Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
In partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the publication-based thesis

Investigating Storage and Retrieval Processes of Intentional Forgetting: A Multinomial Modeling Approach

presented by
Ivan Marevic

year of submission
2018

Dean: Prof. Dr. Dirk Hagemann
Advisor: PD Dr. Jan Rummel

Table of Contents

Acknowledgements

The past years were at times a real pleasure and joy and at other times very challenging and frustrating. During this time I was always surrounded by friends, colleagues and fellow students who went through these phases together with me.

First of all, I would like to thank my first advisor PD Dr. Jan Rummel, who guided and supported me through all the phases of my work. Thank you for all the time you spent with me discussing research, analyzing data, and giving me helpful advice and feedback on my work. Thank you also for giving me the freedom to pursue other research projects and to be involved in several teaching activities.

I am grateful to Prof. Dr. Klaus Fiedler who instantly agreed to supervise this thesis. Thank you also for giving me the opportunity to present my research at your adaptive cognition colloquium. I profited greatly from the feedback and discussions with you during that colloquium.

Part of this work has been shaped through valuable discussions and exchanges with Dr. Beatrice Kuhlmann and Dr. Nina Arnold. Thank you both for sharing your thoughts on my research with me and for providing excellent statistics guidance regarding multinomial modeling techniques.

Many thanks further go to my fellow colleagues Lena S. and Sebastian who I shared office with over the past years. I thank you both for discussing science and life, drinking coffee/chocolate, and playing ping-pong with me. Many thanks also go to Bianca, Katja, Lena L., and Ruth who I shared many moments and activities with. I really enjoyed our joint adventures such as watching pro-league ice hockey or playing 3D minigolf. I am also very thankful to Sabine for helping me with numerous administrative things which eased my academic life.

Many thanks go also to Veronika, Alica, Ulf, Pit, Stefan, Andreas, and Mischa for sharing their experiences and many entertaining and pleasant lunches with me. Especially the discussions regarding predictive analytics, programming, and entrepreneurship were and still are invaluable experiences for me.

The data collection for the studies reported in this thesis would have taken considerably longer without the support from our research assistants. To all research assistants, thank you for collecting the data and providing valuable input in our lab meetings. Moreover, I thank the Landesgraduiertenförderung for supporting my research with an individual doctoral fellowship.

Lastly, I am also very grateful to my parents and my sister who have unconditionally supported and encouraged me to pursue my interests and who have helped me through many challenges. Many thanks also go to Sarah for all her love and support.

List of scientific publications of the publication-based dissertation

Article 1

Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *42*(10), 1526-1543. doi: 10.1037/xlm0000266

Article 2

Marevic, I., & Rummel, J. (submitted). Retrieval-reliant mechanisms of item-method directed forgetting: The effect of semantic cues. *Psychological Research.*

Article 3

Marevic, I., Arnold, N. R., & Rummel, J. (2017). Item-method directed forgetting and working memory capacity: A hierarchical multinomial modeling approach. *Quarterly Journal of Experimental Psychology.* Advance online publication.

1 Introduction

People encode and memorize a vast amount of information on a daily basis with the goal of retaining as much relevant information as possible. For example, when planning a birthday party it is important to memorize the names of friends one has already invited, in order to avoid sending out invitations to the same friends twice. In such a situation, the goal is to retain information rather than forget it. The information that is not relevant to the current goal of memorizing friends' names, and thus is not retained in memory through rehearsal, is assumed to decay over time (Brown, 1958; Peterson & Peterson, 1959; for a review see Ricker, Vergauwe, & Cowan, 2016). In this case forgetting is a passive process that happens over time and does not require any form of intention or action. In other situations, however, it is not only desirable to retain relevant information, but also to forget irrelevant information. For instance, when getting a new phone number, one has to memorize the new phone number, as well as forget the old one. In this case the old phone number is a very strong memory that has been rehearsed for a long period of time, making forgetting a difficult task that involves a higher degree of intentionality than simply letting memories decay over time (Golding & MacLeod, 1998). Thus, forgetting depends highly on the degree of intentionality, with intentional forgetting referring to situations where forgetting is actively attempted through some forgetting strategy (e.g., shifting attention, suppressing or not thinking about information; see Sahakyan, Delaney, Foster, & Abushanab, 2014). Another form of intentional forgetting is the attempt of forgetting information that was just encoded (E. L. Bjork & Bjork, 1996; R. A. Bjork, 1972; MacLeod, 1999). For instance, when having a conversation with another person, one is often prompted by the conversational partner to forget what has just been said or to disregard irrelevant information. In this case, the ability of people to intentionally forget the information that has been deemed irrelevant is crucial to the goal of following and understanding the important parts of the conversation. These examples illustrate that forgetting is not only a passive byproduct of remembering, but that forgetting can serve an adaptive function in people's lives, enabling them to remember relevant while forgetting irrelevant information.

Given the adaptive function of intentional forgetting, one may ask how it is achieved and what the underlying processes are. Considering the above examples of trying to forget a phone number that is stored in memory, and trying to forget information that has just been encoded, one could speculate that different cognitive processes are involved to achieve intentional forgetting in both situations (Basden & Basden, 1996; Basden, Basden, & Gargano, 1993). As the to-be-forgotten (TBF) information is stored very well in the phone number example one could presume that forgetting is best achieved through preventing retrieval of that information (R. A. Bjork, 1989), while in the conversation example it might be best to prevent further storage of the just encoded information (Basden et al., 1993; R. A. Bjork, 1972). Understanding the cognitive processes that contribute to both forms of intentional forgetting is the main goal of the present thesis. In order to achieve this goal and to investigate the cognitive processes behind intentional forgetting, multinomial modeling techniques that allow for the separation of storage and retrieval processes are employed in the present work. Further, it is also clear that people might differ in their ability to forget irrelevant information. Thus, another goal of this thesis is to identify how the engagement of storage and retrieval processes differs between people with differing cognitive control abilities.

In the following chapter I will give a brief overview of intentional forgetting research relevant to the present thesis, followed by an introduction of the multinomial storage–retrieval model (Riefer & Rouder, 1992; Rouder & Batchelder, 1998), before outlining the empirical studies conducted to tackle the aforementioned research goals.

2 Intentional Forgetting

As mentioned earlier, intentional forgetting describes forms of forgetting in which information that is irrelevant to the goal of remembering some other more relevant information is attempted TBF. In this context, intentional forgetting of TBF information has to serve some implicit or explicit goal. In most cases remembering information that is more relevant in the current situation is such a goal (R. A. Bjork, 1972; R. A. Bjork, LaBerge, & Legrand, 1968). Another goal to intentionally forget could also be to prevent unwanted memories from being retrieved, leading to lasting forgetting of these memories (Anderson & Green, 2001; E. L. Bjork, Bjork, & Anderson, 1998). Thus, when describing intentional forgetting in the following sections, I will focus on these examples and research findings where such a personal goal is present and refrain from discussing findings where “disregard” instructions are used that are independent of the goal of remembering to-be-remembered (TBR) information (e.g., disregarding confidential or trial-related information in a court room; see Golding & MacLeod, 1998; Johnson, 1994 for a review of this literature).

2.1 The study of intentional forgetting

Intentional forgetting can be studied in the laboratory using a variety of research paradigms, with each paradigm designed to tackle a different research question related to forgetting. On the one hand, research on intentional forgetting in terms of exerting control over the retrieval of unwanted memories often uses paradigms such as the think/no-think (Anderson & Green, 2001) or the retrieval-induced forgetting (RIF) paradigm (Anderson, Bjork, & Bjork, 1994). In the think/no-think paradigm, participants initially study and then practice successfully recalling the previously studied word-pairs. Later, in a think/no-think phase, for some word-pairs they are asked to prevent a subset of previously studied words to come to mind when presented with the first word of the pairs as cues. In a final memory test, the participants are then asked to recall as many words in a cued recall test as possible. The memory performance on this final test is characterized by lower recall for the words that were intentionally suppressed (no-think words) compared to words for which retrieval was not suppressed (baseline) and words for which retrieval was deliberately enhanced (think words). Thus, these results suggest that intentional control over retrieval can lead to long lasting forgetting. Further, research using the RIF paradigm shows that similar inhibitory effects are obtained when participants study words from varying superordinate categories and are then asked to retrieval practice only a subset of the words from each of the categories. On a final cued recall test, recall is usually lower for the words from the same category that were not retrieval practiced compared to practiced words from the same category and non-practiced baseline words from a different category. This finding suggests that forgetting can also be achieved through intentionally retrieving related information. Thus, both the think/no-think and the RIF paradigms are useful procedures to study intentional forgetting of unwanted memories through blocking or inhibiting retrieval of these memories (see Anderson, 2003; Anderson & Hanslmayr, 2014 for reviews)

On the other hand, research on intentional forgetting in terms of forgetting irrelevant information for the sake of better remembering relevant information uses paradigms such as the list-method and item-method directed forgetting (DF) paradigm. In the list-method DF paradigm participants are studying two lists of items (e.g., individual words or word-pairs) for a later recall test. Each list item is presented individually for a predefined amount of time (e.g., 5 seconds with a 500 ms inter-stimuli-interval). After studying the first list (L1), half of the participants are told to forget L1 and that their memory for L1 will not be assessed in a final test. The other half of participants are told that L1 was just the first half of the list and that their memory for L1 items will be assessed in a final test. Following these memory instructions, all participants study a second list (L2) that they need to remember for a final memory test. After a short distractor task (e.g., solving math problems) which is included to eliminate potential recency effects, participant's memory for L1 and L2 items is assessed. To prevent output order effects (Anderson, 2005) L1 is often tested before L2. The typical pattern of results from this paradigm is twofold: Lower L1 recall is observed in participants who received a forget cue (forget group) compared to participants who received a remember cue (remember group) after L1. However, for L2 higher recall is observed in the forget group compared to the remember group. The former effect is referred to as the costs of DF (Reitman, Malin, Bjork, & Higman, 1973), whereas the latter effect is referred to as the benefits of DF (Sahakyan & Delaney, 2003). Contrary to list-method DF, in the item-method DF paradigm participants are studying items sequentially with each item being followed by either a forget (**FFF**) or remember instruction (**RRR**). Study items and memory instructions are usually presented randomly intermixed at a fixed presentation rate (e.g., 5 seconds with a 500 ms inter-stimulus-interval). In a final memory test, memory for both TBF and TBR-items is assessed. The typical pattern of results for the item-method involves lower recall of TBF compared to TBR-items (Basden et al., 1993; Taylor, 2005). It is important to note that contrary to the list-method, in item-method DF usually no remember group is included against which the costs and benefits of DF can be assessed (Basden & Basden, 1996, 1998; Basden et al., 1993; but see Foster & Sahakyan, 2012).

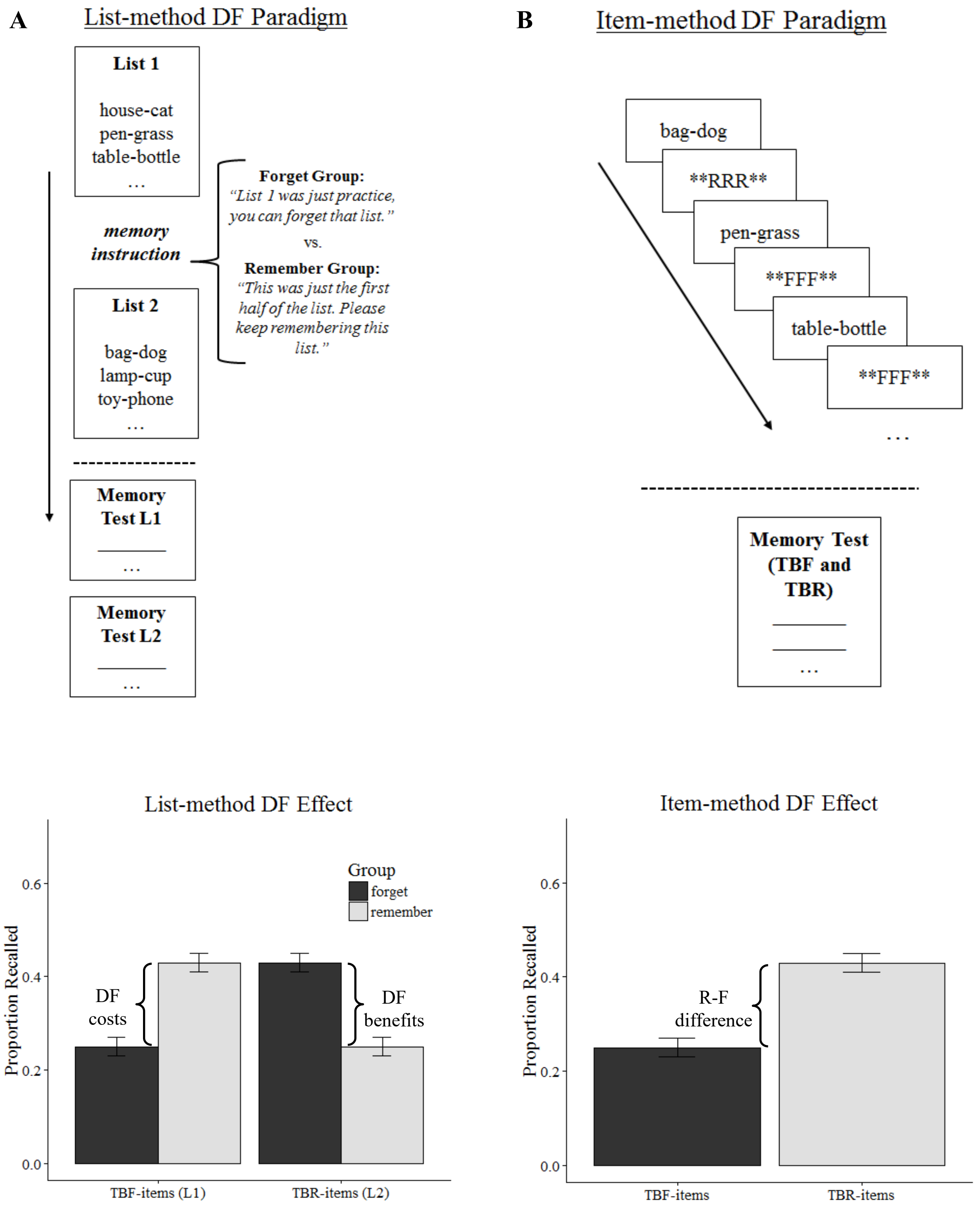


Figure 1. Experimental procedure (top) and typical result pattern (bottom) for the list-method (panel A) and item-method (panel B) DF paradigms.

As the goal of the present thesis is to investigate the latter described form of goal-directed intentional forgetting on the process-level using multinomial modeling techniques, the list-method and item-method DF paradigms were used in the present investigations (see chapter 4).

