt to find a *more general solution for different types of contextual attributes* (- although one could argue that Mulligan simply was not aware of the existence of such a subdivision.) He indirectly argues that the processing account to hold not only for colour and location, background colour and cue colour, as investigated in his studies, but is generally true for other types of contextual details or source attributes as well. Although Mulligan investigated the source memory dimensions word *colour*, spatial *location*, and *background colour* in his studies only, there is no reason not to infer his theory to be applicable to other contextual details as well. This opinion is supported by the fact that Mulligan himself suggested pitch and timbre for auditory stimuli as other forms of contextual detail.

When formulating Riefer et al.'s idea very pointedly, it could be stated that, when *memory for degree of completeness* is concerned, a positive generation effect is expected, whereas, when *memory for target colour* is concerned, a negative generation effect is expected. Bearing Mulligan (2004) in mind, his account is compatible with the latter part of this claim, whereas no argument or allegation is made about *memory for degree of completeness*.

It seems plausible to regard degree of completeness as an external source attribute of the target - just like colour. Moreover, it can be considered an attribute pertaining to the target item itself, instead of being independent of it. Thus, Mulligan's predictions can also be applied to the case of memory for degree of completeness. According to Mulligan's understanding of the dual processing account, a negative generation effect should emerge for memory for degree of completeness as was true for target colour. Hence Riefer et al.'s (2007) and Mulligan's (2004) predictions for memory for degree of completeness are at odds.

In fact, I suggest that *memory for degree of completeness* is not necessarily as clear a concept as suggested by Riefer and colleagues. To reiterate, memory for degree of completeness was described and understood as memory for whether an item had been internally generated by the participant ("Did I conjure this solution myself?") or externally generated and presented to the participant by an outside force (e.g., another person in the room or a face presented on the computer screen); the respective question is similar to the following: "Did someone else conjure the solution and did I just hear it / read it?"

However, memory for degree of completeness does not necessarily have to be defined in this way; instead of asking whether the solution to the stimulus "foot-sh__" was given by oneself or by another person and achieving a certain memory performance on this topic, one can achieve the same degree of correct responses, simply by realising and remembering whether the target "sh__" was presented in a complete or in an incomplete manner. So, instead of remembering the source procuring the object, one could also remember the same information and reach the same performance level by remembering the object's appearance.

Basically, (a) internal-external source and (b) complete-incomplete appearance are like two sides of the same coin. In fact, the dimension of memory for generation (or rather memory for degree of completeness) can be turned from a reality monitoring design (a) into an external source-monitoring design (b).

To rephrase this point, one can say the following: although one could stick with the distinction self-generated versus generated by other, when studying and testing material such as “foot-sh__” and “hand-glove”, there is also the possibility of studying and testing for this type of material in a more perceptual fashion. As illustrated in Figure 2.6, material of this type can be studied both in a reality monitoring way and in an external source monitoring way – to apply Riefer et al.’s (2007) terminology. Thus, in order to study material in the former case, one is to remember whether one has generated an item oneself or whether it was produced by someone else. In contrast, in order to study material in the latter case, one is to remember the perceptual degree of completeness of an item.

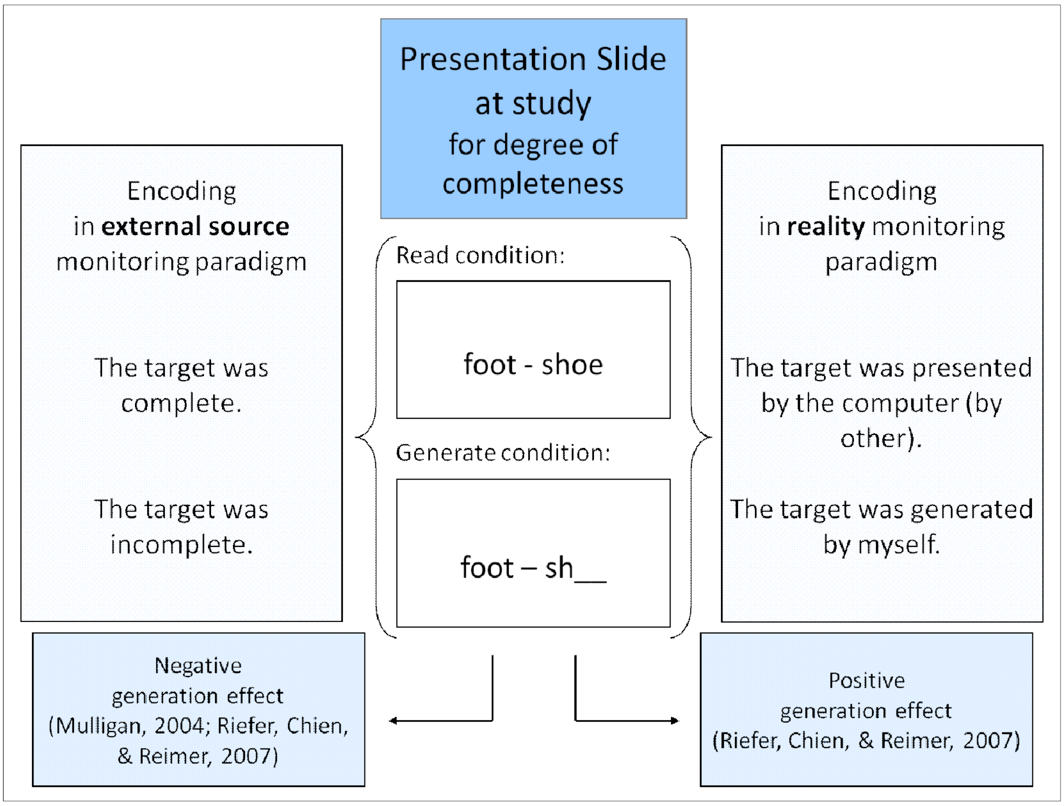

Figure 2.6: Illustration of the two-fold way of encoding the degree of completeness for a target in a generation effect paradigm (encoding in terms of an external source monitoring paradigm vs. encoding in terms of a reality monitoring paradigm) including the predicted outcomes (Mulligan, 2004; Riefer et al., 2007)

This difference is not only a mere discrepancy in *wording*, because, due to this wording, different forms of processing are determined and thus separate experimental *paradigms* are employed. Consequently, due to the existence of these different experimental paradigms, separate *predictions* concerning the experimental *outcome* can be made. According to Riefer et al. (2007), when working within the reality monitoring paradigm, a positive generation effect is expected, whereas, when working within the external source monitoring paradigm, a negative generation effect is expected. According to Mulligan (2004), when looking at the degree of completeness in terms of completeness of the item, predictions are applicable to this dimension as well resulting in expecting a negative generation effect. Still, when continuing within the reality monitoring framework, no straightforward predictions can be derived from Mulligan's hypothesis.

Since Mulligan does not make a prediction on exactly which effect is to emerge under a reality monitoring phrasing, possible predictions cannot easily be inferred from his account. However, since there is a strong conceptual match between memory for target colour and memory for the degree of completeness of the targets (as suggested previously), one can deduce from the processing account that a negative generation effect should emerge. Interestingly, the dual hypothesis by Riefer and colleagues could potentially be applied to both types of phrasing: In case of asking participants to remember whether the item was produced by themselves or by someone else – which essentially is a reality-monitoring paradigm – a positive generation effect is expected. On the other hand, when asking participants to remember whether the item was presented complete or incomplete – which essentially is an external source-monitoring paradigm – a negative generation effect is anticipated.

In short, following this line of argument, one can see that Mulligan's (2004) and Riefer et al.'s (2007) theories are partly at odds. It can be concluded that a mismatch is evident.

Interim Summary

In conclusion, the problem can be characterised as follows: Within a generation effect paradigm, memory for the degree of completeness of an item can be studied in two ways, namely either by employing a reality monitoring design (i.e., produced by self vs. produced by other) or an external source monitoring design (i.e., presented complete vs. presented incomplete) for contextual detail. Depending on which type of design is used, different predictions can be made, according to Mulligan (2004) and Riefer et al. (2007). *These predictions are incompatible*, because a negative generation effect is expected in the external source monitoring design, whereas a positive generation effect is expected in the reality monitoring design. For the degree of completeness, evidence was found when a positive generation effect occurred in a reality monitoring design (Riefer et al., 2007). Moreover, evidence for a negative generation effect in external source monitoring designs emerged, but only for source dimensions such as target colour and target font (Mulligan, 2004; Mulligan et al., 2006). No evidence exists for predictions of a negative generation effect in an external source monitoring design for the dimension degree of completeness.

In short, when considering degree of completeness, no clear picture exists as to which predictions apply to this source attribute. Possible predictions would lead to contradictory outcomes that cannot be reconciled easily.

This contradiction poses a problem in understanding what has been done so far in the literature and it leaves one without clear predictions of memory performance for source attributes of generated items. There is still no consensus on how generation influences source memory. Since Mulligan on the one hand and Riefer and his colleagues on the other hand both have got sound theoretical foundations that have a long tradition in Cognitive Psychology (the dual processing account in the former case and the reality-monitoring account in the latter case), finding a way to reconcile these two approaches would advance our knowledge of the role of self-generation in source memory tasks.

The Suggested Solution and its Implementation

Concerning source memory effects in generation effect paradigms, I see a starting point to overcome existing inconsistencies by taking into account *differences in mental processing* at study and at test. When comparing the experiments of Mulligan (2004) and of Riefer et al. (2007), it can be seen that the way in which participants were *instructed* during the study phase and their ways in which the test questions were worded, were conceptually different and could thus in my view have lead to the divergent result patterns.

In the paper by Mulligan (2004), the instructions emphasised the processing of and thus the memorisation of *perceptual attributes* (PA) of items. Mulligan directed participants' attention to the physical appearance of stimuli and asked them to remember these. The focus lay outside of the self on a specific outside stimulus. Mulligan (2004) found a *negative generation effect* in source memory in his studies.

In Riefer et al. (2007), the instructions highlight the processing of and thus the memorisation of *internal states* (IS) while reading or generating items. Riefer and colleagues directed participants' attention to internal mental processes going on while reading and self-generating items – active internal self-production of items or rather passive internal reading of items – and asked them to remember these. The focus lay on the self and on inner processes. Riefer et al. (2007) found a *positive generation effect* in source memory in their study.

I suggest that in principal, the dichotomy of perceptual attributes versus internal states can be applied to *all* types of source memory variations, such as colour of items, font of items, degree of completeness of items, and so on. Consider three examples that illustrate how each source attribute could be encoded in either of these two ways, i.e., either in terms of processing of perceptual attributes or in terms of processing of internal states: *Font size* could be processed by focussing on the appearance such as "large" versus "small" font (PA) or by focussing on the induced arousal such as "high" or "low" (IS). Likewise, *Font* could be processed by focussing on the specific fonts such as "Times New Roman" versus "Arial" (PA), or by focussing on the induced pleasantness "pleasant" versus "unpleasant" (IS). Finally, *colour* could be processed by focussing on the specific colours such as "warm orange" versus "icy blue" (PA), or by focussing on the induced valence such as "positive" or "negative" (IS).

Whether or not the suggested solution of taking into account differences in mental processing at study and at test is a good remedy for the problem, is best tested in laboratory experiments. Therefore, two studies were designed to tackle the matter and to empirically test the reconciliatory effect of types of processing - processing of perceptual attributes and processing of internal states. The difference between Experiment 1 and Experiment 2 lay in the manipulated strength of the critical variable "instruction-induced types of processing" or simply "instruction". This instruction manipulation was weaker in the former and stronger in the latter experiment.

The critical independent variable "instruction" consisted of two levels: Level 1 "processing of and memory for perceptual attributes" ("PA") versus Level 2 "processing of and memory for internal states" ("IS"). So far, the distinction as operationalised via the variable "instruction" has never been named as precisely or has never been accounted for in any other studies. Another important independent variable was the factor "degree of completeness" which held two levels (Level 1 "complete word" vs. Level 2 "incomplete word"). Dependent variables were measures for old-new item memory, as well as measures for source memory for target colour and source memory for the degree of completeness of items.

Table 2.2: Implementation of the variable "instruction" within Experiments 1 and 2

Variable "instruction"		
Level 1: Perceptual attributes PA	Study phase	Participants were told that some of the words were presented in a complete manner, whereas other words were presented in an incomplete manner.
	Test phase	Response categories were labelled: "The word was complete and red", "The word was complete and green", "The word was incomplete and red", "The word was incomplete and green", and "This is a new word".
Level 2: Internal states IS	Study phase	Participants were told that some of the words could simply be read on the screen, whereas others would have to be generated by the participants themselves from the first letter of the word.
	Test phase	Response categories were labelled: "I could simply read the word on the screen and it was red", "I could simply read the word on the screen and it was green", "I have generated the word myself from the first letter and it was red", "I have generated the word myself from the first letter and it was green", and "This is a new word".

Can a variation in instruction-induced types of processing reconcile the contrasting results found in the effect of generation on source memory performance?

In the following two experiments, the most critical result patterns (in source memory performance) were anticipated to vary according to the type of instructions used. For the experimental group in which the processing of and the memory for perceptual attributes was emphasised, a *negative* generation effect was expected for source memory performance of degree of completeness. In other words, when participants were asked to remember degree of completeness, they should exhibit *inferior* memory for incomplete items compared to that for complete items. For the experimental group in which the processing of and the memory for internal states was emphasised, a *positive* generation effect was expected for source memory performance of degree of completeness. When participants were asked to remember degree of completeness, they should exhibit *superior* memory for incomplete items compared to that for read items.

A significant interaction between the factors "degree of completeness" and "instruction" was anticipated for source memory measures, yielding opposite result patterns: In the PA condition, memory was expected to be better for complete items. In the IS condition, memory was expected to be better for incomplete items. An illustration of the predicted source memory outcomes can be found in Figure 2.7. Note that in the present experiments, the dichotomy (PA vs. IS) was tested for the source memory dimension degree of completeness only. Hence, source memory performance for the dimension colour was theorised not to be affected by this manipulation.

For item memory measures, no significant main or interaction effects were expected, except for a significant main effect of the factor "degree of completeness". A positive generation effect was expected to emerge, resulting in a significantly higher old-new memory performance for self-generated compared to read items.

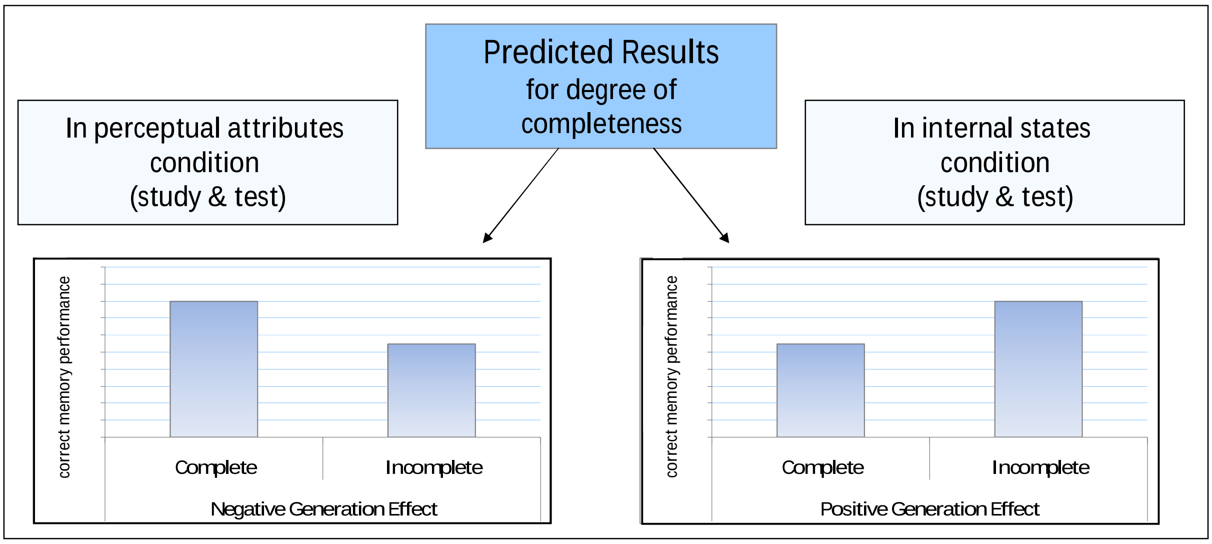

Figure 2.7: Illustration of the predicted results in Experiments 1 and 2 for the source attribute degree of completeness

To reiterate, several predictions were made for Experiment 1 and Experiment 2:

First (Hypothesis 1), regarding item memory (i.e., concerning recognition memory performance of the target word), I expected an advantage of incomplete over complete items. A *positive* generation effect should appear.

Second (Hypothesis 2), regarding source memory for colour, I expected an advantage of complete over incomplete items. A negative generation effect should appear.

Third (Hypothesis 3), regarding source memory for the degree of completeness, I expected an *interaction* between the factors "instruction" and "degree of completeness". For items studied in the perceptual attributes condition (3a), I expected an advantage of complete over incomplete items, i.e., a *negative* generation effect was predicted. In other words, when participants encoded targets by focussing on the targets' perceptual characteristics, I hypothesised a *negative* generation effect to appear. For items studied in the internal states condition (3b), I expected an advantage of incomplete over complete items, i.e., a *positive* generation effect was predicted. In other words, when participants encoded targets by focussing on their own internal states, I hypothesised a positive generation effect to appear.

Fourth (Hypothesis 4), all types of guessing were expected to be at chance level: guessing "old", guessing "red", guessing "complete", and guessing "simply read" versus guessing "new", guessing "green", guessing "incomplete", and guessing "self-generated", respectively.

2.2 Experiment 1

In the present part of the dissertation, Experiment 1 is described in further detail. The primary focus of the experiment lay on the influence of processing type (the processing of perceptual attributes vs. the processing of internal states) on memory performance. Experimental instructions of Experiment 1 followed closely those used in the studies by Mulligan (2004) and Riefer et al. (2007). These wordings were effective in their studies and yielded the above described partly non-compatible result patterns.

2.2.1 Methods

This section provides information on the sample, on the design, on the material, and on the experiment procedure of Experiment 1.

Participants

All participants in this dissertation project were recruited at the University of Mannheim and took part in the studies for monetary compensation or course credit as part of their study requirements. In Experiment 1, 60 persons participated. Since one of the independent variables was the “colour” of the presented stimuli (namely “red” vs. “green”), participants were asked whether they suffered from red-green colour blindness. As a result, one person had to be excluded from the sample. Red-green colour blindness was checked for in all experiments in this way. Of the 59 participants remaining in the sample, 79.66 % were female. Participants’ age ranged from 18 to 29 with a mean of 22.68 years and a median of 22 years. Additionally, participants were asked to judge their command of written and spoken German to rule out problems caused by not understanding the instructions, which were in German. All participants except four confirmed that they were native German speakers. However, this latter group of participants stated that their German proficiency skills were either “very good” (3 of 4) or “good” (1 of 4). Knowledge of German was checked for in all experiments in this way. Fifty-five participants were currently enrolled at the University of Mannheim and primarily were students from the fields of Psychology (50.91 %) or Economic Sciences (20 %). Other fields of study were Language Studies, Political Sciences, Sociology, and Law. Four participants had a degree in Economics, Educational Sciences, or Psychology.

Design

In this laboratory experiment, participants were randomly assigned to the experimental groups, as described below. Four independent variables were employed: First, independent variables for source memory were “colour” and “degree of completeness” and were varied within-subject. They had two levels, namely “red” versus “green” for the independent variable “colour” and “complete” versus “incomplete” for the independent variable “degree of completeness”. Second, the independent variable “instruction” had two levels, namely “the processing of and memory for *perceptual attributes*” (PA) versus “the processing of and memory for *internal states*” (IS), and was manipulated between-subjects. Additionally, study material was subdivided into five stimulus sets, which were counterbalanced across experimental variations. Most important for the current research question were the factors “degree of completeness” and “instruction”. The experimental research design can be seen in Table 2.3.

Material

Stimuli consisted of 55 antonym noun or adjective pairs (e.g., “future-past”, “fast-slow”). Thirty-nine of these pairs were adopted from Masson and MacLeod (1992, Appendix B), the same stimulus source which was also used by Mulligan (2004) and Riefer et al. (2007). Sixteen additional word pairs were developed.

For the presented items, one half was in the read condition and the other half was in the generate condition. In the read condition, words were presented complete. In the generate condition, the first word was presented complete, whereas only the first letter of the second word appeared on the screen and was followed by a four continuous underscore space (e.g., "future-p____"). Items were either displayed in red or in green. These colours were also used by Mulligan (2004) and by Riefer et al. (2007). However, since neither Mulligan nor Riefer and colleagues reported exact colour values, those two colours in the RGB colour value system were used that are explicitly labelled "red" (255-0-0) or "green" (0-255-0); see Hunt (1991).

Table 2.3: Research design of Experiment 1 employing the within-subject variations "colour" and "degree of completeness", and the between-subjects variation "instruction"

Type of instruction	Exp. group	Old / studied items				New / unstudied items
		Red		Green		
		Complete	Incomplete	Complete	Incomplete	
PA	1	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E
	2	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A
	3	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B
	4	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C
	5	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D
IS	6	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E
	7	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A
	8	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B
	9	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C
	10	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D

Consequently, 44 items were displayed at study - each presentation combination consisting of 11 items. In sum, 55 items from the test block were used for analyses. Since 44 of these items (relevant for analysis) had previously been presented in the study phase, 33 additional buffer items were developed to maintain a ratio of 1 : 1 (studied : unstudied items) at test. Additional buffer items were not taken into account for data analysis.

Procedure

The procedure was set up to be very similar to the procedures used by Mulligan (2004) and Riefer et al. (2007). Participants were attending the tasks presented to them via computer in group sessions of up to 15 persons. The experiment was created and executed using the E-

Prime 1.0 software (Schneider, Eschmann, & Zuccolotto, 2002). Participants first signed a consent form, before they were seated in front of the computer.

The experiment comprised of three experimental phases: study phase, distractor phase, and test phase.

Study Phase

The experiment began with the study phase. Participants were told that they would see antonym word pairs and that for half of the trials, only the first letter of the right-hand word (i.e., the target word) was displayed. Moreover, they were also informed that the word pairs additionally were presented either in red or in green colour and would appear against a black background. Participants were told that it was their task to write down the right-hand word of each pair. In cases in which the target word was spelled out, target words could simply be copied onto a response sheet provided to participants (read condition), whereas in cases in which only the first letter was given, participants were instructed to determine what the correct antonym should be and to write their solution down onto the response sheet (generate condition). Consistent with Mulligan (2004), Mulligan et al. (2006), and Riefer et al. (2007), participants were informed that their memory would eventually be tested for the targets themselves, for the words' colour, and for the degree of completeness of the target items.

In the study phase, the between-subjects factor "instruction" was of high importance. Study instructions for the source memory dimension "degree of completeness" largely varied depending on which experimental condition a participant was in: perceptual attributes condition versus internal states condition.

(a) In the perceptual attributes (PA) condition, participants were told that half of the items was presented in a *complete* manner, whereas the other half was presented in an *incomplete* manner. And hence that they should *memorise* whether a target was presented complete or incomplete.

Put briefly, participants were to

1) write down the target words on the response sheet,
2) memorise the target words (item memory),
3) memorise the colour of the target words (source memory "colour"), and

4) memorise whether a target was shown complete or incomplete (source memory "degree of completeness").

(b) In the internal states (IS) condition, participants were told that half of the items could *simply* be *read* on the screen, whereas the other half would have to be *generated* by the participants *themselves* from the first letter of the word. And hence that they should *memorise* whether they could simply read the target on the screen or had to generate a target themselves from the first letter of the word.

Put briefly, participants were to

1) write down the target words on the response sheet,
2) memorise the target words (item memory),
3) memorise the colour of the target words (source memory "colour"), and
4) memorise whether they could simply read the target on the screen or had to generate a target themselves from the first letter of the word (source memory "degree of completeness").

Once it was clear that participants comprehended the task and no questions remained, the study list was presented. For each trial (see Figure 2.8), stimulus pairs were shown sequentially and individually in the centre of the screen, where each pair remained for 7 seconds. A 200 ms blank followed each stimulus pair. Words were presented in 18 point bolt Courier New. During target presentation, participants either copied the target word (read condition) or wrote down their own solution (generate condition) onto the response sheet. Slide display durations and slide order were consistent with Mulligan (2004, Exp. 1). Items in the study blocks were presented randomly.

Distractor Phase

After the study phase, participants worked on the distractor task. Participants completed the names of German cities given a three-letter word stem (e.g., "MAN_____" for "Mannheim"). The logic of this task was adopted directly from Mulligan (2004) and Riefer et al. (2007). Both used this task containing U.S.-cities as stimuli. German city names were obtained from a list of the 188 largest German cities provided by Wikipedia Germany (2008, November 10). Participants worked on a randomised list of these city names. The distractor task lasted for three minutes; the same duration was chosen in the studies by Mulligan and by Riefer and colleagues. Although the city stem completion task was similar to the experimental task in the generate condition to a certain respect, it was used nevertheless, because of the goal to replicate Mulligan's (2004) and Riefer et al.'s (2007) studies as closely as possible. Moreover, Mulligan et al. (2006, Footnote 5) stated that their use of another distractor task in one study did not produce a change in the results. This fact strongly indicates that the use of the city stem completion task constitutes a valid distractor task for the present study.

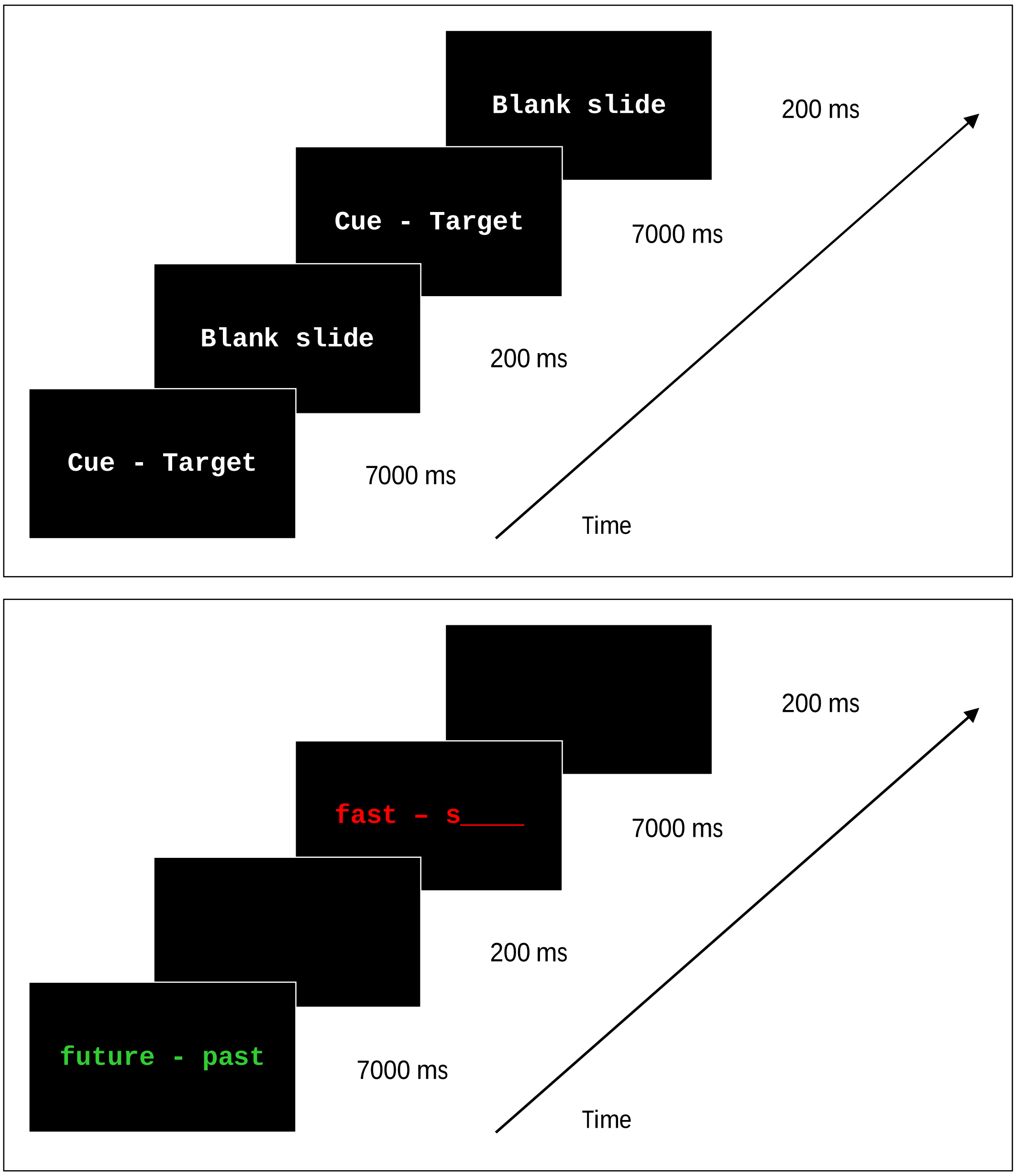

Figure 2.8: Illustration of the sequence of slides for Experiment 1 for two study trials (top) and for specific examples of two study trials (bottom); along with according presentation durations

Test Phase and Post-Experimental Questionnaire

Immediately following the distractor task, participants were tested on their memory for the target items in the test phase. Target item names were presented randomly and sequentially in the middle of the screen and were coloured in white print and displayed against a black background. Participants were informed that half of the items had been presented to them before, while the other half was new and that items would appear for testing randomly. For each trial, participants were required to indicate, which of five categories each test item belonged to and to respond by clicking in one of five fields on the bottom of the screen. An illustration of the test slides can be found in Figure 2.9. Responses in the test phase were self-paced.

At the end of the experiment, participants were given a post-experimental questionnaire and were instructed to report what they had been asked to retain in memory in the experiment. They were probed by the sub-clause "I was asked to remember and was finally tested for ..." before writing down what they recalled.

Participants in the perceptual attributes condition should ideally have answered the following:

"I was asked to memorise and was finally tested to remember...

1. the items themselves
2. whether the items were presented in red or green
3. whether the items were presented complete or incomplete"

Participants in the internal states condition should ideally have answered the following:

"I was asked to memorise and was finally tested to remember...

1. the items themselves
2. whether the items were presented in red or green
3. whether I could simply read the items on the screen or had to generate the words myself from the first letter"

After completion of the experiment, participants were thanked and given their reward. Participants asking to be debriefed, were informed about the purpose of the study.

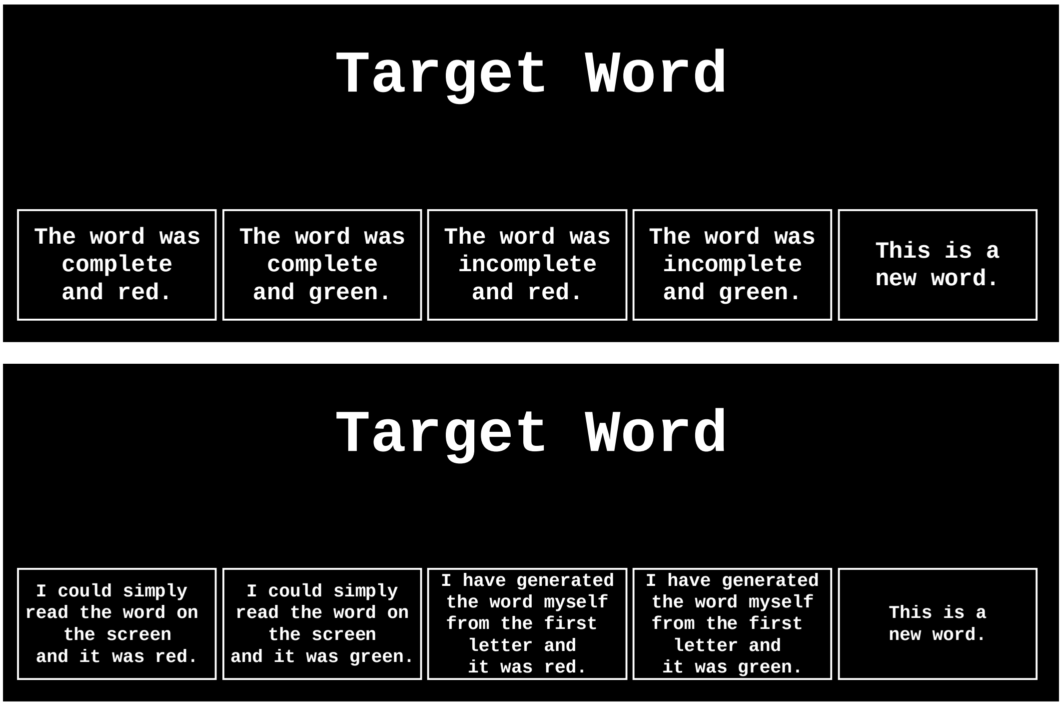

Figure 2.9: Illustration of the test slides for Experiment 1 as displayed in the PA condition (top) and the IS condition (bottom); Note that original sentences were in German and can be found in Appendix C

2.2.2 Results

In the following section, the results of Experiment 1 are described. First, an overview is presented over two possibilities of how the collected data could be analysed: traditional (source memory) measures and the multinomial processing tree model framework. Traditional measures (e.g., corrected hits rates, identification-of-origin scores) of item and source memory have been used most often in the generation effect literature and have typically been tested for statistical significance within the ANOVA framework. These model-free analyses are contrasted with the use of model-based analyses within the multinomial processing tree model framework. Consequently, advantages and disadvantages of both approaches, for example concerning response biases, are discussed. Eventually reasons are presented for why the model-based alternative was preferred in analysing the current data. Then, the presently employed multinomial-processing tree model (i.e., the model for crossed source information) is described along with the data structure found in the experiment. Finally, model fits are reported and commentated, before the experimental hypotheses are subsequently tested.

Traditional Measures and Model-based Alternative

Commonly used traditional measures in generation effect studies are corrected hits rates and identification-of-origin (IDO) scores (cf., Murnane & Bayen, 1996). Corrected hits rates are obtained by subtracting false alarm rates from hits rates. IDO scores represent the percentage of items for which source was remembered correctly, when an item was identified correctly as being old. Scores are typically subjected to (repeated measures multifactorial) ANOVAs. Most papers on the effect of self-generation report these traditional measures.

However, a number of theorists have criticised the validity of these scores and have denoted them as problematic empirical measures for the purpose of accurately assessing source memory performances. Since source memory and the generation effect are the exact topic of this dissertation, high validity of source memory measures is critical.

For example, Batchelder and Riefer (1990) and Murnane and Bayen (1996) have pointed out that IDO scores are conditional on correct item recognition, a fact that may render these scores difficult to interpret for cases in which recognition rates *differ across conditions*. This is of course true in any generation effect study that can actually find the positive generation effect in item memory. Here, recognition rates differ between the read condition and the generate condition. This strong, that is significant, difference in the recognition rates between read and generated items *is per definition* the generation effect. Other researchers (e.g., Rabinowitz, 1990; Riefer, Hu, & Batchelder, 1994) have warned about response biases in the data, which might lead to problems in interpreting IDO scores. Strong response biases could occur when, for example, new items were falsely indicated as being "old", and when the probability of identifying these items as "complete / simply read" was significantly higher than the probability of identifying these items as "incomplete / self-generated", and vice versa. In previous generation effect studies, a strong response bias was found in favour of complete items (e.g., Rabinowitz, 1990; Riefer et al., 1994; Riefer et al., 2007; Voss et al., 1987).

Multinomial processing tree (MPT) models constitute a prominent alternative to traditional memory measures. MPT models are stochastic models that explain categorical data by a sequence of latent states that can be interpreted as (psychological) processes. Moreover, these *models* belong to the family of *confirmatory* methods and comprise a sequence of *latent states*. More technically speaking, MPT models are defined by (1) a vector of S parameters, (2) a set of category probabilities (p_j), and (3) model-implied branch probabilities (p_{ij}). Moreover, they employ the expectation-maximization algorithm for parameter estimation. For more information on the mathematical background of MPT modelling, see Riefer and Batchelder (1988) and Batchelder and Riefer (1990). A thorough description of applications of MPT modelling can be found in Batchelder and Riefer (1999) and in Erdfelder, Auer, Hilbig, Aßfalg, Moshagen, and Nadarevic (2009).

The circumstance that MPT models have been applied to generation effect research only scarcely seems surprising, when considering their advantages for memory research. MPT models offer the possibility of *separately* modelling and illustrating (a) *pure memory performances* and (b) other factors influencing these pure performances such as *response biases* or *guessing strategies*. Separate parameters can be introduced, which reflect biases on the one hand and memory unaffected by these biases on the other hand.

A separation of biased response tendencies and unbiased responses is of course only relevant for cases, i.e. studies or conditions within studies, that are prone to biases or guessing strategies. It is well documented that these types of distortions can have a measurable effect in experiments investigating source memory (Johnson & Raye, 1981; Johnson et al., 1981). A number of researchers have pointed out that identification-of-origin scores are especially problematic, because they are confounded by response biases and other cognitive factors and therefore are no pure measures of source memory (see Batchelder & Riefer, 1990 and Murnane & Bayen, 1996, for a theoretical and empirical discussion on this issue). In the study by Riefer et al. (2007), it was even found that response biases play a crucial role also for the generation effect itself: When Riefer and colleagues calculated statistical significance tests for IDO scores for degree of completeness, they could *not* find a significant difference between self-generated and read items: For IDO scores a *null effect* emerged. In contrast, when Riefer et al. (2007) used an MPT model instead, a *significant* positive generation effect was revealed.

Generally speaking, MPT models are good solutions to disentangle influences of memory processes and response biases, because they are mathematically tractable and are capable of separately measuring diverse cognitive processes and parameters within a model. Thus, the following conclusion can be drawn, which was explicitly stated by Bröder and Meiser (2007):

> Since the machinery of multinomial modelling is well-developed in theory and implemented in comfortable analysis software, we recommend that the common practice of using traditional surface measures be abandoned and that the validated models be used instead or at least as supplemental analyses to disambiguate the interpretation of the descriptive measures. (p. 57)

Data Structure, Actual Model in Use, and its Adaptations to the Current Experiment

Due to potential problems arising from the use of the IDO scores for source memory and thanks to theoretical advantages of this new type of analyses, data were analysed using a multinomial processing tree model. More specifically, the model for crossed source information developed by Meiser and Bröder (2002) was employed. It has been validated empirically as a useful model when two source dimensions are crossed. In the present case the factors “colour” and “degree of completeness” were crossed. This created four types of studied items: red items that were complete, red items that were incomplete, green items that were complete, and green items that were incomplete. Also there were unstudied (i.e., new) items. This resulted in a 5 x 5 response data table. Additionally, for the present experiment, the factor “instruction” played the critical role.

Figure 2.10 shows the basic trees of the Meiser and Bröder (2002) model as applied to the current experiment. For this version of the model, “colour” (red vs. green) is the first source dimension and is crossed with the factor “degree of completeness”, the second source dimension. Source memory or source retrieval processes for the two dimensions are assumed to be stochastically independent. The figure represents the basic trees only, i.e. the basic tree for old items plus the tree structure for new items. However eight separate trees were modelled, one for each of the types of old items. Equally, for the set of parameters described above, a different set of these parameters is needed for each item type. This results in 56 (8x7) old item parameters plus 4 new item parameters. Since there are 36 degrees of freedom (9x4) in the data structure and in the MPT model, such a model would be overidentified. Hence, restrictions have to be set.

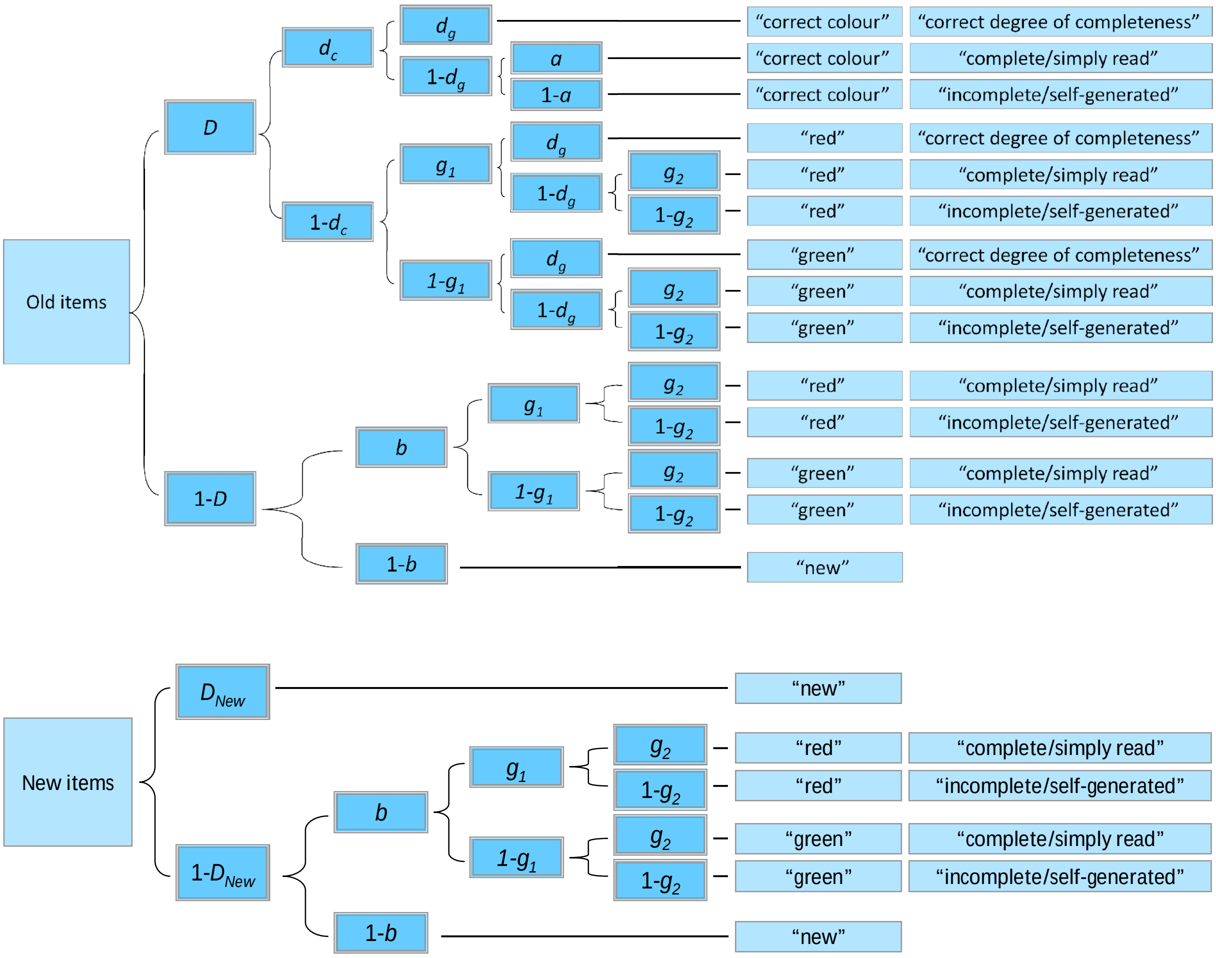

Figure 2.10: Illustration of Meiser and Bröder's (2002) multinomial processing tree model for crossed source information, when implementing colour (red vs. green) crossed with degree of completeness (complete vs. incomplete). Parameters in the model are: D = the probability of correctly detecting a previously presented item as old; D_{New} = the probability of identifying a new item as new; d_c = the probability of correctly discriminating red versus green items; d_g = the probability of correctly discriminating read versus self-generated items; a = the probability of responding "complete / simply read" for detected items, when colour is correctly identified but degree of completeness is not; g_1 = the probability of responding "red", when participants fail to remember the colour of an item; g_2 = the probability of responding "complete / simply read", when participants guess the colour of an item; and b = the probability of responding "old" to non-detected items

In line with Riefer et al. (2007), I applied the following assumptions to make the model identifiable: First, it was assumed that all guessing and response-bias parameters were equal across types of stimuli. Thus, the set of guessing parameters was restricted to guessing that an item was old (*b*), guessing that an item was red (g_1), and guessing that an item was complete (*a*, g_2). Second, it was assumed that there were no a-priori memory differences for the red versus green items; this was argued to be true for the item memory parameter *D* as much as for both source memory parameters: the source memory parameter for colour d_c and the source memory parameter for degree of completeness d_g. Third, it was assumed that the factor "instruction" was only relevant for the source memory parameter for the degree of completeness d_g. Parameters *D*, d_c, and the guessing and response bias parameters were theorised to not be affected by this experimental manipulation. Applying these assumptions, lead to a final set of 13 parameters for the current model (see also Table 2.4).

Fourth, an assumption was made concerning item detection of new items. The parameter D_{New} represents the probability of identifying a new item as new and therefore reflects a high-threshold assumption in the model. Different restrictions can be placed onto this parameter to ensure identifiability of the model. Setting D_{New} equal to 0 or setting D_{New} equal to another item memory parameter in the model ($D_{Complete}$ or $D_{Incomplete}$) constitute possible variants. In their original model, Batchelder and Riefer (1990) set D_{New} equal to 0; which is referred to as the one-high threshold solution. In contrast, Bayen, Murnane, and Erdfelder (1996) have advocated the use of two-high threshold models. For these models D_{New} is non-zero and is rather set equal to the value of *D* for one of the old items. Meiser and Bröder (2002) also used the two-high threshold assumption for their model. However, the question remains whether D_{New} should be set equal to $D_{Complete}$ or to $D_{Incomplete}$ for the current work.

For analyses, all three assumptions ($D_{New} = 0$, $D_{New} = D_{Complete}$, and $D_{New} = D_{Incomplete}$) were explored to determine which one was most appropriate for the according data set and to see which model provided a satisfactory fit to the data. Model fits for the two sets of data (all data vs. data for correctly copied study words) and under the three restrictions of D_{New} ($D_{New} = 0$, $D_{New} = D_{Complete}$, $D_{New} = D_{Incomplete}$) are displayed in Table 2.5. Subsequently, descriptive and statistical hypothesis tests are outlined. Analyses of the data and for all experiments reported in this work were conducted with the "MultiTree" computer program developed by Moshagen (2010).

Table 2.4: List of parameter names, their function within the model, and a description of their interpretation used in Experiment 1

Parameters in the final model used for analyses in Experiment 1		
Name	Type of parameter in model	Description
$D_{Complete}$	Item memory	Probability of correctly detecting a previously presented complete item as old
$D_{Incomplete}$	Item memory	Probability of correctly detecting a previously presented incomplete item as old
D_{New}	Item memory	Probability of identifying a new item as new
a	Guessing	Probability of guessing "complete"/"simply read" when colour is identified correctly
b	Guessing	Probability of guessing "old" for non-detected items
g_1	Guessing	Probability of guessing "red" when colour is not known
g_2	Guessing	Probability of guessing "complete"/"simply read" when colour is guessed
$d_{c\ Complete}$	Source memory colour	Probability of correctly recalling colour of a complete item
$d_{c\ Incomplete}$	Source memory colour	Probability of correctly recalling colour of an incomplete item
$d_{g\ Complete\ PA}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of a complete item under PA instructions
$d_{g\ Complete\ IS}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of a complete item under IS instructions
$d_{g\ Incomplete\ PA}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of an incomplete item under PA instructions
$d_{g\ Incomplete\ IS}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of an incomplete item under IS instructions

Goodness-of-fit Statistics and Estimated Parameter Values

During the study phase, 99.77 % of the complete words were copied correctly, and 98.79 % of the incomplete words were generated correctly (see Appendix D, for further details). Thus, percentages of correctly copied or correctly generated items were high; comparably high and even higher than those reported by Riefer et al. (2007) – 92.5 % generated correctly and 99.9 % copied correctly – and Mulligan (2004) – between 86 % and 99.5 % generated correctly and 100 % copied correctly.

Two types of data sets were analysed applying the log-likelihood ratio statistic G^2. First, all data were considered for MPT analysis and second only correct data, that is data obtained for items that were named correctly in the study phase, were considered for MPT analysis. Model fits were calculated separately for the data sets by setting D_{New} either to 0, to $D_{Complete}$, or to $D_{Incomplete}$. For a summary of the results, see Table 2.5.

Table 2.5: List of G^2 and p values for Experiment 1

Data set global model for	Restriction on D_{New} (df = 24)	G^2	p	Critical G^2 for α = 5 %	Critical G^2 for α = 1 %	Fit on
All data (1)	$D_{New} = 0$	39.99	0.02	36.42	42.98	1 % level
	$D_{New} = D_{Complete}$	39.05	0.02			1 % level
	$D_{New} = D_{Incomplete}$	36.51	0.04			1 % level
Data for items named correctly (2)	$D_{New} = 0$	39.98	0.02	36.42	42.98	1 % level
	$D_{New} = D_{Complete}$	38.99	0.02			1 % level
	$D_{New} = D_{Incomplete}$	36.19	0.05			1 % level

The model fit the data well (on the α = 1 % level). When comparing the goodness-of-fit statistics for the two separate data sets, no considerable difference could be found. This result makes sense when taking into account the low error rates for naming in the study phase. The best G^2 value for both data sets was obtained for setting D_{New} equal to $D_{Incomplete}$. To conclude, since model fit was similar across these data sets as well as for different restrictions put on D_{New}, the following analyses were based on all data by restricting D_{New} to $D_{Incomplete}$. Using the first set instead of the second increased statistical power.

Note that all analyses and hypothesis tests were additionally conducted for the remaining data sets and restrictions listed in Table 2.5. They qualitatively yielded the same results concerning both descriptive data and statistical significance tests.

Hypothesis Testing

As for the goodness-of-fit tests, the G^2 statistic was applied for hypothesis testing; for all of these tests in this dissertation, the 5 % significance level was employed. Parameter tests can be applied using the G^2 difference statistic ΔG^2, which is asymptotically χ^2 distributed. For tests involving one degree of freedom (*df*), which was the case for the majority of parameter tests, the critical value was 3.84. Comparisons of two parameters generate 1 *df*. Note that *p*-values smaller than 0.05 indicate that the imposed restrictions are incompatible with the data. Thus, the hypothesis implied in the restrictions must be rejected. Results are organised around answering the hypotheses and predictions.

According to *Hypothesis 1*, an advantage of incomplete over complete items was expected regarding item memory. Thus, a positive generation effect should have appeared.
As can be seen in Figure 2.11, a strong positive generation effect emerged for item detection; incomplete words were recognised correctly more often than complete words ($D_{Incomplete}$ = 0.69, $D_{Complete}$ = 0.23). This difference was statistically significant (ΔG^2_{1df} = 203.00, $p < 0.001$). Basically, the positive generation effect found in previous studies could be replicated in the current experiment.

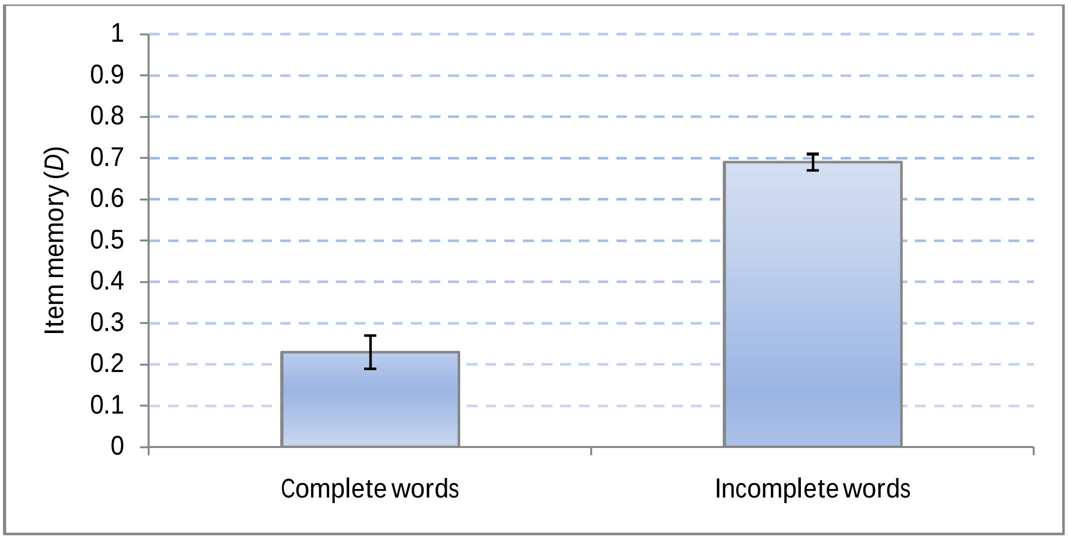

Figure 2.11: Item memory performances in Experiment 1 - error bars represent standard errors

According to *Hypothesis 2*, regarding source memory for colour, an advantage of complete over incomplete items was expected. Thus, a negative generation effect should have appeared.

As can be seen in Figure 2.12, a strong negative generation effect emerged for source discrimination for the source dimension colour; colour of complete words was remembered correctly more often than colour of incomplete words ($d_{c\ Complete}$ = 0.68, $d_{c\ Incomplete}$ = 0.19). This difference was statistically significant (ΔG^2_{1df} = 23.41, p < 0.001). Basically, the negative generation effect found in previous studies for the source memory dimension colour could be replicated in the current experiment.

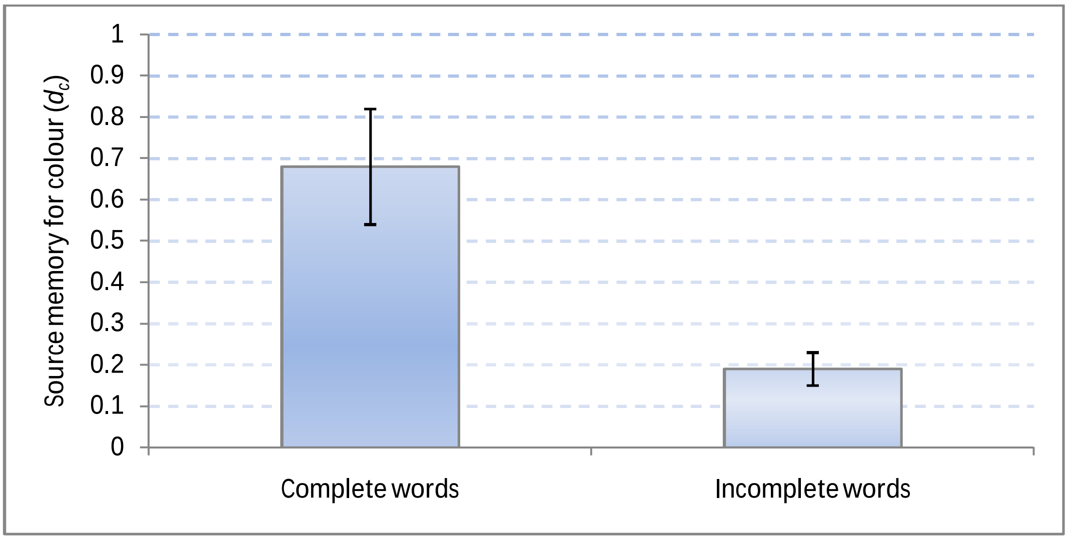

Figure 2.12: Source memory performances for the dimension colour in Experiment 1 - error bars represent standard errors

According to *Hypothesis 3*, regarding source memory for the degree of completeness, an interaction between the factors "instruction" and "degree of completeness" was expected. A positive generation effect should have appeared in the perceptual attributes condition, whereas a negative generation effect should have appeared in the internal states condition.

As can be seen in Figure 2.13, no interaction emerged between degree of completeness and type of instruction. Instead, degree of completeness of complete words ($d_{g\ Complete\ PA}$ = 0.77 and $d_{g\ Complete\ IS}$ = 0.74) was correctly remembered more often than degree of completeness of incomplete words ($d_{g\ Incomplete\ PA}$ = 0.62 and $d_{g\ Incomplete\ IS}$ = 0.56). Nevertheless, the differences between the complete and the incomplete words in the two types of instruction conditions were not statistically significant (ΔG^2_{1df} = 0.05, *p* = 0.83 for $d_{g\ Complete\ PA} = d_{g\ Complete}$

$_{IS}$; $\Delta G^2_{1df} = 1.20$, $p = 0.27$ for $d_{g\ Incomplete\ PA} = d_{g\ Incomplete\ IS}$. Moreover, setting all parameters for the degree of completeness equal ($d_{g\ Complete\ IS} = d_{g\ Complete\ PA} = d_{g\ Incomplete\ IS} = d_{g\ Incomplete\ PA}$) did not produce a significant increase in G^2 ($\Delta G^2_{3df} = 1.67$, $p = 0.64$). Parameters representing source memory performance for degree of completeness did not differ significantly.

Basically, no influence of the experimental factor "instruction" could be found. Instead, values reflecting performance levels for all four parameters were comparable at the descriptive level - they are not significantly different. Neither the variation of complete versus incomplete words nor the experimentally induced processing of perceptual attributes and processing of internal states, lead to a significant variation in the data pattern.

However, statistical power of these tests was low: Power for 1 *df* tests were only at 6 % (test of $d_{g\ Complete\ PA} = d_{g\ Complete\ IS}$, effect size $w < 0.010$, $\alpha = 0.05$), and at 19 % (test of $d_{g\ Incomplete\ PA} = d_{g\ Incomplete\ IS}$, $w = 0.02$, $\alpha = 0.05$), and power for a 3 *df test* was only at 16 % (test of $d_{g\ Complete\ IS} = d_{g\ Complete\ PA} = d_{g\ Incomplete\ IS} = d_{g\ Incomplete\ PA}$, $w = 0.02$, $\alpha = 0.05$). Therefore, it cannot be ruled out that it is actually false to retain the above reported null effect. Results should consequently be treated with caution.

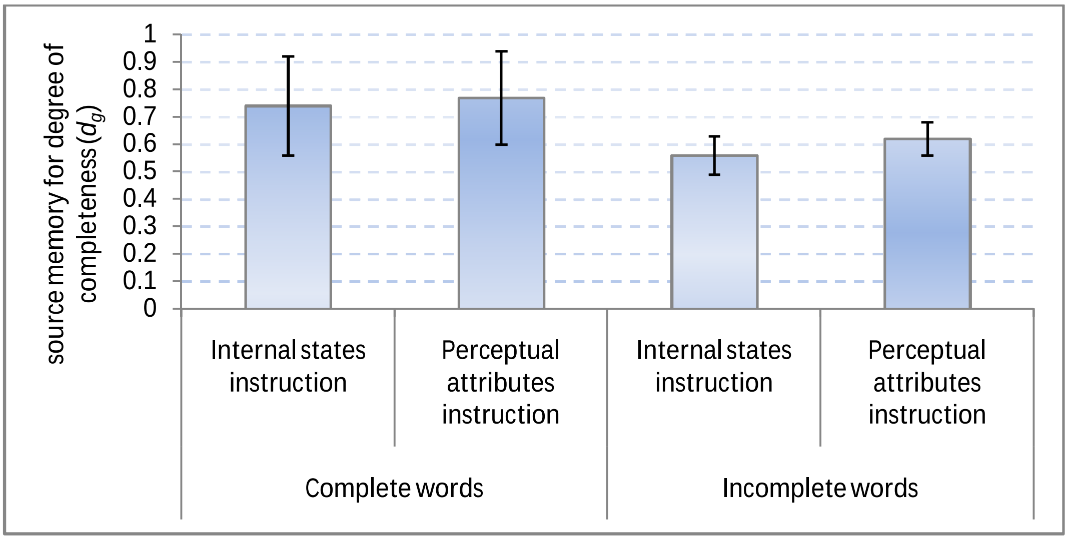

Figure 2.13: Source memory performances for the dimension degree of completeness in Experiment 1 - error bars represent standard errors

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.

Since these parameters would ideally reflect true guessing, it was tested whether the parameters differed significantly from chance (i.e., from $p = 0.50$). Guessing "old", guessing "red", and guessing "complete", when colour was identified correctly, was not significantly different from chance level; Parameters *b* (0.45), g_1 (0.49), and *a* (0.47) could be set equal to 0.50 (*b*: $\Delta G^2_{1df} = 3.10$, $p = 0.08$; g_1: $\Delta G^2_{1df} = 0.38$, $p = 0.54$; *a*: $\Delta G^2_{1df} = 0.01$, p = 0.92;). However, guessing “complete”, when colour was guessed ($g_2 = 0.67$), was significantly above chance ($\Delta G^2_{1df} = 17.29$, p < 0.001).

The Analysis of the Postexperimental Questionnaires

At the end of Experiment 1, participants were given a post-experimental questionnaire and were instructed to write down what they had been asked to retain in memory in the experiment. No errors were made concerning item memory and source memory for colour. Errors were committed only concerning the degree of completeness: Three participants responded in a way which did not match the anticipated responses in the internal states condition. Additionally, 6 participants did not fill in the questionnaire at all (4 in the internal states condition; 2 in the perceptual attributes condition). In other words, participants’ responses matched anticipated responses in 84.75 % of all cases (76.67 % for IS and 93.1 % for PA). However, when excluding (a) participants that did not write an anticipated response or when excluding (b) participants that did not fill in the questionnaire or when excluding (c) both groups, no deviations from the previously described result pattern were found.

2.3 Experiment 2

The non-occurrence of a significant effect for the factor "instruction" in Experiment 1 may be due to a weak manipulation and in consequence due to a small effect size. So far, the goal was to show the expected effect by *replicating* the experiments by Mulligan (2004) and Riefer et al. (2007). In fact, the instructions were modelled after their formulations. Therefore the manipulation for the factor "instruction" could be regarded as subtle – so subtle even that the effect was simply not strong enough to produce a statistically significant difference.

Consequently, I applied several changes to the design and the procedure of Experiment 2, which were aimed at maximising the effect of the independent variable "instruction". I identified three areas, which could be improved to achieve this goal: (1) *relevance* of the suggested type of processing, (2) *acquaintance* with the experimental task (i.e., with the type of processing), (3) *emphasis* of actually intended type of *processing* at test.

First, it is possible that participants in Experiment 1 did not comply with the experimental instructions. This might have been due to their own previously acquired study strategies that had proven to work well and were readily accessible outside of the context of the experiment. Therefore, it seemed important to *decrease* the possibility of using other study strategies than those suggested in the instructions. In contrast, it was very important to *increase* the relevance of actually using the encoding processes suggested in the experimental instructions. This was done in Experiment 2 by telling participants that previous studies had examined how to encode this type of material best and that the instructions they were about to see were confirmed to be most useful. Participants were asked to follow instructions carefully, since the alleged goal of the current study would be to replicate the previous findings. Although these cover-story-like instructions were most relevant at study, they were repeated in the test phase of the experiment.

Second, participants in Experiment 1 had to fulfil the experimental task straight after having read all instructions. It is possible that participants were not fully acquainted with the task from the start and therefore needed practice trials to get used to it. Also, participants could not stop or pause the presentation of study items and pose additional questions once the study phase had begun. Although much time was allocated to reading the original task instructions and although participants had many opportunities to ask questions beforehand, it is still possible that some of them did not fully comprehend the task and that they consequently could not show their full potential at study and at test. Accordingly, a practice phase and a repeated presentation of instructions after the practice trials were introduced in Experiment 2, before starting the actual study phase.

The third improvement pertained to the test phase. At test, participants were presented with five response options. In Experiment 1, these five options were presented on the screen at the same time *in a fixed order*. It could therefore be possible that rephrasing (according to previous general study experience, see first point) could have taken place. This is problematic, because rephrasing might not have matched the actual type of processing intended in the experimental instructions. Instead, participants could have rephrased test options into any other cue that they found helpful for (successful) completion of the task. To avoid rephrasing and to enforce the use of actually intended processing, the position of response options was chosen randomly in Experiment 2. This means that participants had to actively search for and thereby had to actively read through all of the response options for each trial. This manipulation was expected to minimise rephrasing, which would then no longer be an easier but rather a more difficult and more strenuous strategy.

2.3.1 Methods

Design, material, and procedure were the same as in Experiment 1, except for variations that arose from differences in the strength of the factor “instruction”, which have already been mentioned above.

Participants

Forty-two persons participated in Experiment 2. One person had to be excluded from the sample due to red-green colour blindness. Of the 41 participants remaining in the sample, 80.49 % were female. Participants' age ranged from 19 to 32 with a mean of 22.37 years and a median of 22 years. All participants but three confirmed that they were native German speakers. However, this latter group of participants stated that their German proficiency skills were "very good". All participants were currently enrolled at the University of Mannheim and were students from the fields of Psychology (53.66 %), Economic Sciences (21.95 %), Sociology (14.63 %), or Language Studies (9.76 %).

2.3.2 Results

Due to potential problems of traditional model-free measures, the MPT model for crossed source information developed by Meiser and Bröder (2002) was employed. It has been validated empirically when source dimensions are crossed. The actual model in use and its adaptations were the same as in Experiment 1. In the present experiment, two separate source dimensions were present, namely "colour" and "degree of completeness", which created four types of studied items: complete and incomplete red items, complete and incomplete green items. Also there were unstudied (i.e., new) items. Additionally, for the present experiment, the factor "instruction" played the critical role. This between-subjects factor also needed to be taken into account when analysing types of items and responses given to these items.

During the study phase, 99.78 % of the complete words were copied correctly, and 99.46 % of the incomplete words were generated correctly (see Appendix D, for further details). Thus, percentages of correctly copied or generated items were high and comparable to Riefer et al. (2007) and Mulligan (2004). Two types of data sets were analysed. First, all data were considered for MPT analysis and second only correct data, that is data for items that were named correctly in the study phase, were considered for MPT analysis. Model fits were calculated separately for the data sets by setting D_{New} either to 0, to $D_{Complete}$ or to $D_{Incomplete}$. For a summary of the results, see Table 2.6.

The model fit the data well (on the α = 5 % or the α = 1 % level). When comparing G^2 for the two separate sets of data, no considerable differences could be found. This result makes sense when considering the low error rates for naming in the study phase. Different model solutions for the parameter D_{New} also rendered comparable results. The best G^2 values for both data sets (all data and correct data) were obtained for setting D_{New} equal to 0.

Table 2.6: List of G^2 and p values for Experiment 2

Data set global model for	Restriction on D_{New} (df = 24)	G^2	p	Critical G^2 for α = 5 %	Critical G^2 for α = 1 %	Fit on
All data (1)	$D_{New} = 0$	35.77	0.06			5 % level
	$D_{New} = D_{Complete}$	35.93	0.06	36.42	42.98	5 % level
	$D_{New} = D_{Incomplete}$	36.60	0.05			1 % level
Data for items named correctly (2)	$D_{New} = 0$	35.12	0.07			5 % level
	$D_{New} = D_{Complete}$	35.27	0.07	36.42	42.98	5 % level
	$D_{New} = D_{Incomplete}$	35.90	0.06			5 % level

To summarise, since model fits were similar, when comparing all data and correct data only and when comparing the different restrictions put on D_{New}, the following analyses were based on all data restricting D_{New} to 0. Using the first set instead of the second increased statistical power. All analyses were conducted for the second data set and for all other restrictions as well resulting in the same pattern concerning the descriptive data. Statistical significance tests showed slight deviations from the pattern reported. However, these deviations did not influence the general conclusions, which are outlined below.

Hypothesis Testing

According to *Hypothesis 1*, an advantage of incomplete over complete items was expected regarding item memory. Thus, a positive generation effect should have appeared.

As can be seen in Figure 2.14, a strong positive generation effect emerged for item detection; incomplete words were remembered more often than complete words ($D_{Incomplete}$ = 0.78, $D_{Complete}$ = .46). This difference was statistically significant (ΔG^2_{1df} = 160.03, p < 0.001). Basically, the positive generation effect found in previous studies, could be replicated in the current experiment.

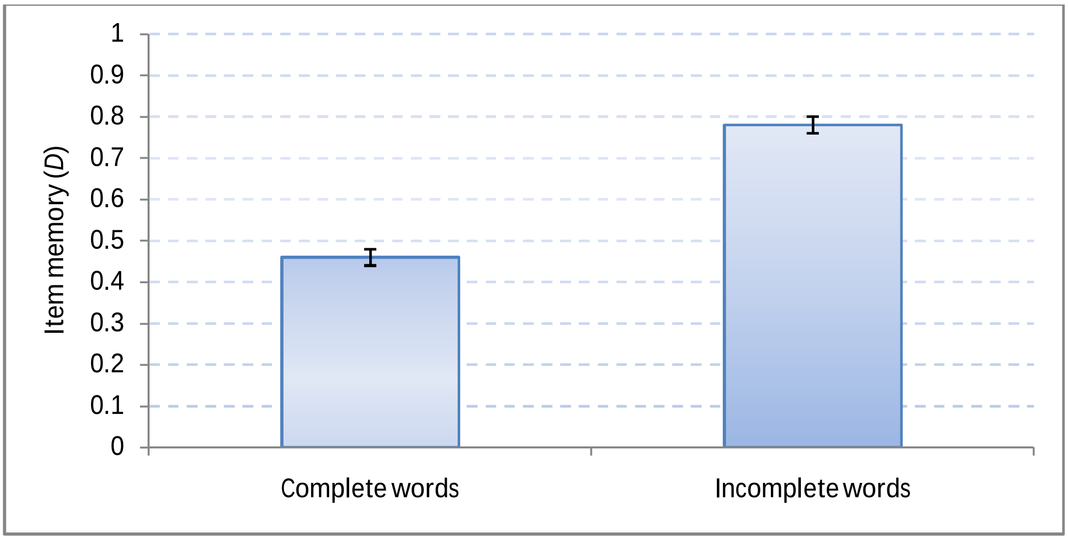

Figure 2.14: Item memory performances in Experiment 2 - error bars represent standard errors

According to *Hypothesis 2*, regarding source memory for colour, an advantage of complete over incomplete items was expected. Thus, a negative generation effect should have appeared.

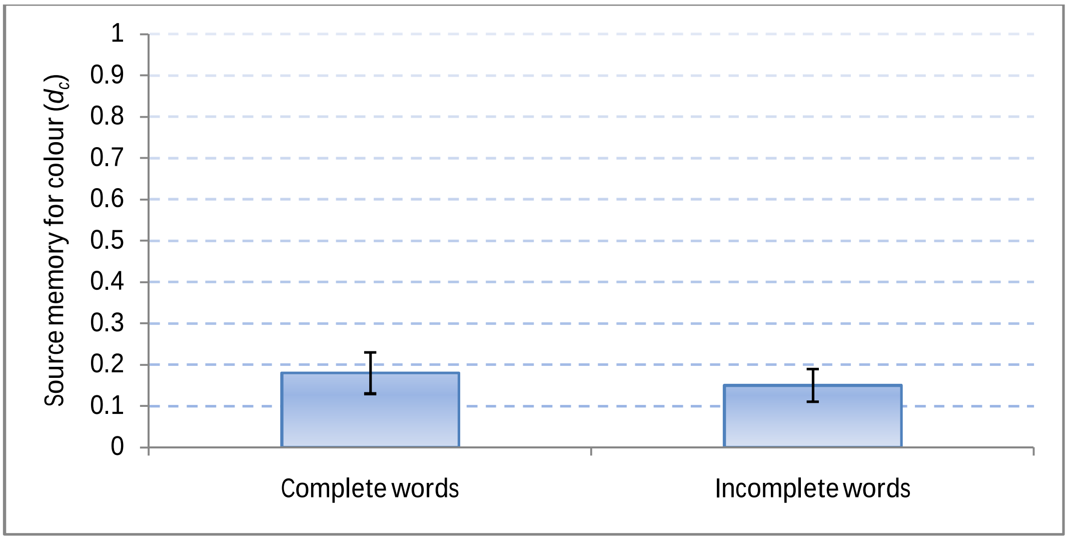

Figure 2.15: Source memory performances for the dimension colour in Experiment 2 - error bars represent standard errors

As can be seen in Figure 2.15, source memory for the source dimension colour was superior for complete words; colour of complete words was remembered correctly more often than colour of incomplete words ($d_{c\ Complete}$ = 0.18, $d_{c\ Incomplete}$ = 0.15). However, this difference was not statistically significant (ΔG^2_{1df} = 0.26, p = 0.61). Basically, the negative generation effect found in previous studies for the source memory dimension colour could only be replicated at the descriptive level in the current experiment.

Statistical power for this test was high (above 90 % for $\alpha = 0.05$). However, the influence of a floor effect, which could obscure a real effect, has to be considered as well. For this specific case, it seems reasonable to assume the existence of a negative generation effect, since a negative generation effect for the source dimension colour has been theorised and found repeatedly in the literature.

According to *Hypothesis 3*, regarding source memory for the degree of completeness, an interaction between the factors “instruction” and “degree of completeness” was expected. A positive generation effect should have appeared in the perceptual attributes condition, whereas a negative generation effect should have appeared in the internal states condition.

As can be seen in Figure 2.16, no interaction emerged between degree of completeness and type of instruction. Instead, for both types of instruction a positive generation effect could be found at the descriptive level; degree of completeness of incomplete words ($d_{g\ Incomplete\ PA}$ = 0.49 and $d_{g\ Incomplete\ IS}$ = 0.47) was correctly remembered more often than degree of completeness of complete words ($d_{g\ Complete\ PA} < 0.001$ and $d_{g\ Complete\ IS}$ = 0.03).

Comparing complete and incomplete words separately for the two types of instructions rendered significant results. The positive generation effects were statistically significant in both conditions ($\Delta G^2_{1df} = 7.14$, p = 0.01 for setting $d_{g\ Complete\ PA}$ equal to $d_{g\ Incomplete\ PA}$; ΔG^2_{1df} = 5.46, $p = 0.02$ for setting $d_{g\ Complete\ IS}$ equal to $d_{g\ Incomplete\ IS}$).

Moreover, there was no significant decrease in model fit when setting parameters reflecting source memory degree of completeness for complete words ($d_{g\ Complete\ PA} = d_{g\ Complete\ IS}$) equal and when setting source memory degree of completeness for incomplete words ($d_{g\ Incomplete\ PA} = d_{g\ Incomplete\ IS}$) equal: $\Delta G^2_{2df} = 0.15$, $p = 0.93$.

Statistical power for setting $d_{g\ Complete\ PA}$ equal to $d_{g\ Complete\ IS}$ and for setting $d_{g\ Incomplete\ PA}$ equal to $d_{g\ Incomplete\ IS}$ was at 16 % ($w = 0.02$; 2 *df*; $\alpha = 0.05$). This low value indicates the possibility of true differences between the parameters for complete and between the parameters for incomplete words.

Overall, these two further restrictions lead to the creation of a model that fit well on the α = 5 % level for a critical value of 38.885: $G^2_{26df} = 35.92$, $p = 0.09$. In this model, the positive generation effect re-emerged showing a memory advantage for source memory degree of completeness for incomplete over complete words ($d_{g\ Incomplete}$ = 0.48; $d_{g\ Complete}$ < 0.001). This difference was statistically significant (ΔG^2_{1df} = 7.12, p = 0.01).

Basically, no influence of the experimental factor "instruction" could be detected. Instead, a positive generation effect for the degree of completeness was present within both types of instructions and when aggregating across types of instructions.

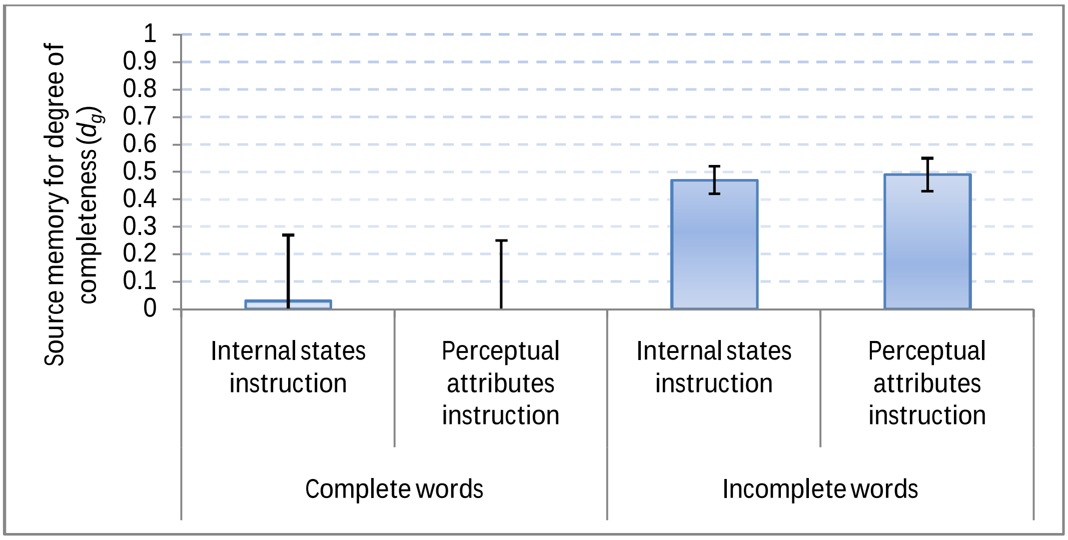

Figure 2.16: Source memory performances for the dimension degree of completeness in Experiment 2 - error bars represent standard errors

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.

Since these parameters would ideally reflect true guessing, it was tested whether the parameters differed significantly from chance (p = 0.50). All guessing parameters were significantly different from chance level. Guessing "red" ($g1$ = 0.54) and guessing "complete" (a = 0.80; g_2 = 0.71) were higher than 0.50; g_1: ΔG^2_{1df} = 5.11, p = 0.02; a: ΔG^2_{1df} = 5.23, p = 0.02; g_2: ΔG^2_{1df} = 17.33, p < 0.001. However, there was no significant difference between the two types of guessing "complete" parameters, that is, when colour was identified correctly and when colour was guessed: $a = g_2$: ΔG^2_{1df} = 0.46, p = 0.45. Guessing "old" (b = 0.14) was significantly below chance level: ΔG^2_{1df} = 257.97, p < 0.001.

Participants in this study showed several biases: First, there was a tendency to judge words to have been complete more often than chance, independent of whether colour was identified correctly or was guessed. Second, there was a slight tendency to judge words to have been red. And third, there was a bias (and thus an underestimation) to judge words to have been old.

The Analysis of the Postexperimental Questionnaires

At the end of Experiment 2, participants were given a post-experimental questionnaire and were instructed to report what they had been asked to retain in memory in the experiment. No errors were made concerning item memory and source memory for colour. Errors were committed only concerning the degree of completeness: Seven participants filled in the questionnaire in a way which did not match perfectly the anticipated responses (5 in the internal states condition; 2 in the perceptual attributes condition). Additionally, 4 participants did not fill in the questionnaire at all (1 in the internal states condition; 3 in the perceptual attributes condition). In other words, participants' responses matched anticipated responses in 73.17 % for IS and 78.26 % for PA). However, when excluding (a) participants that did not write an anticipated response from analyses or when excluding (b) participants that did not fill in the questionnaire or when excluding (c) both groups, no deviations from the previously described result patterns were found.

2.4 Discussion of Chapter 2

Chapter 2 is concerned with the role of the processing of perceptual attributes and the processing of internal states in memory for source, i.e., in memory for the degree of completeness, for studies employing a generation effect paradigm. To reiterate, superior item or source memory for incomplete items as compared to that for complete items is referred to as a positive generation effect, whereas superior item or source memory for complete items as compared to that for incomplete items is referred to as a negative generation effect.

Within a generation effect paradigm, memory for source attributes of an item can be studied in different ways. When considering the processing account by Mulligan (2004) and the dual-hypothesis by Riefer et al. (2007), contradictory predictions can be found for the source memory dimension degree of completeness. According to the processing account, a negative generation effect should occur, whereas according to the dual-hypothesis, a positive generation effect is theorised.

To overcome these contradictions, I suggested to consider two processing modes: processing of and memory for (a) perceptual attributes (PA) and (b) internal states (IS). These types of processing can, in general, be applied to diverse source attributes, but were currently employed for the degree of completeness only.

For this case of degree of completeness, the PA condition matches the external source monitoring paradigm (ESMP) and the IS condition matches the reality monitoring paradigm (RMP). ESMP and RMP were outlined thoroughly by Riefer et al. (2007). In fact, the source memory dimension of the factor “degree of completeness” can be turned from a reality monitoring design into an external source-monitoring design. This can be achieved by manipulating encoding and recall processes via the use of different experimental instructions (that focus either on the internal vs. external dimension, or on the self vs. other dimension). Here, applying either of the two designs, that is differentiating between internal and external sources (RMP or IS) or between different perceptual looks (ESMP or PA), are like two sides of the same coin for the degree of completeness.