2.2 Different methods, different processes?

The list-method and item-method of DF do not only differ in the procedure itself, but past research suggests that different cognitive processes are responsible for the pattern of results described above. In a first comparison of both methods conducted by Basden et al. (1993), the item-method DF effect was present in both recall and recognition memory tests whereas the list-method DF effect was present in recall but not recognition tests. Further, only the item-method produced reliable DF effects in implicit memory tests (Basden & Basden, 1996; Basden et al., 1993). The presence of a DF effect in both implicit and explicit memory tests lead early researchers to conclude that item-method DF seems to be caused by *selective rehearsal* (R. A. Bjork, 1970; R. A. Bjork et al., 1968). That is, if the DF effect is present in tests that provide strong retrieval cues (e.g., the item itself in recognition tests), it is likely that the item was not rehearsed and thus not stored in memory. So the *selective rehearsal account* (R. A. Bjork, 1970; R. A. Bjork et al., 1968) of item-method DF posits that the DF effect is a result of rehearsing TBR-items while passively dropping TBF-items from the rehearsal set. On the other hand, the absence of a DF effect on tests with strong retrieval cues (e.g., recognition) in the list-method lead to the conclusion that TBF-items from L1 seemed to be stored but are *inhibited* (R. A. Bjork, 1989; Geiselman, Bjork, & Fishman, 1983), causing the costs and benefits of DF to emerge on tests that rely heavily on retrieval processes (e.g., free recall). That is, the costs are a direct result of inhibiting TBF-items and the benefits arise because the inhibited TBF-items no longer interfere with TBR-items at retrieval (Geiselman et al., 1983). Importantly, according to this inhibitory view, TBF-items can be released from inhibition if recall tests that provide strong retrieval cues are employed at test (Basden et al., 1993; Geiselman & Bagheri, 1985; MacLeod, 2007).

Even though the *selective rehearsal account* (R. A. Bjork, 1970; R. A. Bjork et al., 1968) and the *retrieval inhibition account* (R. A. Bjork, 1989; Geiselman et al., 1983) are able to account for both list-method and item-method DF findings, recent theoretical arguments and empirical findings pose a challenge to key assumptions of these accounts. Regarding the list-method, Lehman & Malmberg (2009, 2011) made the argument that the retrieval inhibition account provides a valid description of the DF phenomenon, namely reduced TBF-item compared to TBR-item recall, but it does not specify how TBF-items are inhibited in the first place and how they are suddenly released from inhibition when recognition tests are used. Thus, the theoretical conceptualization of this account seems to be lacking critical assumptions about the nature of the inhibitory cognitive processes that are postulated to cause the observed effect. As a result, an alternative theoretical account proposes mental *context-change* between L1 and L2 study together with the later mismatch of study and retrieval context as the underlying cognitive mechanism of list-method DF. According to this *context-change account* (Sahakyan & Kelley, 2002) the forget instruction causes a mental context-change in participants, leading to different study contexts for TBF (L1) and TBR-items (L2). At recall, the current context is then different from the original TBF-item study context, causing the costs of DF as participants lack the context cues necessary for successful retrieval. Additionally, reduced interference of TBF-items together with a reset of encoding strategies following the forget cue is assumed to account for the DF benefits (Sahakyan & Delaney, 2003)[1]. Evidence for the context-change account comes from studies where context-change manipulations between L1 (TBF-items) and L2 (TBR-items) produced the same costs and benefits of DF as traditional forget instructions. Importantly, the reinstatement of L1 study context at test, as well as strong semantic associations between TBF (L1) and TBR-items (L2) eliminated the DF costs completely, speaking for a critical role of study context and the ability to reinstate this context at test as crucial factors of the list-method DF effect (Sahakyan & Goodmon, 2007; Sahakyan & Kelley, 2002). Further, participant self-reports together with experimental manipulations of TBR-item (L2) study strategy (shallow encoding vs. deep encoding) suggest that the DF benefits seem to be caused by both reduced proactive interference of TBF-items (L1) as a result of context-change and a reset of encoding taking place prior to TBR-item encoding (L2) (Pastötter, Tempel, & Bäuml, 2017; Sahakyan & Delaney, 2003). Thus, the context-change account is a dual-process account that emphasizes both context-change and selective rehearsal as the underlying causes of both the costs and benefits of list-method DF (see Sahakyan et al., 2014 for a review).

Because the context-change account proposes two processes, namely context-change and reset of encoding strategies, as causes of the costs and benefits of list-method DF respectively, a recent revival of the selective rehearsal account as a single-process explanation has been proposed (Sheard & MacLeod, 2005). In a series of experiments, Sheard and MacLeod (2005) found that list-method DF costs and benefits are predominantly present in the primacy and recency portions of TBF compared to TBR-item recall. As serial position effects are indicative of selective rehearsal (Rundus, 1971; Rundus & Atkinson, 1970), the authors posit that the selective rehearsal account provides the most parsimonious explanation of the list-method DF effect and should not be dismissed prematurely.

Regarding item-method DF, recent evidence has also questioned selective rehearsal (R. A. Bjork, 1970; R. A. Bjork et al., 1968) as the only process involved in the DF effect. Research by Taylor and colleagues suggests that attention withdrawal from TBF-item representations forms an additional process recruited when encountering a forget cue during the study phase of an item-method DF paradigm. According to this *attention withdrawal account* (Hourihan & Taylor, 2006; Taylor, 2005; Wylie, Foxe, & Taylor, 2008) participants actively withdraw their attention from TBF-items immediately after encountering the forget cue in order to continue the selective rehearsal of TBR-items. Evidence for this account comes from a series of experiments in which a variety of secondary reaction time (RT) tasks were interleaved with the standard item-method DF paradigm. This was done

---

[1] Proponents of the inhibition account may argue that study context might be inhibited instead of changed, rendering a context-change explanation obsolete (Anderson, 2005; Pastötter & Bäuml, 2010; Pastötter, Bäuml, & Hanslmayr, 2008). As such a view reframes the inhibitory theory on a descriptive level without offering an explanation of how inhibition of context might be achieved, I will refrain from reiterating the previously described inhibitory view as a separate account of context-inhibition.

to test whether attention, that is necessary for the detection of probes in RT tasks, is affected differently following forget compared to remember cues. For instance, in one such task the study items and memory cues (TBF vs. TBR) were presented at different locations of the screen, immediately followed by secondary probes that appeared in the same or different locations. The results showed that participants were slower at reacting to secondary probes that appeared in the same location following TBF compared to TBR-items and that this even holds for TBF-items that were later correctly recalled in a final memory test (Fawcett, Lawrence, & Taylor, 2016; Fawcett & Taylor, 2010; Y.-S. Lee, Lee, & Fawcett, 2013; Taylor, 2005; Taylor & Fawcett, 2012; Thompson & Taylor, 2015; Wylie et al., 2008). According to the attention withdrawal account, secondary probes that appear in the same location are reacted to more slowly because attention is initially actively withdrawn from the TBF-item location and thus it takes longer to reallocate attention to the initial location. Accordingly, item-method DF is likely to be caused by both the active withdrawal of attention from TBF-items and the selective rehearsal of TBR-items.

Finally, others have proposed that retrieval inhibition is another mechanisms involved in producing the item-method DF effect. According to this *retrieval inhibition account*, TBF-items are inhibited for the sake of better TBR-item rehearsal during study. At test, due to the persisting inhibition that acts on TBF-items, access to these items in memory is limited, leading to the standard DF effect (Geiselman & Bagheri, 1985; Zacks, Radvansky, & Hasher, 1996). Evidence for this account comes from two investigations: In the first investigation participants completed a standard item-method DF paradigm, followed by a memory test, another study phase in which all words were TBR, and final memory test. The results showed that the DF effect was present in the first but not the second memory test, indicating that TBF-items were initially inhibited and released from inhibition after re-exposure in the second study phase (Geiselman & Bagheri, 1985). The second investigation examined item-method DF in young and old adults using a standard item-method DF paradigm. The results showed that older adults did not forget as many TBF-items as the young adults. This was taken as evidence for retrieval inhibition, as inhibitory capabilities decline with age (Hasher & Zacks, 1988), leading to less inhibition of TBF-items for older compared to young adults.

From the above review of theoretical accounts for the list-method and item-method of DF, it becomes evident that the theoretical conceptualizations and, consequently, assumptions about underlying cognitive processes differ greatly between DF methods. Regarding the involvement of storage and retrieval processes in list-method and item-method DF, the above accounts make very different predictions. For the list-method, the retrieval inhibition and context-change account would predict the DF effect to be mediated by retrieval processes only, as the retrieval of TBF-items from memory is inhibited or access to the context which could act as a retrieval cue is not given. The selective rehearsal account, on the other hand, would predict list-method DF to be mediated by storage processes, as rehearsing TBR-items should lead to better storage of TBR-items compared to TBF-items. For the item-method, the selective rehearsal and attention withdrawal accounts would predict that the DF effect is mediated by storage processes only, as both rehearsing TBR-items and withdrawing attention from TBF-items shortly after their encoding should both lead to better storage of TBR and worse storage of TBF-items. The retrieval inhibition account, on the other hand, would predict the DF effect to be more of a retrieval phenomenon, as inhibition of TBF-items should prevent the retrieval of these items. However, as the item-method DF effect is present in tests that vary in the amount that recollection is engaged (e.g., free recall and recognition) and release from inhibition is possible after restudying the items (Geiselman & Bagheri, 1985), this prediction can only be made with caution. Thus, according to the predominant theoretical conceptualizations (e.g., the selective rehearsal and attention withdrawal account) that are based on a substantial amount of empirical evidence item-method DF has so far been considered a pure storage phenomenon.

2.3 Individual differences in working memory and DF

The DF effect is a robust phenomenon that has been replicated in many studies using different populations (e.g., age groups, psychological disorders) and different study materials (e.g., words, pictures, line drawings) (see Bäuml, 2008; Golding & MacLeod, 1998; Sahakyan et al., 2014 for reviews). Therefore, the current body of research provides good support for the claim that people are able to intentionally forget when instructed to do so in a list-method or item-method DF paradigm. But does the ability to forget differ between people with differing cognitive control abilities? A few studies have investigated individual differences in DF with regard to executive functioning and cognitive control, but all of these studies used the list-method DF paradigm (Aslan, Zellner, & Bäuml, 2010; Delaney & Sahakyan, 2007; Soriano & Bajo, 2007). The results of these studies have shown that individuals with high working memory capacity (WMC) as measured on classical working memory tasks, such as the n-back, memory updating, or complex span tasks (see Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000), display higher DF costs and benefits compared to individuals with low WMC. This evidence has been interpreted as support for both the retrieval inhibition and context change account of list-method DF. As earlier research has shown that WMC is positively related to the ability to inhibit or suppress unwanted information (see Redick, Heitz, & Engle, 2007 for a review), it could be that high WMC individuals are better able to inhibit TBF-items and better able to store TBR-items. However, there is also evidence that suggests people with high WMC are better able at changing mental context compared to low WMC individuals and thus context change might affect the high WMC individuals to a greater extent. That is, high WMC individuals have less access to TBF-items in memory, because they were better able to shift context following the forget instruction, reducing proactive interference from the TBF-items at test (Kane & Engle, 2000). Consequently, the observed negative relation between WMC and TBF-item forgetting could also be viewed as support for the context change account, as high WMC individuals are not able to efficiently rely on context cues after successful context change. Even though both the inhibition and context change account equally well for the negative relation between WMC and TBF-item recall in the list-method, the findings from these investigations pose a challenge to the selective rehearsal account. That is, the selective rehearsal account would predict a positive relation between WMC and the better storage of TBR-items, as high WMC individuals are better at remembering information from long-term memory (Spillers & Unsworth, 2011; Unsworth, 2009), but it would not predict worse recall of TBF-items. Hence, regarding the list-method of DF, individual differences in cognitive control abilities predict both higher DF costs and benefits and thus are viewed as support for the inhibition and context-change account.

Regarding the item-method of DF, to date there is no study that investigated the role of individual differences in cognitive control as indexed by WMC in relation to the item-method. Accordingly, part of the present thesis aimed at investigating the relation between WMC and item-method DF and thus tries to fill this research gap (see chapter 4.3). From a theoretical standpoint such an investigation can test all three item-method DF theories against each other, as each theoretical framework makes different predictions about the aforementioned relation: The selective rehearsal account would predict that WMC should be positively related to storage of TBR-items only, as high WMC individuals should be better at rehearsing the TBR-items compared to low WMC individuals. The retrieval inhibition account, however, would predict a negative relation between WMC and the retrieval of TBF-items, as high WMC individuals should inhibit TBF-items better than low WMC individuals. Finally, the attention withdrawal account would predict a negative relation between WMC and the storage of TBF-items, in that high WMC individuals with higher cognitive control mechanisms should be better able at withdrawing attention from TBF-items compared to low WMC individuals. Thus, investigating item-method DF from an individual difference perspective does not only fill a research gap, but it also allows for a critical test of predictions derived from all three theoretical accounts on item-method DF.

In order to test these different predictions, the DF effect needs to be investigated on the process-level. That is, one needs to obtain estimates for storage and retrieval processes that contribute to the overall DF effect and then test whether WMC predicts better or worse storage and/or better or worse retrieval. In order to be able to test the relation between WMC and estimates of storage and retrieval, the estimates need to be obtained for each individual. One method that allows for separating storage and retrieval processes is the multinomial storage–retrieval model and with its hierarchical version estimates can even be obtained for each individual. Thus, process-pure estimates of the DF effect, for both groups of participants (e.g., experimental groups) and individuals (e.g., participants), can be obtained through applying the multinomial storage–retrieval model to data from an item-method DF paradigm as depicted in the top right panel of Figure 1 and the use of both a free recall and cued recall test as the final memory assessments (Riefer & Rouder, 1992; Rouder & Batchelder, 1998).

3 The multinomial storage–retrieval model

The multinomial storage–retrieval model was developed by Riefer and Rouder (1992) with the goal of measuring the relative contribution of storage and retrieval processes from data derived from a free-then-cued-recall paradigm. The free-then-cued-recall paradigm can be used for assessing memory performance in memory experiments where participants initially study a series of word pairs and during the final test phase, a free recall test followed by a cued recall test is administered. In the free recall test, participants are asked to recall as many words as possible without receiving any cues, whereas in the following cued recall test, they are presented with the first word of a pair and are asked to recall the second word. This memory testing procedure is widely used by memory researchers for testing whether a specific memory effect is present in both free and cued recall or in free recall but not cued recall (Hirshman, 1988; Hirshman, Whelley, & Palij, 1989; Rawson & Kintsch, 2002; Thomson & Tulving, 1970). Whenever a memory effect is present in both the more difficult free recall test and the easier cued recall test, this is then interpreted as evidence for the effect being a pure storage phenomenon. Conversely, whenever an effect is present in free recall but not cued recall, this is then interpreted as evidence for the effect being a pure retrieval phenomenon. However, inferences about the underlying cognitive processes that are made on this behavioral level can be misleading in many cases. For instance, on the behavioral level, the data structure of free recall and cued recall frequencies is evaluated only in terms of the combination of successful free and cued recall or failed free but successful cued recall and inferences about storage and retrieval are drawn from these two possible combinations of events (see Batchelder & Riefer, 1986; Rouder & Batchelder, 1998). But what about the other possible combinations that are present in the data? In fact, in a free-then-cued-recall-paradigm the following combinations are possible, with each combination representing an event (E_n) with n denoting the event number: (E_1) successful free recall of a pair, and successful cued recall; (E_2) successful free recall of a pair, but failed cued recall; (E_3) successful free recall of a single item from a pair, and successful cued recall (E_4) successful free recall of a single item from a pair, but failed cued recall; (E_5) failed free recall, but successful cued recall (E_6) failed free recall, and failed cued recall[2].