Consequently, two experiments were set up, in which memory for the degree of completeness was studied. The crucial independent variable was the factor “instruction”, which had two levels: (a) “ the processing of perceptual attributes” (“PA”) and (b) “the processing of internal states” (“IS”). In Experiment 1, instructions followed closely the wordings employed by Mulligan (2004) and Riefer et al. (2007), whereas, in Experiment 2, instructions of encoding and retrieval conditions were altered to enforce the impact of the independent variable “instruction”.

Predictions were the same for Experiments 1 and 2: First, a positive generation effect was expected in item memory. Second, a negative generation effect was theorised for the source memory dimension colour. Third, an interaction was anticipated between the factors “degree of completeness” and “instruction”: For the PA condition (which parallels the external source monitoring paradigm), a negative generation effect for the source memory dimension “degree of completeness” was expected, whereas for the IS condition (which parallels the reality monitoring paradigm), a positive generation effect for the source memory dimension “degree of completeness” was expected.
Data were analysed within the multinomial processing tree model framework.

Overview of the Most Important Results and Their Interpretations

Item Memory

In both experiments, incomplete words were remembered significantly better than complete words. Hence, the typically found positive generation effect in item memory could be replicated. This finding is in line with the hypothesis phrased for item memory. As stated previously, it is also in line with many previous findings and with the most relevant papers for the present studies, namely with Mulligan (2004) and with Riefer et al. (2007). This is a strong hint that the generation effect design employed in the present experiments is valid.
When comparing item memory performance values for Experiment 1 (complete words = 0.23; incomplete words = 0.69) and Experiment 2 (complete words = 0.46; incomplete words = 0.78), one can see an increased item memory discriminability for the latter experiment.

This increase might have occurred due to the repetition of study instructions and thus due to the stronger emphasis on encoding in Experiment 2. This effect could be regarded as an indirect manipulation check, which proved successful.

Source Memory for Colour

In both experiments, the typically found negative generation effect in source memory for the source dimension colour could be replicated at the descriptive level. However, this effect was only significant in Experiment 1. In Experiment 2, no significant effect emerged, which is likely to be due to a general floor effect, since parameter values were below 0.20 both for complete and for incomplete words. These findings are in line with the hypothesis and with predictions and findings by Mulligan (2004) and Riefer et al. (2007). Also, they are in line with the more general and less elaborate item-source trade-off account.

When comparing performance values for Experiment 1 (complete words = 0.68; incomplete words = 0.19) and Experiment 2 (complete words = 0.18; incomplete words = 0.15), one can see that source memory performance for colour reduces drastically in the latter experiment. The very low overall colour memory performance in Experiment 2 seems illogical at first. However, when taking into account the design differences between Experiments 1 and 2, this decline seems reasonable. In the former experiment, both source dimensions (colour and degree of completeness) were referred to equally with respect to the amount of instructions pertaining to them. The same amount of emphasis was put on both: Participants were instructed equally strongly to encode and later on to recall colour or degree of completeness. In the latter experiment, this previously balanced emphasising of source dimensions was skewed towards a much stronger emphasis on the dimension degree of completeness. Between Experiments 1 and 2, no change in the instructions of memory for colour was made, whereas the crucial variation between the experiments lay exactly in the increased emphasis on degree of completeness. There was an overall quantitative increase (namely more words and sentences used for explaining and for heightening relevance of degree of completeness) along with a qualitative increase (namely increased acquaintance with the experimental task).

This overall comparatively stronger emphasis on degree of completeness and the concomitant comparatively weaker emphasis on colour, could be a possible explanation for the overall decline in source memory performance for colour in Experiment 2.

Source Memory for the Degree of Completeness

No interaction effect between the factors “degree of completeness” and “instruction” could be found for any of the two experiments. These finding are not in line with the hypothesis, according to which an advantage of complete items was expected, when participants had been instructed to encode perceptual attributes of the target words, and according to which an advantage of incomplete items was expected, when participants had been instructed to encode internal states at study. Thus, the interaction hypothesis has to be rejected.

In Experiment 1, no significant differences could be found between any of the parameters, showing no differences in memory performance between complete and incomplete words in any of the two instruction conditions, nor between instruction conditions. Complete words were remembered as well as incomplete words, when participants had been instructed to encode perceptual attributes of the target words. The same thing was true, when participants had been instructed to encode internal states at study. There was a trend at the descriptive level towards an advantage of complete words over incomplete words (i.e., towards a negative generation effect), which would be in line with Mulligan (2004). Furthermore, statistical power of these tests was low. Therefore, it cannot be ruled out that it is actually false to retain the above reported null effect. Results should consequently be treated with caution.

In Experiment 2, no difference between IS and PA condition existed; instead, an overall source memory advantage of incomplete over complete words emerged in the data. The type of instruction did not seem to play a role. In contrast, only a significant influence of the factor “degree of completeness” was present in the data. Source memory for complete words was consistently close to zero, whereas participants reliably exhibited good memory for the degree of completeness of incomplete words (perceptual attributes condition = 0.49; internal states condition = 0.47). This result pattern for the degree of completeness supports Riefer et al. (2007) and the more general and less elaborate item-source enhancement hypothesis. To reiterate, in Riefer et al. (2007), predictions were made for the degree of completeness, while differences in processing of this source attribute were not theorised.

Post-Experimental Questionnaire

To check the critical manipulation for the factor "instruction", namely to check that material was accurately processed, memorised, and remembered in terms of PA or in terms of IS, a post-experimental questionnaire was employed. In this questionnaire, participants were asked to report what they had been instructed to memorise in the experiment.

No errors were made concerning item memory and source memory for colour, whereas errors were only made concerning the degree of completeness. However, due to the inconsistent or highly variable pattern and due to the low overall error rate, no precise conclusions can be drawn from the post-experimental questionnaires. However, it can be pointed out that whether or not data were excluded for participants that committed an error in the questionnaire, did not influence the result pattern. Based on this finding, I assumed that the post-experimental questionnaire in its current form might not be a good indicator of which type of processing was involved. Also, if there was a higher awareness of processing modes brought on by the instruction in Experiment 2, this higher awareness, it seems, was not mirrored in the questionnaire.

However, in the post-experimental questionnaire, participants were asked only to give a self-report. Due to potential problems of self-report data as well as of introspective data (cf. Nisbett & Wilson, 1977), it would have been better to employ a behavioural measure (e.g., to record response times) instead.

Critical Issues and Recommendations for Further Studies

Although it seems reasonable to assume that the present research design fully and unambiguously tested the hypotheses presented, it could have been the case that instructions might have been hard to follow for the participants. To avoid this, one could have excluded instructions for item memory and instructions for source memory for colour altogether. Alternatively, it might have been clearer for participants what to do exactly, when, for instance, each experiment would have been split into three separate experiments or, when (at least) source memory for colour would have been excluded. The decreased values for source memory for colour in Experiment 2 (as compared to values in Experiment 1), for which a stronger emphasis was put on instructions for the degree of completeness, may indicate that this assumption is justified.

The Role of the Processing of Perceptual Attributes and of the Processing of Internal States

In Experiment 1, neither the independent variable “degree of completeness” (“complete” vs. “incomplete”) nor the independent variable “instruction” (“PA” vs. “IS”) seemed to have an effect. Additionally, no interaction between the two variables occurred.

It seems unclear why the independent variable “degree of completeness” did not show an effect for the source dimension degree of completeness, while it produced consistent effects and replications for item memory and for source memory for colour.

Furthermore, it seems unclear why the independent variable “instruction” did not show an effect in the present case, although the employed instructions were modelled to follow the instructions used in Mulligan (2004) and Riefer et al. (2007) – instructions that worked well in their studies. However, it is possible that the instruction manipulation simply was not strong enough. An indication of this may be found in the fact that the independent variable "instruction” produced an effect in memory of degree of completeness in Experiment 2.

In Experiment 1, the possibility of a power problem exists, indicating that truly significant effects might not appear as a result of low statistical power. Evidence for this problem can be seen at the descriptive level in the lower values for incomplete as compared to complete words across instruction conditions, and in smaller standard deviations for incomplete words compared to complete words across instruction conditions. Overall, it has to be stated that the performance level in all conditions is remarkably high. Thus, another interpretation could be an enhanced memory performance due to the overall increased distinctiveness of the dimension degree of completeness. After the experiment, participants consistently orally reported that they found it hardest to remember the degree of completeness as compared to item memory or source memory for colour. Interestingly, the behavioural data do not substantiate this notion. However, there is evidence from Riefer et al. (2007) that performance for remembering whether an item was self-generated or presented by someone else actually exhibited better memory than remembering colour; as is the case in Experiment 1.

In Experiment 2, the independent variable “degree of completeness” had an effect. However, there was no interaction between this independent variable and the independent variable “instruction”. Values for complete words were close to zero, whereas values for incomplete words were at an intermediate level. In both instruction conditions, a positive generation effect appeared – a result that was expected only for items encoded and retrieved under IS instruction. Thus, one can say that irrespective of the actual instruction condition, participants encoded and retrieved words as if participants had been instructed to encode and remember words in terms of whether they had generated the items themselves or they had simply read the items on the screen. Even in the PA condition, processing occurred in line with predictions made for the IS condition.

Nevertheless, why should participants in the PA condition have encoded and retrieved items in terms of their own internal states instead of the items’ perceptual attributes? One possible answer to this question is based on findings gained from research on the self in memory.

The described result pattern could have occurred in this way due to the influence of the self-reference effect in memory (Klein & Loftus, 1993; Rogers et al., 1977; Symons & Johnson, 1997). The self has been used in psychology as a central element when explaning various phenomena (e.g., self-serving bias, medical student syndrome) and in diverse research areas (e.g., emotion, motivation, and memory). The term self-reference effect in memory denotes self-referent encoding strategies yielding superior memory related to, for example, semantic encoding strategies. A common example of a self-referent encoding strategy is judging whether or not an adjective describes oneself. In contrast, an affective self-focus was shown to disrupt source monitoring performance. Destun and Kuiper (1999) argued that affective self-focus reduces the chance that a listener binds features of the speaker to the semantic content. Johnson, Nolde, and DeLeornadis (1996) concluded that focusing on one’s own feelings may help make a statement memorable but it may not necessarily allow one to identify the origin of the information at a later point.

In spite of partly contradictory results in this field of research, one can conclude that the self can be considered an aspect of the human information-processing system and appears to function as a superordinate system, which is deeply involved in processing, interpreting, and memorising information. Researchers “have argued that the self-structure in memory is unique relative to other concepts by virtue of its superior elaborative and organisational properties and its frequent use in information processing” (Symons & Johnson, 1997, p. 371). Symons and Johnson found evidence for this assumption in their meta-analysis. Thus, the self-reference effect seems to occur because this well-developed and often-used construct promotes organised and elaborate encoding.

In the present studies, self might have been activated because in both conditions (PA and IS) participants were asked to do something *themselves*. In both conditions, participants were instructed to write down the target word onto a response sheet and participants were addressed *directly* in the course of Experiments 1 and 2. Changes in the experimental setup of Experiment 2 might have additionally heightened self-awareness as compared to self-awareness in Experiment 1.

Thus, I argue that self-reference was high, both in the IS condition and in the PA condition. This possibly lead to processing of stimuli in terms of internal states in both cases. As a consequence, a positive generation effect, as theorised in the dual-hypothesis by Riefer et al. (2007) occurred.

Conclusions to Chapter 2

The purpose of the research introduced in Chapter 2 was twofold. First, it was aimed at solving the contradictions between the theories of Mulligan (2004) and Riefer et al. (2007) especially concerning the dimension degree of completeness. This solution was aimed for by implementing the processing of perceptual attributes and the processing of internal states for the specific case of degree of completeness. Second, research was aimed to advance current theorising in the field by adding a new and more abstract layer concerning types of processing. Thereby, degree of completeness was regarded, on the one hand, as the impetus for creating a new layer in the first place and, on the other hand, as a suitable means and method for testing it.

It has to be stated that implementing a new and more abstract layer concerning types of processing via instructing participants in terms of IS or in terms of PA did not prove successful in the expected way. Instead of an interaction, a positive generation effect occurred in both conditions. This parallel result pattern was not likely due to weakness of study instructions, but can rather be explained when considering the self-reference effect in memory. Thus, finding a solution to the contradiction between Mulligan (2004) and Riefer et al. (2007) via instructing higher order types of processing has failed in the current experiments. Moreover, advancing current theorising in the field by adding a new and more abstract layer concerning types of processing could not be obtained.

However, self-reference was found to be an important moderator for the effects of self-generation on source memory tasks. To conclude, when self-reference is high, memory for source attributes seems to be processed in a way considering the role of self at encoding. In a reality monitoring paradigm, the self of the learner is involved to a large degree and accordingly self-referent processing is high. Therefore, processing, studying, and recalling items in terms of the reality monitoring paradigm occurs automatically, even for cases, in which instructions suggest a different type of item processing, studying, and recalling. As a consequence, a positive generation effect emerges. An effect of self-reference hence plays a critical role for generation effect studies when attempting to investigate source memory.

The role of the self-reference effect has not been investigated explicitly for generation effect studies so far. Thus, an empirical test of the current interpretations and conclusions is still missing.

Chapter 3:
The Role of Conceptual Processing

The present chapter is concerned with the role of increased conceptual processing for source (i.e., for memory for presentation colour).

3.1 Introduction to Chapter 3

The experiments that are introduced in Chapter 3 were conducted in order to determine if schema-typicality affects source memory performances in generation effect paradigms and in which way or direction it does so, given higher conceptual processing of self-generated items.

Conducting research on this topic is important, for several reasons. First, the role of conceptual activation in generation effect studies can be supported. Second, following Mulligan's processing account, which was brought forward by him in his 2004 paper, leads to the deduction of several hypotheses. Many hypotheses have already been drawn and tested in his paper. However, another extension is directly deducible from his account, but has not yet been tested. Hence, this gap is now filled by the present studies. Third, the roles of conceptual processing versus perceptual processing as the basic mechanisms producing the generation effect are under debate; Mulligan (2004) argues in favour of this account, whereas Mulligan et al. (2006) instead support the role of visual processing versus non-visual processing. Thus, when investigating the dominant role of conceptual processing more closely in the present experiments, one can differentiate between these two accounts and eventually favour one over the other. Additionally, by introducing different types of stimuli (those bearing no typical colour, and those bearing a typical colour and being presented either in this typical colour or not) one can also learn about the effect of colour schema-typicality on memory for the item itself. Although this aspect is not new to the research of memory psychology in general, the issue has not yet been researched within generation effect paradigms. Therefore the possibility of an influence of self-generation remains.

In the following part of the introduction, the basic research idea is outlined in greater detail. First an illustration is given of theories supporting the occurrence of a positive generation effect in item memory, before the account by Mulligan (2004) is repeated. The assumptions of Mulligan et al. (2006) are then discussed, which argue against Mulligan's (2004) processing account suggesting a more general frame of processing of visual versus processing of nonvisual information. Furthermore, theories and studies are offered on presentation colour, concept colour, and concept activation, before the actual research idea is finally revealed.

Explanations for Generation Effects in Item Memory

The lexical activation hypothesis was created to explain the positive generation effect in item memory. It assumes that the benefit of self-generation lies in a stronger activation of the semantic network or semantic lexicon of stimuli as compared to the activation induced when perceiving stimuli. When rephrasing this idea, one could say that self-generation leads to increased conceptual processing. Other important hypotheses concerned with the positive generation effect in item memory, are the two-factor hypothesis and the multifactor hypothesis. According to both ideas, activation of concept is one of the primary factors contributing to increased memory for self-generated items. Further details concerning these hypotheses can be found in Section 1.1. In conclusion, these accounts provide support for the statement that self-generation leads to higher conceptual processing and thus to higher schema activation.

The dual-processing account by Mulligan (2004)

Mulligan (2004) formulated the dual-processing account, which addresses explanations for item memory and source memory equally within generation effect paradigms and which is based on Jacoby (1983). Both scientists theorised that no difference as such exists in the underlying processes of encoding and recalling a stimulus. They rather claim that a difference lies in what is encoded and to which degree it is encoded. Referring to Jacoby (1983), Mulligan pointed out that basically two types of processing take place in generation effect paradigms, namely *perceptual processing* and *conceptual processing*. The former refers to encoding of the actual perceptual attributes (i.e., the appearance) of an item, such as its colour. In contrast, the latter refers to the processing and encoding of the concept of the item itself and is related to priming, semantic activation, conceptual activation or spreading activation (e.g., Anderson, 1976, 1983; Collins & Loftus, 1975; McClelland & Rumelhart, 1986; McNamara, 1992, 1994 Meyer & Schvaneveldt, 1971). When an item is processed conceptually, its meaning is activated primarily, along with information about how the word is spelled, its use, examples of it, what its counterpart is, whether or not it is innate, et cetera. For example, instead of processing and encoding that the word "shoe" was written in lower case Times New Roman (i.e., perceptual processing), the concept SHOE is activated. Note that terms denoting concept names or concept features are henceforth written in uppercase. In terms of modal theories of knowledge, this means that features of the concept SHOE are activated, such as "USED TO PROTECT ONE'S FEET", "BOUGHT IN SHOPS", and "TYPICALLY HAS SHOE-LACES AND A SOLE". Perceptual processing and conceptual processing are theorised to occur simultaneously and competitively, drawing on the same pool of cognitive resources. Consequently, an increase in perceptual processing leads to a decrease in conceptual processing, and vice versa.

According to Mulligan (2004), the type of processing taking place at study determines the memory trace. When *perceptual processing has predominantly taken place at study*, the memory trace that has been created, is rich in perceptual detail. Perceptual source cues have been encoded well. When *conceptual processing has taken place predominantly at study*, the memory trace is rich in meaning and the object's semantic network is encoded well.

Finally, when memory is tested, performance depends on the quality of the memory trace. For cases in which a match exists between memory trace and memory test demands, memory is theorised to be good. This is the case (a) when a perceptually rich memory trace was formed and a test of perceptual attributes of the item follows or (b) when a conceptually rich memory trace was formed and a test of conceptual meaning or conceptual features of the item follows. For cases in which this overlap or match is not perfect, memory performance should be diminished.

Mulligan applied the above described dual processing account to processing taking place within the generation effect paradigm, whereas Jacoby (1983) had devised this theory to apply to any type of task or material. Thus, for complete items contrasted with incomplete items, there is no difference *as such* in encoding processes. There is not *one thing* taking place for complete items (read condition) and a *different thing* taking place for incomplete items (generate condition). Instead, for both types of stimuli, perceptual processing and conceptual processing occur. However, they do so to different degrees, which is induced and enforced by the different task demands brought forward by the differences in item presentation. When complete words are to be studied, their complete form is presented. Here the words can be identified easily and most processing can take place for perceptual attributes of the items ("foot-shoe"). Hence, perceptual processing primarily occurs and a memory trace results that is perceptually rich. When incomplete words are to be studied, only parts of the target word are presented. Here, the word can also be identified easily, due to the co-occurrence of the cue word, but the identification is due to a conceptual deduction from the cue word and the remainders of the target word ("foot-sh__"). Hence, conceptual processing primarily occurs and a memory trace results that is conceptually rich.

For item memory tests, a conceptually rich memory trace is advantageous, because in a typical old-new recognition test, for example, memory for the concept of an item is tested. Hence, incomplete items, which are processed in a predominantly conceptual fashion, are remembered better than complete items. This leads to a positive generation effect.

However, for source memory tests, a perceptually rich memory trace is advantageous, because a perceptual attribute (e.g., colour) has to be remembered. Hence, for complete items, which are processed in a predominantly perceptual fashion, colour is remembered better than for incomplete items. This leads to a negative generation effect.

Mulligan added a specificity sub-clause to his explanations regarding source memory. He thought the type of perceptual attribute to be crucial and he therefore differentiated between attributes pertaining to the target itself and attributes independent of the target. This means, that the same task dimension, e.g. colour, could either pertain to the object (e.g., the colour of the target) or be independent of the object (e.g., the background colour of the screen). Other examples of attributes, which are independent of the target, are cue word colour and target location. It was pointed out that "the encoding of extratarget features are at neither an advantage nor a disadvantage in the generate condition" (Mulligan, 2004, p. 837). Hence, when a perceptual attribute is part of the object, a negative generation effect should occur. However, when a perceptual attribute is independent of the object, memory for this attribute should be equally good for complete and incomplete items.

In the experiments reported by Mulligan (2004), the positive generation effects in item memory could be replicated. Also, a negative generation effect in source memory occurred for the external perceptual attribute employed, namely for target colour. Therefore, a negative generation effect emerged when the to-be-studied *perceptual attribute pertained to the actual object*. Finally, a null effect in source memory resulted for target location and background colour, i.e., when the *perceptual attribute was independent of the actual object*.

In conclusion, Mulligan found full support for his predictions deduced from the processing account and thus provided conclusive evidence for the statement that self-generation leads to higher conceptual processing and thus to higher schema activation. For a more detailed description and evaluation of Mulligan's dual processing account see Chapter 2.

The Studies by Mulligan et al. (2006)

In the following section, the claim that self-generation produces higher conceptual processing (supported by theories on self-generation and item memory, and by Mulligan, 2004) is discussed. In a more recent paper, Mulligan, et al. (2006) criticised the dual-processing account by Mulligan (2004). More precisely, they questioned the appropriateness of the processing account in the presented form and suggested that instead of a *perceptual versus conceptual* dichotomy, applying a more general *visual versus non-visual* dichotomy could be more appropriate. Mulligan et al. (2006) argued that the original processing account might be too specific. They pointed out that the perceptual-conceptual distinction is a specific instantiation of the more general notion of transfer-appropriate processing, which states that the overlap in encoding and retrieval processes determines memory performance (Morris, Bransford, & Franks, 1977). The relationship between how information was encoded and how it is retrieved is central. According to the transfer-appropriate processing idea, the bigger the match between encoding and retrieval processes, the better the memory outcome. The smaller this match, the worse the memory outcome.

Mulligan et al. (2006) argued that not all generation manipulations emphasised conceptual processing (as is true in antonym pairs or semantic associates pairs, e.g., “hot-cold” or “foot-shoe”). Rather, they stated, other forms have been commonly used as well, such as rhyme generation tasks, which require participants to generate a rhyme in response to a cue word (e.g., “coat-boat”). These tasks had also lead to a positive generation effect in previous studies, but no conceptual or semantic analysis of cue and target item was necessarily required to account for the results. Since, from their point of view, tasks that do not imply any type of conceptual processing functioned just the same as those that necessitate conceptual processing, Mulligan et al. (2006) *reframed* the original processing account into a more general visual versus non-visual processing account.

According to this more general account, processing of visual information takes place primarily in the read condition. Therefore, the processing of complete items, which are processed visually and which concomitantly produce a visual memory trace, should lead to the successful completion of tests on the visual attributes of items. In the generate condition, it is hypothesised that the processing of non-visual information would take place primarily. This is true if the task is conceptual as in the case of antonym generation, or if the task is non-conceptual as in the case of rhyme generation. Hence, a positive generation effect is theorised to occur in item memory, for which conceptual, that is non-visual information, is relevant. In source memory, for which visual information is relevant, a negative generation effect is theorised to occur for attributes pertaining to the target item and a null effect is expected to occur for attributes independent of the target item. A short illustration of the account can be seen in Figure 3.1.

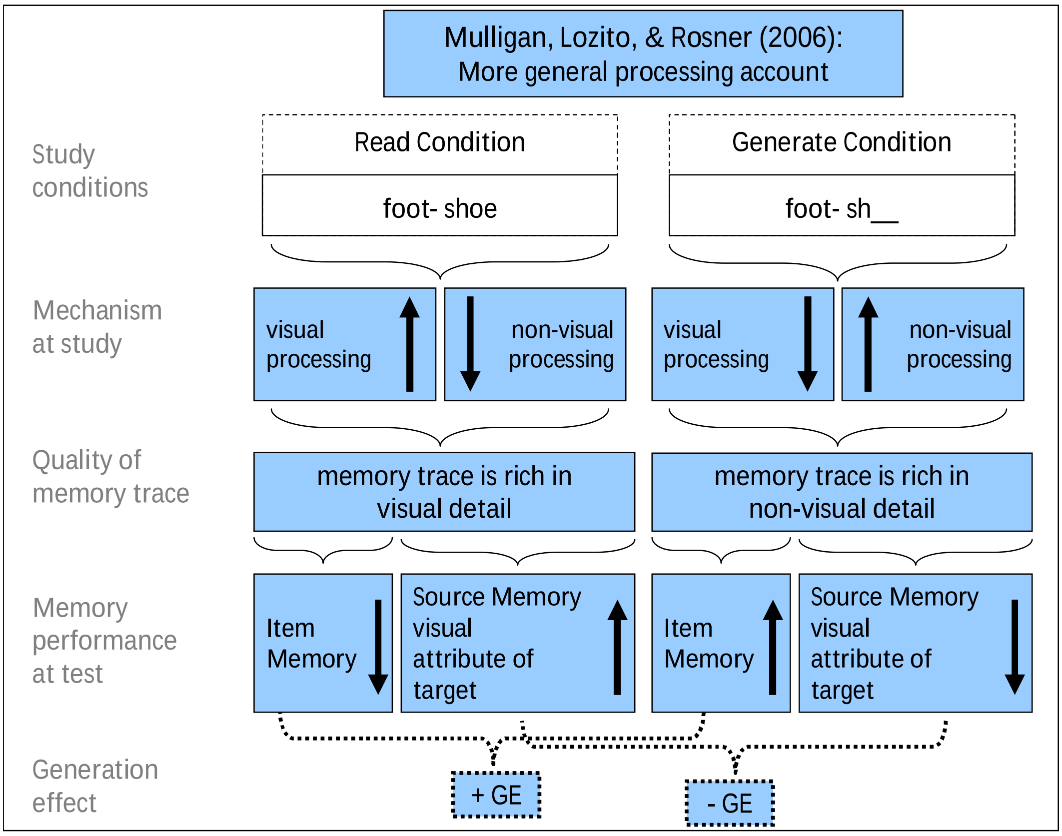

Figure 3.1: Illustration of the *cognitive mechanisms*, the entailed *memory trace*, and the resulting *test performance*, as theorised in the more general processing account by Mulligan et al. (2006) regarding a generation effect paradigm

To summarise, whether or not this non-visual processing is semantic (as in Mulligan, 2004), was considered irrelevant for item memory and source memory outcomes. Put differently, the modified processing account therefore posits that a non-semantic but also non-visual generation task should produce the same pattern of negative generation effects, as traditional semantic generation tasks do. Mulligan et al. (2006) argued that when their idea could be validated experimentally, the modified visual versus non-visual processing account was to be favoured over the original processing account due to its more general applicability.

Moreover, Mulligan et al. (2006) wrote in a footnote that they framed their account in terms of a visual versus non-visual processing account, because vision was the relevant modality for the read-generate manipulations in their article "and, indeed, in virtually the entire literature on the generation effect" (p. 837). They added that other sensory modalities are also imaginable, such as hearing.

Whether the original processing account (i.e., the account contrasting perceptual and conceptual processing) or the modified processing account (i.e., the account contrasting visual and non-visual processing) is more applicable to the question of self-generation and source memory, was tested empirically by Mulligan et al. (2006). The question motivating the experiments was the following: If all aspects that could lead to semantic processing were eliminated from the experimental design, so that the only reasonably remaining dichotomy is visual versus non-visual, would positive generation effects in item memory and negative generation effects in source memory still emerge? To answer this question, Mulligan et al. (2006) implemented nonsemantic generation effect paradigms. An overview of the most relevant studies conducted in Mulligan et al. (2006) can be seen in Table 3.1.

First, Mulligan et al. (2006) employed the *rhyme generation task*, in which a cue word is given that is either followed by a complete rhyme or followed by just one letter and underscore, for example: "coat-boat" or "coat-b___". According to the authors, in this task, complete items trigger and involve greater processing of the visual characteristics of the target word, whereas incomplete items require more non-visual information for successful completion. In their view, rhyme generation does not require conceptual processing, because the task directs attention to the word's phonology rather than to its meaning. This generation task was given to participants in four separate experiments varying four source memory attributes: target colour, target font, spatial location, and background colour. All four experiments demonstrated result patterns consistent with those found in the antonym generation task.

Second, the *rhyme generation task* was employed for *nonwords*. The generation task with nonwords was given to participants in three separate experiments varying target colour, target font, and spatial location. There was no effect of generation in item memory. Also, no significant effect of generation on source memory for location emerged. However, the hypothesised null effect for location was unlikely due to low power ($1-\beta = 0.85$ for $n = 40$, $\alpha = 0.05$, one-tailed). For the source memory dimensions colour and font, a negative generation effect occurred, as found in previous experiments.

Third, Mulligan and colleagues employed the *letter transposition task*. In the applied version of this task, the target word is either presented unaffected by any manipulation or is presented with the first two letters transposed, e.g., “banana” in the read condition; vs. “abnana” in the generate condition. When the first two letters are transposed, participants need to re-arrange the letters to generate the target items. Compared with complete items, *solving* the transposed items typically produces a memory advantage in item memory. Critical is that the amount of visual information in the generate condition matches that in the read condition. This contrasts with other types of generation effect tasks in which part of the information is occluded or erased to produce an incomplete item. What remains the same between letter elimination tasks and letter transposition tasks is that both require mental manipulation of presented visual information. Yet, more critical is the point that in the letter transposition task, attention needs not to be diverted to sources other than the target item itself. *When* letters are missing, conceptual processing involving the cue word takes place immediately and automatically to a specific degree. *In the letter transposition task however*, when all letters of the word are given, participants‘ attention is not diverted to processing other aspects of the target such as meaning or phonology.

Finally, Mulligan et al. (2006) state that the letter transposition task is not supposed to disrupt the encoding of perceptual visual features of the target. Put differently, since visual processing takes place for complete items and for incomplete, i.e., letter transposed items, memory performance should be equal. This idea was tested in two experiments studying target word colour: experiments 6A presenting words and 6B presenting words and nonwords. In both experiments, a positive generation effect emerged in item memory. In source memory, a null effect emerged, which was in line with the predictions made.

To reiterate, the major issue discussed in Mulligan et al. (2006) pertains to the appropriateness of the perceptual and conceptual processing dichotomy concerning the generation effect. Additionally, it was intended to investigate (a) whether the positive effects of generation on item memory can be dissociated from the negative effects on source memory, and (b) whether the negative generation effects found in Mulligan (2004) can be extended to the effect of another perceptual attribute, such as font. Moreover, Mulligan et al. (2006) defined the generation effect more broadly, since they regard other tasks such as the letter transposition task also as appropriate generation tasks. All three goals are sound and constructive for advancing knowledge in the scientific community concerning the generation effect. However, several questions can, in my view, be raised concerning the validity of the specific procedures and tasks employed.

Table 3.1: Selected experiments from Mulligan et al. (2006) that are relevant for the current presentation of evidence for the more general dual processing account. Result patterns include positive generation effects ("+ GE"), negative generation effects ("-GE"), and null effects.

Selected experiments from Mulligan et al. (2006)					
Experiment number	Approach of eliminating semantics	Type of perceptual attribute	Source memory dimension	Results item memory	Results source memory
2A	*Rhyme* generation task	Pertaining to object	Colour of target word	+ GE	- GE
2B	*Rhyme* generation task	Pertaining to object	Font of target word	+ GE	- GE
2C	*Rhyme* generation task	Independent of object	Location of target word	+ GE	Null effect
2D	*Rhyme* generation task	Independent of object	Background colour of screen	+ GE	Null effect
4A	*Rhyme* generation task with *nonwords*	Pertaining to object	Colour of target	Null effect	- GE
4B	*Rhyme* generation task with *nonwords*	Pertaining to object	Font of target	Null effect	- GE
5	*Rhyme* generation task with *nonwords*	Independent of object	Location of target	Null effect	Null effect (reported power at .85 for n = 40, α = .05, one-tailed)
6A	*Letter transposition task*	Pertaining to object	Colour of target word	+ GE	Null effect
6B	*Letter transposition task*	Pertaining to object	Colour of target word	+ GE	Null effect

First, it can be debated whether or not rhyme generation can be considered a truly non-semantic task. Indeed, Slamecka and Graf (1978) argued that rhyme generation was non-semantic only in the sense applied in the LOP idea. However, McElroy (1987) and Taconnat and Isingrini (2004) concluded that generation enhances semantic processing via postgenerative semantic elaboration. They found this to be the case even when semantic processing was not explicitly required through the generation rule itself. I see a possible reason for postgenerative semantic elaboration in participants' need for cognitive closure (Schlink & Walther, 2007; Webster & Kruglanski, 1994), i.e., their desire for solutions for problems as compared to confusion and ambiguity. Cognitive closure was found to be a general human tendency in spite of interindividual differences found. I believe that in Mulligan et al. (2006), participants might have felt a need to make sense of the provided stimuli.

On a note closer related to the actual rhymes employed, it can be debated whether, when generating "found" from "sound", one really generates "found" *purely* phonologically. When using words as rhymes it seems reasonable to assume that semantic activation of the word "found" arises automatically to a certain degree. It may not be a question of *either or*, but instead *both* phonological and conceptual processing could co-occur – possibly even to the same degree.

Second, even though Mulligan and colleagues additionally employed nonwords as targets in rhyme generation tasks, cues were still words. Moreover, no positive generation effect in item memory could be found in Mulligan et al. (2006) for experiments with nonwords. Note that this was not due to a ceiling effect however, since performance scores ranged from 66 % to 72 %, in the read condition, and from 62 % to 71 %, in the generate condition.

Third, whether or not the letter transposition task is close enough in design to the classical generation task can be debated. One of the basic or even defining features of a generation effect task is the fact that a certain amount of information is *missing*. However, this is not the case in the letter transposition task. Consequently, employing the traditional semantic generation tasks and modifying them by eliminating the cue and thus turning them into non-semantic generation tasks, seems a better alternative to me. Note however that from the perspective of the processing account, semantic generation tasks and a non-semantic generation task without cues may differ in important ways. Mulligan (2004) stated that such non-semantic generation tasks likely produce substantial perceptual processing of the target item, since there is no basis from which to generate the target other than the target itself. Thus, the amount of perceptual processing in the generate condition may be equal or may even exceed that in the read condition. Consequently, the processing view would not necessarily predict a negative generation effect for memory for perceptual attributes of the target item. Mulligan et al. (2006) agree with this idea and write the following: "Speculatively, it may be the case that visual generation tasks requiring even more visual processing than the letter transposition task may improve (rather than merely leave unaffected) contextual memory for visual features of the target" (p. 844). Some studies employing non-semantic generation tasks found that both for semantic and for non-semantic generation tasks, a positive generation effect emerges. However, differences have been found, e.g., (1) semantic generation tasks produce hypermnesia (i.e., increased recall of targets after repeated tests), whereas non-semantic generation tasks do not (Mulligan, 2001, 2002b), and (2) semantic generation tasks enhance conceptual implicit memory, whereas non-semantic tasks generally do not (Mulligan, 2002a).

Two things can be concluded here. First, the ideas and goals underlying Mulligan et al. (2006) are well-elaborated and need to be addressed. Second, in spite of this, difficulties can be found with their experimental implementations. In other words, it can be debated whether or not the presented tasks are free of conceptual processing.

Consequently, it remains unclear whether or not the perceptual versus conceptual dichotomy, which proved to be so successful in previous studies, can fully be replaced by the more general visual versus non-visual dichotomy. Further empirical evidence to this end is needed. Thus, no hard evidence seems to exist so far to fully abolish the perceptual and conceptual processing account at this point, although it has to be considered as challenged. Therefore, the basic assumption of self-generation leading to increased conceptual processing was presumed to be sensible for the present purpose.

Presentation Colour, Concept Colour, and Concept Activation

An interesting special case of the relevance of higher conceptual processing for self-generated items can be seen, when source memory is tested for an attribute that is highly associated with or that is part of the concept of the item itself. A plausible example is *colour* of pictures.

As was concluded in the previous paragraphs, self-generation is likely to lead to higher conceptual processing. Higher conceptual processing then leads to a strong activation of the concept of an item. Hence, when colour memory is tested, it is important to consider that for some objects, their concept does include a typical colour. An example of this is the BANANA for which the colour YELLOW is a typical descriptive feature. Other imaginable descriptive features for BANANA are IS CURVED or IS FOOD. Compared to BANANA, the concept TOOTHBRUSH, although containing various other features, does not include any specific colour, because a toothbrush can occur in any colour imaginable. Other imaginable descriptive features for TOOTHBRUSH are USED TO BRUSH TEETH or TYPICALLY FOUND IN BATHROOMS. Therefore, colour is relevant for the concept of a BANANA, but is not relevant for the concept of a TOOTHBRUSH. Both amodal (e.g., feature lists, semantic networks, or frames) and modal (e.g., perceptual symbol system) theories of knowledge and concept representation agree on this issue. For a review, see Barsalou (1999). Examples in this dissertation are offered in terms of amodal knowledge representations, in which perceptual states are transducted into non-perceptual representational formats. Additionally, the activation of an object's concept is theorised to take place rapidly and automatically (e.g., Anderson, 1976, 1983; Collins & Loftus, 1975; McClelland & Rumelhart, 1986; McNamara, 1992, 1994; Meyer & Schvaneveldt, 1971).

It has been shown in many studies that variations in whether or not an item was studied in a complete or in an incomplete manner can lead to differences in experimental outcome. For example, memory performance varies according to the degree of completeness of a target item. This effect was theorised to be deductible from an effect of differing degrees of conceptual processing. Therefore, conceptual processing translates into changes in memory.

However, is the effect of concept activation strong enough to occur in other psychological measures, i.e., in behavioural measures? Support that concept activation impacts behavioural measures, can be gained from a wide range of studies from diverse fields such as the psychology of concepts, the psychology of language, the psychology of perception and the psychology of memory.

First, evidence exists from a multitude of empirical studies that the activation of a concept leads to a measurable and significant effect on behavioural measures in studies on the psychology of concepts. Support can be found for examples when considering the ease with which participants judge category membership of an item dependent on its typicality. For example, participants are quicker to affirm that a robin is a bird, but show increased response times for affirming that a chicken (as a less typical example) is a bird. This was found to be true for verbal and for visual stimuli (Murphy & Brownell, 1985; Rips, Shoben, & Smith, 1973). More typicality effects were found in experiments on item production. Mervis, Catlin, and Rosch (1976) showed that when participants are asked to produce category members, typical members (rated as such by other participants) are produced more often; in fact, the average correlation of typicality and production frequency is .63, across categories. Also in category learning (Rosch, Simpson, & Miller, 1976) and language learning, an advantage of typical items over atypical examples can be seen. Furthermore, Kelly, Bock, and Keil (1986) found that when participants mention two category members of the same category in just one sentence – one member being typical and the other being atypical – typicality plays a role. The authors showed that the more typical member is usually mentioned first (e.g., “apple and star fruit” rather than “star fruit and apple”).

Second, evidence exists that the activation of a concept leads to a measurable and significant effect on behavioural measures in studies on the psychology of perception (e.g., Bartleson, 1960; Bolles, Hulicka, & Hanly, 1959; Bruner, Postman, & Rodrigues, 1951; Duncker, 1939; Fisher, Hull, & Holtz, 1956; Jin, & Shevell, 1996). Among others, Hansen, Olkkonen, Walter, and Gegenfurtner (2006) showed that schema-typicality plays an important role in colour perception. They investigated whether the *known* colour (i.e., prior knowledge) of an object affects colour appearance. To this end, they presented photos of natural fruit on uniform gray backgrounds for which the presentation colour of the fruit could be altered interactively by participants. Participants were asked to adjust the colour of the presented fruit until they appeared to be as grey as their background (i.e., achromatic). As a result, participants perceived objects as grey when this value actually did not match the objective grey colour of the background. Instead, overcompensation occurred in the direction opposite to the typical colour of the according fruit. “In actual fact, subjects adjusted the banana to a slightly bluish hue – its opponent colour – in order for it to appear neutral gray. At the point where the banana was actually achromatic, at the origin of the colour space, it still appeared yellowish" (Hansen et al., 2006, p. 1367).

Third, evidence exists from a wide range of empirical studies that the activation of a concept leads to a measurable and significant effect on behavioural measures in studies on the psychology of memory.

In human memory, *schemata*, *frames,* and *scripts* play important roles. Schemata can be understood as knowledge structures that represent concepts (e.g., events, roles, scenes), specifying the characteristics of the concept and the relations between those characteristics (Tuckey & Brewer, 2003). They are part of a person's semantic memory and knowledge system. Schemata help generate specific expectancies and thus influence perceptions and recollections of information. In fact, schema effects result through an interplay of top-down and bottom-up information. Alba and Hasher (1983) reflected the view of schema theorists by proposing that what is encoded or stored in memory is heavily determined by a guiding schema or knowledge framework. This guiding schema selects and actively modifies pieces of experiences to arrive at a coherent, unified, expectation-confirming, and knowledge-consistent representation of an experience. They suggested four central encoding processes: selection, abstraction, interpretation, and integration. Alba and Hasher emphasised that a schema allows for the encoding, storage, and retrieval of information related for instance to a certain domain. According to Alba and Hasher, a frame is a schema that contains knowledge about the structure of a familiar event, e.g., knowledge about the structure of a short story. This frame does not specify the exact contents of the event, but rather focuses on the general type of information expected in that situation and on the order in which it should be encountered (e.g., in a fairy tale, setting information would be followed by theme information and the final words would be "lived happily ever after"). Compared to this, a script – just as a frame – contains general information about particular, frequently experienced events (e.g., a visit to a restaurant). However, scripts also contain more specific information about the contents of the event (e.g., being seated, ordering a meal, and so on).

In a classical study by Bartlett (1932), participants were to recount uncommon Indian tales. Instead of correctly reproducing the stories, participants showed schema-driven reconstructions by omitting inapprehensible details and by adding information so that the uncommon stories fit into common western story frameworks. Another exemplary study was conducted by Anderson and Pichert (1978) who had participants read stories from the perspective of either a burglar or a house-buyer. At test, participants were to reproduce the stories and Anderson and Pichert counted how many burglar-relevant details and how many house-buyer relevant details were reproduced under either reading perspective. As a result, information which was in line with ones reading perspective was remembered more often. Additionally, when asking participants to switch perspective before a second round of reading and recalling details, more details that were in line with the new perspective were reproduced.

As can be seen from these examples, presentation and processing of schema-typical and schema-atypical, or of schema-consistent and schema-inconsistent information constitute a central issue in research on schemata, frames, and scripts. The question is whether schema-consistent or schema-inconsistent information is better remembered. Three of the most important schema theories are *filter theory* (*FT*), *schema-copy-plus-tag hypothesis* (*SCTH*), and *attention-elaboration* (*AE*) *hypothesis*. See Graesser and Nakamura (1982), for a review of these accounts. According to filter theory, only schema-consistent information is stored in memory, whereas atypical or schema-inconsistent information is filtered out and is thus not stored. According to SCTH, a copy of the schema is stored along with additional *tags* that indicate extraordinary events or deviations from defaults of variables. Highly typical events are not stored, since they can be deduced from the copy of the schema (cf., principle of cognitive economy; Collins & Quillian, 1969; Rosch, 1978). According to AE, schemata guide attention. Information that is under attention is stored; a possible mechanism is increased semantic elaboration or depth of cognitive processing (cf., LOP; Craik and Lockhardt, 1972).

A memory advantage for atypical information has often been found for different age groups, e.g., for adults or young children, as well as for many different stimuli and study tasks, e.g., for face recognition, natural situations, persuasive messages, and person attributes (see Erdfelder & Bredenkamp, 1998, for a review). In contrast, work in social beliefs and stereotypes has found either a memory advantage for schema-consistent information (Rothbart, Evans, & Fulero, 1979; Snyder & Uranowitz, 1978) or a memory advantage for schema-inconsistent information (e.g., Bower, Black, & Turner, 1979; Goodman, 1980). However, for some studies showing an advantage of schema-consistent information, methodological or conceptual shortcomings could be found. In some cases, the recognition hit rates were not properly corrected for response bias. In other cases, typicality was confused with the degree of relatedness of information to a schema. Moreover, Rojahn and Pettigrew (1992) conducted a meta-analysis, in which they resolved these apparent contradictions by evaluating results across studies. They found that when measures of recognition were corrected for false alarm rate, schema-inconsistent information was remembered better than schema-consistent information. For example, for a library schema, a common false alarm might be reporting to have seen a book when in fact a book did not appear in any of the studied scenes (see Sakamoto & Love, 2004). A similar meta-analysis on stereotype research was conducted by Stangor and McMillan (1992), who reached the same conclusion as Rojahn and Pettigrew.

Thus, one can conclude that (a) schemata have an important influence on human memory and (b) more convincing evidence exists in favour of a memory advantage for schema-inconsistent or schema-atypical information.

Note that in the following line of experiments (i.e., Experiments 3, 4, and 5) the concept and the implementation of typicality do not fit perfectly in the schema, script, and frame literature. Usually in those studies, an object or feature is typical or atypical within a certain schema, script, or frame. This object or feature *is part of* something (e.g., ordering ice cream for dessert vs. dancing a tango after dinner). In the present case, a current presentation feature of an object is either consistent or inconsistent with a schema-typical presentation of that feature (e.g., *yellow* presentation of the feature colour of a banana vs. *blue* presentation of the feature colour of a banana). The feature is not part of something (e.g., a sequence of events), but this something (e.g., concept of a banana) and its feature (e.g., typical vs. atypical colour) are intertwined.

Interim Summary

To reiterate, the activation of an object's concept including its semantic features takes place rapidly and automatically. Conceptual processing is theorised to take place to a higher degree or in a more elaborate form for self-generated items as compared to simply read items (cf., theories on item memory and the generation effect and Mulligan, 2004). However, Mulligan et al. (2006) argue against this point. They question the roles of conceptual and perceptual processing as the basic underlying mechanisms and instead prefer a more general version of a processing account involving visual and non-visual processing.

Concerning concept colour, one can see that some everyday objects have got a colour that is a typical feature of them and others have not. In general, when being presented with the picture of an object in a memory experiment and when being asked to remember all items in the study list along with their colour at presentation, two things happen at study: (a) the concept of the object is activated and (b) the actual presentation colour of an object in the picture is encoded. When an item has no schema-typical colour, it is simply presented in any colour which does not have a special relationship to the object's concept. However, when an item has a schema-typical colour, it is either presented in this colour or it is not. It is likely that whether or not an item has a schema-typical colour influences memory performance for presentation colour.

The generation effect paradigm adds another interesting issue to this argumentation, because due to definition, some objects are presented complete, whereas others are presented incomplete. This circumstance is theorised to entail different degrees of conceptual processing. Since self-generation of items is theorised to lead to higher conceptual processing, it is likely that whether or not an incomplete item has a schema-typical colour influences colour memory performance (experimental outcome). Moreover, it does so to a higher degree than is true for simply read items.

Evidence and suggestions that concept activation impacts behavioural measures can be gained from a wide range of studies from psychological research on concepts, language, perception, and memory. Moreover, schemata influence human memory and evidence exists in favour of a memory advantage for schema-atypical information.

3.2 Experiments 3, 4, and 5

Does schema-typicality of object colour affect colour memory performance for complete versus incomplete items and to what degree is this influence? This question was addressed in three laboratory experiments. Experiments 3, 4, and 5 represent concurrent attempts to experimentally operationalise an answer to this research question. These studies were outlined and conducted in a parallel fashion rather than as a sequence of experiments.

Table 3.2: Types of items in Experiments 3, 4, and 5 separated by colour and degree of completeness along with examples

Type of item	Colour	Degree of completeness	Example
Items without schema-typical colour	Colour A	Complete	Yellow toothbrush presented complete
		Incomplete	Yellow toothbrush presented incomplete
	Colour B	Complete	Blue toothbrush presented complete
		Incomplete	Blue toothbrush presented incomplete
Items with schema-typical colour	Typical colour	Complete	Yellow banana presented complete
		Incomplete	Yellow banana presented incomplete
	Atypical colour	Complete	Blue banana presented complete
		Incomplete	Blue banana presented incomplete

Due to this parallel setup, important similarities exist concerning (a) the design, (b) the crucial experimental factors, and (c) the types of items. The primary focus of the experiments lay on several experimental variables, e.g., on the existence of a schema-typical colour (items without schema-typical colour, items with schema-typical colour), on colour of items (Colour A or Colour B, for items without schema-typical colours; typical colour or atypical colour, for items with schema-typical colour), and on degree of completeness (complete, incomplete). A list containing all types of items along with examples in a schematised form bearing these crucial experimental factors in mind can be seen in Table 3.2. All of these experimental factors were varied within-subject.

Note that in the experiments, red and green were used as colours, rather than yellow and blue, which are chosen in the present paragraph only to retain the previous examples involving toothbrush and banana.

Essentially the same predictions can be made for the three experiments. The most relevant aspect of research presented in Chapter 3 lay in source memory for the perceptual attribute presentation colour or, simply, colour.

First (Hypothesis 1), regarding item memory:

In item memory, I expected a positive generation effect, as found in many previous studies; item memory for incomplete items was predicted to exceed that for complete items.

Second (Hypothesis 2), regarding source memory concerning the degree of completeness:

In source memory for the degree of completeness, I expected a positive generation effect, as found in Riefer et al. (2007), that is, source memory for degree of completeness for incomplete items was predicted to exceed that for complete items.

Third (Hypothesis 3a), regarding source memory for presentation colour:

For items without schema-typical colour, I expected colour to be remembered better for complete items than for incomplete items. Thus, a negative generation effect was predicted. This is in line with Mulligan (2004), Mulligan et al. (2006), and Riefer et al. (2007).

In contrast, for *items with schema-typical colour*, I hypothesised two effects: (1) a basic effect of schema-typicality on memory for colour, and (2) an additional effect of increased conceptual processing for self-generated items.

When an item has got a typical colour, I expected an effect of whether or not presentation colour matches concept colour. More precisely, I expected an atypicality effect, i.e., superior memory for items presented in an atypical colour. Moreover, this effect was expected to be influenced by whether or not items are presented complete or incomplete, because incomplete items evoke stronger conceptual processing than complete items. I expected the effect of schema-typicality to be moderated by the degree of conceptual processing as operationalised by the degree of completeness: I hypothesised that superior memory for atypical items was more pronounced for incomplete items.

Fourth (Hypothesis 3b), regarding guessing colour:

However, whether or not an item with a typical colour is presented in this typical colour could not only lead to memory differences, but also to differences in guessing. More specifically, guessing colour, when presentation colour cannot be remembered explicitly, could be biased towards guessing the concept colour. Hence, guessing colour could be significantly different from chance level and could indeed be in favour of the typical colour (i.e., concept colour): typically green items could be guessed primarily to have been presented in green, whereas typically red items could be guessed primarily to have been presented in red. Note that for items without schema-typical colour, guessing colour is expected to be at chance level.

Fifth (Hypothesis 4), regarding guessing rates:

I expected guessing old and guessing the degree of completeness to be at chance level.

In spite of many similarities, several differences exist between Experiments 3, 4, and 5 (see Figure 3.2). First, the degree of naturalness of the items was varied between the first and the latter two experiments. It has been shown that the more life-like and natural an item, the greater the activation of its concept (cf., Barsalou, 1999). Following this logic, it is possible that the hypothesised effect would not be pronounced enough with words. Therefore, in the current work, drawings (Experiment 3) and photos as an even more natural presentation form of the stimuli (Experiments 4 and 5) were employed.

Additionally, different types of study instructions, namely intentional (Experiments 3 and 4) versus incidental (Experiment 5) study instructions, were employed. In the former two experiments, intentional study instructions were implemented, in order to model designs applied in most previous generation effect papers. Until now, studies using incidental study instructions have rather been rare. Although the activation of the concept (along with the activation of its schema-typical colour) takes place automatically, it can be regarded as a rather weak, that is easily suppressed, activation. It seems reasonable to assume that activation of the concept's colour could be suppressed or overcome when using intentional study instructions, which actively force participants to withdraw attention from the colour of the concept and instead to focus attention on the actual colour at presentation. Therefore, the role of types of study instructions for photos was investigated between Experiments 4 and 5.

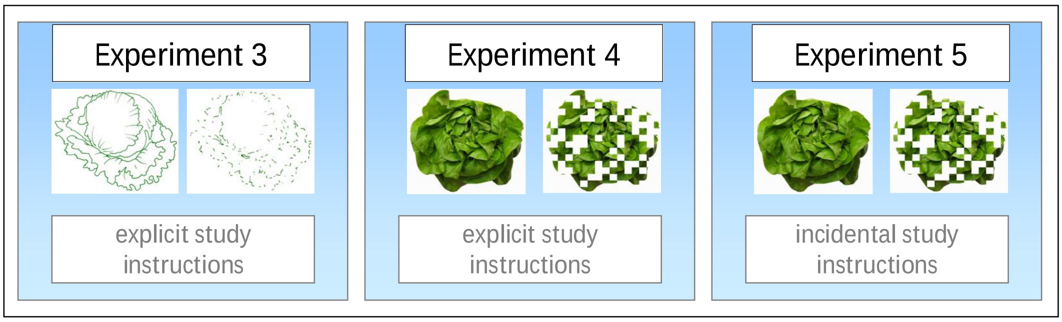

Figure 3.2: Overview of Experiments 3, 4, and 5, in which an emphasis is put on the differences between experimental stimuli and between the types of study instructions employed

As stated previously, material for the three current experiments was pictorial. Pictures rather than words were used as stimuli for two reasons. First, employing pictorial stimuli should be favourable due to the basic cognitive mechanism that the more natural an item, the greater the activation of its concept. Second, and more importantly, in this way, the following studies were closer in design to Hansen et al. (2006), who found an impact of concept colour on performance measures in perception. However, the choice of this type of material (especially in contrast to the predominantly used verbal material) is somewhat uncommon bearing previous research in mind. Is it safe to assume that memory performance in the generation effect design should occur in a parallel manner, independent of whether verbal or pictorial material is considered?

Although the generation effect in item memory and source memory has been studied using a variety of material, the predominant area of research has been conducted with verbal material. So far, three studies exist, which employ pictorial stimuli: Kinjo and Snodgrass (2000), Lohaus and Lachnit (2001), and Peynircioglu (1989). To summarise the conclusions drawn in these papers, it can be stated that all of these studies replicated the positive generation effect in *item memory*, as typically found for studies employing verbal material. Therefore, it can be deduced that whether or not pictorial material is used instead of verbal material, does not result in different generation effect patterns on item memory tasks.

But how about source memory for source attributes of pictorial material? To repeat, a *negative generation effect is expected* and is the theorised effect for an external source dimension pertaining to the target within a generation effect paradigm. This result should emerge according to both Mulligan (2004) and Riefer et al. (2007). Moreover, according to Riefer and colleagues, a positive generation effect is anticipated in a reality monitoring design. Kinjo and Snodgrass (2000) added a source memory task, in which they asked participants to recall whether an item had been presented complete or incomplete. They found a positive generation effect, which is in line with Riefer and colleagues (reality monitoring task). To my knowledge, no paper investigating external source monitoring for pictorial material has been published in a peer-reviewed journal. However, this very issue was one goal of my previous research (Zillig, 2007). In the crucial condition of this experiment, line drawings from the Snodgrass and Vanderwart (1980) set were used as target items. They were presented either in red or green. Participants were asked to remember which items they had been presented with and the colour in which the items had appeared. At test, old-new recognition of the items was tested first, followed by the question of whether items had been presented in red or green. In the analysis of memory performances using a multinomial processing tree model, a positive generation effect was found on item memory. For source memory for colour of the pictures, a negative generation effect emerged. This result was in line with the predictions of both Mulligan (2004) and Riefer et al. (2007) for external source monitoring tasks.

It can be concluded from this evidence, that the generation effect patterns (*item memory* and *external source monitoring*) repeatedly found for verbal (and other) material also appear for the study of pictorial material. It was therefore highly likely to expect a repetition of these effects in the present line of experiments. Note that whether or not generation effects in reality monitoring tasks are different for pictorial material remains unclear when considering previous research.

3.2.1 Experiment 3: Methods

In this section, Experiment 3 is described in detail and information on the sample, on the design employed, on the material presented to the participants, and on the experiment procedure is provided. Experiment 3 is the first of three parallel experiments designed to answer the following question: Does schema-typicality of object colour affect colour memory for complete versus incomplete items and to what degree is this influence? Here, drawings are presented as material and the memorisation of these is intentional.

Participants

Sixty-six persons participated in Experiment 3. Collected data for one person was lost due to technical failure after completion of the experimental session. Of the remaining 65 participants, 70.77 % were female. Participants' age ranged from 19 to 32 with a mean of 23.18 years and a median of 22 years. No participant suffered from red-green colour blindness. All participants except for three confirmed that they were native German speakers. However, this latter group of participants stated that their German proficiency skills were either "very good" (2 out of 3) or "good" (1 out of 3). All participants except five were currently enrolled at the University of Mannheim and primarily were students from the fields of Psychology (36.67 %), Economic Sciences (31.67 %), and Sociology (20 %). Other fields of study were Language and Communication Studies, Political Sciences, and Educational Sciences.

Design and Material

The generation effect was investigated for pictorial material only, i.e., for line drawings, to be more precise. Two colours were used in the experiment: red and green. Note that the same experiment was also conducted for two other colour dichotomies, namely for yellow versus brown and for grey versus skin-coloured. However, since Experiments 4 and 5 were focused on red and green colour only, results obtained for these other colour dichotomies are not reported in the present work.

Table 3.3: Research design for Experiment 3 employing the within-subject variations “colour“, „degree of completeness”, and “item type”; “CB” = counterbalancing group

Item type	CB	Old / studied items				New / unstudied items
		Red		Green		
		Complete	Incomplete	Complete	Incomplete	
(1) Items without schema-typical colour	1	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E
	2	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D
	3	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C
	4	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B
	5	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A
(2) Items with schema-typical colour (red)	1	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E
	2	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D
	3	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C
	4	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B
	5	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A
(3) Items with schema-typical colour (green)	1	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E
	2	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C	Stimulus set D
	3	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B	Stimulus set C
	4	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A	Stimulus set B
	5	Stimulus set B	Stimulus set C	Stimulus set D	Stimulus set E	Stimulus set A

Several independent variables were employed: First, “item type” was varied within-subject and had three levels: “drawings without schema-typical colour”, “drawings with schema-typical colour that were presented in their typical colour”, and “drawings with schema-typical colour that were presented in an atypical colour”. Schema colours were red and green. Second, “degree of completeness” was varied within-subject and had two levels: “complete drawings” versus “incomplete drawings”. Third, “colour” of the picture lines was varied within-subject and had two levels: “red” versus “green”. Moreover, for counterbalancing, five random versions of the study list were created so that each drawing appeared in each of the presentation conditions.

Most important for the current research question were the factors “degree of completeness” and “item type”. Five experimental conditions resulted, to which participants were assigned randomly. An illustration of the research design can be seen in Table 3.3.

One half of the presented items was in the read condition and the other half was in the generate condition. In the read condition, cue words and target pictures were presented complete. In the generate condition, the cue words were again displayed complete, whereas the target pictures were fragmented, i.e. appeared in an incomplete manner.

For picture fragmentation, the picture fragmentation tool "Frag" (Snodgrass, Smith, Feenan, & Corwin, 1987) was used. One half of the target drawings was presented in red and the other half was presented in green.

Experimental Stimuli

The critical study stimuli were taken from the Snodgrass and Vanderwart (1980) picture set, which includes drawings of everyday objects and a wider range of animals and tools; groups were: four-footed animals, kitchen utensils, articles of furniture, parts of the human body, fruit, weapons, carpenter's tools, articles of clothing, parts of buildings, musical instruments, birds, types of vehicles, toys, vegetables, and insects. Since the norms for the 260 pictures obtained in Snodgrass and Vanderwart (1980) are relatively old and based on a sample of Americans (i.e., native *English* speakers), a paper investigating a German sample (Genzel, Kerkhoff, & Scheffter, 1995) was selected instead. The collected norms on name agreement, familiarity, and visual complexity, and the German labels for the pictures, given by participants, were used in the present experiment. Since norms for some of the pictures were not or could not be assessed in Genzel et al. (1995), due to the fact that items were specific for the US-American culture or due to technical reasons, these stimuli had to be assessed anew. This was done in a pilot study (see Appendix E).

The target stimuli consisted of 40 line drawings of common objects (such as bowl, tomato, and lettuce). Twenty of the 40 drawings were pictures of items that have got a schema-typical colour; i.e. items that are either typically red (10 drawings, e.g., tomato) or items that are typically green (10 drawings, e.g., lettuce). The remaining 20 items were objects that have no schema-typical colour (e.g., bowl). The 10 typically red drawings, the 10 typically green drawings and the 20 drawings without schema-typical colour were randomly divided into five stimulus sets and counterbalanced (CB); see Table 3.3, for an illustration.

At study, 16 further drawings were selected from Snodgrass and Vanderwart (1980): Half of these served as a primacy buffer and half served as a recency buffer to eliminate primacy and recency effects. Additionally, three items served as practice items: All of these drawings were pictures of objects that have no schema-typical colour. Primacy buffer items, recency buffer items, and practice items were not considered for analyses. To obtain materials for Experiment 3, three pilot studies were conducted, two of which were aimed at finding appropriate target material; Moreover, a further pilot study was conducted to identify cue words that are semantically associated with the target drawings (see Appendix E). At test, studied items and distractor items were presented. Additionally, 24 new items (i.e., buffer items) were developed for the test, to fulfil the 1 : 1 ratio of studied versus non-studied items. These buffer items were not included in the analyses of the results.

Note that a complete list of all cues and targets used in this experiment and a list of descriptive information on the samples of participants for the pilot studies can be found in the Appendix (Appendices B and F, respectively).

Procedure

Participants were working on the tasks presented to them via computer in group sessions of up to 20 persons. The experiment was created and executed using the E-Prime 1.0 software (Schneider et al., 2002). Participants signed a consent form, before they were seated in front of the computer. The experiment comprised four experimental phases: practice phase, study phase, distractor phase, and test phase.

Practice Phase

Three practice trials allowed the participants to get acquainted with the experimental task. Practice items were displayed sequentially and in a fixed order.

Study Phase

The study phase consisted of one block. Items were presented either red or green and were displayed against a white background. Target items were presented in the centre of the screen. The pertaining instructions preceded the study block. For an illustration of the sequence of presentation slides within a trial, see Figure 3.3. Presentation durations in one trial roughly equal those applied by Mulligan (2004, Exp. 1) and were pretested to ensure sufficient presentation length for each slide. Solutions (i.e., target names) given on the slide following target drawings matched the target names displayed at test. Items in the study blocks were presented randomly. Items introduced in the study phase to account for primacy and recency effects were presented in a fixed order (8 items before the study block and 8 items after the study block). Cue words were displayed in 25 point Arial, whereas the target names were displayed in 35 point bold Courier New. Different appearances (font and size) were employed to increase the ease of discriminability of cue words and target names.

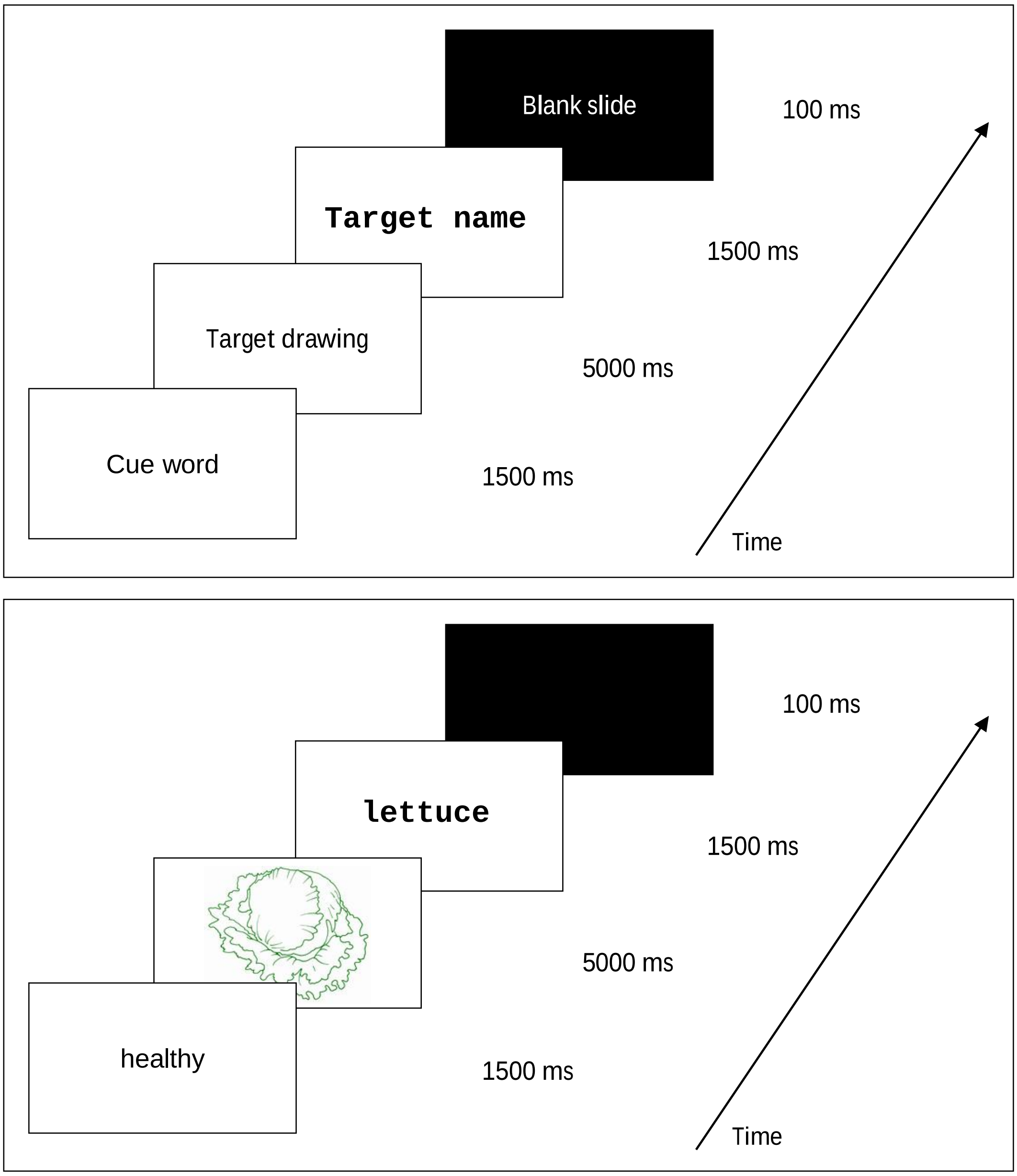

Figure 3.3: Illustration of the sequence of slides for Experiment 3 for one study trial (top) and a specific examples of a study trial (bottom); along with according presentation durations

Concerning memory encoding, intentional study instructions were employed. It was the participants' task to memorise which drawings were presented to them as well as their colour, and whether drawings had been presented complete or incomplete. Study instructions were were intentional as in Mulligan (2004) and Riefer et al. (2007). Participants were informed that they would finally be tested for all three pieces of information.
In addition to encoding, participants were also, of course, asked to solve, that is to actively generate, incomplete drawings. To have reason (i.e., from participants' perspective) and evidence (i.e., from experimenter's perspective) of this active generation, participants were asked to write down the names of the target drawings. Sheets of paper were distributed, on which participants were asked to note down the names of complete and incomplete items, *while* the drawings were shown on the screen. It was pointed out to participants that they should simply make a line, when they could not identify what the object was. I collected response sheets directly after the study block, so that participants could not consult them during distractor or test phase.

To reiterate, in the study phase, participants were to (a) write down the names of the objects on the response sheet while the drawings were presented to them, (b) memorise the objects (i.e., item memory), (c) memorise the presentation colour of the objects (i.e., source memory "colour"), and (d) memorise whether an object was presented complete or incomplete (i.e., source memory "degree of completeness").

Distractor Phase

In the distractor phase, participants were asked to judge solutions of equations for their correctness. Participants were asked to make their choices by pressing keys on the keyboard: k-key with a blue label, when a solution was correct; d-key with an orange label, when a solution was incorrect. Equations were presented for 7 seconds followed by feedback. The feedback indicated a correct response, a wrong response, or urged participants to answer more quickly, when no response was given within the 7 seconds time window. Equations were pretested and could be solved within the given time window. Equations in the distractor phase were presented randomly. The task lasted for three minutes. A numerical task, for which completeness was obvious throughout all trials, was chosen to minimise a conceptual overlap between distractor phase and study phase.

Test Phase

The pertaining instructions preceded the test phase. In the test block, there were 40 items that were used for analyses: 20 of these items have no schema-typical colour, whereas the other 20 items have got a schema-typical colour. Four items within each of the 20 item clusters were new items that were used for analysis. Since 32 items (relevant for analysis) had previously been presented in the study phase, 24 additional buffer items were developed to maintain a ratio of 1 : 1 for studied and unstudied items at test. Additional test buffer items were not taken into account for data analysis. Items in the test blocks were presented randomly.

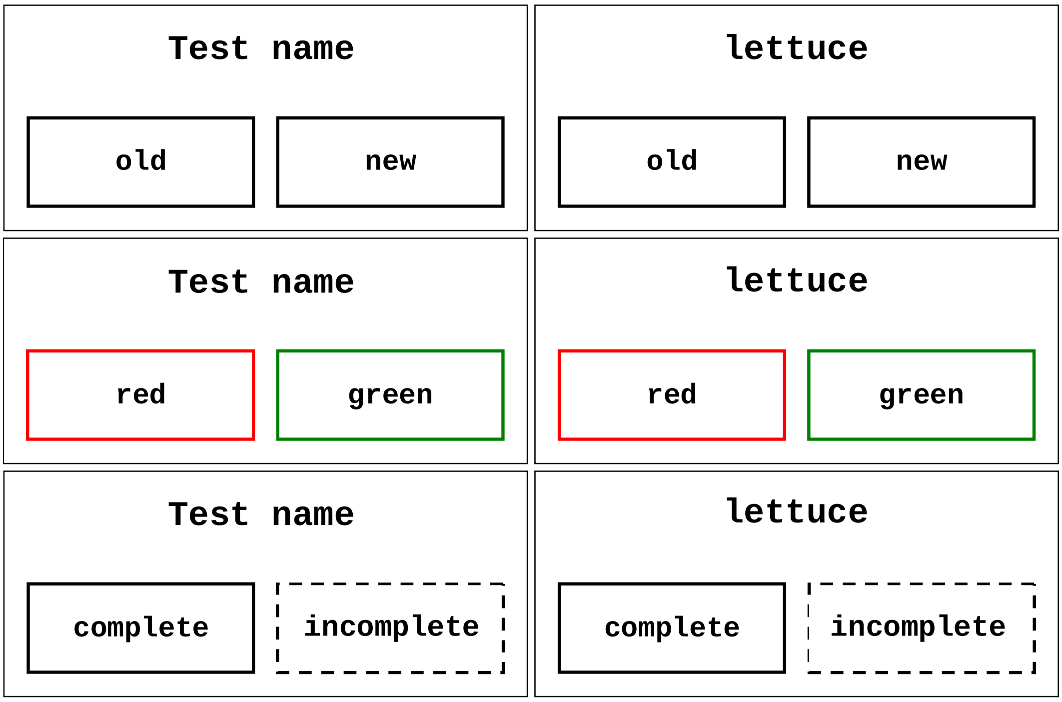

Figure 3.4: Illustration of the presentation slides for Experiment 3 at test (left-hand side) and the presentation slides for a specific example (right-hand side)

For each trial, participants were first presented with the name of the target drawing (if the item was old) or with the name of another *new* object (for (a) distractors or (b) test buffer items). Then, for each name, participants *first* had to decide whether they previously had or had not studied this object (“old” vs. “new”). When they decided that the item was new, another object name was presented to them. In contrast, when they decided that the drawing was an old drawing, participants had to make a decision about its presentation colour. Last, participants had to remember whether the drawing had been presented complete or incomplete. Each of these judgments was made via the computer mouse and each memory question was given on a separate slide. Responses were self-paced. An illustration of this sequence can be found in Figure 3.4.

After they had completed the experiment, participants were thanked and given their reward. Finally, debriefing took place, but only when a participant explicitly asked for it.

3.2.2 Experiment 3: Results

In the following section, the results of Experiment 3 are described. First, the impact of naming errors at study on memory at test is discussed, followed by information on the actual MPT models in use and their adaptations to the current experiment. Then, model fits are reported and commentated, before the experimental hypothesis tests are presented (a) for items without schema-typical colour and (b) for items with schema-typical colour.

As in Chapter 2, responses given by participants in the study phase were assessed. There were different types of responses produced by participants that did not match the pre-experimentally assigned picture name and therefore did not perfectly match the target name, which was presented in the test phase. However, some of these responses were still defined as correct responses, whereas other responses were defined as errors. Answers given by the participants were treated as correct, when participants produced (a) a match with the supposed target name, (b) a synonym, or (c) either a more general or a more specific than expected, though correct, term. All other responses deviating from the pre-determined correct solution were called errors. As can be seen in Appendix I, correct and erroneous responses can be categorised in several ways, bearing just some or bearing all of the experimental variations in mind. For the current section, analyses were run both for a data set including all target drawings (Set 1) and for another data set including only target drawings, which were named correctly at study (Set 2).

As discussed in Section 2.2.2, traditional model-free measures which are nevertheless widely used in generation effect research bear some disadvantages - such as being influenced easily by response biases. Thus, model-based analyses (e.g., with MPT models) are preferable. Especially in the present case, when researching schema-typicality, data may be specifically prone to response biases (e.g., Bayen et al., 1996). Presently, the model for crossed source information developed by Meiser and Bröder (2002) was employed.

In the present case, the factors “colour” (either red vs. green or typical colour vs. atypical colour) and “degree of completeness” were crossed. The source memory or source retrieval processes for the two source dimensions were assumed to be stochastically independent. In line with Riefer et al. (2007) and with Chapter 2 of this dissertation, several assumptions were applied to the basic model offered in Meiser and Bröder (2002).

In the present line of experiments, two separate models were designed to analyse the empirical data: one model for items that have no schema-typical colour and a second model for items that have got a schema-typical colour.

Table 3.4: List of parameter names, their function within the model, and a description of their interpretation used in Experiments 3, 4, and 5, for items without schema-typical colour

Parameters in the final model used for analyses in Experiments 3, 4, and 5; for items *without schema-typical colour*		
Name	Function of parameter in model	Description
$D_{Complete}$	Item memory	Probability of correctly detecting a previously presented complete item as old
$D_{Incomplete}$	Item memory	Probability of correctly detecting a previously presented incomplete item as old
D_{New}	Item memory	Probability of identifying a new item as new
a	Guessing	Probability of guessing "complete", when colour was identified correctly
b	Guessing	Probability of guessing "old" for non-detected items
g_1	Guessing	Probability of guessing "red", when colour is not known
g_2	Guessing	Probability of guessing "complete", when colour is guessed
$d_{c\ Complete}$	Source memory colour	Probability of correctly recalling colour of a complete item
$d_{c\ Incomplete}$	Source memory colour	Probability of correctly recalling colour of an incomplete item
$d_{g\ Complete}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of a complete item
$d_{g\ Incomplete}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of an incomplete item

For this second model, data obtained for items, when this schema-typical colour was red, and data obtained for items with schema-typical colour, when this schema-typical colour was green, were summarised. At first, it was considered whether memory performances for typically red and typically green items differed. To this end, the model was calculated for the types of items separately. However, since neither trends nor parameter values differed to a considerable degree (e.g., when comparing standard errors of parameter estimates), *both* types of items were included for analysis. Lists of parameters are given separately for these two models in Tables 3.4 and 3.5.

Table 3.5: List of parameter names, their function within the model, and a description of their interpretation used in Experiments 3, 4, and 5, for items with schema-typical colour (colour being either red or green)

Parameters in the final model used for analyses in Experiments 3, 4, and 5; for items *with schema-typical colour* (colour being either red or green)		
Name	Function of parameter in model	Description
$D_{Complete}$	Item memory	Probability of correctly detecting a previously presented complete item as old
$D_{Incomplete}$	Item memory	Probability of correctly detecting a previously presented incomplete item as old
D_{New}	Item memory	Probability of identifying a new item as new
a	Guessing	Probability of guessing "complete", when colour was identified correctly
b	Guessing	Probability of guessing "old" for non-detected items
g_1	Guessing	Probability of guessing "typical colour", when colour is not known
g_2	Guessing	Probability of guessing "complete", when colour was guessed
$d_{c\ Complete\ Atypical\ colour}$	Source memory colour	Probability of correctly recalling colour of a complete item, when the item was presented in atypical colour
$d_{c\ Complete\ Typical\ colour}$	Source memory colour	Probability of correctly recalling colour of a complete item, when the item was presented in typical colour
$d_{c\ Incomplete\ Atypical\ colour}$	Source memory colour	Probability of correctly recalling colour of an incomplete item, when the item was presented in atypical colour
$d_{c\ Incomplete\ Typical\ colour}$	Source memory colour	Probability of correctly recalling colour of an incomplete item, when the item was presented in typical colour
$d_{g\ Complete}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of a complete item
$d_{g\ Incomplete}$	Source memory degree of completeness	Probability of correctly recalling degree of completeness of an incomplete item

Items Without Schema-Typical Colour

The model fit the data for items that have no schema-typical colour well on the α = 5 % level (all G^2s < 5.55, all *p*s > 0.85). When comparing the goodness-of-fit statistics for the two separate sets of data (all data and data for drawings that were named correctly at study), no considerable differences could be found. This result is reasonable, when taking into account the low error rates for naming in the study phase. Since model fit was similar across both data sets and for different restrictions put on D_{New} ($D_{New} = D_{Complete}$, $D_{New} = D_{Incomplete}$, D_{New} = 0), the following analyses are based on all data by restricting D_{New} to $D_{Incomplete}$. Using the first data set instead of the second increased statistical power: G^2_{19df} = 5.55, p = 0.85. Note that analyses were additionally conducted for all of the remaining data sets and restrictions on D_{New}. They yielded the same results both concerning descriptive trends and statistical significance tests.

The following results are organised around illuminating the hypotheses and predictions.

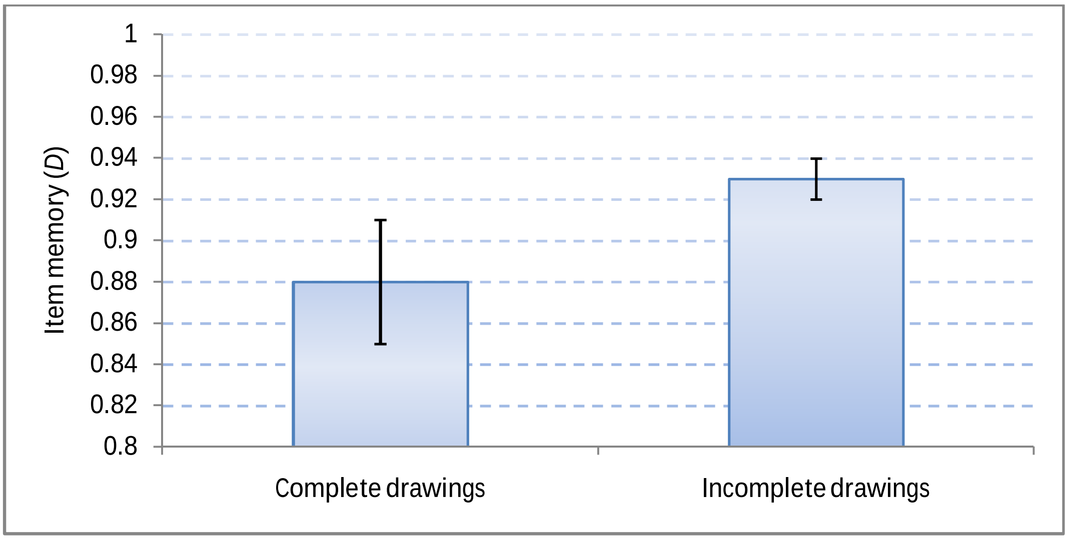

Figure 3.5: Item memory performances for items without schema-typical colour in Experiment 3 - error bars represent standard errors

According to *Hypothesis 1*, an advantage of incomplete over complete items was expected regarding item memory. Thus, a positive generation effect should have appeared.