Thus, fitting linear models such as analyses of variance (ANOVA) to data from free recall and cued recall separately and then drawing inferences from a combination of the resulting effects does not model the complex data structure adequately (Batchelder & Riefer, 1986). In other words, the processes that lead to the above recall events cannot be inferred from analyzing free and cued recall data on the behavioral level separately. Therefore, a more complex model is needed that accounts for all of the possible recall combinations ($E_1 – E_6$). The multinomial storage–retrieval model can account

---

[2] Note that this set of combinations does not distinguish between singleton recall of the first and second word of a pair, as for the current investigation the processes that govern recall order of individual items from a pair were not central to the research question. Further, as such a distinction would increase the number of event categories from six to eight and thus could lead to possibly sparse event frequency counts for these two additional categories, it was not considered in the present investigation.

for all event combinations by specifying the probability p of an event falling in one of the aforementioned event categories through a set of equations:

$$p(E_1) = a\, r\, (1 - l) + a\, (1 - r)\, s^2\, (1 - l)$$
$$p(E_2) = a\, r\, l + a\, (1 - r)\, s^2\, l + (1 - a)\, u^2$$
$$p(E_3) = 2\, a\, (1 - r)\, s\, (1 - s)\, (1 - l)$$
$$p(E_4) = 2\, a\, (1 - r)\, s\, (1 - s)\, l + 2\, (1 - a)\, u\, (1 - u)$$
$$p(E_5) = a\, (1 - r)\, (1 - s)^2\, (1 - l)$$
$$p(E_6) = a\, (1 - r)\, (1 - s)^2\, l + (1 - a)\, (1 - u)^2$$

The parameters of these equations reflect the probability of latent cognitive processes or states that lead to each of the six event categories and thus these estimates always fall in the range [0; 1]. Accordingly, each of the model parameters measures the probability of a latent cognitive process occurring. The above equations can also be visualized in the form of a multinomial processing tree (MPT) with each branch representing a series of latent processes leading to one of the six recall events (see Figure 2).

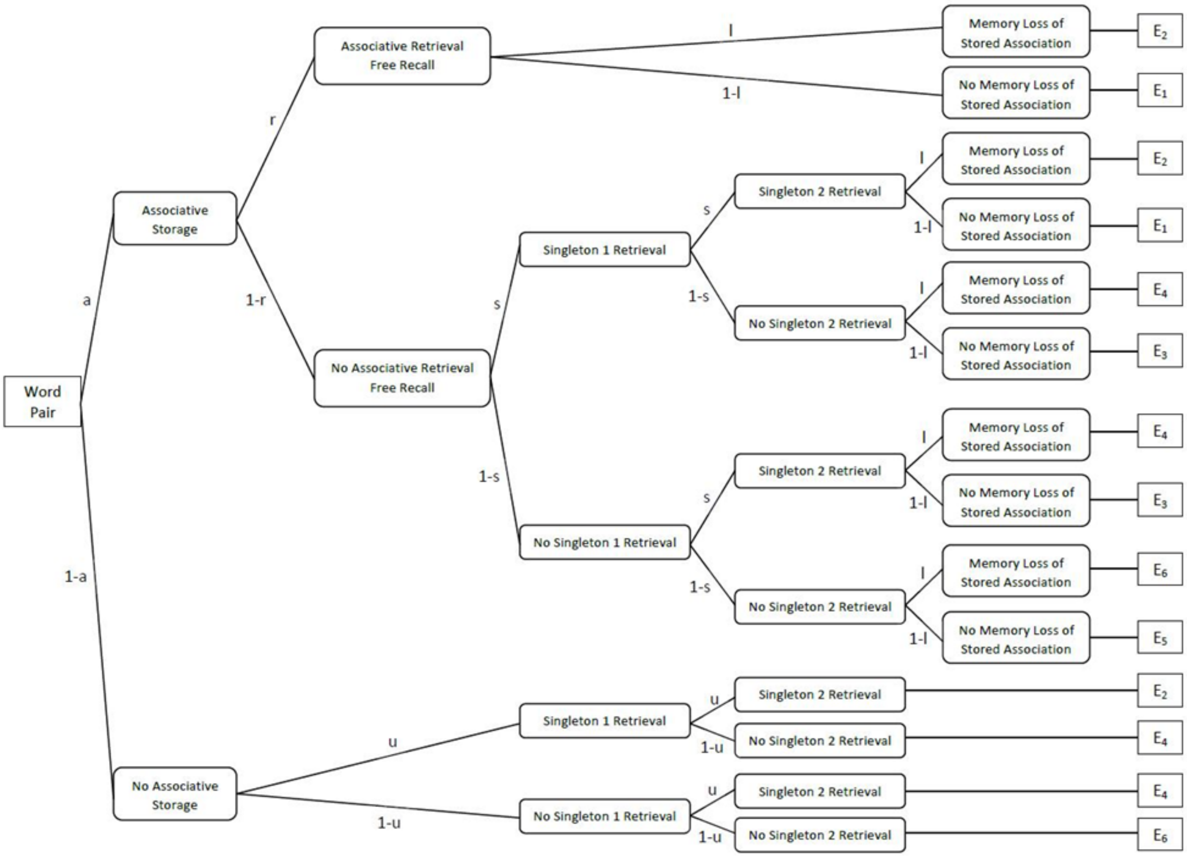


Figure 2. Multinomial processing-tree model (MPT) for the storage–retrieval model (adopted from Rouder & Batchelder, 1998; see also Rummel, Marevic, & Kuhlmann, 2016).

Note that mathematically, the order of parameters in the model is not relevant (commutative property), but as the model is developed on the basis of psychological theories, reversing the order would not be reasonable in many cases. For instance, assuming that retrieval is engaged before storage would not be possible, because one cannot retrieve an item that has not been stored in the first place. For the present investigation the parameters measuring storage (*a* parameter) and retrieval (*r* parameter) were of most theoretical interest and thus the other parameters are necessary for model identifiability but can be regarded as nuisance parameters (Riefer & LaMay, 1998; Riefer & Rouder, 1992). Descriptions of all model parameters are provided in Table 1.

Table 1

Overview of Model Parameters of the Storage–retrieval Model

Parameter	**Description**
a	The probability of associatively storing both items of the pair and maintaining that association until free recall.
r	The probability of associatively retrieving both items of the pair during free recall, contingent on successful storage of the association. Separate and sequential singleton retrieval of both item pairs are also considered associative retrieval, as the retrieval of one item of the pair may act as a cue and thus lead to the retrieval of the second item of the pair. These instances are also considered associative retrieval, because retrieval is mediated through the formed association.
s	The probability of singleton retrieval contingent on successful storage of the association. Even though the association is stored, each item of the pair might only be retrievable as a singleton, without retrieving the association.
u	The probability of retrieving each item pair as a singleton, given that the pair was not stored associatively.
l	The probability of memory loss from free to cued recall, resulting in successful free but failed cued recall. This memory loss is usually infrequent, but the model allows for it to occur contingent on successful free recall.

The just described storage–retrieval model has been used to disentangle storage and retrieval processes in a variety of research areas, including the bizarreness effect (Riefer & LaMay, 1998; Riefer & Rouder, 1992), lag effects (Küpper-Tetzel & Erdfelder, 2012), memory deficits in clinical populations (Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002), or age-related memory differences (Riefer & Batchelder, 1991). The validity of the model has also been evaluated in both simulation and empirical investigations. In these investigations, different data patterns are simulated or data is collected from different experimental paradigms from which researchers are certain that they affect only storage or only retrieval processes and it is tested how well the model parameters capture these isolated effects. The results of such validation attempts suggest that the model accounts for storage and retrieval processes very well, even when extreme variations in frequency counts are present in one of the event categories (e.g., E_5) (see Rouder & Batchelder, 1998 for an overview and a specific simulation example). Thus, it can safely be assumed that the model is capable to reliably estimate storage and retrieval processes, given data from a free-then-cued-recall paradigm, in a process-pure way.

3.1 Parameter estimation and goodness-of-fit testing

Given an identifiable multinomial model[3] such as the storage–retrieval model, the parameters of the model can be estimated by minimizing some goodness-of-fit statistic. In the case of multinomial models, the log-likelihood ratio statistic $G^2(df)$ is minimized using maximum-likelihood (ML) estimation such as the expectation-maximization algorithm suggested by Hu and Batchelder (1994) or the general purpose optimization algorithm (Kaufman & Gay, 2003) that are implemented in most MPT modeling software packages (*multiTree*, Moshagen, 2010; *MPTinR*, Singmann & Kellen, 2013; *HMMTree*, Stahl & Klauer, 2007). Once G^2 is minimized, it is compared against a $\chi^2(df)$ distribution in order to determine model fit. If the G^2 statistic falls below the $(1 - \alpha)$ percentage of the distribution, the model is retained and otherwise it is rejected at a predefined α-level (e.g., $\alpha = .05$). In the investigations described in this thesis, a model with only one l-parameter was always estimated in order to achieve identifiability. This approach is very common as l is measuring infrequent occurrence of memory loss from free to cued recall and is assumed to be equal across item types (e.g., TBF, TBR) and experimental conditions. The restricted model fitted the data best in all studies described later (see chapter 4) and thus the restriction was employed consistently.

3.2 Parameter comparison

Once the parameters of a model are estimated and the model fits the data, parameters can be compared for the different item types (TBF, TBR) and experimental conditions. Such comparisons are achieved by imposing restrictions on the parameters of interest and comparing model fits of the restricted model with the superordinate model through the G^2 difference statistic ΔG^2. For instance if one wants to compare whether the storage estimate (a parameter) for TBF-items is lower compared to

---

[3] Identifiability refers to the situation where the parameter values of a model uniquely determine the distribution of the data and vice versa (cf. Erdfelder et al., 2009). Most common is nonidentifiability when the number of unique model parameters exceeds the number of independent observations. For the storage–retrieval model, identifiability is given when there are less unique model parameters than unique event categories.

the storage estimate for TBR-items, we would restrict *a* estimates to be equal for TBF and TBR-items and test whether the model fit for the restricted sub model differs significantly from the superordinate model. That is, if ΔG^2 falls below $\alpha = .05$, this is indicative of the restricted model differing significantly from the superordinate model. Consequently, in the example of setting *a* estimates equal for TBF and TBR-items, this would mean that there is a significant difference in storage of TBF and TBR-items. Further, the same procedure applies to testing whether a parameter of interest differs significantly from a predefined value. For instance, if one is interested if the probability of memory loss differs from a value of 0.2, a restricted model with $l = .2$ could be compared to the superordinate model. Setting parameters to predefined values is often used in multinomial models in which a guessing parameter is measuring guessing processes and thus testing whether estimates of this parameter differ from 0.5 could be a reasonable approach to identify any guessing bias (e.g., Bayen, Murnane, & Erdfelder, 1996; Kuhlmann, Vaterrodt, & Bayen, 2012; Meiser, Sattler, & Weisser, 2008). In the current investigation, such procedures were not applied, as the main goal of applying the storage–retrieval model was in comparing estimates for item types and conditions.

3.3 Hierarchical storage–retrieval model

In order to investigate relations between parameter estimates of the storage–retrieval model and external measures such as WMC (e.g., through correlation analyses), parameter estimates need to be obtained for each individual. To measure model fit using G^2 when the model is applied to data from each individual separately, the frequencies in each of the event categories ($E_1 – E_6$) need to exceed the count 5 (Hays, 1994). However, in the case of the storage–retrieval model this is not given in many situations as there is only a limited amount of item pairs that participants can be expected to study and also later recall in a free-then-cued recall test. Further, aggregating data over participants for item type and experimental conditions assumes that estimates do not vary greatly (homogeneity assumption). Even though, in many cases violations of the homogeneity assumption are not problematic, as researchers are interested in the aggregated mean values for the conditions of interest, in some cases however such a violation can result in biased parameter estimates and consequently lead researchers to wrong conclusions (Klauer, 2006; Smith & Batchelder, 2008). In order to obtain reliable parameter estimates on the individual level even with sparse event category counts, hierarchical Bayesian estimation procedures can be used (M. D. Lee, 2011). Currently, there are two hierarchical approaches, namely the beta-MPT (Smith & Batchelder, 2010) and the latent-trait approach (Klauer, 2010). Both approaches rely on Bayesian hierarchical modeling techniques, but only the latent-trait approach was used in the present thesis.

A central part of the latent-trait approach is the assumption that participants' parameters are drawn from a multivariate normal distribution of probit transformed parameters. That is, parameters are transformed from values in the range [0, 1] to real numbers. Importantly, as participants' parameters are drawn from a multivariate normal distribution and are not estimated individually for each participant, the latent-trait approach combines both group-level and individual information, leading to more robust parameter estimates for each individual (Rouder & Lu, 2005). When estimating parameters of the model on the individual level, Bayesian modeling is used: In Bayesian statistics prior beliefs are incorporated by treating parameter values as random variables and specifying prior parameter distributions before starting the analysis. The specification depends on the knowledge one has about the parameters a priori and prior distributions can thus be either vague or very concrete. Given the prior distribution and the observed data, a posterior distribution is then calculated using Bayes' theorem. Bayesian credibility intervals (BCIs) reflect the range of values in the posterior distribution in which the true estimates for each parameter are to be found. Thus, BCIs can be interpreted similar to frequentist confidence intervals (CIs). As for complex models, the posterior distribution cannot be computed from the Bayes' theorem in closed form, estimation of summary statistics (e.g., posterior mean) is achieved through the use of Markov Chain Monte Carlo (MCMC) sampling methods. In MCMC sampling, a large amount of draws is obtained from the posterior distribution for each parameter of the model. For MPT models, such sampling procedures are implemented in the R package *TreeBUGS* (Heck, Arnold, & Arnold, 2017). When fitting a model using MCMC sampling it is important to check whether the MCMC chains have reached a stationary distribution. This can be done by either inspecting time series plots across all iterations or through calculating and assessing the $\hat{R}$ statistic which compares the variance within, to the variance between the chains (Gelman & Rubin, 1992)[4]. An $\hat{R}$ close to 1 is then indicative of good convergence.

Notably, *TreeBUGS* does not only allow for estimating parameters of a MPT model, but external predictors such as WMC can also be included in the model (Klauer, 2010). The resulting regression coefficient that is estimated for the external predictor summarizes the relation between that predictor (e.g., WMC) and the parameter of interest (e.g., *a* parameter). In our case, considering storage estimates (*a* parameter) for TBR-items as an example, a positive regression coefficient for WMC as the external predictor would imply a higher probability of storage occurring as WMC increases (cf. Heck et al., 2017). Hence, applying the storage–retrieval model hierarchically to data from an item-method DF paradigm in combination with including WMC as a predictor in the model, allowed us to tackle the previously mentioned research goal of investigating the relation between cognitive control abilities and cognitive processes in item-method DF (see chapter 2.3).

4 The present process-level investigation

The aim of this thesis was to investigate storage and retrieval processes of intentional forgetting using the previously described multinomial storage–retrieval model. From the broad research goals that were outlined in the introduction section of chapter 1, together with the theoretical

---

[4] Because multiple chains with different starting points are run in MCMC sampling, early draws from the posterior often have poor convergence and thus these draws are not used in assessing convergence.

background described in chapter 2, specific research questions can be derived that I tried to answer in the three articles summarized in this chapter: First, I aimed at identifying cognitive processes involved in both the list-method and item-method of DF and tried to answer the question whether these processes differ between methods (chapter 4.1). The second research question aimed at exploring candidate mechanisms that could account for a retrieval deficit in item-method DF (chapter 4.2). Finally, I aimed at answering the question of how individual differences in WMC relate to the cognitive processes involved in item-method DF and what specific relations with storage and/or retrieval imply for current DF theories (chapter 4.3).

4.1 Process-level comparison of DF methods (Article 1)

In this article, we addressed several of the previously mentioned questions regarding the two directed forgetting paradigms. The first goal if this investigation was to compare the list-method and item-method of DF and to identify to what extent storage and retrieval processes are involved in both methods. Earlier comparisons of both methods found that list-method DF seems to be retrieval-based whereas item-method DF seems to be storage-based (Basden & Basden, 1996; Basden et al., 1993). These conclusions rely on the finding that item-method but not list-method DF is present in memory tests that rely heavily on retrieval, such as recognition tests or implicit memory tests. However, in recent years, new theoretical and methodological developments emerged in the field (i.e. differentiation between DF costs and benefits). Thus, we aimed at comparing both methods again by considering the theoretical and methodological advances that were made in rent years. Further, we applied the storage–retrieval model as a more sensitive measure of storage and retrieval in order to be able to measure their relative contribution to the costs and benefits of DF. A second goal of the present investigation was to ensure that any effects we found in storage and retrieval would not be an artifact of participants' output order strategies in the final memory test. That is, when left to recall items in any order, participants usually start recalling items that have high memory strength (TBR-items). These items can then proactively interfere with the later recalled weak items (TBF-items), reducing the recall performance of TBF-items even more (Anderson, 2005). As in list-method DF, TBF-items recall order is controlled in most studies, with TBF-items being recalled prior to TBR-items, output order is unlikely to play a major role in the list-method. However, in the item-method, controlling for output order is not very common and thus in order to rule out any output order effects we aimed at replicating effects we find on storage and retrieval with open output order, also in a setting where we controlled output order.

To address these two questions, we ran two experiments. In Experiment 1, participants were presented with word-pairs for study. The way these items were instructed to be TBF or TBR and the DF method that was used varied between four experimental conditions: In the list-method-forget condition, participants studied two lists and were instructed to forget L1 (TBF-items) and remember L2 (TBR-items); in the list-method-remember condition, both L1 and L2 were post-cued as TBR; in the item-method-forget condition half of the word-pairs were randomly followed by a TBF and the other half by TBR instructions; and in the item-method-remember condition, all word-pairs were post-cued as TBR. Thus, the list-method-forget and item-method-forget conditions were designed to induce DF whereas the list-method-remember and item-method-remember conditions were designed as remember-all baselines against which the DF costs and benefits could be assessed for both methods. In the list-method conditions, each study list was followed by a task in which participants had to solve math problems. This was done to eliminate any recency memory effects. To keep the design as parallel as possible, we introduced a break in which the same math problems were solved in the item-method conditions as well. In all conditions, participants completed a free-then-cued recall test for both TBF and TBR-items. For free recall in the list-method, L1 recall always preceded L2 recall, whereas in the item-method output order was left open. For cued recall in the list-method, the first words of the word-pairs served as cues within the L1 and L2 cued recall blocks and cued recall always followed the free recall test of the respective list (L1, L2). The cues were thus randomly presented within each list and participants had to recall the second word on each cued recall trial. For cued recall in the item-method, cued recall followed the free recall of all word-pairs and the presentation order of cues was also random. In order to increase participants trust in the forget instructions, all participants completed a practice block at the beginning of the experiment in which they were asked to recall TBR-items only in both free and cued recall.

In Experiment 2, we re-ran the item-method conditions from Experiment 1 using the same procedure, but this time we controlled output order in the final free recall test. That is, we designed the following four experimental conditions: In the TBF-first-forget condition, participants recalled TBF-items first, followed by TBR-items; in the TBR-first-forget condition, they recalled TBR-items first, followed by TBF-items; in the unrestricted-forget condition, recall order was left open (replication of Experiment 1); and in the remember-all condition, all items were post-cued as TBR and thus, as in Experiment 1, this condition served as a baseline to assess the costs and benefits of DF. Accordingly, Experiment 2 was designed to test whether the findings from Experiment 1 generalize to different output order forget conditions.

The results from Experiment 1 showed that in free recall, DF costs and benefits were observed in both methods. That is, for the DF costs, items that were instructed as TBF in the forget conditions were recalled worse than the same items in the remember-all control groups of both methods. For the DF benefits, items that were instructed as TBF in the forget conditions were recalled better than the same items in the remember-all control groups. In cued recall, on the other hand, DF costs and benefits were obtained with the item-method, but not the list-method. These results were in line with previous findings that found no list-method but reliable item-method DF effects on tests that facilitate retrieval (Basden & Basden, 1996, 1998; Basden et al., 1993; E. L. Bjork & Bjork, 1996, 2003; Gottlob & Golding, 2007; Racsmany & Conway, 2006; Zellner & Bäuml, 2006).

For the model-based analyses, we applied the storage–retrieval model to the aggregated data for each item type (TBF, TBR) and experimental condition (list-method-forget, list-method-remember, item-method-forget, item-method-remember). The model with one *l* parameter fitted the data well and the results revealed that DF costs were present in storage estimates (*a* parameter) of the item-method but not the list-method. These results indicated that only item-method DF costs were driven by storage processes. The benefits of DF, however, were present in storage estimates of both the item-method and list-method of DF. This finding indicated that storage processes drive the benefits for both methods. Regarding retrieval estimates (*r* parameter), reliable DF costs were observed in both the list-method and item-method. This finding indicated that retrieval processes do not only drive list-method but also item-method DF. We found no benefits for either the list-method or the item-method in retrieval estimates, indicating that retrieval processes were not involved in producing the DF benefits. In sum, the behavioral and model-based results were in line with current theorizing regarding the list-method of DF. As predicted by the retrieval inhibition (R. A. Bjork, 1989; Geiselman et al., 1983) and context change account (Sahakyan & Kelley, 2002), DF costs were driven by retrieval processes only, suggesting that TBF-items retrieval is hampered through either inhibition or context change. In line with dual-process views of list-method DF that attribute the benefits to a reset of encoding strategy (Pastötter et al., 2017; Sahakyan & Delaney, 2003), the benefits were driven by storage processes only. Regarding the item-method, however, the present results were not in line with current theorizing. According to the predominant theoretical views such as the selective rehearsal (R. A. Bjork, 1970; R. A. Bjork et al., 1968) and attention withdrawal accounts (Hourihan & Taylor, 2006; Taylor, 2005), item-method DF costs and benefits should be a pure storage phenomenon as they are present in less retrieval-reliant memory tests. Even though our results replicated the prominent behavioral finding of DF costs being present in both free recall and the less retrieval-reliant cued recall, the model-based results indicated that both storage and retrieval processes contribute to the DF costs. For the benefits, the model-based results were in line with prominent theorizing as they were driven by storage processes only. Our model-based findings regarding DF costs thus posed a challenge to the predominant theories on item-method DF. In the article, we discussed that the presence of costs on retrieval estimates could have been due to the fact that we did not control for output order in the final memory test. Thus, early recalled TBR-items could have interfered with the later recall of TBF-items, causing the effect in retrieval.

As mentioned earlier, Experiment 2 was designed to rule out such an output order explanation. The results of Experiment 2 revealed on the behavioral-level, that the three forget conditions for which output order was varied (TBF-first-forget condition, TBR-first-forget condition, unrestricted-forget condition) all produced reliable DF costs when compared to the remember-all condition. Importantly, the three forget conditions did not differ in TBF or TBR free and cued recall rates. Consequently, the behavioral analyses suggest that output order did not seem to affect the DF effect in a different way than in Experiment 1. The model-based results further confirmed that output order is not a viable explanation for the involvement of retrieval processes in DF costs: Retrieval estimates did not differ between the forget conditions for which output order was varied, but all the forget conditions showed lower retrieval estimates compared to the remember-all condition.

In the article we discussed and related our method comparison to the comparison that was conducted by Basden and colleagues in the early 1990s (Basden & Basden, 1996; Basden et al., 1993). We argued that even though most of our behavioral results converged with Basden and colleagues findings, the inclusion of remember-all conditions for both methods allowed us to differentiate between costs and benefits of the DF effect. Further, in employing a remember-all condition for the item-method–which is rarely done in item-method DF studies, we were able to dissociate costs and benefits (see also Foster & Sahakyan, 2012) and future research on item-method DF could benefit from such a dissociation. We further discussed the importance of employing the storage–retrieval model to DF paradigms as it requires no major modifications of the paradigms as such, except for the use of word-pairs as study items and a free-then-cued-recall test as the final memory tests. The finding that item-method DF is driven by both storage and retrieval processes provided a good example of how the use of such process pure measures can reveal new insights into an effect that has long been assumed to be caused by storage processes only (see Nowicka, Jednoróg, Wypych, & Marchewka, 2009; Van Hooff, Whitaker, & Ford, 2009, for further neuropsychological evidence suporting this view). Regarding the use of a free recall test that is followed by a cued recall test, we discussed the possibility that preceding free recall could have influenced subsequent cued recall performance. Even though some studies provide evidence for such a claim (Cull, 2000), research by Riefer and Rouder (1992), in which the storage–retrieval model was also used, showed that a preceding free recall test does not influence subsequent cued recall (see also pretest of the article summarized in chapter 4.2). Therefore, even though such carry-over effects could be possible, we found them unlikely to be responsible for the observed effects in the just described investigation. Finally, the use of the storage–retrieval model does not allow for testing predictions of theoretical frameworks that are assumed to affect the same processing stage (e.g., retrieval), such as the retrieval inhibition and context-change account. However, even when interested in phenomena that affect the same processing stage, the application of the storage–retrieval model could provide process-pure estimates of the processes of interest while controlling for other processes. We thus suggested combining our model-based approach with careful experimental manipulations targeting phenomena that involve the same processing stage.

4.2 Retrieval-reliant mechanisms in item-method DF (Article 2)

In this article we addressed the question of what the underlying mechanisms could be that are responsible for the retrieval deficit in item-method DF and how such a deficit could be reduced. In our earlier research (Rummel et al., 2016), we found that item-method DF is driven by both storage and retrieval processes. The fact that retrieval-mediated processes also drove the DF effect were rather surprising as the predominant theories on item-method DF (i.e. selective rehearsal, attention withdrawal) did not predict such an involvement. Thus, the goal of the present article was to replicate our earlier finding and to identify candidate mechanisms that could account for these retrieval-reliant mechanisms. First evidence from other studies suggests that the loss of item-specific features for TBF-items could hamper the retrieval of these items at recall. For instance, Fawcett, Lawrence, and Taylor (2016) showed that TBF-item fidelity is impoverished compared to TBR-item fidelity in an item-method DF. That is, feature details such as colors are not present with lower details for TBF compared to TBR-items and the authors argue that such item feature or fidelity loss is a result of attention withdrawal from TBF-item memory representations. Further evidence for this claim comes from studies in which item-specific features were reinstated during a final recognition test of an item-method DF paradigm. For example, in one study, participants studied words in different screen locations and in the final recognition test, some of the words appeared in the same and others in a different location. The results showed that the DF effect was eliminated when the location was reinstated and this elimination was driven by increased recognition rates for TBF-items only (Hourihan, Goldberg, & Taylor, 2007). In another study, in which study background colors were reinstated in a final recognition test, however, both TBF and TBR-items profited from feature reinstatement. As both studies provide mixed evidence on the issue, our goal was to resolve this conflict by applying more process-pure measures that disentangle retrieval and storage processes. In order to apply the storage–retrieval model, we needed data from a free-then-cued-recall test paradigm. We decided to reinstate item features in the form of superordinate categories for some of the word-pairs, because such a reinstatement can also be applied prior to the free recall test. The goal of the reinstatement manipulation was to restore TBF-item fidelity which should lead to a reduction of the DF effect. In order to test whether category reinstatement affects TBF-items alone or both TBF and TBR-items, we reinstated superordinate categories for TBF-items in one experiment and for both TBF and TBR-items in another experiment. To that end, we conducted a pretest and two experiments in which we used the just described feature reinstatement manipulation.

In the pretest, we wanted to rule out the earlier mentioned hypothetical influence of free recall on cued recall in our DF paradigm. Even though such spillover effects from free to cued recall have been shown to be negligible (Riefer & Rouder, 1992), we wanted to generalize this to the item-method DF paradigm. Thus, in the pretest, we had participants study word pairs in a standard item-method DF paradigm. For the final memory test, half of the participants completed a free and cued recall test, whereas the other half completed only a cued recall test. To equate the time for both groups regarding the critical manipulation (presence vs. absence of free recall), the participant who did not receive the free recall test, completed a math task for the duration the first half engaged in free recall. The results of the pretest showed that the preceding free recall test did not influence subsequent cued recall performance in any way as both groups (preceding free recall, no preceding free recall) did not differ in terms of their cued recall performance. Thus, we were confident that carry-over effects from free to cued recall did not bias our model estimates in our earlier investigation (Rummel et al., 2016) and also should not play a major role in the present study.

In Experiment 1, participants studied word pairs in a standard item-method DF paradigm. In order to test whether feature reinstatement reduces the DF effect, the first word of some of the word pairs was a member of a superordinate category (e.g., clothes, furniture). Thus, some of the word pairs were related in that the first word of the pair came from the same superordinate category. We chose only the first word to be part of that category, because this way the superordinate category was reinstated prior to free recall but the category member was presented as a cue in cued recall. This procedure ensured that no extra cues guided retrieval during cued recall. For the study phase, one third of the pairs were word pairs post-cued as TBF and were members of a superordinate category. Another third of the word pairs were also post-cued as TBF and were not members of a superordinate category. The final third of the word pairs were post-cued as TBR and were not members of a superordinate category. Thus, these TBR-items served as a baseline against which the other related and unrelated TBF-items were compared to. Prior to the final memory test, the superordinate category was reinstated for half of the participants (i.e. they were told that some of the word pairs belonged to a superordinate category such as clothes). For the other half of participants the superordinate category was not reinstated. Following this reinstatement manipulation, all participants were asked to recall as many word pairs as possible in a free-then-cued-recall memory test.

In Experiment 2, we extended the design of the first experiment to allow for reinstatement of category features for the TBR-items as well. Consequently, we had four experimental conditions: A first condition in which related study items were post-cued as TBF and for which the superordinate category was reinstated prior to free recall, a second condition in which related study items were post-cued as TBF but for which the superordinate category was not reinstated prior to free recall, a third group in which related study items were post-cued as TBR and for which the superordinate category was reinstated prior to free recall, and a fourth group in which related study items were also TBR-items but for which the superordinate category was not reinstated prior to free recall. So, for each group there were unrelated TBF and TBR-items and related items from a superordinate category. The memory instruction (TBF, TBR) of these related items varied across conditions—that is in two conditions the related items were TBF-items and in the other two conditions they were TBR-items. This design allowed us to test whether related TBF-items for which category features are reinstated benefit from the reinstatement only, or whether this effect generalizes to TBR-items as well. For the final memory test, all participants' memory was assessed in a free-then-cued-recall test paradigm.

The results from Experiment 1 revealed that the DF effect was present for both related and unrelated TBF-items relative to the TBR-items in free and cued recall. However, in free recall the DF effect was reduced for the related TBF-items in the condition feature reinstatement took place compared to the condition were no feature reinstatement was present. In cued recall, this pattern was not observed. These results support the view that TBF-item recall benefits from feature reinstatement (Hourihan et al., 2007).