As predicted, a positive generation effect for item detection emerged; incomplete drawings were remembered more often than complete drawings ($D_{Incomplete}$ = 0.93, $D_{Complete}$ = .88). This difference was statistically significant (ΔG^2_{1df} = 7.94, p < 0.01). Basically, the positive generation effect found in previous studies could be replicated in the current experiment; see Figure 3.5.

According to *Hypothesis 2*, regarding source memory for the degree of completeness, an advantage of incomplete over complete items was expected. A positive generation effect should have appeared.

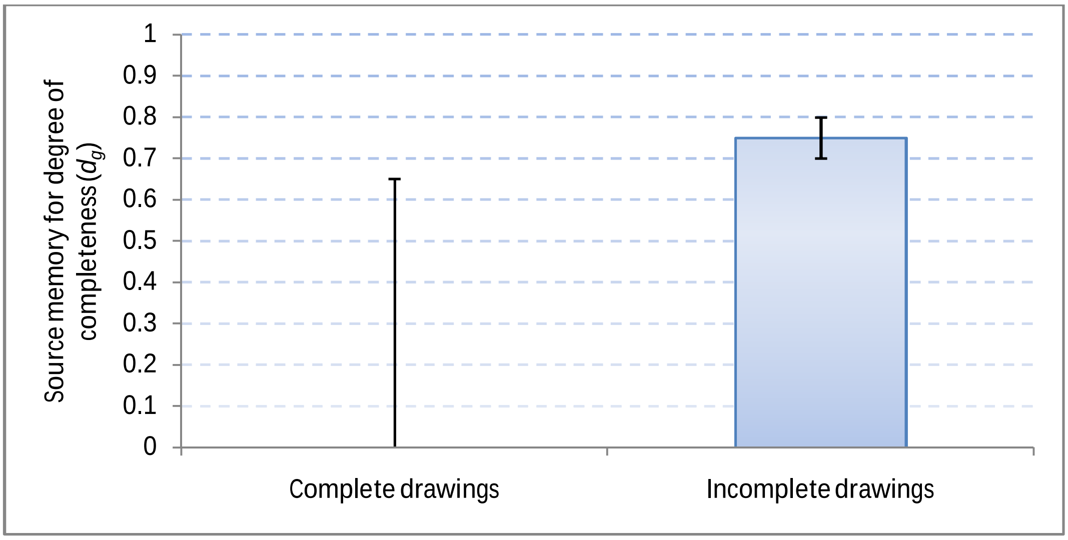

Figure 3.6: Source memory performances for the dimension degree of completeness for items without schema-typical colour in Experiment 3 - error bars represent standard errors

A positive generation effect emerged in source memory for the dimension degree of completeness; degree of completeness was remembered more often for incomplete drawings than for complete drawings ($d_{g\ Incomplete}$ = 0.75, $d_{g\ Complete}$ < 0.001). In spite of the large standard error for complete drawings, this difference was statistically significant (ΔG^2_{1df} = 4.67, p = 0.03). Basically, the positive generation effect for degree of completeness as found in previous studies and as theorised in the dual hypothesis by Riefer et al. (2007) could be replicated in the current experiment; see Figure 3.6.

According to *Hypotheses 3a and 3b*, regarding source memory for colour, an advantage of complete over incomplete items was expected. That is, a negative generation effect should have appeared. Guessing colour was anticipated to be at chance level.

As can be seen in Figure 3.7, there was a negative generation effect in source memory for the colour dimension; colour of complete drawings was remembered correctly more often than colour of incomplete drawings ($d_{c\ Complete}$ = 0.41, $d_{c\ Incomplete}$ = 0.28). This difference was statistically significant (ΔG^2_{1df} = 5.09, p = 0.02). Basically, the negative generation effect for presentation colour as found in previous studies and as theorised by Mulligan (2004) and Riefer et al. (2007) could be replicated in the current experiment.

All guessing parameters in an MPT model are in most cases expected to be at chance level (p = 0.50), however, due to the way in which colour was manipulated in the present experiment, it could be possible that participants assumed a strategy of guessing items to have been presented either in red or in green, respectively. Nevertheless, guessing "red" as the presentation colour (g_1 = 0.50) was not significantly different from chance level: ΔG^2_{1df} = 0.01, p = 0.94. Thus, guessing presentation colour was not biased.

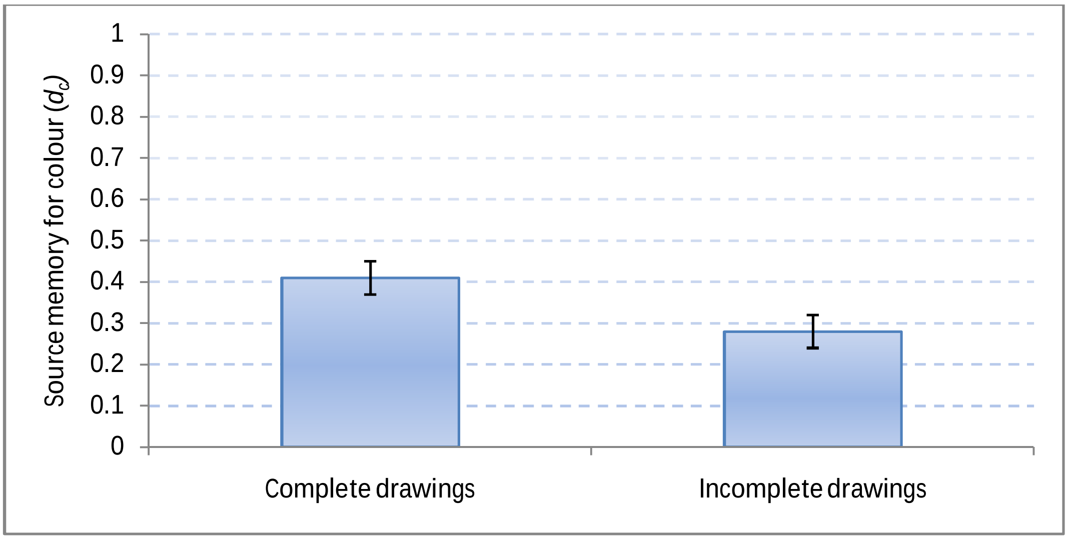

Figure 3.7: Source memory performances for the dimension colour for items without schema-typical colour in Experiment 3 - error bars represent standard errors

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.

Since these parameters would ideally reflect true guessing, it was tested whether parameters differed significantly from chance (p = 0.50). Guessing "old", and guessing "complete" when colour was identified correctly could not be set equal to 0.50; guessing "old" was significantly lower than chance level (b < *0.001;* ΔG^2_{1df} = 31.05, p < 0.001), whereas guessing "complete" when colour was identified correctly was significantly above chance level (a = 0.96; ΔG^2_{1df} = 8.39, p < 0.01). Guessing "complete" when colour was guessed (g_2 =

0.73) was not significantly different from chance (ΔG^2_{1df} = 2.40, p = 0.12). To summarise, participants had a bias to judge items to be new and to be complete, when colour was identified correctly, more often than average.

Items With Schema-Typical Colour

For the current analyses, data that have got a schema-typical colour were analysed by data set (all data vs. data for drawings named correctly at study) and for different restrictions put on D_{New} ($D_{New} = D_{Complete}$, $D_{New} = D_{Incomplete}$, D_{New} = 0). Additionally, data were analysed separately for items that typically appear in red and for items that typically appear in green colour. Instead of presenting all results in detail here, only the best solution concerning goodness-of-fit value and statistical power, is presented. The model fit the data well on the α = 5 % level for items that have got a schema-typical colour when using all data and when setting D_{New} equal to $D_{Complete}$: G^2_{8df} = 9.55; p = 0.30; the critical value for α = 0.05 % equalled 15.51. Note that analyses were additionally conducted for all of the remaining data sets and restrictions on D_{New}. They yielded the same results concerning descriptive trends and statistical significance tests.

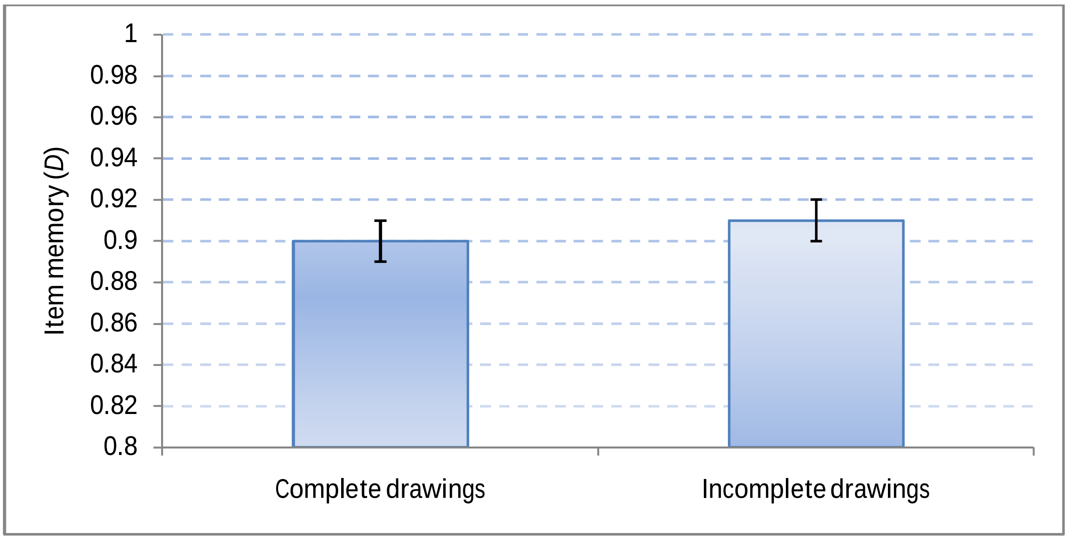

Figure 3.8: Item memory performances for items with schema-typical colour in Experiment 3 - error bars represent standard errors

According to *Hypothesis 1*, an advantage of incomplete over complete items was anticipated regarding item memory; i.e. a positive generation effect should have appeared.
Contrary to this prediction, item detection was about the same for incomplete drawings and complete drawings ($D_{Incomplete}$ = 0.91, $D_{Complete}$ = .90). Moreover, this difference was not statistically significant (ΔG^2_{1df} = 0.44, p = 0.51) and could be explained by the high overall values resulting in a ceiling effect; see Figure 3.8.

According to *Hypothesis 2*, regarding source memory for the degree of completeness, an advantage of incomplete over complete items was expected. A positive generation effect should have appeared.
Indeed, a trend of a positive generation effect emerged in source memory for the dimension degree of completeness at the descriptive level; completeness was remembered more often for incomplete drawings than for complete drawings ($d_{g\ Incomplete}$ = 0.73, $d_{g\ Complete}$ = 0.16). However, this difference was not statistically significant (ΔG^2_{1df} = 0.69, p = 0.41). It can be noted that there was a large standard error for complete drawings. Basically, the positive generation effect for the degree of completeness as found in previous studies and as theorised in the dual hypothesis by Riefer et al. (2007) could be replicated in the current experiment at the descriptive level only; see Figure 3.9.

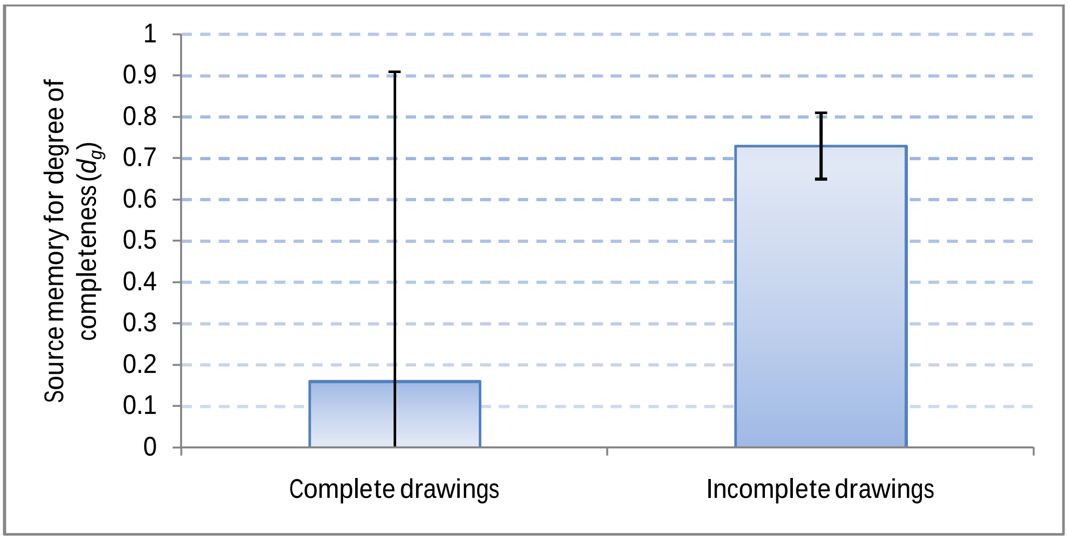

Figure 3.9: Source memory performances for the dimension degree of completeness for items with schema-typical colour in Experiment 3 - error bars represent standard errors

According to *Hypotheses 3a and 3b*, regarding source memory for colour, an effect was expected of whether items were presented in their typical or in their atypical colour, moderated by the degree of completeness. Guessing colour could potentially have been biased in favour of the typical colour.

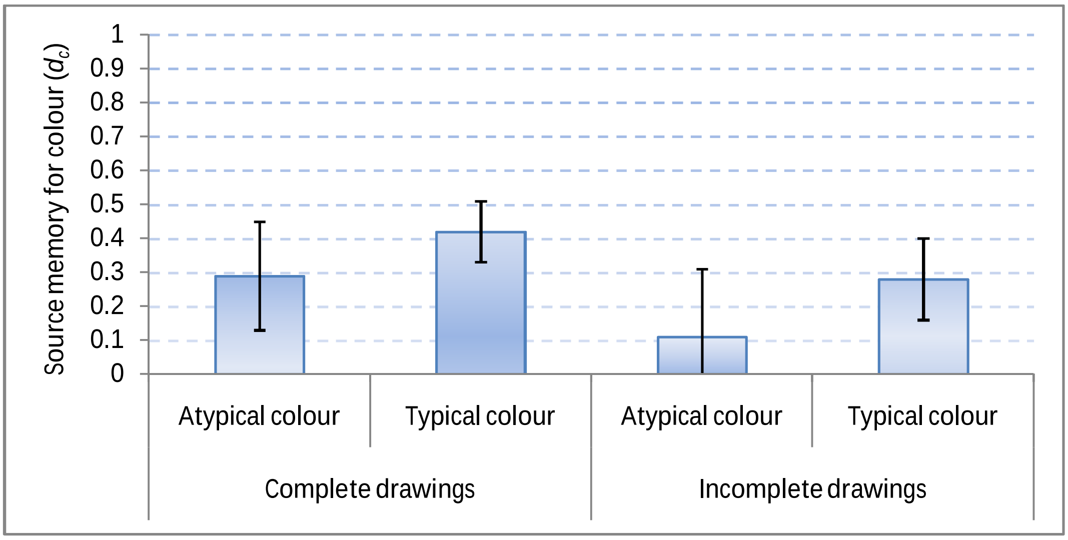

Figure 3.10: Source memory performances for the dimension colour for items with schema-typical colour in Experiment 3 - error bars represent standard errors

As can be seen in Figure 3.10, there was a trend at the descriptive level towards a negative generation effect of colour for drawings presented in a typical colour ($d_{c\ Complete\ Typical\ colour}$ = 0.42; $d_{c\ Incomplete\ Typical\ colour}$ = 0.28) and for drawings presented in an atypical colour ($d_{c\ Complete\ Atypical\ colour}$ = 0.29; $d_{c\ Incomplete\ Atypical\ colour}$ = 0.11). That is, the colour of complete drawings was remembered more often than the colour of incomplete drawings. However, these trends were not statistically significant: ΔG^2_{1df} = 3.16, p = 0.08 in the former and ΔG^2_{1df} = 2.77, p = 0.10 in the latter case. Additionally, there was a trend at the descriptive level towards a memory advantage of typical items over atypical items. However, no significant difference could be found, when comparing colour memory of complete and incomplete drawings when aggregating across presentation colour: ΔG^2_{1df} = 0.24, p = 0.62 for complete drawings; ΔG^2_{1df} = 0.25, p = 0.61 for incomplete drawings. Also, when setting all source memory parameters equal, model fit did not decrease significantly (ΔG^2_{3df} = 6.53, p = 0.09).

Basically, regarding source memory for colour, no significant effect of whether items were presented in their typical or in their atypical colour could be found. Moreover, no moderating effect of degree of completeness was obvious.

Guessing colour, i.e., guessing that presentation colour did or did not match the typical colour (g_1 = 0.40), was not significantly different from chance level: ΔG^2_{1df} = 0.61, p = 0.44. Thus, guessing colour was not biased in favour of the typical colour.

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.
Since the parameters would ideally reflect true guessing, it was tested whether the parameters differed significantly from chance (p = 0.50). Guessing "old" was significantly below chance level (b = 0.08; ΔG^2_{1df} = 26.49, $p < 0.001$). In comparison, guessing "complete" when colour was guessed and guessing "complete" when colour was identified correctly were not significantly different from chance level (g_2 = 0.73; ΔG^2_{1df} = 0.62, p = 0.43; a = 0.88; ΔG^2_{1df} = 0.75, p = 0.39). Thus, participants had a bias to judge items to be new more often than average.

Note that a comprehensive discussion of the results and conclusions obtained from Experiment 3 (involving drawings and an intentional study design) can be found in Section 3.3.

3.2.3 Experiment 4: Methods

In this section, Experiment 4 is described in detail. Experiment 4 is the second of three parallel experiments designed to answer the following question: Does schema-typicality of object colour affect colour memory for complete versus incomplete items and to what degree is this influence? Here, photos are presented as material and the memorisation of these is intentional.

Participants

Sixty-five persons participated in Experiment 4. One person in the sample did not fill in the demographic questionnaire; nevertheless, his or her data were considered for analysis, because the pattern of results remained the same whether or not these data were excluded from the sample. Moreover, introducing this person's data into the set increased statistical power. However, the description of the demographic data can of course only refer to the sample of 64 participants. Of these 64 participants, 87.50 % were female. Participants' age ranged from 19 to 35 with a mean of 21.62 years and a median of 20 years. No participant suffered from red-green colour blindness. All participants except five confirmed that they were native German speakers. However, the latter group of participants stated that their German proficiency skills were either "very good" (3 out of 5) or "good" (2 out of 5). All participants were currently enrolled at the University of Mannheim and were primarily students from the fields of Psychology (71.87 %) or Sociology (17.19 %). Other fields of study were Language Studies, Economic Sciences, and Law.

Design and Material

The generation effect was investigated for pictorial material, i.e. for photos, which were either red or green. In Experiment 4, the same independent variables were employed as were described in the methods section of Experiment 3. One half of the presented photos was assigned to the read condition and the other half to the generate condition. In the read condition, cue words and the target pictures were presented complete. In the generate condition, cue words were shown complete, whereas the photos were fragmented, i.e. appeared in an incomplete manner.

The picture fragmentation tool “Frag” (Snodgrass et al., 1987) could unfortunately not be applied; due to technical requirements of the programme, that could not be fulfilled by the present stimuli. However, the same logic implemented in this computer programme was followed when manipulating study stimuli. One half of the photos was presented complete and the other half was presented incomplete. As in Experiment 3, one half of the target photos was presented red and the other half was presented green.

The target stimuli consisted of 60 photographs of common objects (e.g., raspberry, lettuce, or armchair). Most of these photos could be adopted from Naor-Raz, Tarr, and Kersten (2003) - they are provided by Tarrlab (2009, April 15). Some equivalent photos were additionally developed by myself. Thirty of the 60 photos were pictures of items that have got a schema-typical colour; i.e., items that are either typically red (15 photos, e.g., raspberry) or items that are typically green (15 photos, e.g., lettuce). The remaining 30 items were objects that have no schema-typical colour (e.g., armchair). The 15 typically red items, the 15 typically green items and the 30 items without schema-typical colour were randomly divided into five stimulus sets. These sets served as counterbalancing groups (i.e., CB); see Table 3.7, for an illustration.

Thirty-two further photos were selected from Naor-Raz et al. (2003). One half of these served as primacy buffer items and one half served as recency buffer items. Additionally, all practice items were pictures of objects that have no schema-typical colour. Neither primacy buffer items nor recency buffer items or practice items were considered for analyses.

The cues consisted of words that are semantically associated with the target photos. Cues were developed by myself and validated by a set of 15 participants (Pilot V, see Appendix F).

At test, all studied photos and the distractor items were presented. Additionally, 36 new items (i.e., buffer items) were developed for test, to fulfil the 1 : 1 ratio of studied versus unstudied items. These buffer items were not included in the analyses of the results.

A complete list of all cues and targets used in this experiment can be found in Appendix H.

Table 3.6: Research design for Experiment 4 employing the within-subject variations "colour", „degree of completeness", and "item type"

Item type	CB	Old / studied items				New / unstudied items
		Red		Green		
		Complete	Incomplete	Complete	Incomplete	
(1) Items without schema-typical colour	1	6 items	6 items	6 items	6 items	6 items
	2	6 items	6 items	6 items	6 items	6 items
	3	6 items	6 items	6 items	6 items	6 items
	4	6 items	6 items	6 items	6 items	6 items
	5	6 items	6 items	6 items	6 items	6 items
(2) Items with schema-typical colour (red)	1	3 items	3 items	3 items	3 items	3 items
	2	3 items	3 items	3 items	3 items	3 items
	3	3 items	3 items	3 items	3 items	3 items
	4	3 items	3 items	3 items	3 items	3 items
	5	3 items	3 items	3 items	3 items	3 items
(3) Items with schema-typical colour (green)	1	3 items	3 items	3 items	3 items	3 items
	2	3 items	3 items	3 items	3 items	3 items
	3	3 items	3 items	3 items	3 items	3 items
	4	3 items	3 items	3 items	3 items	3 items
	5	3 items	3 items	3 items	3 items	3 items

Procedure

The procedure was similar to that used in Experiment 3 and exhibited only minor changes as compared to it: First, participants were working on the tasks presented to them via computer in group sessions of up to 4 persons, instead of groups of up to 20 persons. Second, four instead of three items were used in the practice phase. Furthermore, differences between Experiments 3 and 4 arose due to the use of other pictorial material: In the current study, photos of objects were used instead of line drawings of objects. This main difference entailed essential changes which are outlined below:

At study, 48 items were presented that were used for analysis; 24 of these have no schema-typical colour, whereas the other 24 items have got a schema-typical colour. Since 48 items within each study block seemed to be too easy to remember and also in order to account for primacy and recency effects on memory performance, these items were preceded and followed by 16 additional items, which were not included in the data analysis. Consequently, 80 items were displayed at study.

At test, 60 items were used for analyses; 30 of these have no schema-typical colour, whereas the other 30 items have got a schema-typical colour. Six items within each of the 30 item clusters were new items that were used for analysis. Since 48 items (relevant for analysis) had previously been presented in the study phase, 36 additional test buffer items were developed to maintain a ratio of 1 : 1 (studied : unstudied) items at test. These additional test buffer items were not considered for data analysis.

3.2.4 Experiment 4: Results

In the following section, the results of Experiment 4 are described. Model fits are reported and commentated, before experimental hypothesis tests are presented (a) for items without schema-typical colour and (b) for items with schema-typical colour.

Items Without Schema-Typical Colour

When attempting to employ the same model as it was applied in Experiments 3 and 5 to items that have no schema-typical colour, no MPT analyses could be conducted. The model fit the data neither on the α = 5 % nor on the α = 1 % level, as can be seen in Figure 3.7.

Table 3.7: List of G^2 and p values for the more restricted model applied in Experiment 4

Data set global model for	Restriction put on D_{New} (df = 10)	G^2	p	Critical G^2 for α = 5 %	Critical G^2 for α = 1 %	Fit
All data (1)	$D_{New} = 0$	35.23	< 0.001	18.31	23.21	No fit
	$D_{New} = D_{Complete}$	34.84	< 0.001			No fit
	$D_{New} = D_{Incomplete}$	34.70	< 0.001			No fit
Data for items named correctly (2)	$D_{New} = 0$	33.56	< 0.001	18.31	23.21	No fit
	$D_{New} = D_{Complete}$	33.16	< 0.001			No fit
	$D_{New} = D_{Incomplete}$	33.03	< 0.001			No fit

Therefore, the model was slightly altered by allowing for parameter differences (in D, d_c, and d_g) for presentation colour of items. Consequently, these parameters were doubled to reflect memory performances for items presented in red colour and items presented in green colour, respectively (see Table 3.8).

This new model fit the data well on the α = 5 % level. When comparing the goodness-of-fit statistics for the two separate sets of data, no considerable difference could be found. This result can be explained, when taking into account the low error rates for naming in the study phase (see Appendix I). Since model fit was similar across both data sets and for different restrictions put on D_{New}, the following analyses were based on all data by restricting D_{New} to

$D_{Complete\ red}$. Using the first data set instead of the second increased statistical power: G^2_{4df} = 6.74, p = 0.15. Note that analyses were additionally conducted for all of the remaining data sets and restrictions listed in Table 3.8. They yielded the same results concerning descriptive trends and statistical significance tests.

Table 3.8: List of G^2 and p values for the less restricted model applied in Experiment 4

Data set global model for	Restriction on D_{New} (df = 4)	G^2	p	Critical G^2 for α = 5 %	Critical G^2 for α = 1 %	Fit on
All data (1)	$D_{New} = D_{Complete\ green}$	7.26	0.12	9.49	13.28	5 % level
	$D_{New} = D_{Complete\ red}$	6.74	0.15			5 % level
	$D_{New} = D_{Incomplete\ green}$	6.91	0.14			5 % level
	$D_{New} = D_{Incomplete\ red}$	7.04	0.13			5 % level
	$D_{New} = 0$	7.54	0.11			5 % level
Data for items named correctly (2)	$D_{New} = D_{Complete\ green}$	8.37	0.08	9.49	13.28	5 % level
	$D_{New} = D_{Complete\ red}$	7.80	0.10			5 % level
	$D_{New} = D_{Incomplete\ green}$	7.99	0.09			5 % level
	$D_{New} = D_{Incomplete\ red}$	8.13	0.09			5 % level
	$D_{New} = 0$	8.68	0.07			5 % level

According to *Hypothesis 1*, an advantage of incomplete over complete items was expected regarding item memory, i.e., a positive generation effect should have appeared.

A positive generation effect for item detection emerged for photos presented in green; incomplete photos were remembered more often than complete photos ($D_{Incomplete\ green}$ = 0.95, $D_{Complete\ green}$ = .93). This difference was statistically significant (ΔG^2_{1df} = 6.93, p < 0.01). However, for pictures presented in red, item detection did not differ significantly for complete ($D_{Complete\ red}$ = 0.96) and incomplete ($D_{Incomplete\ red}$ = 0.93) photos: ΔG^2_{1df} = 1.86, p = 0.17. Presentation colour led to significant differences, when comparing item memory for complete photos (ΔG^2_{1df} = 11.11, p = 0.01), but not, when comparing item memory for incomplete photos (ΔG^2_{1df} = 0.42, p = 0.52); see Figure 3.11.

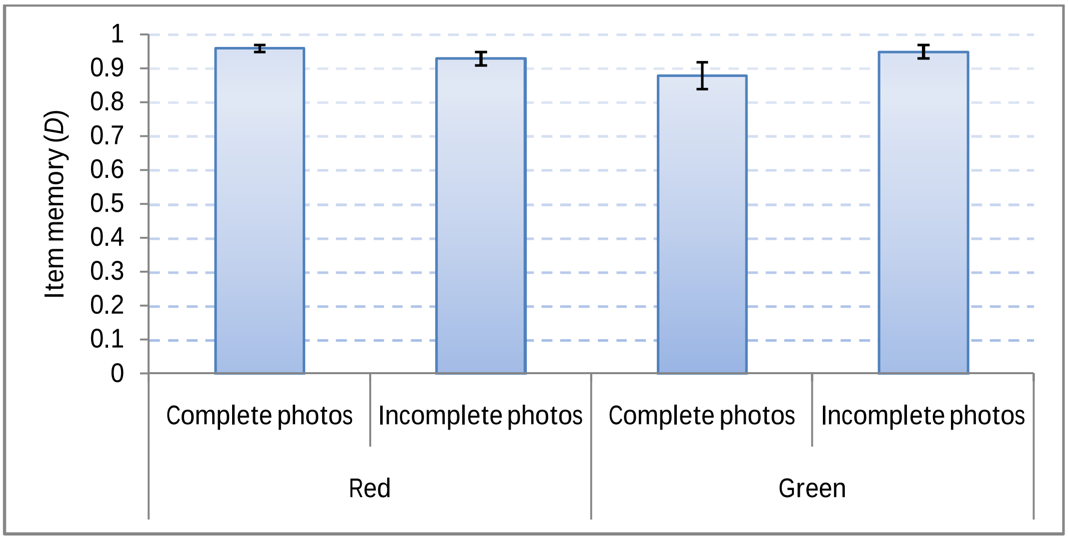

Figure 3.11: Item memory performances for items without schema-typical colour in Experiment 4 - error bars represent standard errors

According to *Hypothesis 2*, regarding source memory for the degree of completeness, an advantage of incomplete over complete items was expected. That is, a positive generation effect should have appeared.

There was a trend of a positive generation effect in source memory for the dimension degree of completeness at the descriptive level for red and green photos. For pictures presented in red, completeness of incomplete photos was remembered more often than of complete photos ($d_{g\ Incomplete\ red} = 0.54$, $d_{g\ Complete\ red} < 0.001$); However, this difference was not statistically significant ($\Delta G^2_{1df} = 1.50$, $p = 0.22$). For photos presented in green, completeness of incomplete photos was remembered more often than of complete photos ($d_{g\ Incomplete\ green} = 0.56$, $d_{g\ Complete\ green} = 0.35$); However, this difference became statistically significant only, when first setting guessing parameters that were not significantly different from chance level to 0.50: $\Delta G^2_{1df} = 15.19$, $p < 0.001$. Otherwise, this difference was not statistically significant ($\Delta G^2_{1df} = 0.51$, $p = 0.48$).

Presentation colour did not lead to significant memory differences for complete items ($\Delta G^2_{1df} = 2.47$, $p = 0.12$) or for incomplete items ($\Delta G^2_{1df} = 0.13$, $p = 0.72$). However, in a model in which no difference for presentation colour concerning the source memory dimension degree of completeness was included, no significant difference emerged when comparing complete photos ($d_{g\ Complete} = 0.45$) and incomplete photos ($d_{g\ Incomplete} = 0.51$): $\Delta G^2_{1df} = 0.02$, $p = 0.88$.

Basically, the positive generation effect for the degree of completeness as found in previous studies and as theorised in the dual hypothesis by Riefer et al. (2007) could be replicated in the current experiment at the descriptive level only; see Figure 3.12.

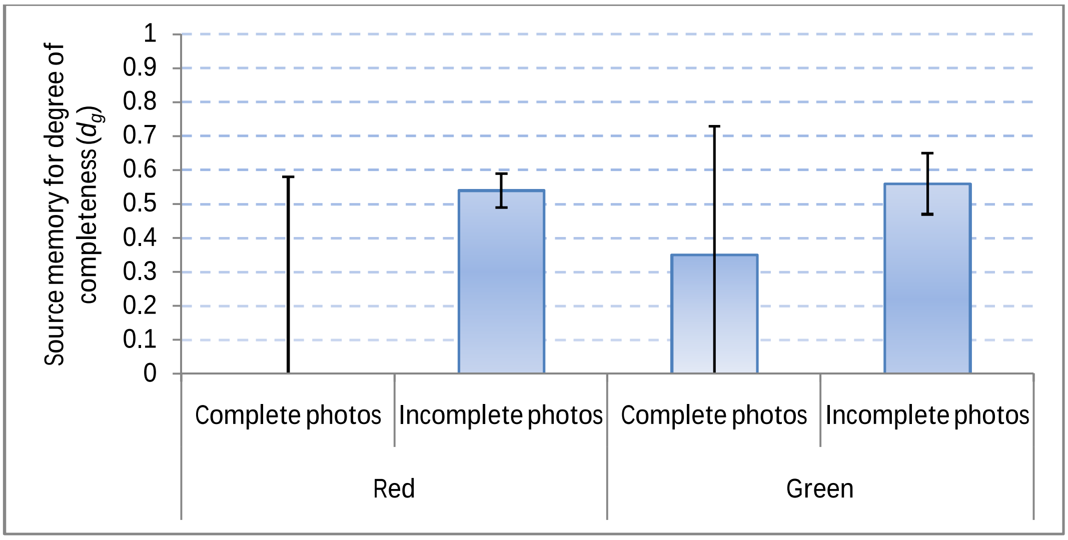

Figure 3.12: Source memory performances for the dimension degree of completeness for items without schema-typical colour in Experiment 4 - error bars represent standard errors

According to *Hypotheses 3a and 3b*, regarding source memory for colour, an advantage of complete over incomplete items was expected. That is, a negative generation effect should have appeared. Guessing colour was expected to be at chance level.

As can be seen in Figure 3.13, there was a negative generation effect in source memory for the colour dimension; colour of complete photos was remembered correctly more often than colour of incomplete photos when they were presented in red or in green ($d_{c\ Complete\ red}$ = 0.84, $d_{c\ Incomplete\ red}$ = 0.69; $d_{c\ Complete\ green}$ = 0.32, $d_{c\ Incomplete\ green}$ < 0.001). This difference was statistically significant for red pictures (ΔG^2_{1df} = 10.26, p < 0.01) and for green pictures (ΔG^2_{1df} = 12.01, p < 0.001). Including presentation colour lead to significant differences when comparing complete photos (ΔG^2_{1df} = 10.97, p < 0.01), and when comparing incomplete photos (ΔG^2_{1df} = 9.33, p < 0.01).

Basically, the negative generation effect for colour found in previous studies and as theorised by Mulligan (2004) and Riefer et al. (2007) could be replicated in the current experiment.

All guessing parameters in an MPT model are generally suggested to be at chance level. However, due to the way colour was manipulated in the present study, it could be possible that participants assumed a strategy of guessing items to have been presented either in red or in green. As a result, guessing “red” as the presentation colour (g_1 = 0.36) was significantly below chance level: ΔG^2_{1df} = 4.39, p = 0.04. This result indicates that guessing presentation colour was biased in favour of the green colour.

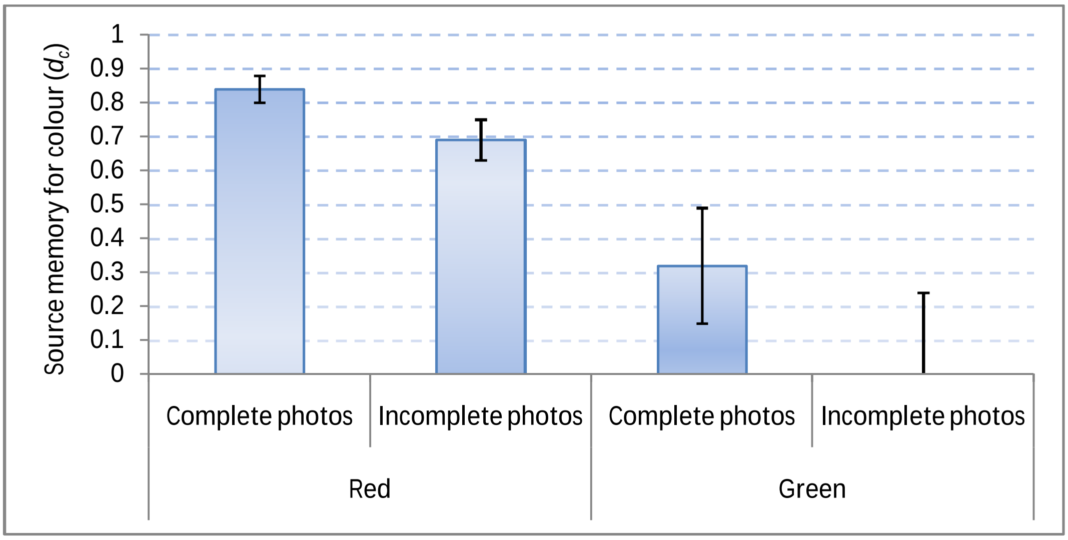

Figure 3.13: Source memory performances for the dimension colour for items without schema-typical colour in Experiment 4 - error bars represent standard errors

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.

Since the parameters would ideally reflect true guessing, it was tested whether parameters differed significantly from chance (p = 0.50). Guessing “old” (b = 0.42), and guessing “complete”, when participants guessed the colour of the items (g_2 = 0.73), could be set equal to 0.50: ΔG^2_{1df} = 0.38, p = 0.54 for b and ΔG^2_{1df} = 3.73, p = 0.05 for g_2. Guessing “complete” when participants correctly remembered the colour of the items could not be set equal to 0.50 and was significantly above chance level (a = *0.90*; ΔG^2_{1df} = 4.80, p = 0.03). Participants had a bias to judge items to be complete when colour was identified correctly.

Items With Schema-Typical Colour

For the current analyses, data for items that have got a schema-typical colour were analysed by data set (all data vs. data for photos named correctly at study) and for different restrictions put on D_{New} ($D_{New} = D_{Complete}$, $D_{New} = D_{Incomplete}$, D_{New} = 0). Additionally, data were analysed separately for items that typically appear in red and for items that typically appear in green. Instead of presenting all results in detail here, only the best solution concerning goodness-of-fit value and statistical power is presented: The model fit the data well on the α = 5 % level for items with a schema-typical colour, when using all data and when setting D_{New} equal to $D_{Complete}$: G^2_{8df} = 12.24; p = 0.14; critical value for 5 % equals 15.51. Note that analyses were additionally conducted for all of the remaining data sets and restrictions placed on D_{New}. They yielded the same results concerning descriptive trends and statistical significance tests.

According to *Hypothesis 1*, an advantage of incomplete over complete items was expected regarding item memory, i.e., a positive generation effect should have appeared.
Item detection was about the same for incomplete photos and complete photos ($D_{Incomplete}$ = 0.94, $D_{Complete}$ = .95). Moreover, this difference was not statistically significant (ΔG^2_{1df} = 0.07, p = 0.79) and could be explained by the very high overall values resulting in a ceiling effect; see Figure 3.14.

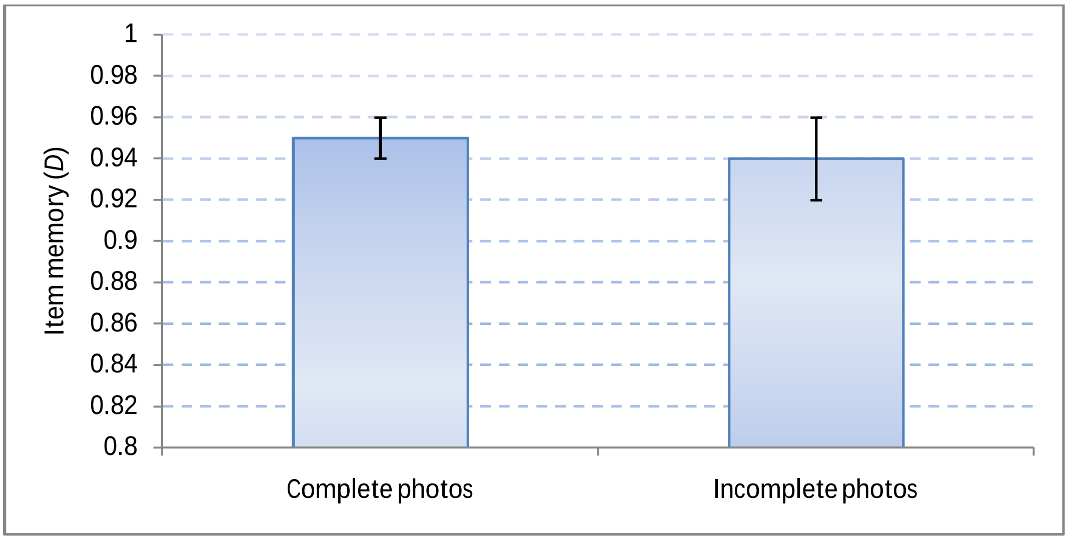

Figure 3.14: Item memory performances for items without schema-typical colour in Experiment 4 - error bars represent standard errors

According to *Hypothesis 2*, regarding source memory for the degree of completeness, an advantage of incomplete over complete items was expected, i.e., a positive generation effect should have appeared.

In line with this prediction, a positive generation effect emerged in source memory for the dimension degree of completeness at the descriptive level; memory for degree of completeness was better for incomplete photos than for complete photos ($d_{g\ Incomplete}$ = 0.51, $d_{g\ Complete}$ < 0.001). However, this difference was not statistically significant (ΔG^2_{1df} = 2.70, p = 0.10). It can be noted that there was a large standard error for complete pictures. Basically, the positive generation effect for degree of completeness as found in previous studies and as theorised in the dual hypothesis by Riefer et al. (2007) could be replicated in the current experiment at the descriptive level only; see Figure 3.15.

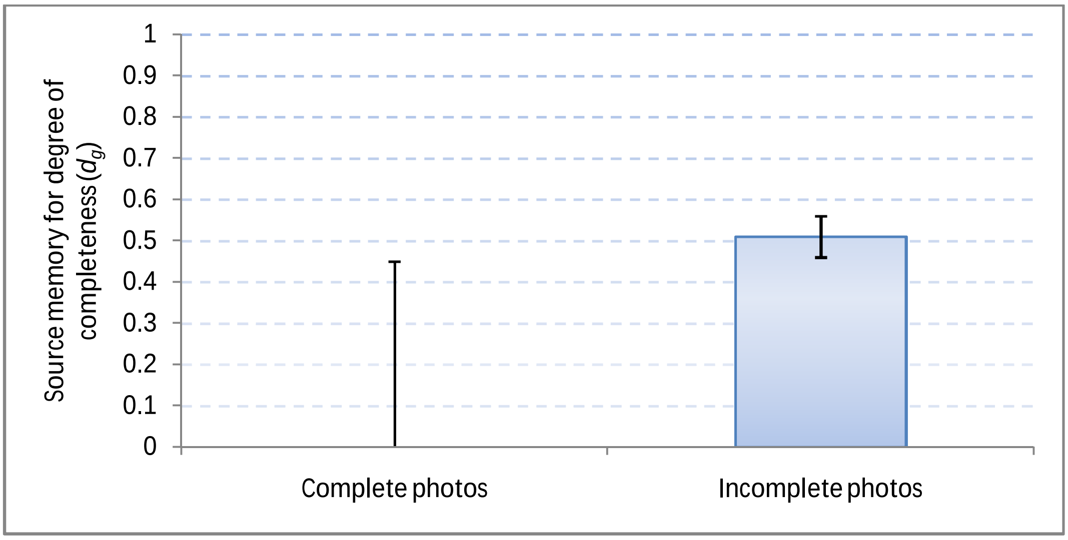

Figure 3.15: Source memory performances for the dimension degree of completeness for items without schema-typical colour in Experiment 4 - error bars represent standard errors

According to *Hypotheses 3a and 3b*, regarding source memory for colour, an effect was expected of whether items were presented in their typical or their atypical colour, moderated by the degree of completeness. Guessing colour could potentially be biased in favour of the typical colour.

As can be seen in Figure 3.16, there was a trend at the descriptive level towards a negative generation effect for items presented in a typical colour ($d_{c\ Complete\ Typical\ colour}$ = 0.53; $d_{c\ Incomplete\ Typical\ colour}$ = 0.35) and for items presented in an atypical colour ($d_{c\ Complete\ Atypical\ colour}$ = 0.79; $d_{c\ Incomplete\ Atypical\ colour}$ = 0.47). That is, colour memory for incomplete photos was better than colour memory for complete photos. These trends were statistically significant: ΔG^2_{1df} = 7.53, p = 0.01 in the former and ΔG^2_{1df} = 27.56, p < 0.001 in the latter case.

Additionally there was a trend at the descriptive level towards an advantage of atypical photos over typical photos. A significant difference could be found, when comparing complete and incomplete pictures irrespective of colour for complete items (ΔG^2_{1df} = 6.48, p = 0.01). For incomplete items, a significant effect existed, when restricting all guessing parameters that were not significantly different from 0.50 to chance level: ΔG^2_{1df} = 7.03, p = 0.01. Note that without these restrictions put on guessing parameters, no significant effect occurred: ΔG^2_{1df} = 0.43, p = 0.51.

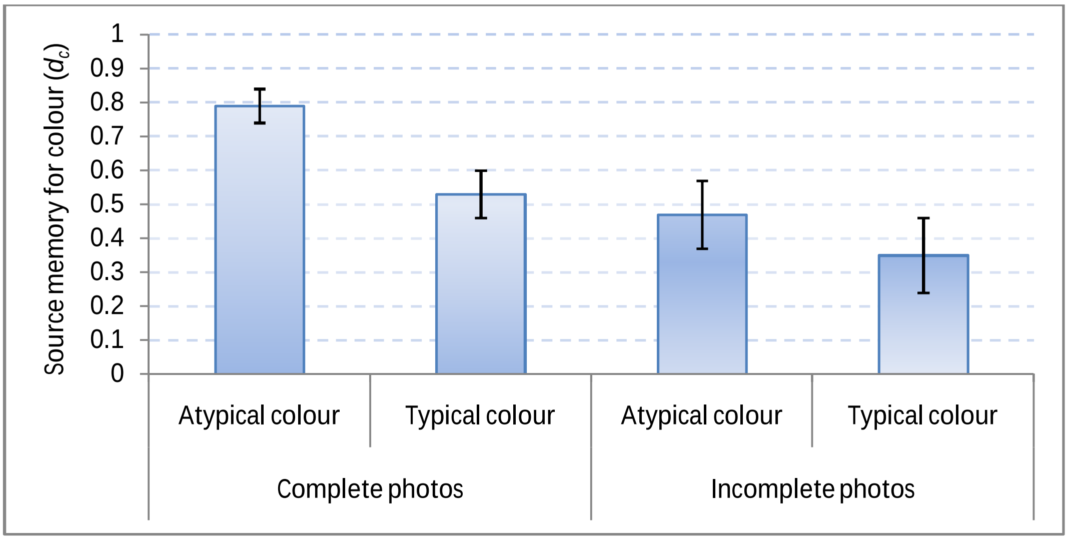

Figure 3.16: Source memory performances for the dimension colour for items with schema-typical colour in Experiment 4 - error bars represent standard errors

When considering the differences between parameters for items presented in their typical colour and items presented in their atypical colour for complete and incomplete items separately, one can see that these differences existed at the descriptive level. Thus, one could test whether the difference between two parameters (photos presented in typical colour and photos presented in atypical colour) differs between complete and incomplete items. However, with MPT models, equality or inequality of differences cannot be tested in this way – instead it can be tested whether or not a relative increment is equal across groups. To do so, the model needs to be reparameterised such that an effect of one parameter is decomposed into a main effect and an increment, which is then tested by restricting the incremental parameters to be equal across groups. However, when comparing the increments for incomplete items and complete items, no significant difference could be found: ΔG^2_{1df} = 0.09, p = 0.77.

Basically, a negative generation effect emerged for memory for colour. Regarding source memory for colour, a significant effect of whether photos were presented in their typical or in their atypical colour occurred. However, no moderating effect of degree of completeness could be found.

Guessing colour, i.e., guessing that presentation colour did or did not match the typical colour, (g_1 = 0.47), was not significantly different from chance level: ΔG^2_{1df} = 0.14, p = 0.71. Thus, guessing colour was not biased in favour of the typical colour.

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.

Since the parameters ideally reflect true guessing, it was tested whether the parameters differed significantly from chance (i.e., from p = 0.50). Guessing "old" could be set equal to 0.50, and was thus not significantly different from it (b = *0.49;* (ΔG^2_{1df} = 0.01, p = 0.92). Guessing "complete" when participants guessed the colour of the items and guessing "complete" when participants correctly remembered the colour of the items were significantly above chance level (g_2 = 0.76; ΔG^2_{1df} = 9.42, p < 0.01; a = 0.94; ΔG^2_{1df} = 15.15, p < 0.001).

Thus, participants had a bias to judge items to be complete.

Note that a comprehensive discussion of the results and conclusions obtained from Experiment 4 (involving photos and an intentional study design) can be found in Section 3.3.

3.2.5 Experiment 5: Methods

In this section, Experiment 5 is described in detail and information on the sample, on the design, on the material, and on the experiment procedure is provided. Experiment 5 is the last of three parallel experiments designed to answer the following question: Does schema-typicality of object colour affect colour memory for complete versus incomplete items and to what degree is this influence? Here, photos are presented as material and the memorisation of these is incidental.

Participants

Forty-nine persons participated in Experiment 5. Two persons had to be excluded from the sample due to red-green colour blindness. Of the remaining 47 participants, 76.60 % were female. Participants' age ranged from 18 to 28 with a mean of 22.25 years and a median of 21 years. All participants except two confirmed that they were native German speakers. However, these latter two participants stated that their German proficiency skills were either "very good" or "good". One of the participants had already finished her studies, whereas the other 46 participants were currently enrolled at the University of Mannheim. They primarily were students from the field of Psychology (45.65 %). Other fields of study were Language Studies (19.56 %), Sociology (15.22 %), Economic Sciences (15.22 %), and Law (4.35 %).

Design and Material

Design and material were the same as in Experiment 4.

Procedure

The procedure of Experiment 5 was almost the same as for Experiment 4. Differences were caused by the one critical variation between the two experiments, namely by the intentionality of study: In Experiment 4, *intentional study instructions* were employed, whereas *incidental study instructions* were used for the present study.

For Experiment 5, participants were recruited by offering to take part in a study on human perception. No mention was made of memory or learning. During study, participants were to read or self-generate target names and to report them on a response sheet (as in Experiment 4). It was explained to the participants that they should simply make a line when they could not identify what the object was. To increase the credibility of the cover story that the task was about human perception, participants were asked to place a finger (left index finger, if right-handed; right index finger, if left-handed) on a specific key of the keyboard (d-key with an orange label, if right-handed; k-key with a blue label, if left-handed). Upon identification of the target object, participants were to press the according key as quickly as they could. Only after they had pressed the key were they allowed to write down the name of the object. Keys used for this task were the same as those used in the distractor phase.

Distractor Phase and Test Phase

After the study phase, participants worked on the distractor phase (arithmetic task). Finally, the surprise test phase followed. As in Experiments 3 and 4, participants were shown names of objects and asked to judge (a) whether photos of these objects had been presented to them in the course of this study, (b) whether these photos had been in red or green, and (c) whether these photos had been presented complete or incomplete.

Note

In previous generation effect studies concerned with source memory (e.g., Mulligan, 2004; Mulligan et al., 2006; Riefer et al., 2007), participants were informed that their memory would eventually be tested for the targets and for source attributes of the target. Some other studies have investigated the impact of generation on indirect measures of memory. However, these studies did not implement incidental study instructions, but were rather focused on effects of generation on subsequent word completion or word identification (e.g., Schwartz, 1989). Therefore, strictly speaking, no generation effect studies exist of the effect of an incidental study task. Nevertheless, the same hypotheses as tested for Experiments 3 and 4 were assumed for Experiment 5. Thus, alls results might be affected by employing an incidental study design.

3.2.6 Experiment 5: Results

In the following section, the results of Experiment 5 are described. Model fits are reported and commentated, before experimental hypotheses are tested (a) for items without schema-typical colour and (b) for items with schema-typical colour.

Items Without Schema-Typical Colour

The model fit the data for items that have no schema-typical colour well on the α = 5 % level (all G^2s < 17.48, all *p*s > 0.06). When comparing the goodness-of-fit statistics for the two separate sets of data (all data and data for photos that were named correctly at study), no considerable difference could be found. This result can be explained, when considering the low error rates for naming in the study phase (see Appendix I). Since model fit was similar across both data sets and for different restrictions put on D_{New} ($D_{New} = D_{Complete}$, $D_{New} = D_{Incomplete}$, $D_{New} = 0$), the following analyses were based on all data by restricting D_{New} to 0. Using all data instead of only data for photos that were named correctly at study increased statistical power: $G^2_{10df} = 16.80$, $p = 0.08$. Note that analyses were additionally conducted for all of the remaining data sets and restrictions on D_{New}. They yielded the same results concerning trends at the descriptive level and statistical significance tests.

According to *Hypothesis 1*, an advantage of incomplete over complete items was expected regarding item memory, i.e., a positive generation effect should have appeared.
In line with this prediction, a positive generation effect emerged for item detection at the descriptive level; incomplete photos were remembered more often than complete photos ($D_{Incomplete} = 0.96$, $D_{Complete} = .94$). However, this difference was marginally not significant ($\Delta G^2_{1df} = 3.72$, $p = 0.05$). Basically, the positive generation effect in item memory as found in previous studies could be replicated in the current experiment at the descriptive level only. Nevertheless, this not significant effect could potentially be due to a ceiling effect; see Figure 3.17.

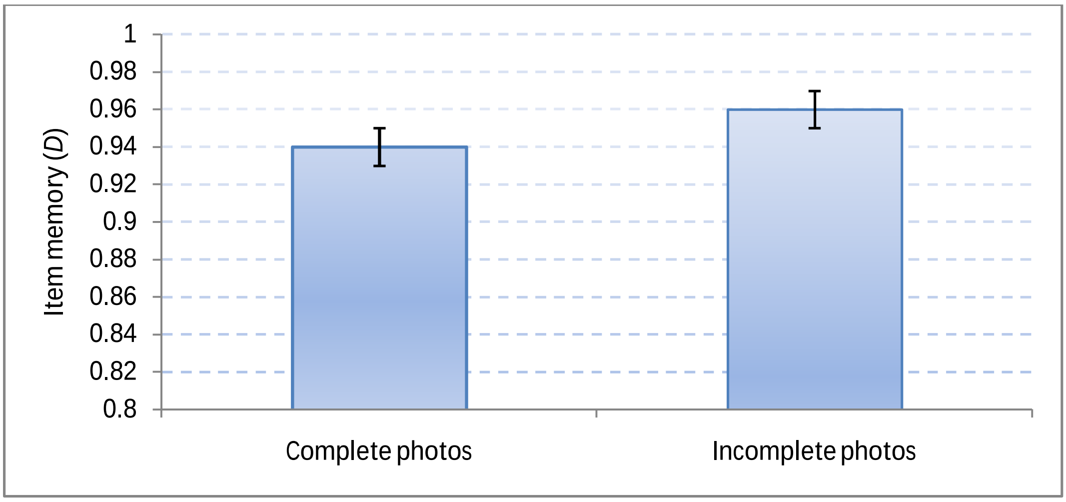

Figure 3.17: Item memory performances for items without schema-typical colour in Experiment 5 - error bars represent standard errors

According to *Hypothesis 2*, regarding source memory for the degree of completeness, an advantage of incomplete over complete items was expected. That is, a positive generation effect should have appeared.

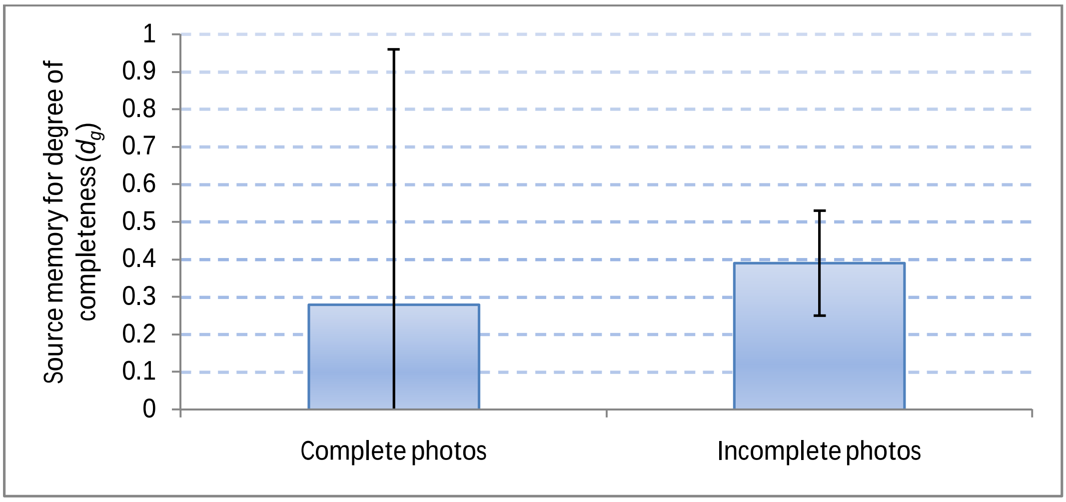

Figure 3.18: Source memory performances for the dimension degree of completeness for items without schema-typical colour in Experiment 5 - error bars represent standard errors

Indeed, a positive generation effect emerged in source memory for the dimension degree of completeness at the descriptive level; completeness was remembered more often for incomplete photos than for complete photos ($d_{g\ Incomplete}$ = 0.39, $d_{g\ Complete}$ = 0.28). However, this difference was not statistically significant (ΔG^2_{1df} = 0.02, p = 0.88). It can be noted that there was a large standard error for complete pictures. Basically, the positive generation effect for degree of completeness as found in previous studies and as theorised in the dual hypothesis by Riefer et al. (2007) could be replicated in the current experiment at the descriptive level only; see Figure 3.18.

According to *Hypotheses 3a and 3b*, regarding source memory for colour, an advantage of complete over incomplete items was expected. That is, a negative generation effect should have appeared. Moreover, guessing colour was expected to be at chance level.

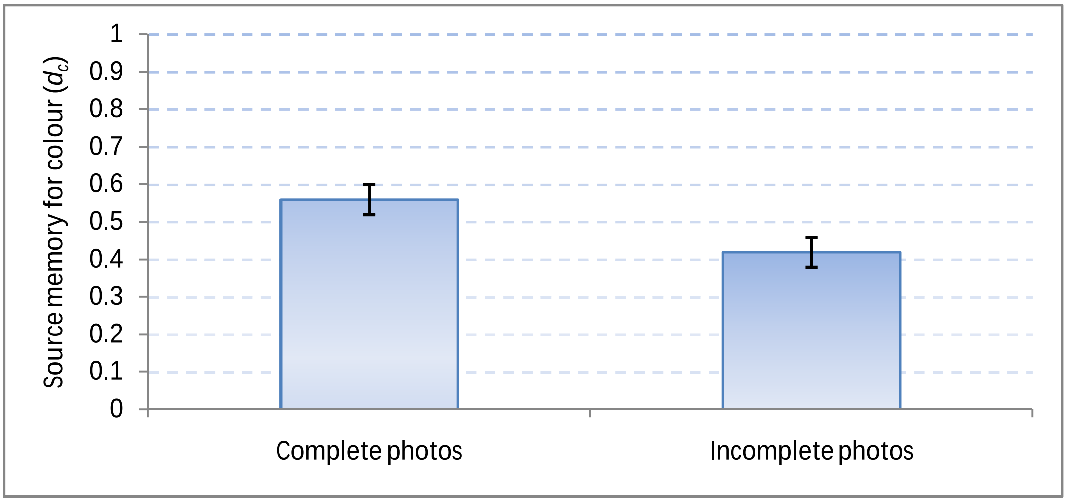

Figure 3.19: Source memory performances for the dimension colour for items without schema-typical colour in Experiment 5 - error bars represent standard errors

As can be seen in Figure 3.19, there was a negative generation effect in source memory for the colour dimension; colour memory for complete photos was better than colour memory for incomplete photos ($d_{c\ Complete}$ = 0.56, $d_{c\ Incomplete}$ = 0.42). This difference was statistically significant (ΔG^2_{1df} = 6.47, *p* = 0.01). Basically, the negative generation effect for colour as found in previous studies and as theorised by Mulligan (2004) and Riefer et al. (2007) could be replicated in the current experiment.

Guessing parameters in an MPT model are mostly suggested to be at chance level. However, due to the way colour was manipulated, it could be possible that participants assumed a strategy of guessing items to have been presented either in red or in green. Guessing "red" as the presentation colour (g_1 = 0.57) was significantly different from chance level: ΔG^2_{1df} = 7.03, *p* = 0.01. Thus, participants had a tendency of guessing items to have been presented in red.

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.
Since the parameters would ideally reflect true guessing, it was tested whether the parameters differed significantly from chance (p = 0.50). Guessing "old" and guessing "complete" when participants correctly remembered the colour of the photos, could not be set equal to 0.50; Guessing "old" was significantly lower than chance level ($b = 0.01$; ΔG^2_{1df} = 367.58, p < 0.001), whereas guessing "complete" when participants correctly remembered the colour was significantly above chance level (a = 0.88; ΔG^2_{1df} = 4.61, p = 0.03). Guessing "complete" when participants guessed colour (g_2 = 0.76) was not significantly different from chance (ΔG^2_{1df} = 1.61, p = 0.21). Thus, participants had a bias to judge items to be new and to be complete when colour was identified correctly.

Items With Schema-Typical Colour

For the current analyses, data for items that have got a schema-typical colour were analysed by data set (all data vs. data for photos named correctly at study) and for different restrictions put on D_{New} ($D_{New} = D_{Complete}$, $D_{New} = D_{Incomplete}$, D_{New} = 0). Additionally, data were analysed separately for items that typically appear in red and for items that typically appear in green colour. Instead of presenting all results in detail here, only the best solution concerning goodness-of-fit value and statistical power is presented: The model fit the data well on the α = 5 % level for items with a schema-typical colour when using all data and when setting D_{New} equal to $D_{Complete}$: G^2_{8df} = 13.58; p = 0.09; the critical value for α = 5 % equals 15.51. Note that analyses were additionally conducted for all of the remaining data sets and restrictions on D_{New}. They yielded the same results concerning descriptive trends and statistical significance tests.

According to *Hypothesis 1*, an advantage of incomplete over complete items was expected regarding item memory, i.e., a positive generation effect should have appeared.
Contrary to this prediction, item detection was about the same for incomplete photos and complete photos ($D_{Incomplete}$ = 0.93, $D_{Complete}$ = .94). Moreover, this difference was not

statistically significant (ΔG^2_{1df} = 0.14, p = 0.71) and could be explained by the very high overall values resulting in a ceiling effect; see Figure 3.20.

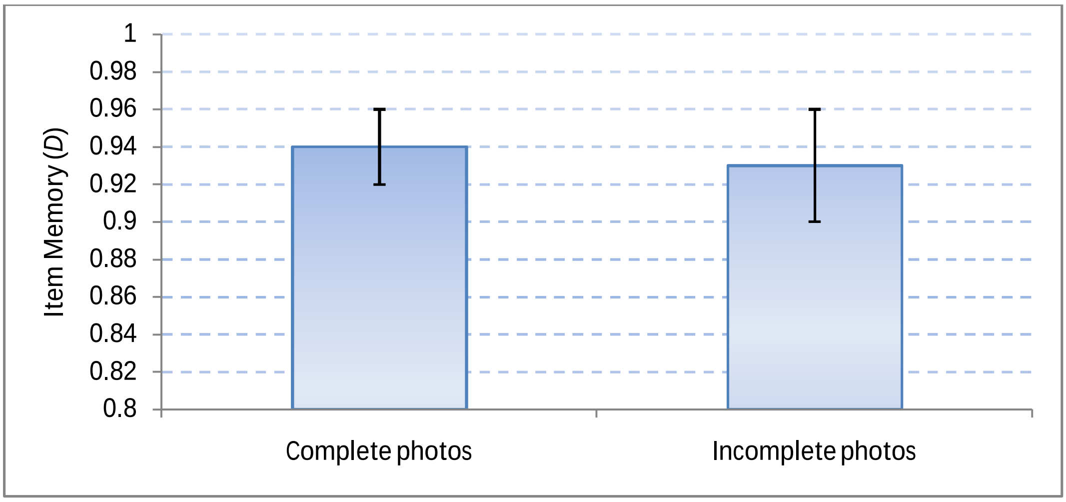

Figure 3.20: Item memory performances for items with schema-typical colour in Experiment 5 - error bars represent standard errors

According to *Hypothesis 2*, regarding source memory for the degree of completeness, an advantage of incomplete over complete items was expected, i.e., a positive generation effect should have appeared.

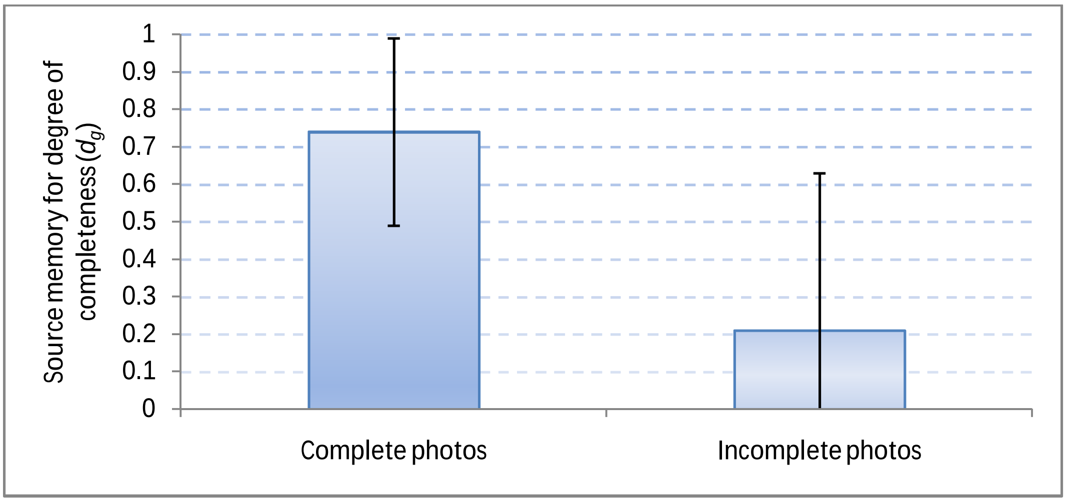

Figure 3.21: Source memory performances for the dimension degree of completeness for items with schema-typical colour in Experiment 5 - Error bars represent standard errors

Unexpectedly, a negative generation effect emerged in source memory for the dimension degree of completeness at the descriptive level; memory for degree of completeness was better for complete photos than for incomplete photos ($d_{g\ Complete}$ = 0.74, $d_{g\ Incomplete}$ = 0.21).

However, this difference was not statistically significant (ΔG^2_{1df} = 1.78, p = 0.18). Standard errors for these parameters were rather large. When restricting guessing parameters that were not significantly different from chance level to 0.50 before conducting this analysis, the memory difference between complete and incomplete pictures became significant: ΔG^2_{1df} = 150.39, p < 0.001. Basically, the negative generation effect for degree of completeness as found in Experiment 5 is at odds with previous findings and with the dual hypothesis by Riefer et al. (2007); see Figure 3.21.

According to *Hypotheses 3a and 3b*, regarding source memory for colour, an effect was anticipated of whether items were presented in their typical or in their atypical colour, moderated by the degree of completeness. Guessing colour could potentially have been biased in favour of the typical colour.

As can be seen in Figure 3.22, there was a trend at the descriptive level towards a negative generation effect for colour for photos presented in a typical colour ($d_{c\ Complete\ Typical\ colour}$ = 0.21; $d_{c\ Incomplete\ Typical\ colour}$ = 0.14) and for photos presented in an atypical colour ($d_{c\ Complete\ Atypical\ colour}$ = 0.72; $d_{c\ Incomplete\ Atypical\ colour}$ = 0.49). Thus, colour memory for incomplete photos was better than colour memory for complete photos. This trend was not statistically significant in the former case (ΔG^2_{1df} = 0.29, p = 0.59), but was statistically significant in the latter case (ΔG^2_{1df} = 10.32, p < 0.01).

Additionally, there was a trend at the descriptive level towards an advantage of atypical items over typical items. A significant difference could be found when comparing complete and incomplete pictures irrespective of colour for complete items (ΔG^2_{1df} = 8.89, p < 0.01). For incomplete items, a significant effect existed when restricting all guessing parameters that were not significantly different from chance level: ΔG^2_{1df} = 4.61, p = 0.03. Note that without these restrictions put on guessing parameters, no significant effect was found: ΔG^2_{1df} = 2.76, p = 0.10.

When considering the differences between parameters for items presented in their typical colour and items presented in their atypical colour for complete and incomplete photos separately, one can see that these differences were different at the descriptive level. Thus, one could test whether the difference between two parameters (items presented in typical colour and items presented in atypical colour) differs between complete and incomplete pictures. However, when comparing the increments for incomplete photos and complete photos, no significant difference could be found: $\Delta G^2_{1df} < 0.01$, $p = 0.96$.

Basically, a negative generation effect emerged for memory for colour. Regarding source memory for colour a significant effect of whether items were presented in their typical or their atypical colour could be found. Moreover, no moderating effect of degree of completeness occurred.

Guessing colour, i.e., guessing that presentation colour did or did not match the typical colour, ($g_1 = 0.55$) was not significantly different from chance level: $\Delta G^2_{1df} = 0.41$, $p = 0.52$. Thus, guessing colour was not biased in favour of the typical colour.

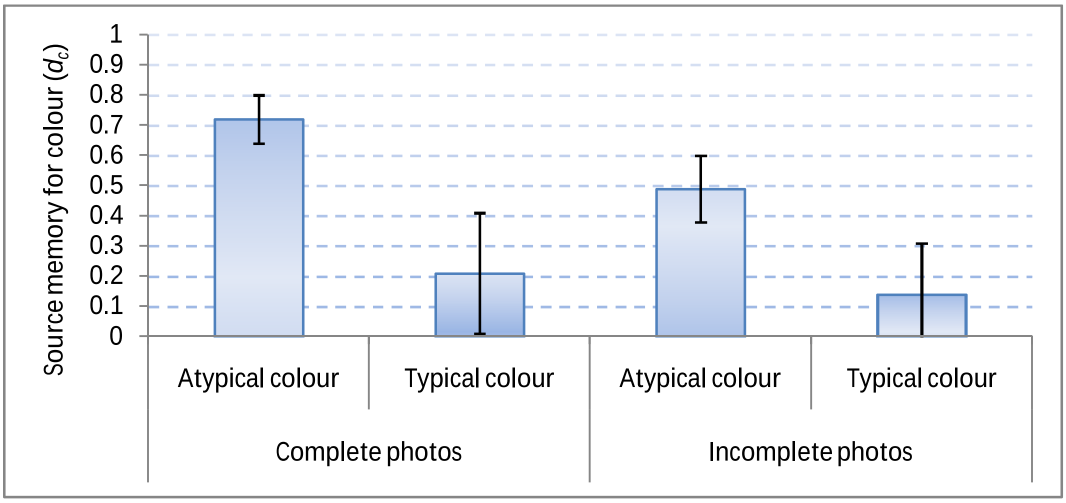

Figure 3.22: Source memory performances for the dimension colour for items with schema-typical colour in Experiment 5 - error bars represent standard errors

According to *Hypothesis 4*, all types of guessing were expected to be at chance level.

Since the parameters would ideally reflect true guessing, it was tested whether the parameters differed significantly from chance ($p = 0.50$). Guessing "old", guessing "complete" when colour was identified correctly, and guessing "complete" when colour was guessed were at chance level. All three parameters could be set equal to 0.50: $b = 0.53$ ($\Delta G^2_{1df} = 0.06$, $p = 0.81$); $a = 0.74$ ($\Delta G^2_{1df} = 2.08$, $p = 0.15$); $g_2 = 0.48$ ($\Delta G^2_{1df} = 0.02$, $p = 0.90$). Thus, participants were not prone to any guessing bias.

A comprehensive discussion of the results and conclusions obtained from Experiment 5 (involving photos and an incidental study design) can be found in the next section.

3.3 Discussion of Chapter 3

Chapter 3 is concerned with the role of increased conceptual processing for source – or more precisely for memory for presentation colour – for studies employing a generation effect paradigm.

Within a generation effect paradigm, typically a positive generation effect, namely better memory for self-generated items, occurs in item memory. For memory for contextual information (i.e., for source memory) the picture is less clear. However, for memory for presentation colour, a negative generation effect has been found consistently (e.g., Mulligan, 2004; Mulligan et al., 2006; Riefer et al., 2007). That is, superior source memory has occurred for complete items. Some theories state that for self-generated items, conceptual processing takes place to a higher degree (Donaldson & Bass, 1980; Graf, 1980; Mulligan, 2004; Slamecka & Graf, 1978); some theories additionally claim that perceptual processing takes place to a lesser degree (Mulligan, 2004). Nevertheless, this point was called into question by Mulligan et al. (2006). They challenge the appropriateness of a processing account, which emphasises perceptual and conceptual processing, and consider it too specific. Instead, Mulligan and colleagues suggest a more general processing account, which is concerned with visual and non-visual processing instead.

In general, when being presented with the picture of an object in a memory experiment and asked to remember all items in the study list along with their colour at presentation, I argue that two things occur at study concerning colour: (a) the concept colour of the object is activated and (b) the actual presentation colour of the object is encoded. The activation of an object's concept comprising its many semantic features was shown to take place rapidly and automatically (cf., Barsalou, 1999). Concerning concept colour, one can see that some (everyday) objects have got a colour that is an important part of their concept, whereas other objects have not.

When an item has no schema-typical colour, it is simply presented in any colour, which does not have a special relationship to the object's concept. However, when an item has a schema-typical colour, it is either presented in this colour or it is not. I argue thatwhether or not an item has a schema-typical colour is likely to influence memory performance for colour at presentation (i.e., experimental outcome).

In my view, the generation effect paradigm adds another interesting issue to this, since due to the definition of the generation effect, some objects are presented complete, whereas others are presented incomplete. This circumstance of complete versus incomplete presentation is theorised to entail different degrees of conceptual processing (see Section 3.1, for more details). Since self-generation of items is strongly indicated to lead to higher conceptual processing, I argue that it is likely that whether or not an incomplete item has a schema-typical colour, colour memory performance (experimental outcome) is influenced; moreover, I theorise that it does so to a higher degree for self-generated items than is true for simply read items.

Experiments 3, 4, and 5 represented concurrent attempts to tackle this research idea and to find a solid experimental operationalisation of it. I outlined and conducted the studies in a parallel fashion rather than as a sequence of experiments. Therefore many similarities existed concerning the crucial experimental factors “item type”, presentation “colour”, and “degree of completeness” as well as concerning the basic types of items employed. Differences lay in the naturalness of the stimulus material (i.e., drawings vs. photos) and in the intentionality of study instructions (i.e., intentional vs. incidental study instructions). These manipulations were implemented to maximise the impact of concept colour by boasting the degree of naturalness of stimuli and by reducing the supposed impact of intentional study instructions as a way of suppressing the activation of concept colour.

The basic hypotheses and predictions were the same for the three experiments. For items without schema-typical colour, I expected colour to be remembered better for complete pictures than for incomplete pictures. Consequently, a negative generation effect was hypothesised to occur. For items with schema-typical colour, I anticipated an effect of whether or not presentation colour matched concept colour, that is, an effect of whether presentation colour was the typical or the atypical colour. Moreover, I hypothesised this effect to be influenced by whether pictures were presented complete or incomplete. The effect of schema-typicality was expected to be moderated by the degree of conceptual processing as operationalised by the degree of completeness.

Data were analised within the multinomial processing tree model framework.

Overview of the Most Important Results and Their Interpretations, Critical Issues, Further Research and Recommendations

Item Memory

For items *without schema-typical colour,* there was a tendency and trend for a positive generation effect in all studies. In Experiment 3, this trend became significant. For Experiments 4 and 5, a ceiling effect emerged. However, despite this ceiling effect, a significant positive generation effect could be found for green stimuli in Experiment 4. Thus, the typically found positive generation effect in item memory could be replicated and was either statistically significant or clearly present at the descriptive level (one comparison rendered unclear results most likely due to a ceiling effect). This finding is in line with the hypothesis phrased for item memory. As stated previously, it is also in line with many previous findings and with the most relevant papers for the present studies, namely with Mulligan (2004) and with Riefer et al. (2007). This is a strong hint that the generation effect design employed in the present experiments is valid.

When comparing item memory performance values for Experiment 3 to those for Experiments 4 and 5, one can find an equivalent increase of values in the latter two studies as compared to the former study. This increase might have occurred due to a change in experimental stimuli from line drawings to photos and thus due to the enhanced naturalness of the latter type of stimuli. However, the stimulus sets are not perfectly comparable. Fewer drawings (n = 40) as compared to photos (n = 60) were employed and, moreover, studied objects did not match perfectly. However, the difference in values can be considered as evidence that the manipulation of increased naturalness worked.

A presumed decrease in values, due to the switch from intentional study instructions to incidental study instructions, could not be found in the data. This may also be due to a ceiling effect; item memory was very high in Experiments 4 and 5. Hence, study instructions do not really seem to be of much consequence in this case.

For items *with schema-typical colour*, memory values ranged between 0.90 and 0.95. Thus, a ceiling effect is probable. None of the differences within any of the experiments rendered statistically significant results. Thus, the typically found positive generation effect in item memory could not be replicated. When comparing item memory performance values across experiments, no difference was detected. No effect of manipulations between experiments could be found.

When comparing memory for items without schema-typical colour and items with schema-typical colour, one can see that items with schema-typical colour were remembered better – actually, performance was nearly perfect.

Source Memory for the Degree of Completeness

For items *without schema-typical colour*, there was a trend of a positive generation effect in all studies. For Experiment 3, this trend became significant. For Experiments 4 and 5, the positive generation effect found at the descriptive level did not become statistically significant. For items *with schema-typical colour*, there was a trend towards a positive generation effect at the descriptive level of all experiments. However, this trend did not become significant in any of the studies. Thus, the typically found positive generation effect in source memory for the degree of completeness could be replicated and comparisons were either statistically significant or at least present at the descriptive level. This finding is in line with the hypothesis. As stated previously, it is also in line with previous findings (e.g., Riefer et al., 2007). This is again a strong indication that the generation effect design employed in the present experiment is valid.

When comparing memory for items without schema-typical colour and that for items with schema-typical colour, one cannot find a clear deviation between experiments. As such, no remarkable differences were obvious. Thus, no effect of manipulations of study intentionality or naturalness of stimuli between experiments was present for source memory for the degree of completeness.

Source Memory for Colour and Guessing of Colour

For items *without schema-typical colour*, guessing presentation colour generally matched the base rate. In Experiment 5, there was a slight tendency for guessing “red”. Moreover, a trend for a negative generation effect emerged in all studies. For Experiments 3 and 5, this trend became significant across presentation colours. For Experiment 4, this trend became significant for red and green items, when considering them separately.

For items *with schema-typical colour*, no response bias (towards guessing the typical colour) appeared. Instead, guessing colour was at chance level. When drawings were employed as study material (Experiment 3), no effect of schema-typicality of presentation colour or of completeness of the items could be found on the source memory parameters for colour. Instead, all source memory parameters for colour could be set equal leading to a model fit that did not deteriorate significantly.

This effect fits well with results reported in Olkkonen, Hansen, and Gegenfurtner (2008). They investigated the colour appearance of familiar objects and tested for effects of object shape, texture, illuminance changes, and, most importantly, for effects of naturalness of stimuli. Fruit photographs, painted fruit, and fruit outline shapes served as stimuli. As Hansen et al. (2006), Olkkonen and colleagues investigated whether the *known* colour (i.e., prior knowledge) of an object affected colour perception by asking participants to adjust the colour of presented fruit until they appeared to be grey (i.e., achromatic). Olkkonen et al. (2008) generally showed that participant-produced achromatic settings for fruit were systematically biased away from the gray point towards the opposite direction of a fruit's memory colour. They additionally found that "the magnitude of the compensation for the perceived memory color was greatest with the most natural stimuli and decreased monotonically with decreasing stimulus realism, being absent for fruit outline shapes." (Olkkonen et al., 2007; p. 7).

In the present dissertation, all values for drawings with schema-typical colour were equally low – and lower than values found for photos. In line with previous research, there was a trend at the descriptive level towards an advantage of complete stimuli and thus a trend towards the typically found negative generation effect, when considering degree of completeness only.