The model-based results replicated our earlier finding that the DF effect was present in both storage estimates (*a* parameter) and retrieval estimates (*r* parameter) (Rummel et al., 2016). Further, as expected, the reduction of the DF effect for related TBF-items in the feature reinstatement condition as apparent in retrieval estimates only. This supports the view that access to some item features in memory can be restored through feature reinstatement, leading to retrieval enhancement for TBF-items.

In Experiment 2, we wanted to test whether the effect of feature reinstatement benefited TBF-items only, as postulated by Hourihan et al. (2007), or whether both TBF and TBR-items benefited from feature reinstatement, as postulated by Burgess et al. (2017). On the behavioral level, the results from Experiment 2 revealed that in free recall feature reinstatement reduced the DF effect for related TBF-items but related TBR-items did not profit from such feature reinstatement. In cued recall, no feature reinstatement effects were observed for TBF and TBR-items. Thus, on the behavioral level, these results support the view that feature reinstatement is effective for TBF-items only (Hourihan et al., 2007).

The model-based analyses, however, showed that retrieval estimates (*r* parameter) of both related TBF and TBR-items increased as a result of feature reinstatement. That is, the retrieval portion of the DF effect was reduced when features of related TBF-items were reinstated and TBR-item recall increased when features of related TBR-items were reinstated. Therefore, in contrast to the behavioral results, the model-based results support the view by Burgess et al. (2017) in that both TBF and TBR-items profited from feature reinstatement.

In the article we discussed that feature reinstatement aided the recall of TBF-items for which the reinstatement took place and that this process was retrieval-mediated. This finding speaks against a pure selective rehearsal view of directed forgetting, because TBF-items that are not selectively rehearsed and thus are not stored should not be retrievable when item features are reinstated. Importantly, in contrast to list-method DF findings, the reinstatement of TBF-item features did not eliminate the item-method DF effect completely. A possible explanation could be that different cuing mechanisms are at work in both methods. In the item-method, item memory status (TBF, TBR) changes item-by-item, whereas in the list-method the TBF-items form a coherent study episode. Hence, in the item-method, it could be that items are only cued by the reinstated feature and not by other items of the episode, leading to a reduction instead of complete elimination of the DF effect. As the goal of this article was to identify possible mechanisms that could be causing the retrieval deficit for TBF-items in item-method DF, we further discussed that attention withdrawal mechanisms could be responsible for such a retrieval deficit. As Fawcett et al. (2016) found that the fidelity of TBF-items is greatly impoverished for both correctly and incorrectly recalled TBF-items, we argued that this loss of item fidelity together with the inability to restore the relevant item features at test could be responsible for the retrieval deficit in item-method DF. The findings from our investigation supported this idea, as feature reinstatement for a subset of TBF-items lead to a reduction of the DF effect. Regarding the question of whether feature reinstatement also enhances the retrieval of TBR-items, our investigation supported the view that this certainly was the case, but only after these retrieval processes were measured in isolation through the application of the storage–retrieval model. Thus, we further demonstrated the benefits of using multinomial techniques with which more process-pure measures of the processes of interest can be obtained. Additionally, employing multinomial modeling techniques is superior to traditional general linear models, because an effect that is influenced by both storage and retrieval can also produce a behavioral data pattern form which one would conclude that the effect is only retrieval mediated (e.g., effect is present in free but not cued recall). Finally, two limitations of our investigation should be acknowledged: The first limitation was that the category reinstatement could have influenced participant's output order strategies in that they may have started recalling the related items first, causing output interference on later recalled items. A post-hoc output order analysis however revealed that this was not the case. Participants were similarly likely at recalling related items in early as well as later recall output positions. The second limitation was that the separate manipulation of relatedness across conditions resulted in unequal distributions of items for each item type (TBF, TBR). For instance, in each condition only one third of items were related and this could have caused facilitated recall for these items. The fact that we did not observe increased recall for related items even in the conditions where category features were not reinstated, however, spoke against this alternative explanation. As a final point, reinstating only one type of feature may not be enough to eliminate the DF effect completely and reinstating more item features may have led to an even greater reduction of the DF effect. Nonetheless, causing a reduction in the DF with a weak reinstatement manipulation of only one item feature seemed even more impressive in this context and we recommended that future research should focus on testing reinstatement manipulations that involve multiple features.

4.3 Item-method DF and working memory capacity (Article 3)

In this article, we addressed the question of how item-method DF is related to individual differences in WMC. For the list-method this relation has been investigated extensively (Aslan et al., 2010; Delaney & Sahakyan, 2007; Soriano & Bajo, 2007), whereas for the item-method it has not yet been investigated, and with the present investigation we aimed to fill this research gap. As mentioned in chapter 2.3, investigating this relation was of particular interest, because item-method DF theories make very different predictions about the relation of WMC and storage and retrieval processes in item-method DF. WMC is a measure that is highly correlated with different features of cognitive control. For instance, individuals with high WMC tend to have a higher short-term memory capacity, better attention control, better updating and inhibition abilities, as well as better shifting abilities (Miyake & Friedman, 2012; Unsworth, 2016). Thus, WMC seems to be a good proxy for cognitive control abilities with individuals high in WMC being better at remembering relevant information while also being better able to get rid of irrelevant information (Unsworth, 2017). For the item-method of DF this could imply different relations depending on the theoretical conceptualization: According to the selective rehearsal account (R. A. Bjork, 1970; R. A. Bjork et al., 1968), high WMC individuals should be better able at rehearsing and storing TBR-items compared to low WMC individuals. As WMC is highly related to attention control, the attention withdrawal account (Hourihan & Taylor, 2006; Taylor, 2005) would predict that high WMC individuals should be better able at withdrawing their attention from TBF-items during study compared to low WMC individuals. Finally, the retrieval inhibition account (Geiselman & Bagheri, 1985; Zacks et al., 1996) would predict that high WMC individuals should be better able at inhibiting the retrieval of TBF-items compared to low WMC individuals, because of the aforementioned positive relation between WMC and inhibitory control. It becomes clear that in order to test these different predictions, estimates for storage and retrieval of TBF and TBR-items need to be obtained and related to WMC. To achieve this, we applied the storage–retrieval model hierarchically to data from an item-method DF paradigm and related the resulting estimates of storage and retrieval to WMC. Specifically, we made the following predictions regarding the relation of WMC and the model parameters: (1) s*elective rehearsal*: WMC should predict better storage of TBR-items (*a* parameter estimate for TBR-items); (2) *attention withdrawal*: WMC should predict worse storage of TBF-items (*a* parameter estimate for TBF-items); (3) *retrieval inhibition*: WMC should predict worse retrieval of TBF-items (*r* parameter for TBF-items).

We ran a study with a decent number of participants to test these predictions. Participants completed two WMC span tasks, one at the beginning and one at the end of the experimental session. The first WMC task was the operation span task (OSpan; Unsworth, Heitz, Schrock, & Engle, 2005). In this task, participants studied a series of letter of varying length (three to seven letters) over 15 study blocks. After each letter presentation, they had to solve a math problem. Once all letters of a block were presented, participants had to select the previously studied letters from a set of letters. The second task was the running span task (RunSpan; Harrison et al., 2013). Here, participants had to a series of letters over five blocks. At the beginning of each series they were told to remember a specific amount of the last letters that they were about to learn. So participants knew how many of the last presented letters they had to recall, but they did not know the amount of letters that they were about to study. The amount of last letters varied between blocks. After each letter series they had to select the last letters from a set of letters on the screen[5]. The German versions of these tasks have been recently validated by our research group (Rummel, Steindorf, Marevic, & Danner, in press). In between the two WMC assessments, participants completed an item-method DF paradigm in which they studied

---

[5] For the analysis we z-standardized and combined both WMC measures to obtain a more reliable WMC estimate.

word pairs, with each pair being followed by either a forget or remember cue. For the final memory test, they were asked to recall as many pairs as possible in a free-then-cued recall paradigm.

On the behavioral level, the results of the experiment revealed a standard DF effect characterized by lower TBF than TBR-item recall in both free and cued recall. Importantly, WMC interacted with this DF effect. Specifically, WMC correlated positively with TBR-item free and cued recall but did not correlate with TBF-item free and cued recall. These results were in line with the selective rehearsal account of DF, as individuals with higher WMC were better able to rehearse TBR-items, resulting in higher recall rates.

In order to obtain estimates for storage and retrieval for each individual and relate these estimates to WMC, we applied the storage–retrieval model hierarchically to the data and included WMC as a predictor in the model (see chapter 3.3). The model-based results revealed that across all participants, the DF effect was present in both storage (*a* parameter) and retrieval estimates (*r* parameter), replicating our earlier findings (Marevic & Rummel, submitted; Rummel et al., 2016). Further, WMC predicted storage estimates for TBR-items but not storage estimates for TBF-items nor retrieval estimates for TBF and TBR-items. This indicated that high-WMC individuals were better at selectively rehearsing and storing TBR-items but were not better at withdrawing their attention from TBF-items or inhibiting the retrieval of TBF-items than low-WMC individuals. Thus, our findings supported the selective rehearsal account of item-method DF.

In the article, we discussed the importance and usefulness of using the hierarchical latent-trait version of the storage–retrieval model in the context of relating individual parameter estimates to external measures such as WMC. Furthermore, using such an approach can also advance theory greatly. In the present investigation, the hierarchical approach allowed us to contrast the different item-method DF theories against each other and test them in one study. In this regard, our results support the selective rehearsal account, as WMC was only relate to better storage of TBR-items. However, there are also alternative explanations. For example, recent evidence has shown, that the speed of removal of outdated information from working memory is not related to WMC as measured by span tasks (Ecker, Lewandowsky, & Oberauer, 2014; Ecker, Oberauer, & Lewandowsky, 2014). Thus, it could have been that the processes we measured with the OSpan and RunSpan tasks were not engaged in attention withdrawal from TBF-item representations during storage. In fact, the compelling amount of evidence in favor of attention withdrawal made such an explanation very plausible (Fawcett et al., 2016; Fawcett & Taylor, 2010; Hourihan & Taylor, 2006; Thompson, Hamm, & Taylor, 2014; Wylie et al., 2008). Finally, we also discussed the differences between list-method and item-method DF findings in relation to WMC. For the list-method, it seems to be the case that cognitive control as indexed by WMC is mainly involved at the retrieval stage. That is, individuals with high WMC seem to be better able to either inhibit or prevent context access to TBF-items (Aslan et al., 2010; Delaney & Sahakyan, 2007; Soriano & Bajo, 2007). In the item-method, on the other hand, cognitive control processes seem to be mainly engaged in the storage of TBR-items. Nevertheless, because of the possibility that attention withdrawal might tap into different processes than measured by our WMC span tasks, we suggested that future research should focus on delineating these mechanisms in item-method DF.

5 Discussion

The work presented in the present thesis aimed at unraveling the cognitive processes involved in intentional forgetting when studied with two variants of the directed forgetting paradigm, namely the list-method and item-method of DF. From current theorizing on DF, one would assume that the list-method DF effect (costs of DF) is a retrieval driven phenomenon, whereas the item-method DF effect is a storage phenomenon. In the present process-level investigation, we found that retrieval processes are responsible for the list-method DF costs, but both storage and retrieval processes drive the item-method DF effect (article 1). We further showed that the retrieval-reliant DF effect in the item-method can be reduced if item features are reinstated prior to recall. Thus, in our view, a loss of item fidelity might be responsible for the retrieval portion of the DF effect (article 2). Finally, using hierarchical multinomial modeling techniques, we found that variations in cognitive control as indexed by WMC predict better storage of TBR but not better or worse storage of TBF and/or retrieval of TBF and TBR-items (article 3).

The findings summarized in the present thesis thus clearly challenge the assumption that item-method DF is a pure storage phenomenon, as the DF effect on both storage and retrieval seems to be a robust finding that was replicated in all of our studies. Importantly, effects that are not apparent on the behavioral level, such as the enhanced recall of related TBR-items following feature reinstatement, can be observed if the processes contributing to the effect are isolated. As was the case in article 2, such an isolated effect for the retrieval process can also speak to conflicting evidence that were derived on the behavioral level only (Burgess et al., 2017; Hourihan et al., 2007). As most of the study specific findings and limitations have already been summarized for each article in chapter 4, I will take a cross article perspective in discussing the major results in the following sections. Specifically, I will relate these major findings to the research goals that were set at the beginning of my thesis (chapter 1).

The first goal of my thesis was to understand the cognitive processes that contribute to the two major forms of intentional forgetting, namely forgetting of information one has learned in the past and forgetting of information one has just encoded. To use the examples from the Introduction, the former type of intentional forgetting could be trying to forget an old phone number, whereas the latter form of forgetting could be trying to forget information a conversational partner has just said. Our findings suggest that forgetting of information one has learned in the past (costs of DF), as studied with the list-method paradigm, is mainly driven by retrieval processes. This finding is in line with theoretical frameworks that attribute the effect to such a retrieval deficit, such as the retrieval inhibition account (R. A. Bjork, 1989; Geiselman et al., 1983) and the context change account (Sahakyan & Kelley, 2002). Further, our findings challenge the selective rehearsal account (R. A. Bjork, 1970; R. A. Bjork et al., 1968), because intentional forgetting did not involve storage processes. Given the recent revival of the selective rehearsal view in this regard (Sheard & MacLeod, 2005), our findings clearly speak against the involvement of selective rehearsal processes in causing forgetting of irrelevant information that has been encoded in the past. So according to our view, simply stopping rehearsing the outdated information (e.g., old phone number) does not alone lead to successful forgetting of that information. Instead, trying to actively inhibit and/or change ones mental context in order to prevent the retrieval of that information seem to be more promising strategies in this regard. Importantly, our findings further suggest that successful intentional forgetting of outdated information also enables people to better store new information (benefits of DF). Therefore, from our process-level findings it becomes clear that intentional forgetting of outdated information is a highly adaptive mechanism that prevents the retrieval of that information and also enables better storage of new information.

The major finding of my thesis is the involvement of both storage and retrieval processes in intentional forgetting of information that one has just encoded (e.g., irrelevant information of an ongoing conversation). The involvement of retrieval processes challenges theoretical frameworks that attribute the effect to storage processes, such as the selective rehearsal account (R. A. Bjork, 1970; R. A. Bjork et al., 1968) and possibly the attention withdrawal account (Hourihan & Taylor, 2006; Taylor, 2005; Wylie et al., 2008), while possible supporting an inhibitory view (Geiselman & Bagheri, 1985; Zacks et al., 1996). Even though an inhibitory view may seem a plausible account for the involvement of retrieval processes, given the weak theoretical conceptualization of the inhibitory view (see chapter 2.2) together with recent evidence that stresses the loss of item fidelity in this context, I find it more likely that a loss of item-specific features and the later inability to retrieve these features is responsible for the retrieval portion of intentionally forgetting just encoded information (cf. Fawcett et al., 2016). Further, the finding from article 3 that WMC does not predict worse retrieval of TBF information further speaks against the involvement of inhibitory processes in this regard, as WMC is highly related to inhibitory abilities (Brewin & Beaton, 2002; Wessel, Overwijk, Verwoerd, & de Vrieze, 2008). Hence, when attempting to intentionally forget information that one has just encoded, simply terminating rehearsal of that information seems not to be the only mechanism involved, but some retrieval-reliant mechanism that hampers the access to features of the TBF content seems to be responsible for parts of the phenomenon as well.