In contrast to this, when photos, i.e., when more natural stimuli, were employed, an effect could be found for the degree of completeness (complete vs. incomplete) and for schema-typicality of presentation colour (typical colour vs. atypical colour). When considering degree of completeness only, a negative generation effect could be found. This negative generation effect persisted, when items presented in typical colour and items presented in atypical colour were considered separately. For items in typical colour in Experiment 5, this trend was visible only at the descriptive level. The present negative generation effect can be seen as evidence in favour of Mulligan (2004), Mulligan et al. (2006) and Riefer et al. (2007). Moreover, memory was superior for items presented in atypical colour to that for items presented in typical colour. These effects were significant in Experiments 4 and 5, but were more pronounced in the latter study. As described previously, I suggested that although the activation of a concept (along with the activation of its schema-typical colour) takes place automatically, it can be regarded as a rather weak, that is easily suppressed, activation. It seemed reasonable to me that activation of the concept's colour could be suppressed or overcome, when using intentional study instructions, which actively force participants to withdraw attention from the colour of the concept and to focus attention instead on the actual colour at presentation. Therefore, incidental study instructions were employed in Experiment 5. The finding of a more pronounced atypicality effect in Experiment 5 fits well with my assumption.

Incidental Study Instructions

In general, true presentation colour was remembered better when items were presented in a schema-atypical colour (Experiments 4 and 5). Moreover, when considering only Experiment 5, it can be stated that good colour memory encoding of items in atypical colour remained for incidental study instructions. In addition to that, colour memory for items in typical colour deteriorated. The decrease was to a considerable degree, whereas values for items in atypical colour remained about equal. As described previously, several explanations were offered for how and why the atypicality effect emerges, such as the script-copy-plus-tag hypothesis and the attention-elaboration hypothesis (see Section 3.1, for further details). The present findings seem to suggest the script-copy-plus-tag hypothesis as a reasonable explanation. It could be argued that in Experiment 5, tags indicating deviations from default values (i.e., tags indicating colouring deviating from the schema-typical colour) were created nevertheless, that is in spite of the incidental study instructions. These tags could thus have lead to stable memory for schema-atypical information, whereas schema-typical information, for which no tags were encoded, was forgotten easily.

The Role of Increased Conceptual Processing

When an item has got a typical colour, I expected an effect of whether or not presentation colour matched concept colour. I anticipated an atypicality effect, i.e., superior memory for items presented in an atypical colour. Moreover, this effect was theorised to be influenced by whether items were presented complete or incomplete, because incomplete items are theorised to evoke stronger conceptual processing than complete items. I expected the effect of schema-typicality to be moderated by the degree of conceptual processing as operationalised by the degree of completeness; I hypothesised that superior source memory or superior colour memory for atypical items is more pronounced for incomplete items.

For the present case, no additional effect of incompleteness of pictures could be found in Experiments 3 to 5. Increments of effects (between complete and incomplete pictures) were not statistically significant. These results can be interpreted as hints against a significant influence or as hints against the existence of increased conceptual processing in self-generated items, which would lead to a positive generation effect in item memory and to a negative generation effect in source memory for external attributes (i.e., for external source monitoring designs).

I conclude that it would be false to hold conceptual (and perceptual) processing responsible for generation effects in item memory and source memory as hypothesised in Donaldson & Bass (1980), Graf (1980), and Slamecka and Graf (1978), and as claimed by Mulligan (2004). In contrast, the present results could be interpreted as evidence in favour of Mulligan et al. (2006) who questioned the appropriateness of a processing account, which emphasises perceptual and conceptual processing, and considered it too specific. Instead, they suggest a more general processing account concerned with visual and non-visual processing.

Assuming the correctness of Mulligan et al.'s (2006) broader processing account, increased conceptual processing is missing in the current experiments. Thus, no effect of increased conceptual processing is visible and thus, increments of effects (between complete and incomplete pictures) were not statistically significant.

An alternative explanation would be that conceptual processing was not strong enough in Experiments 3 to 5. However, this idea seems less reasonable due to the appropriate operationalisation and setup used in the studies (see Chapter 4, for a detailed discussion).

Conclusions to Chapter 3

Conducting the current research was important, for several reasons. First, the role of conceptual activation in generation effect studies was aimed to be supported. While second, several hypotheses can be deduced from Mulligan's (2004) account, some of which have already been drawn and tested. The extension presented in Chapter 2 was deduced directly from Mulligan's account, but had not been tested. Hence, this gap was filled. Third, the roles of conceptual processing versus perceptual processing as the basic mechanisms producing the generation effect are under debate; Mulligan (2004) argues in favour of this account, whereas Mulligan et al. (2006) instead support the fundamental role of visual processing versus non-visual processing. Thus, when investigating the role of conceptual processing in the present experiments more closely, I aimed to differentiate between these accounts and eventually aimed to favour one over the other. Additionally, by introducing different types of stimuli (those bearing no typical colour, those bearing a typical colour and being presented either in this typical colour or not) I intended to learn more about the effect of colour schema-typicality on memory for the items themselves. Although this aspect is not new in research of memory psychology in general, the issue had not yet been researched within generation effect paradigms. Therefore, the possibility of an influence of self-generation remained.

From my research, one can see that also within generation effect paradigms, an atypicality effect emerges, i.e., a memory advantage of items presented in atypical colour appears over that for items presented in typical colour. The special conditions present in generation effect paradigms did not influence the general effect of schema-typicality described in other studies. In addition, there is also a hint that the script-copy-plus-tag hypothesis can be regarded as a more valid explanation of atypicality effects than the attention-elaboration hypothesis.

Furthermore, I conclude that it would be false to hold conceptual (and perceptual) processing responsible for generation effects in item memory and source memory as hypothesised in Donaldson & Bass (1980), Graf (1980), and Slamecka and Graf (1978), and as claimed by Mulligan (2004). In contrast, the present results could be interpreted as evidence in favour of Mulligan et al. (2006) – they question the appropriateness of a processing account, which emphasises perceptual and conceptual processing, and consider it too specific. Instead, they suggest a more general processing account concerned with visual and non-visual processing.

Since the current studies do not explicitly address and thus do not thoroughly test Mulligan et al.'s (2006) broader processing account, and since the validity of evidence provided by Mulligan and colleagues can be regarded debatable, a detailed empirical test of the current conclusions is still missing.

Chapter 4: General Discussion

In the previous chapters, new findings are presented and discussed, which are shedding light on source memory performances within generation effect paradigms. The role of the processing of perceptual attributes and of the processing of internal states was examined in Chapter 2. Whereas in Chapter 3, the role of conceptual processing was investigated. In the present chapter, I summarise the crucial findings and conclusions once more and discuss them in a broader context, before I give a description of their comprehensive meaning. Then, I address limitations and delimitations of the current studies and discuss critical issues concerning generation effect paradigms for Experiments 1 to 5. Eventually, I present remaining research questions, highlight contributions to the scientific field, and suggest further research opportunities.

Item Memory and Effects of Self-Generation

On the basis of the present findings, it appears that an advantage exists for incomplete stimuli over complete stimuli concerning item memory. The reverse pattern could only be found at the descriptive level for cases, in which a ceiling effect can be assumed to account for the results. Thus, the present findings are in line with previous studies that found a positive generation effect in item memory and can hence be regarded as additional evidence. Influencing factors that might lead results to deviate from this standard outcome were considered when designing Experiments 1 to 5. Consequently, the current work supports the suggestions made by Steffens and Erdfelder (1998) and Bertsch et al. (2007). They argued in favour of certain factors concerning design and underlying processes during self-generation that play an important role and that can thereby influence the memory outcome.

Concerning item memory and effects of self-generation, the scope of the current research was rather broad. The present studies were not aimed to support a specific theory on generation effects in item memory. Therefore, the results can be explained by all of the previously described theories, such as the lexical activation hypothesis, effort hypothesis, multifactor account, two factor theory, and three factor theory (see Section 1.1).

Source Memory and Effects of Self-Generation

Note that the present studies are not appropriate to test Mulligan's specificity assumption, which provides a way of differentiating the processing account from the item-source trade-off hypothesis. For cases in which the relevant feature is not part of the target (such as for background colour), the accounts differ. As highlighted previously, Mulligan's account is more specific and accommodates findings well. However, since this critical condition is not introduced in the experiments for this dissertation, findings for the source memory dimension colour can be regarded as evidence in favour of Mulligan (2004) and the item-source trade-off hypothesis, likewise.

Summary of Chapter 2 and Chapter 3

Chapter 2 is concerned with *the role of the the processing of perceptual attributes and the processing of internal states* regarding memory for source, i.e., regarding memory for the degree of completeness, for studies employing a generation effect paradigm.

Within a generation effect paradigm, memory for source attributes of an item can be studied in different ways. When considering the processing account by Mulligan (2004) and the dual-hypothesis by Riefer et al. (2007), contradictory predictions can be found for the source memory dimension degree of completeness. According to the processing account, a negative generation effect should occur, whereas according to the dual-hypothesis, a positive generation effect is theorised.

To overcome these contradictions, I suggested to consider two processing modes: the processing of and memory for (a) perceptual attributes (PA) and (b) internal states (IS). Processing modes can be triggered by manipulating encoding and recall processes via the use of different experimental instructions (that focus either on the internal vs. external dimension or on the self vs. other dimension).

Consequently, two experiments were set up, in which memory for the degree of completeness was studied. The crucial independent variable was the factor "instruction" which had two levels: (a) "the processing of perceptual attributes" ("PA"), and (b) "the processing of internal states" ("IS"). In Experiment 1, instructions followed closely those employed by Mulligan (2004) and Riefer et al. (2007), whereas instructions of encoding and retrieval conditions were altered to enforce the impact of the independent variable "instruction" in Experiment 2. For both studies, an interaction was anticipated between the factors "degree of completeness" and "instruction". For the PA condition, a negative generation effect for the source memory dimension "degree of completeness" was expected, whereas for the IS condition, a positive generation effect for the source memory dimension "degree of completeness" was anticipated.

In Experiment 1, neither the independent variable "degree of completeness" ("complete" vs. "incomplete") nor the independent variable "instruction" ("PA" vs. "IS") seemed to have an effect. Additionally, no interaction between the two predicted effects occurred. However, data could probably be at the verge of a ceiling effect.

In Experiment 2, the independent variable "degree of completeness" had an effect. There was no interaction however between this independent variable and the independent variable "instruction". Values for complete items were close to zero, whereas values for incomplete items were at an intermediate level. In both instruction conditions (PA and IS), a positive generation effect appeared – a result that was theorised only for items encoded and retrieved under IS instruction. Thus, one can say that irrespective of the actual instruction condition, participants encoded and retrieved items as if participants had been instructed to memorise and remember whether they had self-generated the items or had simply read the items on the screen. Even in the PA condition, processing occurred in line with predictions made for the IS condition.

I conclude that when self-reference (cf., Klein & Loftus, 1993; Rogers et al., 1977; Symons & Johnson, 1997) is high, memory for source attributes seems to be processed in a way considering the role of self at encoding. In a reality monitoring paradigm, the self of the learner is involved to a large degree and thus matches needs induced by increased self-reference. Thus, processing, studying, and recalling items in terms of the reality monitoring paradigm occurs automatically, even for cases in which instructions suggest a different type of item processing, studying, and recalling. As a consequence, a positive generation effect emerges. An effect of self-reference plays a critical role for generation effect studies, when attempting to investigate source memory.

Chapter 3 is concerned with *the role of increased conceptual processing* for source – or more precisely for memory for presentation colour – for studies employing a generation effect paradigm.

Within a generation effect paradigm, typically a positive generation effect, namely better memory for self-generated items, occurs in item memory. For memory for contextual information (i.e., for source memory) the picture is less clear. Yet, for memory for presentation colour, a negative generation effect has been found consistently (e.g., Mulligan, 2004; Mulligan et al., 2006; Riefer et al., 2007). That is, superior source memory has occurred for complete items. Some theories state that conceptual processing takes place to a higher degree for self-generated items than for complete items (Donaldson & Bass, 1980; Graf, 1980; Mulligan, 2004; Slamecka & Graf, 1978), while other theories additionally claim that perceptual processing takes place to a lesser degree for incomplete as compared to complete items (Mulligan, 2004). These points were called into question by Mulligan et al. (2006). They challenge the appropriateness of a processing account, which emphasises perceptual and conceptual processing and consider it too specific. Instead, Mulligan and colleagues suggest a more general processing account, which is concerned with visual and non-visual processing instead.

Experiments 3, 4, and 5 represent concurrent attempts to tackle the research idea and to find a solid experimental operationalisation of it. Differences lie in the naturalness of the stimulus material (i.e., drawings vs. photos) and in the intentionality of study instructions (i.e., intentional vs. incidental study instructions) so as to maximise the impact of concept colour.

When an item has got a typical colour, I anticipated an effect of whether or not presentation colour matched concept colour. I expected an atypicality effect, i.e., superior memory for items presented in an atypical colour. Moreover, this effect was expected to be influenced by whether or not items were presented complete or incomplete, because incomplete items are theorised to evoke stronger conceptual processing than complete items. I expected the effect of schema-typicality to be moderated by the degree of conceptual processing as operationalised by the degree of completeness. I hypothesised that superior source memory or superior colour memory for atypical items is more pronounced for incomplete items.

For items *without schema-typical colour*, a trend for a negative generation effect emerged in all studies. This trend became significant across presentation colours in Experiments 3 and 5, whereas, in Experiment 4, this trend became significant for red and green items, when considering them separately.

For items *with schema-typical colour*, no effect of schema-typicality of presentation colour or of completeness of the items could be found on the source memory parameters for colour, when drawings were employed as study material (Experiment 3). Instead, all source memory parameters for colour could be set equal leading to a model fit that did not deteriorate significantly. In contrast to this, when photos, i.e., when more natural stimuli were employed, an effect could be found for the degree of completeness (complete vs. incomplete) and for schema-typicality of presentation colour (typical colour vs. atypical colour). When considering degree of completeness only, a negative generation effect could be found. This negative generation effect persisted, when items presented in typical colour and items presented in atypical colour were considered separately. For items in typical colour in Experiment 5, this trend was visible only at the descriptive level. Moreover, memory was superior for items presented in atypical colour to that for items presented in typical colour. These effects were significant in Experiments 4 and 5, but were more pronounced in the latter study. Furthermore, no additional effect of incompleteness of items could be found in Experiments 3 to 5. Increments of effects (between complete and incomplete items) were not statistically significant. These results can be interpreted as indicators against a significant influence, or otherwise as suggestions against the emergence

of increased conceptual processing in self-generated items, which would lead to a positive generation effect in item memory and to a negative generation effect in source memory for external attributes (i.e., for external source monitoring designs).
I conclude that it would be false to hold conceptual (and perceptual) processing responsible for generation effects in item memory and source memory as hypothesised in Donaldson & Bass (1980), Graf (1980), Slamecka and Graf (1978), and as claimed by Mulligan (2004). In contrast, the present results could be interpreted as evidence in favour of Mulligan et al. (2006) who questioned the appropriateness of a processing account, which emphasises perceptual and conceptual processing, and considered it too specific. Instead, they suggest a more general processing account concerned with visual and non-visual processing. Assuming the correctness of Mulligan et al.'s (2006) broader processing account, an increased conceptual processing is missing in the current experiments. Thus, no effect of increased conceptual processing is visible and thus, increments of effects (between complete and incomplete items) were not statistically significant.

However, the role of the self-reference effect has not been investigated for generation effect studies so far, and the current studies do not explicitly address and thus do not thoroughly test Mulligan et al.'s (2006) broader processing account. Moreover, the validity of evidence provided by Mulligan and colleagues can be regarded debatable. Thus, an empirical test of the current interpretations and conclusions is still missing.

Limitations and Delimitations of the Present Work

Limitations refer to restrictions in studies over which the researcher has no control as well as to restrictions the researcher is not aware of. They may affect the validity and generalisability of results. In contrast to this, delimitations are restrictions which were set deliberately by the researcher, such as restrictions on the recruiting area or on the selection of a certain age group (e.g., children vs. adults). One possible limitation in the present studies results from sampling. For example, more women than men constituted the samples. Moreover, most participants in the studies and in the pilot studies were students. Thus, strictly speaking, the samples were not truly representative of the population. Furthermore, it could be considered challenging to generalise the present findings to the population. The present research, however, was conducted in the field of general psychology (namely memory psychology). For this area, cognitive processes are expected to be very similar among members of the population. Note that this assumption does not include deviations due to developmental (e.g., babies, children, old age), pathological (e.g., ADS, dementia), or medical (e.g., under medication) changes. Apart from this point, many variables outside the control of the researcher could have impacted participants' achievements. These variables may, for instance, include time of day or health status of the participants. However, due to randomisation in the experiments, these influences were aimed to be minimal.

In Chapters 2 and 3, I already provided support for the claim that the research ideas, analyses, and pilots can reasonably be assumed as being appropriate to address and to test the current research questions. However, several issues remain that can be considered critical when assessing effects of self-generation on item and source memory tasks. I therefore address these issues in the following paragraphs and thereby illuminate how validity and reliability were ensured or at least approximated in Experiments 1 to 5.

Was item fragmentation appropriate?

In *Experiments 1 and 2*, targets were fragmented as in Mulligan (2004) and in Riefer et al. (2007).

For picture fragmentation in *Experiment 3*, the picture fragmentation tool "Frag" (Snodgrass et al., 1987) was employed. Applying this computer programme, a black and white bitmap file can be fragmented: A grid of 16x16 blocks is laid out and blocks containing black pixels are identified. The locations of these critical blocks are then stored. Afterwards, "the program randomly selects increasing proportions of critical blocks to be erased according to an exponential function to produce eight levels of fragmentation per picture" (Snodgrass et al., 1987, p.271). The mathematical function can be found in the programme: proportions of fragments left = total fragments * ($0.7^{(8\text{-level})}$).

For each drawing in Experiment 3, several fragmented versions were produced until an appropriate version was found, for which none of the descriptive object features was eliminated completely (as could be the case due to *random* fragmentation); see also Murray and Kinnison, 1989. Complete pictures are at Level 8, whereas most fragmented pictures are at Level 1. For the current experiment, Level 5 (i.e., an intermediate level of fragmentation) was used to fragment drawings.

Finke, Johnson, and Shyi (1988) selected an alternative way to fragment objects. They either eliminated the left or the right side of an object along a vertical axis or eliminated the upper or the lower part of an object along a horizontal axis. However, this type of fragmentation works only when objcts are symmetrical, which was hardly ever the case in the present experiments. Using the fragmentation option offered by Finke et al. (1988), might thus have lead to relatively low identification rates for the current stimuli. In contrast to this, the presently applied form of fragmentation along with presenting a cue, lead to nearly perfect correct identification or perfect correct naming; it was thus considered a more appropriate form of fragmentation.

To fragment photos in *Experiments 4 and 5*, the picture fragmentation tool "Frag" could unfortunately not be applied due to technical requirements of the programme, that could not be fulfilled by the experimental stimuli. However, the same logic as implemented in "Frag" was followed when manipulating study stimuli. A grid of 16x16 blocks was laid out over each picture. Then, I determined which blocks contained (coloured) pixels. Afterwards, increasing proportions of critical blocks were randomly selected and erased. I calculated the amount of erased blocks according to Snodgrass et al. (1987).

Was appropriate self-generation of items ensured?

In generation effect studies, it is critical to apply measures to ascertain true self-generation of the targets. In order to argue convincingly that the experimental setups employed in Experiments 1 to 5 fulfilled this goal, it was necessary to ensure that participants indeed generated the experimental stimuli in the first place. Therefore, participants were instructed to write down their responses instead of just assuming self-generation in a setup, in which participants never gave an answer overtly, but simply (allegedly) produced responses internally. It was pointed out to participants that they should simply make a line when they could not identify what the target was. Note that response sheets were collected directly after the study block, so that participants could not consult them during the distractor or test phases. Moreover, naming errors were low (see Appendix D and Appendix I).

Were semantic framing and the selected cues appropriate?

In all studies, semantic framing was achieved by presenting a semantically related cue along with the target. In *Experiments 1 and 2*, antonym cues were either taken from Masson and MacLeod (1992) as employed in Mulligan (2004) and Riefer et al. (2007) or were achieved in a pilot study. Additional extensive pilot studies were run to obtain valid semantic cue words for target drawings (*Experiment 3*) and for target photos (*Experiments 4 and 5*). How these pilot studies were conducted and how cues were selected from the collected data can be found in Appendix E.

Was a high probability of self-generating the correct solution ensured?

In *Experiments 1 and 2*, all word pairs from Masson and MacLeod (1992) were translated into German and pretested for their appropriateness for a sample of German participants. To illustrate what appropriateness means for these studies and why it had to be ensured, I reiterate two important points about the generation effect paradigm as employed by Mulligan (2004) and Riefer et al. (2007): (a) In the generate condition of a generation effect experiment, only the cue name is given complete, whereas the target word is presented incomplete. (b) Moreover, target items should be *generated* almost as quickly as they can be *read*. Owing to these two basic assumptions, it was essential to guarantee both, namely quick correct reading *and* quick correct generation of the target items. This could be done by making sure that the conceptual link between cue and target word was strong and that this strength persisted from a US-American sample to a German sample. Pilot studies were run with several student assistants (N = 5). Eventually, 39 useful pairs from Masson and MacLeod (1992) were identified. Sixteen additional word pairs had to be developed, for which appropriateness was ascertained. Eventually, 55 item pairs were randomly divided into five stimulus sets and were counterbalanced across groups.

To ensure high probability of correct identification of the presented pictures in *Experiments 3 to 5*, a pilot study was conducted. Additionally, the correct solution was presented directly after picture display at study. In comparison, no solutions were given at study for word antonym pairs (e.g., "fast – s___") in Experiments 1 and 2. This was not necessary at that point, because only those word antonym pairs were included in the experiment for which a high association between cue and target word was apparent. However, a similarly high association cannot necessarily be assumed for the semantic association between cues and targets in Experiments 3 to 5. Although a comprehensive pilot study was run to identify good semantic cues for the drawings and photos, the relation between an object and even the best imaginable semantic cue can hardly ever be as strong as the relation between the members of an antonym pair (e.g., "black and white" vs. "cat and dog") due to a stronger influence of inter-individual differences.

Since target pictures could be misidentified, it was necessary to present the names of the objects at study, to ensure that participants remembered the *correct* study episode. Without the presentation of the target names at study, several potential errors could occur. To further illustrate this point, consider the following exemplary scenarios. *Scenario A*: Both object A and object B were presented at study. A was presented in red and in an incomplete manner, whereas B was presented in green and in a complete manner. At study, A and B were confused (e.g., due to high perceptual similarity). Consequently, given perfect memory performance, at test, A would be judged to have been presented in green and in a complete manner, whereas B would be judged to have been presented in red and in an incomplete manner. *Scenario B*: Object A was misidentified at study; it was identified as an item that is not at all in the list. Correct knowledge of colour or degree of completeness is then lost for Object A. *Scenario C*: Object A was misidentified at study; it was identified as one of the distractors. Consequently, a false alarm is reported, which is due only to misidentification of the object at study.

Was item colouring appropriate?

In all studies, I selected the colour dichotomy red versus green for three reasons. (1) The same colours were used in the studies by Mulligan (2004) and Riefer et al. (2007) and were effective in these cases. (2) Red and green appear to be natural colour categories that have contrasting qualities; this contrast is mirrored also in the setup of our visual system (Gegenfurtner & Kiper, 2003; Goldstein, 2002). Further evidence of this natural dichotomy in human perception can be found in the design of colour systems or theories on how to categorise colour, which were stated by early artists and scientists (Irtel, 2000). (3) Other common colour pairs, such as black and white, or blue and yellow did not appear to be appropriate for use in the present experiments in which pictures of *objects* were implemented. In our everyday world, it seems that items with a typical colour mostly are those that are not man-made (e.g., boxes, folders, toys) but natural (e.g., plants, animals, fruits, vegetables). While searching for appropriate objects and colours, it became more and more clear that simply more objects that are typically either red or green could be found and conveniently implemented than objects that typically appear in any other colour.

Words (Experiments 1 and 2) as well as drawings (Experiment 3) and photos (Experiments 4 and 5) were coloured in by using a standard graphics software. Note that neither Mulligan (2004) nor Riefer et al. (2007) reported RGB-values to describe the exact shades of red and green that they employed. Therefore, in Experiments 1 and 2, those two colours in the RGB colour value system were used that are explicitly labelled "red" (255-0-0) or "green" (0-255-0); see Hunt (1991).

Note that the values used for Experiments 1 and 2 were considered good but not best for all presently described experiments. Thus, in Experiments 3, 4, and 5, in contrast, much more care in colour selection was indicated due to the predominant status of colour. In a pilot study (additional task in Pilot IV), participants were given a questionnaire comprising of various colour shades of the colours red and green. Previously, 30 shades had been judged most relevant for the task in a short pre-pilot study. In Pilot IV, participants were to rate on a seven-point scale how typical the green colour shades were of the colour green and how typical the red colour shades were of the colour red. Response options were "extremely typical", "typical", "mildly typical", "neither typical nor atypical", "mildly atypical", "atypical", and "extremely atypical". Consequently, two colour shades could be identified, for which the highest ratings were achieved (regarding typicality of red vs. typicality of green, respectively). Most importantly, these computer-based questionnaires (and the pre-pilot study) were administered via the same computer screens, which were later on used for running the experiment. Therefore, biased perception induced by types of screens or by differences in screen setups was eliminated.

Was testing appropriate?

In *Experiments 1 and 2*, target names were presented sequentially and individually. For each trial, participants were required to indicate which of five categories each test item belonged to, and to respond by clicking one of five fields on the bottom of the screen. This test format was equivalent to Mulligan (2004), Mulligan et al. (2006), and Riefer et al. (2007).

In *Experiments 3 to 5*, target names were again presented sequentially and individually: For each trial, participants were first presented with the name of the target drawing (if the item was old) or the name of another new object. Then, for each name, participants *first* had to judge whether they previously had or had not studied this item. When they decided that the item was new, another object name was presented to them. In contrast, when they decided that the item was an old item, participants had to choose the colour of the item. Last, participants were to remember whether the object had been presented complete or incomplete. Each of these judgements was made via the computer mouse and each memory question was given on a separate slide. This form of a staged memory judgment process was selected to make the task clearer and easier for participants. Compared to memory tests in the previously described generation effect studies, it is rather uncommon. However, an effect of this presentation design on memory performance seems implausible.

Moreover, object *names* instead of the actual *pictures* of the objects, which were displayed at study, were provided at test. This was done to avoid transfer appropriate processing at test. Transfer appropriate processing describes the phenomenon of memory being best, when retrieved under circumstances identical to the original experience. Since drawings had either been studied complete or incomplete, they could potentially be shown at test either complete or incomplete. All four combinations (complete at study and complete at test, complete at study and incomplete at test, incomplete at study and complete at test, incomplete at study and incomplete at test) could be problematic, when attempting to test for memory irrespective of transfer appropriate processing effects. There could be (a) a match of perceptual features and (b) a match of cognitive task (Morris et al., 1977; Schendan & Kutas, 2007). In order to avoid both, target names were presented.

Because of the present arguments, I conclude that item fragmentation, semantic framing, the selected cues, item colouring, and testing can be considered appropriate, and that self-generation of items and a high probability of self-generating the correct solution were reasonably well ensured.

Open Questions, Further Directions, and Future Research for Effects Similar to the Generation Effect

In the introduction section of this dissertation, I state clearly that the present work is concerned only with the generation effect and that definitions and empirical evidence pertaining to only this effect are delineated.

However, other effects exist in memory research that bear some resemblance to the generation effect. Examples are *effects of bizarreness* and *effects of visual masking*, or *effects of perceptual interference*. Researchers found that bizarre stimuli usually facilitate recall compared to common stimuli (e.g., Macklin & McDaniel, 2005; Marchal & Nicolas, 2000; McDaniel, Dornburg, & Guynn, 2005). Marchal and Nicolas (2000), created bizarreness in pictures by multiplying specific elements of objects, such as presenting an axe with three axe blades instead of one. In backward masking, a target stimulus is typically displayed very briefly and is followed by another stimulus, such as a row of Xs. It was shown that stimuli that are masked during encoding are remembered better than stimuli that are unmasked (see also Westerman & Greene, 1997). This effect of visual masking is also called the perceptual-interference effect and generally denotes the phenomenon that interfering with stimulus identification (during study) can enhance later recall of this stimulus (see also Hirshman & Mulligan, 1991; Mulligan, 1996). Additionally, stimuli presented under rapid viewing conditions are remembered better than stimuli presented at a pace which allows for perfect recognition (see Nairne, 1988).

The common denominator in these manipulations can be seen in the facts that (a) participants encode and retrieve information and that (b) all stimuli are in a way altered and thus have to be (mentally) changed either automatically or effortfully to comply with their original (i.e., typical) form of presentation.

Since these effects are similar to the generation effect, it seems interesting to investigate the current issues (the processing of perceptual attributes and the processing of internal states, self-reference effect, conceptual processing, visual and nonvisual processing, schema-typicality) for these effects as well.

Thoughts on What Was Learned and a Short Outlook on the Generation Effect

In the current dissertation project contributions to the scientific field were obtained concerning several areas and effects as well as concerning several levels of specificity.

First, conclusions can be drawn for other psychological effects. These conclusions refer to a rather broad level of specificity, namely they refer to general psychological laws, theories, and factors. In this dissertation, it was shown that the more natural the stimuli are, the better item memory for these stimuli is. It was furthermore demonstrated that the more natural the stimuli are, the stronger the effect of prior knowledge, concepts, schemata, and typicality of features is. Concerning typicality, it was revealed that an atypicality effect exists for drawings and photos of objects which have got a schema-typical colour. Presentation colour was remembered better when it was atypical of an object than when it was typical of an object. Moreover, further evidence was provided that the script-copy-plus-tag hypothesis seems to serve as a reasonable explanation for why the atypicality effect emerges. Also, a more pronounced atypicality effect was revealed, when incidental study instructions are employed. Thus, it can be concluded that the activation of a concept (along with the activation of its schema-typical colour) indeed is a rather weak - that is easily suppressed - activation that can be diminished or overcome when using intentional study instructions.

Second, conclusions can be drawn for the effect of self-generation in item memory tasks. These conclusions refer to a narrower level of specificity, namely they refer to effects of self-generation on item memory tasks only. In this dissertation, it was again shown that for semantic generation tasks, generally a tendency and trend exists for a positive generation effect in item memory tasks, equally for words, drawings, and photos. This result endorses the robustness of the effect and fits well with previous empirical evidence.

Third, conclusions can be drawn for the effects of self-generation in source memory tasks. These conclusions refer to another topic, which is at a level of specificity that is as narrow as the previous one, namely they refer to effects of self-generation on source memory tasks only. In this dissertation, it was again shown that for the source memory dimension target colour, a negative generation effect generally results. Moreover, for the source memory dimension degree of completeness, a positive generation effect generally emerges. These findings support Mulligan (2004), Mulligan et al. (2006), and most essentially they corroborate Riefer et al. (2007).

Most importantly, specific conclusions can be drawn for a specified research level. It can be concluded that the two lines of research investigated in the course of this dissertation, which deal with the generation effect and source memory, considerably help to shed light on present inconsistencies, contradictions, and help to illuminate unanswered questions and hypotheses. These conclusions refer to the narrowest level of specificity. Apart from (1) illuminating the role of the processing of perceptual attributes and the processing of internal states, as well as the role of conceptual processing, the present research findings furthermore (2) help to identify a need for more specific data on the role of the self-reference effect in generation effect studies, for more data on the role of schema-typicality, for more data on the role of conceptual and perceptual processing, and finally a need for more data on the role of visual and non-visual processing.

To reiterate, Bertsch et al. (2007) concluded their review paper on generation effects in *item memory* tasks by stating that "regardless of what the underlying cognitive mechanisms may be, the generation effect appears to be a real phenomenon that deserves further empirical study" (p. 207). The current dissertation strongly indicates that the same claim can be stressed for the effect of self-generation on *source memory* tasks: Self-generating stimuli results in significant source memory effects that deserve further empirical study.

Eventually, it would be useful for the future, to attain a meta-analytic review for source memory and the generation effect, and ultimately to achieve a coherent framework for effects of self-generation on source memory tasks. It is furthermore desirable to obtain an overarching theory that includes explanations and fundamental factors (possibly emerging from general psychological laws), which are ready to account for effects of self-generation on source memory and on item memory likewise.

Chapter 5:
References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203-231.

Anderson, J. R. (1976). Language, memory, and thought. Hillsdale. NJ: Erlbaum.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1-12.

Baddeley, A. D. (1978). The trouble with levels: A re-examination of Craik and Lockhart's framework for memory research. *Psychological Review*, 85, 139-152.

Barrett, A. M., Crucian, G. P., Schwartz, R. L., & Heilman, K. M. (2000). Testing memory for self-generated items in dementia: Method makes a difference. *Neurology*, *54*, 1258-1264.

Barsalou, L. W. (1999). Perceptual symbol system. *Behavioral and Brain Sciences*, *22*, 577-660.

Bartleson, C. J. (1960). Memory colors of familiar objects. *Journal of the Optical Society of America*, *50*, 73-77.

Bartlett, F. C. (1932, reprint 1995). Remembering: A study in experimental psychology. New York, NY, US: Cambridge University Press.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57-86.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197-215.

Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 977-989.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*, 201-210.

Bolles, R. C., Hulicka, I. M., & Hanly, B. (1959). Colour judgment as a function of stimulus conditions and memory colour. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *13*, 175-185.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177-220.

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*, 52-60.

Brown, A. S., Jones, E. M., & Davis, T. L. (1995). Age differences in conversational monitoring. *Psychology & Aging*, *10*, 111-122.

Bruner, J. S., Postman, L., & Rodrigues, J. (1951). Expectation and the perception of color. *The American Journal of Psychology*, *64*, 216-227.

Burns, D. J. (1990). The generation effect: A test between single- and multifactor theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1060-1067.

Chechile, R. A., & Soraci, S. A. (1999). Evidence for a multiple-process account of the generation effect. *Memory*, *74*, 483–508.

Colllins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Colllins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 669-675.

Destun, L. M., & Kuiper, N. A. (1999). Phenomenal characteristics associated with real and imagined events: The effects of event valence and absorption. *Applied Cognitive Psychology*, *13*, 175-186.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*, 945-955.

Diagnostic colour objects provided by Tarrlab. (n.d.). Retrieved April 15, 2009, from http://www.cnbc.cmu.edu/tarrlab/stimuli/objects/object-stroop.zip/view.html

Donaldson, W., & Bass, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior*, *19*, 26-35.

Duncker, K. (1939). The influence of past experience upon perceptual properties. *The American Journal of Psychology*, *52*, 255-265.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 588-598.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 671-682.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, *217* Special issue: New directions in multinomial modeling, 108-124.

Erdfelder, E., & Bredenkamp, J. (1998). Recognition of script-typical versus script-atypical information: Effects of cognitive elaboration. *Memory & Cognition, 26*, 922-938.

Finke, R. A., Johnson, M. K., & Shyi, G. C. (1988). Memory confusion for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, *16*, 133-137.

Fisher, C., Hull, C., & Holtz, P. (1956). Past experience and perception: Memory color. *The American Journal of Psychology*, *69*, 546-560.

Foss, P. W., Mora, J. J., & Tkacz, S. (1994). Student Study Techniques and the Generation Effect. *Journal of Educational Psychology*, *86*, 567-576.

Gardiner, J. M., Dawson, A. J., & Sutton, E. A. (1989). Specificity and generality of enhanced priming effects for self-generated study items. *The American Journal of Psychology*, *102*, 295-305.

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 732-741.

Gardiner, J. M., & Rowley, J. M. C. (1984). A generation effect with numbers rather than words. *Memory & Cognition*, *12*, 443-445.

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, *26*, 181-206.

Gegham, K. D., & Multhaup, K. S. (2004). How generation affects source memory. *Memory & Cognition*, *32*, 819-823.

Genzel, S., Kerkhoff, G., & Scheffter, S. (1995). PC-gestützte Standardisierung des Bildmaterials von Snodgrass & Vanderwart (1980) I. Deutschsprachige Normierung. *Neurolinguistik*, *9*, 41-53.

Goldstein, B. (2002). *Wahrnehmungspsychologie*. Heidelberg – Berlin: Spektrum-Verlag.

Goodman, G. S. (1980). Picture memory: How the action schema affects retention. *Cognitive Psychology*, *12*, 473-495.

Graesser, A. C., & Nakamura, G. V. (1982). *The impact of schemas on comprehension and memory*. In: Bower, G. H. (Ed.), The Psychology of Learning and Motivation, Vol. 16. New York: Academic Press.

Graf, P. (1980). Two consequences of generating: Increased inter- and intraword organization of sentences. *Journal of Verbal Learning and Verbal Behavior*, *19*, 316-327.

Greene, R. L. (1992). *Human memory: Paradigms and Paradoxes.* Hillside, NJ: Erlbaum.

Grosofsky, A., Payne, D. G., & Campbell, K. D. (1994). Does the generation effect depend upon selective displaced rehearsal. *American Journal of Psychology*, *107*, 53-68.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*, 1367-1368.

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 484-494.

Hirshman, E., & Mulligan, N. (1991). Perceptual Interference Improves Explicit Memory but Does Not Enhance Data-Driven Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 507-513.

Horton, K. D. (1987). The incongruity effect in memory for generated targets: Fact or artifact? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 172-174.

Hunt, R. W. G. (1991). *Measuring Colour*. Chichester, UK: Ellis Horwood Series in Applied Science and Industrial Technology.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421-445.

Irtel, H. (2000). *Farbatlanten*. Spektrum der Wissenschaft Spezial: Farben, Bd. 4/2000, pp. 22-27.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*, 485-508.

Jacoby, L. L., & Craik, F. I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, *18*, 585-600.

James, W. (1890). The principles of psychology. New York: Holt.

Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America, A, Optics, Image Science & Vision*, *13*, 1981-1991.

Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 180-190.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.

Johnson, M. K., Nolde, S. F., & De Leonardis, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory and Language*, *35*, 135-156.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67-85.

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, *94*, 37-64.

Johnson, M. K., Raye, C. L., Foley, M. A., & Kim, J. K. (1982). Pictures and images: Spatial and temporal information compared. *Bulletin of the Psychonomic Society*, *19*, 23-26.

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition*, *24*, 648-656.

Kane, J., & Anderson, R. C. (1978). Depth of processing and interference effects in the learning and remembering of sentences. *Journal of Educational Psychology*, *70*, 626-635.

Kelly, M. H., Bock, J. K., & Keil, F. C. (1986). Prototypicality in a linguistic context: Effects on sentence structure. *Journal of Memory and Language*, *25*, 59-74.

Kinjo, H., & Snodgrass, J. G. (2000). Does the generation effect occur for pictures. *American Journal of Psychology*, *113*, 95-121.

Klein, S. B., & Loftus, J. (1993). Behavioral experience and trait judgments about the self. *Personality and Social Psychology Bulletin*, *19*, 740-745.

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and output events in memory. *Psychological Research/Psychologische Forschung*, *53*, 260-270.

Listing of the largest German cities provided by Wikipedia German. (n.d.). Retrieved November 10, 2008, from http://de.wikipedia.org/wiki/Liste_der_gr%C3%B6%C3%9Ften_St%C3%A4dte_Deutschlands.

Lockhart, R. S. (2002). Levels of processing, transfer-appropriate processing, and the concept of robust encoding. *Memory*, *10*, 397-403.

Lohaus, D., & Lachnit, H. (2001). Generierungseffekte mit Bildmaterial. *Psychologische Beiträge*, *43*, 787-802.

Lubinsky, T., Rich, J. B., & Anderson, N. D. (2009). Errorless learning and elaborative self-generation in healthy older adults and individuals with amnestic mild cognitive impairment: Mnemonic benefits and mechanisms. *Journal of the International Neuropsychological Society*, *15*, 704-716.

Macklin, C. B., & McDaniel, M. A. (2005). The bizarreness effect: Dissociation between item and source memory. *Memory*, *13*, 682-689.

Marchal, A., & Nicolas, S. (2000). Is the picture bizarreness effect a generation effect? *Psychological Reports*, *87*, 331-340.

Marsh, E. J. (2006). When does generation enhance memory for location? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1216-1220.

Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition*, *29*, 798-805.

Masson, M. E. J., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General*, *121*, 145-176.