The second goal of the present thesis was to find out whether individuals differ in their ability to intentionally forget outdated or irrelevant information. In terms of intentionally forgetting outdated information studied in the past, recent research suggests that individuals with better cognitive control abilities are better able to forget the outdated information compared to individuals with lower cognitive control abilities (Aslan et al., 2010; Delaney & Sahakyan, 2007; Soriano & Bajo, 2007). Regarding intentional forgetting of information one has just encoded, there is no study to date that has investigated this issue. Filling this gap, our findings from article 3 suggest, that individuals with high cognitive control abilities are better able to store relevant information that they try to remember, but are not better able to forget irrelevant information compared to individuals with low cognitive control abilities. Notably, variations in cognitive control do not predict worse storage of irrelevant information or the retrieval of either relevant or irrelevant information. This finding is in line with the selective rehearsal account (R. A. Bjork, 1970; R. A. Bjork et al., 1968) and clearly poses a challenge to the retrieval inhibition account (Geiselman & Bagheri, 1985; Zacks et al., 1996). Regarding the attention withdrawal account (Hourihan & Taylor, 2006; Taylor, 2005; Wylie et al., 2008), at first glance our findings might be challenging as well, because WMC did not predict worse storage of irrelevant TBF information. However, recent findings suggest that the processes engaged during attention withdrawal might be independent of individuals' cognitive control abilities (Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014). Thus, one needs to be cautious in dismissing attention-mediated mechanisms on the basis of our evidence. In sum, regarding the second research goal of investigating the role of individual differences in cognitive control and intentional forgetting of just encoded information, it seems to be the case that individuals with better cognitive control are better able to store relevant information, while they are not better able to withdraw their attention or inhibit the retrieval of irrelevant information.

Regarding the multinomial modeling approach that was used throughout the investigations presented in this thesis, it becomes evident that the use of such techniques proves to be highly useful when investigating memory phenomena compared to classical behavioral general linear modeling approaches. That is, processes that are central to how people remember and forget information can be investigated in a process-pure way without having to worry about possible confounds with other cognitive processes. In the present thesis, the *storage* and *retrieval* processes were of major interest in the context of investigating two forms of intentional forgetting. When referring to the measurement of storage and retrieval processes in a process-pure way, it is important to note that process-pure does not mean theoretically independent. For instance, one major advantage of the multinomial storage–retrieval model over other mathematical memory models is that it is grounded in psychological theory. That is, retrieval processes are assumed to occur after successful storage (see Figure 2) to just name one theoretically sound example (Batchelder & Riefer, 1986). But what about situations where model-based and traditional behavioral results diverge? In these cases, one could argue that the model is not valid in measuring the proposed processes. In case of the storage–retrieval model, which has been empirically validated in many research areas, however, behavioral and model-based results converge very often (see Rouder & Batchelder, 1998). This was also the case in our investigations. Only in rare cases, such as the retrieval enhancement of reinstated TBR-items of article 2, did we observe diverging results. Notably, these diverging results were also obtained on the behavioral level by independent studies that investigated this issue in the past (Burgess et al., 2017; Hourihan et al., 2007), and finding divergent results between behavioral and model-based results within one study consequently reflects the heterogeneous nature of the phenomenon itself without posing any threat to the model-based approach. Thus, I believe that supplying traditional analyses with the multinomial modeling approach can greatly benefit researchers at minimal additional effort and costs (use of word pairs as study material and use of a free-then-cued-recall paradigm for the final memory test).

In the next section, I will discuss the present work from a meta-level perspective that emphasizes the organizational structure of human memory[6]. According to organization theory (Katona, 1940), which was adopted to explain major memory phenomena with regard to storage, retrieval, depth of processing, and chunking, memories are organized in hierarchical structures. According to Mandler (1979), there are three types of structures: *Coordinate structures* that link memory representations so that access to one representation cues the entire episode (e.g., memories of a party), *subordinate structures* that describe tree-like hierarchical relations (e.g., members of superordinate categories), and *proordinate structures* that are represented by serial and propositional relations (e.g., acquisition of fear). According to organization theory, encoding information always involves organization of the encoded information during storage and whatever factors guide such organization will also aid later recall of the stored information (Mandler, 2002, 2011). Specifically, when memories are stored into a coordinate structure, as is the case in list-method DF, reinstating parts of that coordinate structure prior to retrieval will foster retrieval of that information, even if the information was intentionally attempted TBF. In fact, the context-change hypothesis supports such a view, in that reinstatement of the study context that was present during the initial storage episode of TBF information leads to the elimination of the DF effect compared to when study context is not reinstated (Sahakyan & Kelley, 2002). Further, a recent study in which participants were asked to study related items from a category in a list-method DF paradigm showed that contextual reinstatement but not temporal reinstatement lead to the elimination of the DF effect (Lehman & Malmberg, 2011). Thus, the list-method DF seems to be a good example of intentional forgetting with the goal of forgetting episodic information that is stored in a coordinate structure and thus the reinstatement of the same structure (or context) leads to an elimination of the DF effect. In the case of the present work, feature reinstatement also reduced the DF effect in the item-method paradigm to some degree. Here, however, as item encoding and storage are guided at an item-by-item basis, memories are not stored in a coordinate way in that they form a coherent study episode. Instead, they are stored as individual items. In our article 2, the related items might even be stored in subordinate structures, as some of these items were members of a superordinate category (i.e. clothes or furniture). In this case, reinstating the superordinate category cue should lead to enhanced recall of related information but leave unrelated information unaffected and our findings summarized in chapter 4.2 support this view. Thus, from an organizational perspective of memory, intentional forgetting should only be possible if the processes that cause such forgetting prevent the reinstatement of the organizational structure that guided storage in the first place. In the list-method, context-change

---

[6] I thank Klaus Fiedler for pointing me to organizational theory as an overarching theoretical framework.

(Sahakyan & Kelley, 2002) seems to be such a process, whereas in the item-method the withdrawal of attention (Taylor, 2005) from individual memory representations seems to describe that process. That is, attention withdrawal could lead to item-specific feature loss that causes parts of the DF effect and specifically the retrieval deficit for TBF information. Consequently, an organizational view of intentional forgetting as studied with the list-method and item-method of DF is well in line with the present results and could provide a good overarching framework for DF theories that nowadays are focused on individual effects of the two different paradigms instead of providing a global theory (but see Lehman & Malmberg, 2009).

Finally, the present work could also be relevant for applied research. For example, past research has shown that DF is impaired in some clinical populations such as schizophrenics (Racsmány et al., 2008), addicts (Zou, Zhang, Huang, & Weng, 2011), and patients with depression and anxiety (Wingenfeld, Terfehr, Meyer, Lwe, & Spitzer, 2012) to just name a few. The major theoretical views from these studies state that inhibitory, attention and/or context-change abilities are impaired and thus DF is not observed to the same degree as in healthy individuals. That is, in these studies the inability to forget is attributed to a dysfunction of the mechanisms that are postulated by the respective theoretical frameworks of the DF effect. However, no DF study has considered differences between clinical and healthy populations with regard to the underlying processes that drive the DF effect. In other areas, the use of multinomial modeling techniques such as the application of the storage–retrieval model has proven to be highly effective in identifying cognitive process deficits in clinical populations. For instance, Riefer et al. (2002) used the pair-clustering MPT model to measure storage and retrieval processes in patients suffering from schizophrenia or alcoholism and found that both storage and retrieval processes are hindered in these two clinical populations, but that retrieval deficits were more pronounced compared to storage deficits. Further, they found that alcoholics' storage improved in this specific memory paradigm, but retrieval deficits remained the same. This example illustrates that storage and retrieval processes can be affected in completely opposite directions in clinical populations. Thus, applied studies that investigate DF abilities in clinical populations should consider employing multinomial modeling techniques to be better able at identifying what cognitive processes are affected by a specific disorder.

To conclude, in the present thesis I showed that intentional forgetting is driven by different combinations of cognitive processes, depending on whether intentional forgetting of information encoded in the past or intentional forgetting of just encoded information is attempted. Retrieval processes seem to be driving intentional forgetting of information encoded in the past, whereas both storage and retrieval processes seem to be driving intentional forgetting of just encoded information. Importantly, the involvement of retrieval processes in forgetting just encoded information is likely a result of item-specific feature loss and variations in cognitive control abilities predict the storage of relevant information only in this regard. I hope the process-level approach of studying intentional forgetting presented in this thesis not only offers interesting ideas and methods for future research in the area of intentional forgetting but also for other areas of cognitive psychology.

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. http://doi.org/10.1016/j.jml.2003.08.006

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In C. M. MacLeod & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 159–190). Tokyo: Springer-Verlag.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087. http://doi.org/10.1037//0278-7393.20.5.1063

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

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*(6), 279–282. http://doi.org/10.1016/j.tics.2014.03.002

Aslan, A., Zellner, M., & Bäuml, K.-H. (2010). Working memory capacity predicts listwise directed forgetting in adults and children. *Memory*, *18*(4), 442–450. http://doi.org/10.1080/09658211003742698

Basden, B. H., & Basden, D. R. (1996). Directed forgetting : Further comparisons of the item and list methods. *Memory*, *4*(6), 633–653. http://doi.org/10.1080/096582196388825

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139–172). Mahwah, NJ: Erlbaum.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603–616. http://doi.org/10.1037/0278-7393.19.3.603

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology*, *39*, 129–149.

Bäuml, K.-H. (2008). Inhibitory processes. In H. L. Roediger III (Ed.), *Cognitive psychology of memory* (pp. 195–220). Oxford: Elsevier.

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*(1), 197–215. http://doi.org/10.1037/0278-7393.22.1.197

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition*, *5*(1), 176–196. http://doi.org/10.1006/ccog.1996.0011

Bjork, E. L., & Bjork, R. A. (2003). Intentional forgetting can increase, not decrease, residual influences of to-be-forgotten information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 524–531. http://doi.org/10.1037/0278-7393.29.4.524

Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103–137). Mahwah, NJ: Erlbaum.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*(3), 255–268. http://doi.org/10.1016/S0022-5371(70)80059-7

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Washington, DC: Winston.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, *10*(2), 55–56.

Brewin, C. R., & Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behaviour Research and Therapy*, *40*(8), 923–930. http://doi.org/10.1016/S0005-7967(01)00127-9

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*(1), 12–21. http://doi.org/10.1080/17470215808416249

Burgess, N., Hockley, W. E., & Hourihan, K. L. (2017). The effects of context in item-based directed forgetting: Evidence for "one-shot" context storage. *Memory & Cognition*. http://doi.org/10.3758/s13421-017-0692-5

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215–235. http://doi.org/10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Delaney, P. F., & Sahakyan, L. (2007). Unexpected costs of high working memory capacity following directed forgetting and contextual change manipulations. *Memory & Cognition*, *33*(5), 1074–1082.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language*, *74*, 77–90. http://doi.org/10.1016/j.jml.2013.09.003

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language*, *74*, 1–15. http://doi.org/10.1016/j.jml.2014.03.006

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*(3), 108–124. http://doi.org/10.1027/0044-3409.217.3.108

Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology: General*, *145*(1), 56–81. http://doi.org/10.1037/xge0000128

Fawcett, J. M., & Taylor, T. L. (2010). Directed forgetting shares mechanisms with attentional withdrawal but not with stop-signal inhibition. *Memory & Cognition*, *38*(6), 797–808. http://doi.org/10.3758/MC.38.6.797

Foster, N. L., & Sahakyan, L. (2012). Metacognition influences item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1309–1324. http://doi.org/10.1037/a0027868

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition*, *13*(1), 57–62. http://doi.org/10.3758/BF03198444

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, *112*(1), 58–72. http://doi.org/10.1037/0096-3445.112.1.58

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Golding, J. M., & MacLeod, C. M. (1998). *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ US: Erlbaum.

Gottlob, L. R., & Golding, J. M. (2007). Directed forgetting in the list method affects recognition memory for source. *The Quarterly Journal of Experimental Psychology*, *60*(11), 1524–1539. http://doi.org/10.1080/17470210601100506

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*(12), 1–11. http://doi.org/10.1177/0956797613492984

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Hays, W. L. (1994). *Statistics* (5th editio). Forth Worth, TX: Harcour Brace College.

Heck, D., Arnold, N. R., & Arnold, D. (2017). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, 1–21. http://doi.org/10.3758/s13428-017-0869-7

Hirshman, E. (1988). The expectation-violation effect: Paradoxical effects of semantic relatedness. *Journal of Memory and Language*, *27*(1), 40–58. http://doi.org/10.1016/0749-596X(88)90047-2

Hirshman, E., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory and Language*, *28*(5), 594–609. http://doi.org/10.1016/0749-596X(89)90015-

6

Hourihan, K. L., Goldberg, S., & Taylor, T. L. (2007). The role of spatial location in remembering and forgetting peripheral words. *Canadian Journal of Experimental Psychology*, *61*(2), 91–101. http://doi.org/10.1037/cjep2007010

Hourihan, K. L., & Taylor, T. L. (2006). Cease remembering: Control processes in directed forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1354–1365. http://doi.org/10.1037/0096-1523.32.6.1354

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. http://doi.org/10.1007/BF02294263

Johnson, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin*, *116*(2), 274–292. http://doi.org/10.1037/0033-2909.116.2.274

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 336–358. http://doi.org/10.1037/0278-7393.26.2.336

Katona, G. (1940). *Organizing and memorizing*. New York, NY: Columbia University Press.

Kaufman, L., & Gay, D. (2003). The PORT Library - Optimization. Murray Hill, NJ.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models - A latent-class approach. *Psychometrika*, *71*(1), 7–31. http://doi.org/10.1007/S11336-004-1188-3

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. http://doi.org/10.1007/S11336-009-9141-0

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1365–1376. http://doi.org/10.1037/a0028147

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding , maintenance , and retrieval processes in the lag effect : A multinomial processing tree analysis. *Memory*, *20*(1), 37–48. http://doi.org/10.1080/09658211.2011.631550

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7. http://doi.org/10.1016/j.jmp.2010.08.013

Lee, Y.-S., Lee, H.-M., & Fawcett, J. M. (2013). Intentional forgetting reduces color-naming interference: evidence from item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 220–36. http://doi.org/10.1037/a0028905

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 970–88. http://doi.org/10.1037/a0015728

Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition*, *39*(2), 335–47. http://doi.org/10.3758/s13421-010-0025-4

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*(1), 123–129. http://doi.org/10.3758/BF03210819

MacLeod, C. M. (2007). The concept of inhibition in cognition. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 3–23). Washington, DC: American Psychological Association.

Mandler, G. (1979). Organization, memory, and mental structures. In C. R. Puff (Ed.), *Memory organization and structure*. New York, NY: Academic Press.

Mandler, G. (2002). Organisation: What levels of processing are levels of. *Memory*, *10*(5–6), 333–338. http://doi.org/10.1080/09658210244000153

Mandler, G. (2011). From association to organization. *Current Directions in Psychological Science*, *20*(4), 232–235. http://doi.org/10.1177/0963721411414656

Marevic, I., & Rummel, J. (submitted). Retrieval-reliant mechanisms of item-method directed forgetting: The effect of semantic cues. *Psychological Research*.