McClelland, J. L., & Rumelhart, D. E. (1986). *A distributed model of human learning and memory*. In: McClelland, J.L., & Rumelhart, D.E. (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models (pp. 170-215). Cambridge, MA: MIT Press.

McDaniel, M. A., Dornburg, C. C., & Guynn, M. J. (2005). Disentangling encoding versus retrieval explanations of the bizarreness effect: Implications for distinctiveness. *Memory & Cognition*, *33*, 270-279.

McDaniel, M. A., Wadill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, *27*, 521-536.

McElroy, L. A. (1987). The generation effect with homographs: Evidence for postgeneration processing. *Memory & Cognition*, *15*, 148-153.

McFarland, C. E., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior*, *19*, 210-225.

McNamara, T. P. (1992). Priming and constraints in places on theories of memory and retrieval. *Psychological Review*, *99*, 650–662.

McNamara, T. P. (1994). Priming and theories of memory: a reply to Ratcliff and McKoon. *Psychological Review*, *101*, 185–187.

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language*, *43*, 652-679.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 116-137.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*, 283-284.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*, 225-229.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology*, *19*, 743-768.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227-234.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Mitchell, K. J., & Johnson, M. K. (2000). *Source monitoring: Attributing mental experiences*, In: Tulving, E., & Craik, F. I. M. (Eds.), The Oxford handbook of memory (pp. 179-195). New York, NY, US: Oxford University Press.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519-533.

Moshagen, M. (2010). Multitree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42-54.

Mulligan, N. W. (1996). The Effects of Perceptual Interference at Encoding on Implicit Memory, Explicit Memory, and Memory for Source. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1067-1087.

Mulligan, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 436-450.

Mulligan, N. W. (2002a). The effects of generation on conceptual implicit memory. *Journal of Memory and Language*, *47*, 327-342.

Mulligan, N. W. (2002b). The emergent generation effect and hypermnesia: Influences of semantic and nonsemantic generation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 541-554.

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 838-855.

Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 836-846.

Multhaup, K. S., & Balota, D. A. (1997). Generation effects and source memory in healthy older adults and in adults with dementia of the alzheimer type. *Neuropsychology*, *11*, 382-391.

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition*, *24*, 417-428.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 70-84.

Murray, F. S., & Kinnison, E. L. (1989). Degree of fragmentations and number of distinctive features in the recognition of pictured objects by children and adults. *Bulletin of the Psychonomic Society*, *27*, 121-124.

Nairne, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 248-255.

Nairne, J. S., Pusen, C., & Widner, R. L. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition*, *13*, 183-191.

Nairne, J. S., & Widner, R. L. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 164-171.

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception*, *32*, 667-680.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*, 231-259.

O'Brien, A., Chiaravalloti, N., Arango-Lasprilla, J. C., Lengenfelder, J., & DeLuca, J. (2007). An investigation of the differential effect of self-generation to improve learning and memory in multiple sclerosis and traumatic brain injury. *Neuropsychological Rehabilitation*, *17*, 273-292.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, *8*, 1-16.

Pesta, B. J., Sanders, R. E., & Nemec, R. J. (1996). Older adults' strategic superiority with mental multiplication: A generation effect assessment. *Experimental Aging Research*, *22*, 155-169.

Peynircioglu, Z. F. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica*, *70*, 153-160.

Rabinowitz, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology and Aging*, *4*, 259-268.

Rabinowitz, J. C. (1990). Effects of repetition of mental operations on memory for occurrence and origin. *Memory & Cognition*, *18*, 72-82.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318-339.

Riefer, D. M., Chien, Y., & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *Quarterly Journal of Experimental Psychology*, *60*, 1389-1405.

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 680-693.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, *12*, 1-20.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677-688.

Rojahn, K., & Pettigrew, T. F. (1988). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, *31*, 81-109.

Rosch, E. (1978). *Principles of categorization*. In: Rosch, E., & Lloyd, B.B. (Eds.), Cognition and Categorization (pp. 27-48). New Jersey: Lauwrence Erlbaum Associates.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 491-502.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology*, *15*, 343-355.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*, 534-553.

Schacter, D. L., Harbluk, J. L., & McLachlan, D. R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior*, *23*, 593-611.

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for transfer appropriate processing of memory: Processing versus feature similarity. *Psychonomic Bulletin & Review*, *14*, 612-619.

Schlink, S., & Walter, E. (2007). Kurz und gut: Eine duetsche Kurzskala zur Erfassung des Bedürfnisses nach kognitiver Geschlossenheit. *Zeitschrift für Sozialpsychologie*, 38, 153-161.

Schmidt, S. R., & Cherry, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition*, *17*, 359-369.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Schwartz, B. L. (1989). Effects of generation on indirect measures of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1119-1128.

Shimamura , A. P., & Squire, L. R. (1987). Memory dysfunction and word priming in dementia and amnesia. *Behavioral Neuroscience*, *101*, 347-351.

Shimamura , A. P., & Squire, L. R. (1991). The relationship between fact and source memory: Findings from amnesic patients and normal subjects. *Psychobiology*, *19*, 1-10.

Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior*, *22*, 153-163.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592-604.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language*, *26*, 589-607.

Snodgrass, J. G., Smith, B., Feenan, K., & Corwin, J. (1987). Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications. *Behavior Research Methods, Instruments, & Computer*, *19*, 270-274.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for naming agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174-215.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology*, *36*, 941-950.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*, 42-61.

Steffens, M. C., & Erdfelder, E. (1998). Determinants of positive and negative generation effects in free recall. *Quarterly Journal of Experimental Psychology*, *51*, 705-733.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*, 371-394.

Taconnat, L., Baudouin, A., Fay, S., Clarys, D., Vanneste, S., Tournelle, L., & Isingrini, M. (2006). Aging and implementation of encoding strategies in the generation of rhymes: The role of executive functions. *Neuropsychology*, *20*, 658-665.

Taconnat, L., Froger, C., Sacher, M., & Isingrini, M. (2008). Generation and associative encoding in young and old adults: The effect of the strength of association between cues and targets on a cued recall task. *Experimental Psychology*, *55*, 23-30.

Taconnat, L., & Isingrini, M. (2004). Cognitive Operations in the Generation Effect on a Recall Test: Role of Aging and Divided Attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 827-837.

Troyer, A. K., Häfliger, A., Cadieux, M. J., & Craik, F. I. M. (2006). Name and Face Learning in Older Adults: Effects of Level of Processing, Self-Generation, and Intention to Learn. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *61*, 67-74.

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*, 101-118.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 607-617.

Voss, J. F., Vesonder, G. T., Post, T. A., & Ney, L. G. (1987). Was the item recalled and if so by whom. *Journal of Memory and Language*, *26*, 466-479.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*, 1049-1062.

Westerman, D. L., & Greene, R. L. (1997). The Effects of Visual Masking on Recognition: Similarities to the Generation Effect. *Journal of Memory and Language*, *37*, 584-596.

Zillig, S. D. (2007). *Generation Effects on Item and Source Memory for Pictorial Material*. In: Grainger, J., Alario, F.-X., Burle, B, & Janssen, N. (Eds.), Proceedings of the XVth meeting of the European Society for Cognitive Psychology (p. 111). Marseille: European Society for Cognitive Psychology.

Chapter 6:

Appendix

Appendix A: Listing of studies implementing either a reality monitoring paradigm or an external source monitoring paradigm. Names of authors and years of publication, implementations of the paradigms (read condition vs. generate condition), reported results, and expected results according to Riefer et al. (2007) are presented. Result patterns include positive generation effects ("+ GE"), negative generation effects ("- GE"), and null effects.

Studies implementing a reality monitoring paradigm (RMP)

Author(s) and year	Implementation of RMP: read condition (RC) versus generate condition (GC)	Reported results for source	Expected results according to Riefer et al. (2007)
Geghman & Multhaup (2004)	Participants generated aloud answers (GC) or read solutions (RC) given by either one (Exp. 1) or two (Exp. 2) external sources. Sources were faces on the computer screen	+ GE	+ GE
Voss, Vesonder, Post, & Ney (1987)	Joked pairs of participants recalled items of word lists. Subsequently, participants judged (a) which items were recalled and (b) which person (self vs. other) recalled items judged as recalled.	Null effect	+ GE
Riefer, Hu, & Batchelder (1994)	Re-analysis of data collected in Voss, Vesonder, Post, & Ney (1987) by employing the two-source multinomial model (Riefer & Batchelder, 1988)	+ GE	+ GE
Jurica & Shimamura (1999)	Participants generated answers (GC) or read statements (RC) given by one of three sources (i.e., 3 faces on the computer screen)	- GE	+ GE

Studies implementing an external source monitoring paradigm (ESMP)

Author(s) and year	Implementation of ESMP: read condition (RC) versus generate condition (GC)	Reported results for source	Expected results according to Riefer et al. (2007)
Mulligan (2004); Exps. 1 & 3	Participants either read target words (RC) or generated them from a cue word (GC). Recollection of target colour was tested.	- GE	- GE
Mulligan (2004); Exps. 2 & 4	Participants either read target words (RC) or generated them from a cue word (GC). Recollection of target location on screen was tested.	Null effect	- GE
Mulligan (2004); Exp. 6	Participants either read target words (RC) or generated them from a cue word (GC). Recollection of background colour was tested.	Null effect	- GE
Mulligan et al. (2006) Exps. 2A, 4A, 6A, & 6B	Participants either read targets (RC) or generated them from a cue (GC). Recollection of target colour was tested.	- GE (2A & 4A) Null effect (6A & 6B)	- GE
Mulligan et al. (2006) Exps. 2B & 4B	Participants either read targets (RC) or generated them from a cue (GC). Recollection of target location on screen was tested.	- GE	- GE

Appendix A: continued

Studies implementing an external source monitoring paradigm (ESMP)			
Author(s) and year	Implementation of ESMP: read condition (RC) versus generate condition (GC)	Reported results for source	Expected results according to Riefer et al. (2007)
Mulligan et al. (2006) Exps. 2C, 2D, & 5	Participants either read targets (RC) or generated them from a cue (GC). Recollection of background colour was tested.	Null effect	- GE
Marsh et al. (2001) Exp. 1	Participants either read targets (RC) or generated them from a cue (GC) and were asked to remember the study context ; contexts were two different rooms.	+ GE	- GE
Marsh et al. (2001) Exp. 2A	Participants either read targets (RC) or generated them from a cue (GC) and were asked to remember the study context; contexts were two different computer screens.	+ GE	- GE
Marsh et al. (2001) Exp. 2B	Participants either read target words (RC) or generated them from a cue word (GC). Recollection of target colour and font (orange and uppercase versus blue and lowercase) was tested.	+ GE	- GE

Appendix B: Stimuli employed in Experiments 1 and 2: cue words in German and in English, target names in German and in English, counterbalancing group (CB), and source. Pictures were obtained from Masson and McLeod (1992, Appendix B) or were produced by myself ([1]).

#	Cue word German	Cue word English	Target name German	Target name English	CB
1	Gewinner	winner	Verlierer	loser	1
2	Zukunft	future	Vergangenheit	past	1
3[1]	Feuer	fire	Wasser	water	1
4	jünger	junior	älter	senior	1
5	spät	late	früh	early	1
6	lang	long	kurz	short	1
7	reich	rich	arm	poor	1
8	Freund	friend	Feind	enemy	1
9[1]	hoch	high	tief	deep	1
10[1]	Licht	light	Schatten	shade	1
11	drücken	push	ziehen	pull	1
12	schnell	fast	langsam	slow	2
13	Leben	life	Tod	death	2
14[1]	leeren	(to) empty	füllen	(to) fill	2
15	größer	major	kleiner	minor	2
16	Frage	question	Antwort	answer	2
17	kaufen	(to) buy	verkaufen	(to) sell	2
18	weinen	(to) cry	lachen	(to) laugh	2
19[1]	Salz	salt	Zucker	sugar	2
20	Krieg	war	Frieden	peace	2
21	höher	higher	niedriger	lower	2
22	hübsch	pretty	hässlich	ugly	2
23	voll	full	leer	empty	3
24	weich	soft	hart	hard	3
25	eng	tight	weit	loose	3
26	Einzahl	singular	Mehrzahl	plural	3
27	Tag	day	Nacht	night	3
28[1]	hell	light	dunkel	dark	3
29	glatt	smooth	rau	rough	3
30	betrunken	drunk	nüchtern	sober	3
31[1]	dick	thick	dünn	thin	3
32[1]	alt	old	jung	young	3
33[1]	männlich	male	weiblich	female	3
34	zuerst	first	zuletzt	last	4
35	stark	strong	schwach	weak	4
36	schlafend	asleep	wach	awake	4
37[1]	Berg	mountain	Tal	valley	4
38[1]	Essig	vinegar	Öl	oil	4
39	heiß	hot	kalt	cold	4
40	sitzen	(to) sit	stehen	(to) stand	4
41[1]	vorher	earlier	nachher	later	4
42	mehr	more	weniger	less	4
43	Oberteil	top	Unterteil	bottom	4
44[1]	rund	round	eckig	square-cut	4

Appendix B: continued

#	Cue word German	Cue word English	Target name German	Target name English	CB
45	Eingang	entrance	Ausgang	exit	5
46[1]	wichtig	important	unwichtig	unimportant	5
47	Sommer	summer	Winter	winter	5
48	Gewinn	profit	Verlust	loss	5
49	Norden	north	Süden	south	5
50	offen	open	geschlossen	close	5
51	Rückseite	back	Vorderseite	front	5
52[1]	rechts	right	links	left	5
53	unschuldig	innocent	schuldig	guilty	1
54	schlecht	bad	gut	good	1
55[1]	flüssig	fluid	fest	solid	1

Appendix C: Original wordings for test options in Experiments 1 and 2

Instruction	Original wordings
The processing of and memory for perceptual attributes (i.e., PA condition)	Das Wort war vollständig und rot. Das Wort war vollständig und grün. Das Wort war unvollständig und rot. Das Wort war unvollständig und grün. Dies ist ein neues Wort
The processing of and memory for internal states (i.e., IS condition)	Ich konnte das Wort einfach auf dem Bildschirm lesen und es war rot. Ich konnte das Wort einfach auf dem Bildschirm lesen und es war grün. Ich habe das Wort selbst aufgrund des ersten Buchstabens hergestellt und es war rot. Ich habe das Wort selbst aufgrund des ersten Buchstabens hergestellt und es war grün. Dies ist ein neues Wort.

Appendix D: More detailed description of correct and erroneous responses given during study in Experiment 1 (upper table) and Experiment 2 (lower table)

Instruction	Type of item	% correct responses	Number of absolute errors	% correct responses	Number of absolute errors	% correct responses	Number of absolute errors
Perceptual attributes	Red complete	100 %	-	100 %	-	99.47 %	7
	Green complete	100 %	-				
	Red incomplete	98.79 %	4	98.94	7		
	Green incomplete	99.09 %	3				
Internal states	Red complete	99.39 %	2	99.55 %	3	99.09 %	10
	Green complete	99.70 %	1				
	Red incomplete	98.48 %	3	98.64 %	7		
	Green incomplete	98.79 %	4				

Instruction	Type of item	% correct responses	Number of absolute errors	% correct responses	Number of absolute errors	% correct responses	Number of absolute errors
Perceptual attributes	Red complete	99.49 %	1	99.49 %	2	99.24 %	6
	Green complete	99.49 %	1				
	Red incomplete	98.99 %	2	98.99 %	4		
	Green incomplete	98.99 %	2				
Internal states	Red complete	100 %	-	100 %	-	99.90 %	1
	Green complete	100 %	-				
	Red incomplete	100 %	-	99.80 %	1		
	Green incomplete	99.60 %	1				

Appendix E: Descriptions of how Pilots I, II, IIIa, and IIIb were conducted and of how stimuli were selected for use in Experiment 3

In Pilots I, II, IIIa, and IIIb, five different versions of the original questionnaire were designed, each displaying the stimuli in a separate individually randomised order. The paper-pencil questionnaires were sent via mail or handed to participants, who were instructed to fill in their questionnaire at home and to return it afterwards; there was a range of 1 to 2 weeks till return of questionnaires.

Pilot I was aimed to obtain *semantic cues* and contained all 260 pictures from Snodgrass and Vanderwart (1980). The black and white drawings were displayed sequentially along with their names, to rule out misunderstandings and difficulties regarding correct identification of the objects. Participants were instructed to closely consider each drawing and to write down on a line next to each picture the word which came to mind spontaneously. It was emphasised that all types of words were allowed as responses (i.e., nouns, verbs, adjectives, etc.). The response that was most frequently produced to an item was selected as the picture's cue word. This was true, except under two conditions: (1) The most frequently produced word was either the colour word "red" or the colour word "green" – this was problematic due to the possibility of confounding and confusing the cue with the independent variable "colour*".* (2) The most frequently produced word matched the most frequent response for another item. In case of (1) and (2), the response with the second highest frequency was selected. As a result, cue words were unique for target pictures.

Pilot II was aimed at finding *German picture names for objects not included in Genzel Kerkhoff, and Scheffter (1995)*, and thus comprised these remaining drawings. The black and white drawings were displayed sequentially. Participants were asked to write down the names of the objects in the pictures and to rate the drawings on familiarity and visual complexity (on a 5-point scale). The most frequent naming was chosen to serve as the object's label in Experiment 3; means and standard deviation values were calculated across participants for each of the items for the dimensions name agreement, familiarity, and visual complexity.

Appendix E: continued

Then, *name agreement* was assessed with a range of 0 % (no name agreement) to 100 % (perfect name agreement). Name agreement reflects the percentage of all participants that gave the most frequent name when asked to name a drawing. A *high score of name agreement* was important, because participants were to write down the name of the pictures in the experiments in a fixed period of time. Problems could arise in the course of the experiment, if participants could not name the pictures correctly and would afterwards be prompted with a name inconsistent with what they had written down. The same is true when no name could be generated at study. Problems could arise (a) from poor compliance in study completion or (b) from changes in mood or cooperation, or (c) from differences in participants' encoding or retrieval abilities as compared to that for participants, who had named the items correctly. Across all stimuli, mean name agreement was at 81.83 % and median name agreement was at 89.5 %. For typically red items, mean and median name agreement were at 79.3 % and 92 %, respectively. For typically green items, mean and median name agreement were at 84.3 % and 88 %, respectively.

Pilots IIIa and *IIIb* were aimed at identifying items that have either red or green as a *typical colour*. Participants were given questionnaires consisting of all 260 objects in the Snodgrass and Vanderwart (1980) picture set. The black and white drawings were displayed sequentially along with their names. Participants were instructed to closely consider each drawing and to state whether or not an item typically appeared in a certain colour and what this colour was. Then, *colour agreement* was assessed with a range of 0 % (no colour agreement) to 100 % (perfect colour agreement). For Experiment 3, 20 items were selected with a mean colour agreement of 89.57 % and a median colour agreement of 93.33 % (10 red items, 10 green items). Items for which colour denotation was less pronounced or for which other colour names had been given, were not considered further.

Name and colour agreement values for Experiment 3 can be found in Appendix G.

Appendix F: Sample descriptions for Pilots I, II, IIIa, IIIb, IV, and V

	Pilot I
Total number of participants	35
Gender	82.86 % female (29 out of 35 participants)
Age	Ranging from 20 to 29 years; mean of 24.93 years; median of 25 years
Native speakers	97.14 % native German speakers (34 of 35 participants); 1 non-native speaker: "very good" German proficiency skills (self-report)
Study & work	18 currently enrolled (12 Psychology, 2 Sociology, 3 Economics, 1 Language Studies); 20 alumni; 2 were employees
	Pilot II
Total number of participants	36
Gender	68.57 % female (24 out of 35 participants)
Age	Ranging from 20 to 29 years; mean of 24.77 years; median of 25 years
Native speakers	97.22 % native German speakers (35 of 36 participants); 1 non-native speaker: "very good" German proficiency skills (self-report)
Study & work	23 currently enrolled (10 Psychology, 5 Sociology, 5 Economics, 2 Language Studies, 1 Electrical Engineering); 11 alumni; 2 were employees
	Pilot IIIa and Pilot IIIb
Total number of participants	Stimulus-set A: 15 Stimulus-set B: 14
Gender	Stimulus-set A: 73.33 % female (11 out of 15 participants) Stimulus-set B: 78.57 % female (11 out of 14 participants)
Age	Stimulus-set A: Ranging from 20 to 30 years; mean of 25.78 years; median of 26 years Stimulus-set B: Ranging from 21 to 30 years; mean of 25.85 years; median of 26 years
Native speakers	Stimulus-set A: 86.67 % native German speakers (13 out of 15 participants); 2 non-native speakers: "very good" and "satisfactory" German proficiency skills (self-report) Stimulus-set B: 92.86 % native German speakers (13 out of 14 participants); 1 non-native speaker: "very good" German proficiency skills (self-report)
Study & work	Stimulus-set A: 4 currently enrolled (3 Psychology, 1 Economics); 10 alumni; 1 were employees Stimulus-set B: 3 currently enrolled (2 Psychology, 1 Economics); 10 alumni, 1 were employees
	Pilot IV and Pilot V
Total number of participants	15
Gender	86.66 % female (13 out of 15 participants)
Age	Ranging from 20 to 28 years; mean of 23.93 years; median of 24 years
Native speakers	93.33 % native German speakers (14 out of 15 participants); 1 non-native speaker: "very good" German proficiency skills (self-report)
Study & work	11 currently enrolled (9 Psychology, 1 Sociology, 1 Economics); 4 alumni (Psychology)

Appendix G: Stimuli used in Experiment 3: cue words in German and in English, target names in German and in English, typical colour denotation, colour agreement value, name agreement value, and counterbalancing group (CB). Items were obtained from Snodgrass and Vanderwart (1980); colour and name agreement values were taken from Genzel, Kerkhoff, and Scheffter (1995) or were obtained in Pilots II, IIIa, and IIIb.

#	Cue word German	Cue word English	Target name German	Target name English	Typical colour	Colour agree-ment	Name agree-ment	CB
1	Apfel	apple	Birne	pear	green	71.43 %	100	1
2	gesund	healthy	Salat	lettuce	green	100 %	37	1
3	Panzer	tortoise shell	Schildkröte	turtle	green	86.67 %	100	2
4	Ahorn	maple	Blatt	leaf	green	93.33 %	59	2
5	quaken	(to) croak	Frosch	frog	green	100 %	96	3
6	Plage	plague	Heuschrecke	grasshopper	green	93.33 %	61	3
7	Reptil	reptile	Krokodil	alligator	green	100 %	93	4
8	Pizza	pizza	Artischocke	artichoke	green	93.33 %	63	4
9	Natur	nature	Baum	tree	green	86.67 %	93	5
10	krabbeln	(to) crawl	Raupe	caterpillar	green	100 %	91	5
11	süß	sweet	Erdbeere	strawberry	red	100 %	100	1
12	Meer	sea	Hummer	lobster	red	60 %	49	1
13	Liebe	love	Herz	heart	red	93.33 %	100	2
14	Kuss	kiss	Mund	lips	red	86.67 %	69	2
15	schlau	clever	Fuchs	fox	red	80 %	99	3
16	fliegen	(to) fly	Luftballon	balloon	red	86.67 %	71	3
17	Ketchup	ketchup	Tomate	tomato	red	100 %	93	4
18	Obst	fruit	Kirsche	cherry	red	100 %	86	4
19	Gemüse	vegetable	Paprika	pepper	red	93.33 %	89	5
20	Mädchen	girl	Schleife	bow	red	66.67 %	87	5
21	sehen	(to) see	Auge	eye	none		100	1
22	Richtung	direction	Pfeil	arrow	none		94	1
23	Spielzeug	toy	Puppe	doll	none		81	1
24	Teig	dough	Schüssel	bowl	none		51	1
25	Wachs	wax	Kerze	candle	none		99	2
26	Flughafen	airport	Flugzeug	airplane	none		96	2
27	spielen	(to) play	Ball	ball	none		84	2
28	Tiere	animals	Bauernhof	barn	none		33	2
29	Winter	winter	Schneemann	snowman	none		97	3
30	Sport	sports	Tennisschläger	tennisraket	none		90	3
31	Holland	Netherlands	Windmühle	windmill	none		89	3
32	Kinder	children	Leiterwagen	wagon	none		30	3
33	Wiese	meadow	Blume	flower	none		97	4
34	Verkehr	traffic	Ampel	traffic light	none		91	4
35	Zeit	time	Uhr	watch	none		89	4
36	Hunger	hunger	Sandwich	sandwich	none		56	4
37	Frau	woman	Kleid	dress	none		97	5
38	China	China	Vase	vase	none		94	5
39	Herbst	autumn	Drachen	kite	none		89	5
40	schick	chic	Rock	skirt	none		80	5

Appendix H: Stimuli used in Experiments 4 and 5: cue words in German and in English, target names in German and in English, counterbalancing group (CB), and source. Items were adopted from Naor-Raz et al. (2003) provided by Tarrlab (2009, April 15) or were produced by myself ([1]).

#	Cue word German	Cue word English	Target name German	Target name English	Typical colour	CB
1	Kerne	stone	Kirschen	cherries	red	1
2	bitter	bitter	Grapefruit	grapefruit	red	1
3	Weihnachten	Christmas	Weihnachtsmannmütze	Santa Clause hat	red	1
4	brennen	(to) burn	Feuerlöscher	fire extinguisher	red	2
5	Brötchen	rolls	Würstchen	sausage	red	2
6	Eis	ice cream	Erdbeere	strawberry	red	2
7	fruchtig	fruity	Apfel	apple	red	3
8	schlau	clever	Fuchs	fox	red	3
9	Sommer	summer	Tomate	tomato	red	3
10	Brand	fire	Feuerwehrauto	fire truck	red	4
11	Mais	corn	Paprika	pepper	red	4
12	Meer	sea	Hummer	lobster	red	4
13	Senf	mustard	Ketchup	ketchup	red	5
14	süß	sweet	Himbeere	raspberry	red	5
15[1]	Urlaub	holiday	Reisepass	passport	red	5
16	Pizza	pizza	Artischocke	artichoke	green	1
17	quaken	(to) croak	Frosch	frog	green	1
18	langsam	slow	Schildkröte	tortoise	green	1
19	Creme	cream	Avocado	avocado	green	2
20	stachelig	thorny	Kaktus	cactus	green	2
21	Salat	salad	Sellerie	celery	green	2
22	Gemüse	vegetable	Bohne	bean	green	3
23	Obst	fruit	Kiwi	kiwi fruit	green	3
24	Advent	Advent season	Tannenkranz	wreath	green	3
25	gesund	healthy	Brokkoli	broccoli	green	4
26	Amphibien	amphibians	Leguan	iguana	green	4
27	saftig	juicy	Wassermelone	watermelon	green	4
28	wässrig	watery	Gurke	cucumber	green	5
29[1]	Dressing	salad dressing	Kopfsalat	lettuce	green	5
30	Italien	Italy	Zucchini	zucchini	green	5
31	Luft	air	Luftballon	balloon	none	1
32	Licht	light	Kerze	candle	none	1
33	Foto	picture	Bilderrahmen	picture frame	none	1
34	Boden	floor	Teppich	rug	none	1
35	Hose	trousers	Gürtel	belt	none	1
36	Garten	garden	Blume	flower	none	1
37	Kopf	head	Kappe	baseball cap	none	2
38	gemütlich	comfortable	Sessel	armchair	none	2
39	kalt	cold	Handschuh	glove	none	2
40	sauber	clean	Seife	soap	none	2
41	Flügel	wing	Schmetterling	butterfly	none	2
42	Winter	winter	Stiefel	boot	none	2
43	Schule	school	Hefter	binder	none	3
44	Kleiderschrank	wardrobe	Kleiderbügel	hanger	none	3
45	Weste	vest	Jacke	jacket	none	3

Appendix H: continued

#	Cue word German	Cue word English	Target name German	Target name English	Typical colour	CB
46	Fuß	foot	Schuh	shoe	none	3
47	säubern	(to) clean	Zahnbürste	toothbrush	none	3
48	wegwerfen	(to) throw away	Mülltonne	garbage can	none	3
49	lesen	(to) read	Buch	book	none	4
50	weich	soft	Kissen	cushion	none	4
51	Abendessen	dinner	Serviette	serviette	none	4
52	spitz	pointed	Pin-Nadel	tack	none	4
53	trocknen	(to) dry	Föhn	hairdryer	none	4
54	lang	long	Klebeband	tape	none	4
55	putzen	(to) clean	Eimer	bucket	none	5
56	Computer	computer	Diskette	disk	none	5
57	anrufen	(to) call	Telefon	telephone	none	5
58	Nadel	needle	Wolle	yarn	none	5
59	trinken	(to) drink	Becher	cup	none	5
60	Büro	office	Tacker	stapler	none	5

Appendix I: Naming errors at study in Experiments 3, 4, and 5. Instances of correct naming are given in percentages and are categorised in different ways, bearing just some experimental variations in mind (see Section 3.2.2, for more details).

Experiment 3

Type of item	Colour	Presentation	Correct %
Items *without* schema-typical colour	Red	Complete	97.3
		Incomplete	95.8
	Green	Complete	97.7
		Incomplete	96.5
Items *with* schema-typical colour (*green*)	Red	Complete	93.8
		Incomplete	89.2
	Green	Complete	93.8
		Incomplete	91.5
Items *with* schema-typical colour (*red*)	Red	Complete	96.2
		Incomplete	91.5
	Green	Complete	96.2
		Incomplete	89.2

Type of item	Correct %	Interpretation
Items *without* schema-typical colour	96.8	When considering type of item, one can see that error rates were lowest for items without schema-typical colour; and about equal for the other types of items. However, percentages of errors were well below 10 % in all cases.
Items *with* schema-typical colour (*green*)	92.1	
Items *with* schema-typical colour (*red*)	93.3	

Type of item	Presentation	Correct %	Interpretation
Items presented in *typical* colour	Complete	95.0	When considering items presented in their typical colour (Typ) and items presented in their atypical colour (Atyp) and the experimental variation "degree of completeness", one can see that more errors were made for Atyp items. Moreover, for each type of item, more errors were made for incomplete items.
	Incomplete	91.5	
Items presented in *atypical* colour	Complete	95.0	
	Incomplete	89.2	

Typical colour	Correct %	Interpretation
Red	94.0	When considering responses given to items (typical colour either red or green), one can see that about the same amount of errors was made.
Green	93.0	

Presentation colour	Correct %	Interpretation
Red	94.62	When considering presentation colour (red vs. green), one can see that about the same amount of errors was made.
Green	94.9	

Presentation	Correct %	Interpretation
Complete	97.34	When considering the experimental variation "degree of completeness", one can see that, during the study phase, complete items were named correctly equally often as incomplete items. Moreover, values were high; cf. Mulligan (2004) and Riefer et al. (2007).
Incomplete	94.46	

Appendix I: continued

Experiment 4			
Type of item	Presentation colour	Presentation	Correct %
Items *without* schema-typical colour	Red	Complete	95.1
		Incomplete	96.4
	Green	Complete	94.4
		Incomplete	98.2
Items *with* schema-typical colour (*green*)	Red	Complete	89.2
		Incomplete	70.3
	Green	Complete	90.8
		Incomplete	83.6
Items *with* schema-typical colour (*red*)	Red	Complete	93.8
		Incomplete	98.5
	Green	Complete	99.5
		Incomplete	94.4

Type of item	Correct %	Interpretation
Items *without* schema-typical colour	95	When considering type of item, one can see that error rates for items with schema-typical colour (green) were lower than that for the other types of items. However, error rates were below 15 % in all cases.
Items *with* schema-typical colour (*green*)	86.8	
Items *with* schema-typical colour (*red*)	97.2	

Type of item	Presentation	Correct %	Interpretation
Items presented in *typical* colour	Complete	92.3	When considering items presented in their typical colour (Typ) and items presented in their atypical colour (Atyp) and the experimental variation "degree of completeness", one can see that more errors were made for Atyp items. Moreover, for each type of item, more naming errors were made for incomplete items.
	Incomplete	91.0	
Items presented in *atypical* colour	Complete	94.4	
	Incomplete	82.3	

Typical colour	Correct %	Interpretation
Red	96.5	When considering responses given to items (typical colour either red or green), one can see that more errors were made for items that are typically green. Considering the present error rates, it seems reasonable to check for the influence of typical colour. This analysis can be found in the results section of Experiment 4.
Green	83.5	

Presentation colour	Correct %	Interpretation
Red	96.9	When considering presentation colour (red vs. green), one can see that about the same amount of errors was made.
Green	97	

Presentation	correct %	Interpretation
Complete	97.8	When considering the experimental variation "degree of completeness", one can see that, during the study phase, complete items were named correctly equally often as incomplete items. Moreover, values were high; cf. Mulligan (2004) and Riefer et al. (2007).
Incomplete	96.2	

Appendix I: continued

Experiment 5			
Type of item	Presentation colour	Presentation	Correct %
Items *without* schema-typical colour	Red	Complete	94.6
		Incomplete	97.3
	Green	Complete	95.6
		Incomplete	90.5
Items *with* schema-typical colour (*green*)	Red	Complete	91.8
		Incomplete	71.4
	Green	Complete	94.6
		Incomplete	85.7
Items *with* schema-typical colour (*red*)	Red	Complete	93.2
		Incomplete	98.6
	Green	Complete	100
		Incomplete	99.3

Type of item	Correct %	Interpretation
Items *without* schema-typical colour	95.6	When considering type of item, one can see that error rates for items with schema-typical colour (green) are lower than that for the other types of items. However, error rates were below 15 % in all cases.
Items *with* schema-typical colour (*green*)	88.7	
Items *with* schema-typical colour (*red*)	98.2	

Type of item	Presentation	Correct %	Interpretation
Items presented in *typical* colour	Complete	93.9	When considering items presented in their typical colour (Typ) and items presented in their atypical colour (Atyp) and the experimental variation "degree of completeness", one can see that more errors were made for Atyp items. Moreover, for each type of item, more naming errors were made for incomplete items.
	Incomplete	92.2	
Items presented in *atypical* colour	Complete	95.9	
	Incomplete	85.4	

Typical colour	Correct %	Interpretation
Red	97.8	When considering responses given to items (typical colour is either red or green), one can see that more errors were made for items that are typically green. Considering these error rates, it seems reasonable to check for the influence of typical colour. This analysis can be found in the results section of Experiment 5.
Green	85.9	

Presentation colour	Correct %	Interpretation
Red	97.1	When considering presentation colour (red vs. green), one can see that about the same amount of errors was made.
Green	97.7	

Presentation	Correct %	Interpretation
Complete	98.1	When considering the experimental variation "degree of completeness", one can see that, during the study phase complete items were named correctly equally often as incomplete items. Moreover, values were high; cf. Mulligan (2004) and Riefer et al. (2007).
Incomplete	96.7	

Ehrenwörtliche Erklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig verfasst sowie sämtliche Belege deutlich gemacht und korrekt angegeben habe.

Mannheim, den 16. November 2010

____

(Sandra Daniela Mattern)

Danksagung / Acknowledgements

An dieser Stelle möchte ich mich gerne bei allen Menschen bedanken, die in besonderer Weise am Gelingen dieser Arbeit beteiligt waren.

Zuerst möchte ich mich gerne bei meinem Mentor Prof. Dr. Edgar Erdfelder sowie bei Prof. Dr. Rüdiger Pohl für die Übernahme der Betreuung meiner Dissertation bedanken. Beide gaben mir Hinweise und zeigten Kritikpunkte und Verbesserungsmöglichkeiten meiner Arbeit auf. Dafür vielen Dank! Vor allen Dingen möchte ich aber Edgar Erdfelder dafür Danke sagen, dass er durch seinen Enthusiasmus, seinen Blick fürs Ganze und sein begeisterndes Wesen mein Interesse an der kognitiven Psychologie in besonderem Maße vertieft und meinen Blick auf wissenschafliches Arbeiten und Forschen positiv geprägt hat. Außerdem möchte ich Dir danken, dass Du mich seit meiner Zeit als studentische Hilfskraft unterstützt und mir die Promotion ermöglicht hast. Vielen Dank, Edgar!
Für die spontane Zusage zur Teilnahme an der Prüfungskommission möchte ich mich außerdem ganz herzlich bei Thomas Gschwend bedanken.

Darüber hinaus gilt mein Dank allen Mitarbeitern und studentischen Hilfskräften des Lehrstuhls Psychologie III (Universität Mannheim) sowie allen Dozenten und Studenten der CDSS (Center for Doctoral Studies in Social and Behavioral Sciences, Universität Mannheim) für ihre wertvollen Tipps und konstruktive Kritik. Thanks also to Dr. Marek Nieznanski (University of Warsaw) who listened closely to my generation effect research and discussed parts of the results with me.

Vielen Dank des weiteren an alle studentischen Hilfskräfte, die mich im Laufe der Dissertation bei der Datenerhebung und der praktischen Umsetzung meiner Promotionsidee unterstützt haben. Besonders hervorzuheben sind hierbei Maika Nellen und Alena Senft sowie Marie-Luise Back und Anca Lazar.

Ein herzlicher Dank gilt natürlich auch allen Studienteilnehmerinnen und -teilnehmern ohne welche die präsentierte Forschung gar nicht erst möglich gewesen wäre: Vielen Dank für eure Zeit und eure Anstrengung!

Neben mündlichem Feedback haben mich auch „meine“ Probeleser mit ihren vielen Anmerkungen, Hinweisen, Korrekturen, Verbesserungsvorschlägen (etc.) und ihrer Geduld auf dem Weg zur Abgabe ein Stück weitergebracht. Daher danke ich Kristina Tanz, Bernadette Proffert, Liang Sheng, Janine Heißler, Meike Kroneisen, Lena Nadarevic, Andre Aßfalg, Annegret Krause-Utz und Hayley Michelson.

Primär für emotionale Unterstützung aber auch für den fachlichen Austausch möchte ich darüber hinaus besonders danken...
Carolina Küpper-Tetzel (mit der ich zwei Jahre lang ein Büro sowie viele Erfahrungen und Drücker teilen durfte) und Lena Nadarevic (deren Tür – nicht nur metaphorisch gesprochen – immer offen stand). Mit euch beiden habe ich eine einzigartige Zeit (die vielen Mittags- und Kaffeepausen, unsere WWM-Runden, etc.) verbracht, an die ich mich immer sehr gerne zurückerinnern werde!
Auch mit Meike Kroneisen und Andre Aßfalg, die erst später nach L13, 15 gezogen sind, konnte ich immer wieder lachen – euch auch dafür vielen Dank!
Liang Sheng and Janine Heißler: I so enjoyed our coffee / tea meetings, chats, laughters, Pfälzisch lessons, as well as our deep and meaningful conversations about life, work, and love. Dear Liang, thanks also for your positive views as well as for sharing and exchanging thoughts, worries, and troubles concerning our special p****** project. ^_^
To all six: Thanks for everything!

***

Ein ganz besonderes Dankeschön gilt – wie bereits zu Zeiten der Diplomarbeit – meiner Familie (allen Zilligs und Matterns). Ihr möchte ich danken für die praktische Unterstützung und Entlastung (z.B. für das gefragte und ungefragte Übernehmen des Wäschens und Bügelns ^_^), für das ehrlich gemeinte Interesse am Fortschritt meiner Doktorarbeit und dem facettenreichen Fach Psychologie (und Neuropsychologie) inklusive Debatten über Sinn und Unsinn im „wirklichen Leben“, für willkommene Ablenkungen in jeglicher Form, sowie für viel Verständnis für schlechte Laune, wenig Zeit und die andauernde Wochenend-, Abend- und teilweise Nachtarbeit. Für immer die richtigen Worte (zur Aufmunterung, zur Ermutigung, zum Dampf-Ablassen, einfach zum Lachen, zum Trösten, zum Anfeuern, etc.) möchte ich mich auch herzlich bedanken bei meinen drei Mädels Annegret Krause-Utz, Anke Lauth und Bernadette Proffert (Reihenfolge alphabetisch)! Danke! ^_^

Da vor allen Dingen mein Mann Matthias bis zum Schluss mit mir gelitten, gelacht und gekämpft hat - ohne Verständnis, Rücksichtnahme, Weitblick und Realitätssinn zu verlieren -, möchte ich ihm diese Arbeit widmen und sagen: Vielen Dank für Alles!