Meiser, T., Sattler, C., & Weisser, K. (2008). Binding of multidimensional context information as a distinctive characteristic of remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 32–49. http://doi.org/10.1037/0278-7393.34.1.32

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. http://doi.org/10.1177/0963721411429458

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. http://doi.org/10.3758/BRM.42.1.42

Nowicka, A., Jednoróg, K., Wypych, M., & Marchewka, A. (2009). Reversed old/new effect for intentionally forgotten words: An ERP study of directed forgetting. *International Journal of Psychophysiology*, *71*(2), 97–102. http://doi.org/10.1016/j.ijpsycho.2008.06.009

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity — facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045. http://doi.org/10.1016/S0191-8869(99)00251-2

Pastötter, B., & Bäuml, K.-H. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 54–65. http://doi.org/10.1037/a0017406

Pastötter, B., Bäuml, K.-H., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change: Evidence for a reset of encoding processes. *NeuroImage*, *43*(1), 173–181. http://doi.org/10.1016/j.neuroimage.2008.07.005

Pastötter, B., Tempel, T., & Bäuml, K.-H. (2017). Long-term memory updating: The reset-of-encoding hypothesis in list-method directed forgetting. *Frontiers in Psychology*, *8*(2076), 1–6. http://doi.org/10.3389/fpsyg.2017.02076

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. http://doi.org/10.1037/h0049234

Racsmany, M., & Conway, M. A. (2006). Episodic inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 44–57. http://doi.org/10.1037/0278-7393.32.1.44

Racsmány, M., Conway, M. A., Garab, E. A., Cimmer, C., Janka, Z., Kurimay, T., … Szendi, I. (2008). Disrupted memory inhibition in schizophrenia. *Schizophrenia Research*, *101*(1–3), 218–224. http://doi.org/10.1016/j.schres.2008.01.002

Rawson, K. A., & Kintsch, W. (2002). How does background information improve memory for text content? *Memory & Cognition*, *30*(5), 768–778. http://doi.org/10.3758/BF03196432

Redick, T. S., Heitz, R. P., & Engle, R. W. (2007). Working memory capacity and inhibition: Cognitive and social consequences. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 125–142). Washington, DC: American Psychological Association.

Reitman, W., Malin, J. T., Bjork, R. A., & Higman, B. (1973). Strategy control and directed forgetting. *Journal of Verbal Learning and Verbal Behavior*, *12*(2), 140–149. http://doi.org/10.1016/S0022-5371(73)80003-9

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology*, *69*(10), 1969–1995. http://doi.org/10.1080/17470218.2014.914546

Riefer, D. M., & Batchelder, W. H. (1991). Age differences in storage and retrieval: A multonomial modeling analysis. *Bulletin of the Psychonomic Society*, *29*(5), 415–418.

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*(2), 184–201. http://doi.org/10.1037/1040-3590.14.2.184

Riefer, D. M., & LaMay, M. L. (1998). Memory for common and bizarre stimuli: A storage-retrieval analysis. *Psychonomic Bulletin & Review*, *5*(2), 312–317. http://doi.org/10.3758/BF03212957

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*(6), 601–611. http://doi.org/10.3758/BF03202710

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 195–225). New York: Psychology Press.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604. http://doi.org/10.3758/BF03196750

Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *42*(10), 1526–1543. http://doi.org/10.1037/xlm0000266

Rummel, J., Steindorf, L., Marevic, I., & Danner, D. (2017). A validation study of the German complex-span tasks and some general considerations on task translation procedures in cognitive psychology. *European Journal of Psychological Assessment*, *advance on*. http://doi.org/10.1027/1015-5759/a000444

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*(1), 63–77. http://doi.org/10.1037/h0031185

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, *9*(1), 99–105. http://doi.org/10.1016/S0022-5371(70)80015-9

Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language*, *48*(1), 195–206. http://doi.org/10.1016/S0749-596X(02)00524-7

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2014). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 59, pp. 131–190). Elsevier Inc.: Academic Press. http://doi.org/10.1016/b978-0-12-407187-2.00004-6

Sahakyan, L., & Goodmon, L. B. (2007). The influence of directional associations on directed forgetting and interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1035–1049. http://doi.org/10.1037/0278-7393.33.6.1035

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1064–1072. http://doi.org/10.1037//0278-7393.28.6.1064

Sheard, E. D., & MacLeod, C. M. (2005). List method directed forgetting: Return of the selective rehearsal account. In N. Ohta, C. M. MacLeod, & B. Uttl (Eds.), *Dynamic cognitive processes*.

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560–575. http://doi.org/10.3758/s13428-012-0259-0

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review*, *15*(4), 713–731. http://doi.org/10.3758/PBR.15.4.713

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167–183. http://doi.org/10.1016/j.jmp.2009.06.007

Soriano, M. F., & Bajo, M. T. (2007). Working memory resources and interference in directed forgetting. *Psicológica*, *28*, 63–85.

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1532–1539. http://doi.org/10.1037/a0024852

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, *39*(2), 267–73. http://doi.org/10.1017/CBO9781107415324.004

Taylor, T. L. (2005). Inhibition of return following instructions to remember and forget. *The Quarterly Journal of Experimental Psychology*, *58*(4), 613–629. http://doi.org/10.1080/02724980443000115

Taylor, T. L., & Fawcett, J. M. (2012). Does an instruction to forget enhance memory for other presented items? *Consciousness and Cognition*, *21*(3), 41–6. http://doi.org/10.1016/j.concog.2012.05.002

Thompson, K. M., Hamm, J. P., & Taylor, T. L. (2014). Effects of memory instruction on attention and information processing: Further investigation of inhibition of return in item-method directed forgetting. *Attention, Perception & Psychophysics*, *76*(2), 41–6. http://doi.org/10.3758/s13414-013-0584-0

Thompson, K. M., & Taylor, T. L. (2015). Memory instruction interacts with both visual and motoric inhibition of return. *Attention, Perception & Psychophysics*, *77*(3), 804–818.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*(2), 255–262. http://doi.org/10.1037/h0029997

Unsworth, N. (2009). Examining variation in working memory capacity and retrieval in cued recall. *Memory*, *17*(4), 386–96. http://doi.org/10.1080/09658210902802959

Unsworth, N. (2016). The Many Facets of Individual Differences in Working Memory Capacity. In *Psychology of Learning and Motivation* (Vol. 65, pp. 1–46). Elsevier Ltd. http://doi.org/10.1016/bs.plm.2016.03.001

Unsworth, N. (2017). Examining the dynamics of strategic search from long-term memory. *Journal of Memory and Language*, *93*, 135–153. http://doi.org/10.1016/j.jml.2016.09.005

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. http://doi.org/10.3758/BF03192720

Van Hooff, J. C., Whitaker, T. A., & Ford, R. M. (2009). Directed forgetting in direct and indirect tests of memory: Seeking evidence of retrieval inhibition using electrophysiological measures. *Brain and Cognition*, *71*(2), 153–164. http://doi.org/10.1016/j.bandc.2009.05.001

Wessel, I., Overwijk, S., Verwoerd, J., & de Vrieze, N. (2008). Pre-stressor cognitive control is related to intrusive cognition of a stressful film. *Behaviour Research and Therapy*, *46*(4), 496–513. http://doi.org/10.1016/j.brat.2008.01.016

Wingenfeld, K., Terfehr, K., Meyer, B., Lwe, B., & Spitzer, C. (2012). Memory bias for emotional and illness-related words in patients with depression, anxiety and somatization disorders: An investigation with the directed forgetting task. *Psychopathology*, *46*(1), 22–27. http://doi.org/10.1159/000338609

Wylie, G. R., Foxe, J. J., & Taylor, T. L. (2008). Forgetting as an active process: An fMRI investigation of item-method-directed forgetting. *Cerebral Cortex*, *18*(6), 670–682. http://doi.org/10.1093/cercor/bhm101

Zacks, R. T., Radvansky, G. A., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 143–156. http://doi.org/10.1037/0278-7393.22.1.143

Zellner, M., & Bäuml, K.-H. (2006). Inhibitory deficits in older adults: List-method directed forgetting revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 290–300. http://doi.org/10.1037/0278-7393.32.3.290

Zou, Z., Zhang, J. X., Huang, X., & Weng, X. (2011). Impaired directed forgetting in abstinent heroin addicts. *Memory*, *19*(1), 36–44. http://doi.org/10.1080/09658211.2010.532806

List of tables

List of figures

Appendix A1 – Article 1

Note:

This is a previous version of an article that has been published in its final version in the *Journal of Experimental Psychology: Learning, Memory, and Cognition* [copyright American Psychological Association], doi: 10.1037/xlm0000266

Investigating Storage and Retrieval Processes of Directed Forgetting: A Model-based Approach

Jan Rummel & Ivan Marevic
Heidelberg University

Beatrice G. Kuhlmann
University of Mannheim

Author Note

Jan Rummel & Ivan Marevic, Department of Psychology, Heidelberg University, Heidelberg, Germany. Beatrice G. Kuhlmann, Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany.

Correspondence concerning this article should be addressed to Jan Rummel, Department of Psychology, Heidelberg University, Hauptstrasse 47-51, D-69117 Heidelberg, Germany. E-mail: jan.rummel@psychologie.uni-heidelberg.de

Abstract

Intentional forgetting of previously learned information is an adaptive cognitive capability of humans but its cognitive underpinnings are not yet well understood. It has been argued that it strongly depends on the presentation method whether forgetting instructions alter storage or retrieval stages (Basden, Basden, & Gargano, 1993). In Experiment 1, we compared the processes underlying the directed-forgetting effect in the two mosts widely used presentation methods, namely the list-method and the item-method, and also differentiated between costs (i.e., poorer memory for to-be-forgotten information) and benefits (i.e., better memory for to-be-remembered information) of directed forgetting within both methods. Using a multinomial modeling approach (Riefer & Rouder, 1992; Rouder & Batchelder, 1998), our results showed that directed-forgetting benefits were due to better storage of to-be-remembered information in both methods. In line with current theorizing, list-method directed-forgetting costs occurred due to reduced retrieval of to-be-forgotten information. Item-method costs, however, occurred not only due to reduced storage, which is the dominant current view, but also due to reduced retrieval. In Experiment 2, we replicated the novel finding that retrieval processes contribute to item-method directed forgetting independent of recall-output order. Implications of these findings for current directed-forgetting theories are discussed.

Keywords: directed forgetting; list-method; item-method; multinomial modeling; storage–retrieval model

Investigating Storage and Retrieval Processes of Directed Forgetting: A Model-based Approach

Forgetting is usually seen as a passive process of memory loss due to decay or interference from newer information (see Dempster & Brainerd, 1995, for an overview). Sometimes, however, people intend to forget certain information, for example, because it is no longer relevant (e.g., an expired cell phone number) or turned out wrong (e.g., the date one falsely assumed to be a friend's birthday). As evident from these examples, intentional forgetting can be adaptive and thus it is of interest to understand its underlying processes.

A straightforward way to initiate intentional forgetting is to consciously form the intention to forget some information after it has been learned. Mimicking this, cognitive psychologists use forgetting instructions to investigate deliberate forgetting. This directed forgetting (DF) technique (R. A. Bjork, 1970), has been widely used to study intentional forgetting in the laboratory but the cognitive mechanisms underlying this phenomenon are still under debate (for reviews see Bäuml, 2008; Sahakyan, Delaney, Foster, & Abushanab, 2014). Additionally, different methods of DF have been suggested to affect different cognitive processes (Basden, Basden, & Gargano, 1993). In order to gain a better understanding of the cognitive underpinnings of DF, we used multinomial modeling (Erdfelder et al., 2009; Riefer & Batchelder, 1988) to disentangle effects on memory storage and memory retrieval processes across different DF methods.

Investigating Directed Forgetting in the Laboratory

In the standard DF paradigm, participants are instructed to forget (some of the) information previously learned (Bäuml, 2008; R. A. Bjork, LaBerge, & Legrand, 1968; R. A. Bjork & Woodward, 1973; R. A. Bjork, 1970; Sahakyan et al., 2014). Two variants of this paradigm have been established: (a) the *list-method* and (b) the *item-method*.

In list-method DF (R. A. Bjork et al., 1968; Eppstein, 1972; Sahakyan & Kelley, 2002), one group of participants (i.e., the forget group) studies two lists of items and is instructed to forget the first list before studying the second one. Although participants were told to forget the first list, they are then asked to recall items from both lists in a subsequent memory test. Memory for each list of the forget group is then compared with performance of a so-called remember group that was not instructed to forget the first list. Typical findings are worse memory for the first list and better memory for the second list in the forget than in the remember group. The former effect has been termed *costs* (Reitman, Malin, Bjork, & Higman, 1973) and the latter *benefits* of DF (Sahakyan & Delaney, 2003).

In item-method DF (R. A. Bjork, 1972; Foster & Sahakyan, 2012; Golding & MacLeod, 1998; Woodward & Bjork, 1971), there are also to-be-remembered (TBR) items and to-be-forgotten (TBF) items. However, the two item types are not presented list-wise but randomly intermixed. Instead of one list-wise "forget" cue as employed in the list-method, the item-method employs item-wise cues indicating the item type (TBR or TBF) directly after each item. Thus, both DF procedures ensure that the TBF-items are initially encoded before the "forget" cue occurs, which is crucial when studying forgetting. After being presented with all items, item-method participants are also asked to recall *all* (i.e., TBR and TBF) items in a final memory test. Better memory for TBR than for TBF-items is interpreted as evidence for intentional forgetting (Basden et al., 1993; R. A. Bjork, 1972; Hourihan & Taylor, 2006). It is noteworthy, however, that the comparison between TBR and TBF-items does not allow distinguishing between DF costs (i.e., reduced memory for TBF-items) and DF benefits (i.e., increased memory for TBR-items). In fact, both costs and benefits probably contribute to the standard item-method DF effect. It is possible to also differentiate between costs and benefits in this method by employing a remember group that receives "remember" cues for all items. However, unlike in list-method DF studies, such a remember group is not standard procedure of item-method DF studies (Basden et al., 1993; Basden & Basden, 1996; R. A. Bjork, 1972; Hourihan & Taylor, 2006; but see Foster & Sahakyan, 2012).

Different Methods, Different Processes?

Not only do the two most widely used DF methods differ in key procedural details, the DF effects produced with the two methods are also assumed to affect different memory processes (Basden et al., 1993; Basden & Basden, 1996, 1998; Johnson, 1994; Sahakyan & Foster, 2009). In the item-method, item type varies item-by-item. Here, TBF-items are unlikely to be rehearsed because one knows the item type right after encoding an item and can (and should!) thus terminate rehearsal immediately. Consequently, Basden et al. (1993) argue that item-method DF is due to *selective rehearsal* of TBR-items. List-method forgetting, however, is less likely due to selective rehearsal of TBR-items, because the "forget" cue does not appear until the complete TBF-list (first list) has been studied right before studying the TBR-list (second list). Consequently, TBF-items in the list-method are likely to be rehearsed until the "forget" cue occurs and their forgetting is thus assumed to be due to *retrieval inhibition* (Basden et al., 1993; but see Sheard & MacLeod, 2005).

Evidence that DF processes differ between methods comes from findings that DF effects were observed with both methods in recall tests but only consistently with the item-method in recognition tests (Basden et al., 1993; but see Lehman & Malmberg, 2009). Further, only with the item-method, DF effects were found with implicit memory tests (Basden & Basden, 1996). Finally, there was a stronger release from list-method than from item-method DF in a second recall test that was applied after a recognition test or after priming TBF-items (Basden et al., 1993) .

Notably, Basden and colleagues' research was conducted in the 90s and thus their interpretation built on the theories of DF that existed at that time, that is, the *selective rehearsal account* (R. A. Bjork, 1970) and the *retrieval-inhibition account* to DF (R. A. Bjork, 1989). Since then, however, new theories have emerged that assume somewhat different cognitive mechanisms of DF and also consider that DF costs and benefits may rely on different processes.

The *context-inhibition account* of list-method DF (Pastötter, Bäuml, & Hanslmayer, 2008; Pastötter & Bäuml, 2010) proposes that the context in which the TBF-list has been studied is inhibited, rather than the TBF-list itself. Because context information is needed to reinstate the learning episode

(Godden & Baddeley, 1975; Smith, 1979), its inhibition renders retrieval of the TBF-list more difficult. Further, the proponents of this account argue that DF benefits arise from a reset of encoding initiated by the "forget" cue as well as from an interference reduction for the TBR-list due to the inhibition of prior context (Pastötter, Kliegl, & Bäuml, 2012). Similarly, the *context-change account* of list-method DF (Sahakyan et al., 2014; Sahakyan & Kelley, 2002) assumes that a (temporarily) impoverished memory for context information from the TBF-study episode underlies the DF costs. Notably, Sahakyan and colleagues assume that list-method DF is an active strategic decision to change the (mental) context in order to comply with the "forget" cue. Support for this idea comes from findings that instructions encouraging a mental context change (e.g., think about your parents' house) produce effects that parallel those from standard DF instructions (Sahakyan & Kelley, 2002; Sahakyan & Smith, 2014; but see Pastötter et al., 2008 for neuropsychological evidence that DF and mental context change produce distinguishable activation patterns) as well as from forgetting-strategy self-reports (Foster & Sahakyan, 2011; Sahakyan et al., 2014). Sahakyan and colleagues (Sahakyan et al., 2014; Sahakyan & Kelley, 2002) further argue that proactive interference from TBF-items is reduced due to the context change, contributing to the DF benefits. The larger part of the benefits, however, is supposed to result from a change in post-cue encoding processing, that is, the disengagement from rehearsal of TBF-items. Thus, the processes assumed to cause benefits in list-method DF by Sahakyan et al. (2014) are relatively similar to the ones assumed by Pastötter and Bäuml (2010). The conceptualization of the forgetting mechanism, however, differs.

Context inhibition and context change accounts are both dual-process accounts that assume that selective rehearsal plays a critical role for list-method DF benefits and both accounts do not easily translate to item-method DF where TBF and TBR contexts fluctuate unpredictably. Because any single-process account that is able to explain DF will outmatch context accounts in parsimony, Sheard and MacLeod (2005) proposed a revival of the selective rehearsal account. In a series of studies, they showed that list-method DF costs result from lower recall in the primacy and recency parts of the TBF-list compared to the TBR-list. Such serial position effects are indicative of selective rehearsal (Rundus & Atkinson, 1970; Rundus, 1971). The authors concluded that the selective rehearsal account—as the most parsimonious account of both list and item-method DF—should not be dismissed from the theoretical framework of DF (but see Sahakyan et al., 2014).

This brief literature review shows that DF mechanisms are still an issue of debate. Since the initial work by Basden and colleagues (1993; Basden & Basden 1996; see also Basden & Basden, 1998, for a summary), the two methods have not been compared systematically any more (but see Sahakyan & Foster, 2009 for a method comparison for forgetting of actions). However, results from different studies suggest that the cognitive underpinnings of both methods differ: The item-method typically shows stable DF effects in both recall and recognition memory tests (Basden et al., 1993; MacLeod, 1999; for a review see Golding & MacLeod, 1998). In several studies, the list-method did not produce reliable recognition DF effects (Basden et al., 1993; E. L. Bjork & Bjork, 2003; Gottlob &

Golding, 2007; Zellner & Bäuml, 2006; but see also Sahakyan & Delaney, 2005; Lehman & Malmberg, 2009). Further, a DF effect on implicit memory has been usually found with the item-method (Basden et al., 1993; Basden & Basden, 1996) but not with the list-method (Basden et al., 1993; Basden & Basden, 1998; E. L. Bjork & Bjork, 1996; Racsmány & Conway, 2006; but see Koppel & Storm, 2012). Therefore, and because new theoretical accounts have been developed since the last DF-method comparison by Basden and colleagues, we aimed to re-examine potential differences between the cognitive underpinnings of list-method and item-method DF in the present studies.

To this end, we designed two experiments. The first experiment aimed to investigate the cognitive underpinnings of list-method and item-method DF and to compare them. For this purpose we realized list-method and item-method conditions that were as similar as possible to ensure that method differences can be traced back only to the different methods of presenting TBF-items. In the second experiment, we replicated the central findings of Experiment 1 and further investigated effects of recall-output order in item-method DF (Anderson, 2005) that were not controlled for in Experiment 1. Notably, Basden and colleagues did not differentiate between DF costs and benefits in their studies. However, a recent study showed isolated list-method DF benefits in recognition-memory suggesting that the processes underlying costs and benefits may differ (Benjamin, 2006). To differentiate between DF-cost and benefit mechanisms within both methods, we also realized control groups for both methods that had to remember all study items. In both Experiments, we made use of established multinomial modeling techniques which allow precise measurement of distinct cognitive processes (Batchelder & Riefer, 1999; Erdfelder et al., 2009). Specifically, we applied the storage–retrieval model (Rouder & Batchelder, 1998) to quantify the relative contribution of storage and retrieval processes to DF costs and benefits.

Measuring Storage and Retrieval Processes with Multinomial Models

Multinomial processing tree (MPT) models are a class of mathematical models for categorical data (Batchelder & Riefer, 1999) that can be used to disentangle the cognitive processes underlying observable behavioral effects (see Erdfelder et al., 2009 for a recent review). Importantly, a class of MPT models has been used to disentangle storage and retrieval components of various memory effects, such as bizarreness effects (Riefer & LaMay, 1998; Riefer & Rouder, 1992), age-related memory differences (Riefer & Batchelder, 1991a), memory deficits in clinical populations (Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002), and lag effects (Küpper-Tetzel & Erdfelder, 2012).

In the present study, we used the storage–retrieval MPT model for data from a free-then-cued-recall paradigm investigating memory for word pairs (Riefer & Rouder, 1992; Rouder & Batchelder, 1998) to compare the contribution of storage and retrieval processes to DF between different methods. This model requires study of item pairs to be followed by a free-recall test and then a cued-recall test using the first word of each pair as a cue. There are six possible observable behavioral events in this paradigm: (E_1) successful free recall of the pair and successful cued recall; (E_2) successful free recall of the pair but failed cued recall; (E_3) successful free recall of a single item from a pair (singleton) and successful cued recall; (E_4) successful free recall of a singleton but failed cued recall; (E_5) failed free recall but successful cued recall; (E_6) failed free recall and failed cued recall. The outcomes E_1-E_6 are modelled as a function of the following five latent cognitive states which are reflected by the five model parameters. Parameters reflect the probability ([0; 1]) of cognitive states.

- The *storage parameter a* indicates the probability of associative storage of an item pair (i.e., of both items and their association) and its maintenance in memory until the test.
- The *retrieval parameter r* indicates the probability of successful retrieval of a stored item pair during free recall.
- *Parameters s*, *u*, and *l* are additional parameters needed to ensure model validity, but they do not differentiate storage and retrieval processes but rather cover nuisance effects (Riefer & Batchelder, 1988). Parameter *s* reflects the probability that both items of a pair are retrieved as singletons even though they had been stored as a pair. Parameter *u* reflects the probability that both items of a pair had not been stored associatively and were retrieved as singletons. Even though psychologically plausible, the cognitive states reflected by parameters *s* and *u* occur rather infrequently in standard free-then-cued recall paradigms (Rouder & Batchelder, 1998). Finally, the *parameter l* accounts for the test order and the delay between tests by measuring the probability that some items are lost from memory during the delay between the free (first) and the cued (second) memory test.[7] Given that this delay is very short, the probability of memory loss is typically low (i.e., parameter estimates for *l* are usually close to zero).

Figure 1 illustrates the storage–retrieval MPT model adapted from Rouder and Batchelder (1998). The model can be formalized in terms of the following equations that express the probabilities for each recall event as a combination of the underlying cognitive states:

$$P(E_1) = a\,r\,(1-l) + a\,(1-r)\,s^2\,(1-l)$$

$$P(E_2) = a\,r\,l + a\,(1-r)\,s^2\,l + (1-a)\,u^2$$

$$P(E_3) = 2\,a\,(1-r)\,s\,(1-s)\,(1-l)$$

$$P(E_4) = 2\,a\,(1-r)\,s\,(1-s)\,l + 2\,(1-a)\,u\,(1-u)$$

$$P(E_5) = a\,(1-r)\,(1-s)^2\,(1-l)$$

$$P(E_6) = a\,(1-r)\,(1-s)^2\,l + (1-a)\,(1-u)^2$$

The parameters of the storage–retrieval model at hand have been validated in previous empirical and simulation-based investigations, in which specific manipulations were applied, so that only one parameter (e.g., storage) was affected, while leaving other parameters (e.g., retrieval) unaffected (Riefer & Batchelder, 1991b; Rouder & Batchelder, 1998).

---

[7] In the original model (Rouder & Batchelder, 1998), this parameter was termed forget *(f)* but it was termed loss-from-memory *(l)* for the present study to be not confused with the forgetting phenomenon of investigation.

Application of the storage–retrieval model to DF data for the present study required minimal modification of the standard DF procedures: The to-be-learned (and, if instructed to, to-be-forgotten) items must be word pairs rather than single words and the usual free recall test must be followed by an additional cued-recall test. Notably, adding this additional test afterwards cannot affect DF costs and benefits observed on the standard recall test but is required to meet the model assumptions.

Experiment 1

Experiment 1 was designed to investigate differences in the cognitive underpinnings of DF across different methods. Basden et al. (1993; Basden & Basden, 1996) were the first to argue that the two DF methods differently affect storage (i.e., post-cue encoding and retention) and retrieval processes. Since then, a growing body of research, as reviewed earlier, has support their idea that list-method DF is retrieval-based whereas item-method DF is storage-based by demonstrating that item-method but not list-method DF occur in memory tests that facilitate retrieval (i.e., recognition tests, implicit-memory tests). However, these newer studies did not directly compare item-method and list-method DF within one study. Therefore, other factors such as methodological differences, procedure and design differences, as well as insufficient differentiations between DF costs and benefits may have influenced the differences between these studies. Further, Basden et al.'s (1993) one-to-one mapping of processes and methods may well be an oversimplification of the DF-method differences, because they are based on a null-effect, that is, the mere observation that item-method but not list-method DF affects recognition and implicit memory. Alternatively, these tests might just not be sensitive enough to catch the—typically somewhat smaller—list-method DF effects. Finally, Basden et al. did not differentiate between DF costs and benefits, although they can be assumed to be driven by different processes. For these reasons, we argue that more systematic and experimentally controlled comparisons of DF methods and a more precise measurement of the underlying processes are required to better understand method differences of DF costs and benefits.

Based on theorizing by Basden and colleagues one would expect that retrieval-stage processes play a crucial role for list-method DF whereas storage-stage processes play a crucial role for item-method DF. Alternatively, processes at both stages could be generally affected by DF instructions within both methods to some extent. The more process-pure estimates of storage (parameter *a*) and retrieval (parameter *r*) processes from the multinomial storage–retrieval model should help to evaluate these ideas.

Method

Participants and Design. One hundred and twelve Heidelberg University students (90 women; $M_{\text{age}} = 22.69$ years, range: 18-33 years) participated for course credit or monetary compensation. There were four experimental groups (i.e., list-method–forget, list-method–remember, item-method–forget, and item-method–remember) with $n = 28$ participants each. All groups were presented with two different item types. For forget groups, one type of items was post-cued as TBF, whereas the other was post-cued as TBR. For remember groups, both types of items were post-cued as

TBR. Items that are TBF in the forget and TBR in the remember groups will be referred to as TBF/R. Items that are TBR for all groups will be referred to as TBR/R items. For list-method groups, item type was manipulated list-wise; for item-method groups, item type was manipulated item-by-item.

Material. A set of 120 German nouns of medium frequency were selected from the *dlex* Database (Heister et al., 2011). For a practice phase, 48 of these words were randomly combined to 24 cue-target pairs. For all groups, the same 12 pairs were selected to be TBF or TBR, respectively. For the actual learning phase, another 48 words were again combined to two sets of 12 cue-target pairs each. For the study phase, one set was determined to be TBF/R-pairs, the other to be TBR/R-pairs (sets were randomly determined and counterbalanced within each list-method group). To control for material-specific effects, we used the remaining words to generate a second material set. Therefore, each cue-word from the first set was combined with a new target-word from the remaining 24 words (see Küpper-Tetzel & Erdfelder, 2012, for a similar method). The two material sets were used equally often within each group.

Procedure. After providing informed consent, participants were randomly assigned to one of the four experimental groups, ensuring an equal number of participants within each group. All participants were informed that they were going to study word pairs for a later recall test and that some word pairs needed to be remembered and others not. Participants of the list-method groups were further informed that a cue after each study list would indicate whether the preceding word-pair list had to be remembered for a later recall test or not. Participants of the item-method groups were informed that a cue after each studied word pair would indicate whether the word pair had to be remembered for a latter recall test or not. As a primacy buffer (cf. Lehman & Malmberg, 2009) and to foster the trustworthiness of the DF instructions, all participants first performed a practice phase consisting of 12 TBR-pairs and 12 TBF-pairs. Word pairs were presented one at a time on the screen. The presentation method during the practice phase was identical with the one during the actual learning phase: In list-method groups, pairs were presented list-wise (first TBF, then TBR-pairs) with the instruction to forget all preceding word-pairs applied after the first list. Participants solved math problems for 30s between learning the two lists. In the item-method groups, TBF and TBR-pairs were presented randomly intermixed. Each pair was followed by a cue (i.e., ***EEE*** to instruct remembering or ***VVV*** to instruct forgetting —the first letters of the German words for "remember" [erinnern] and "forget" [vergessen]). To equate procedures between the two methods, there was a 30s-break between the presentation of the first and the second half of pairs during which participants solved math problems. After being presented with all 24 practice pairs, all participants solved further math problems for 30s and then completed a free and a cued recall test for TBR-items only. For the former test, participants were asked to recall as many TBR-words as possible; for the latter test, participants were presented with the first word of a pair (cue) and were asked to recall the second word of that pair (target).

For the actual learning phase, participants were again shown two word-pair sets (TBF/R and TBR/R-pairs) consisting of 12 pairs each. The presentation of word pairs from both sets was either blocked list-wise (list-method) or varied item-by-item (item-method). Again, there was a 30s-break after the first list/ first half of the items during which participants solved math problems. In the two forget groups, the TBF/R-pairs were cued as TBF and the TBR/R-items were cued as TBR. There was either a "forget" cue presented after the first study list (list-method) or for half of the studied items (item-method; half before and half after the break). Remember-group participants were cued to remember all items (i.e., TBF/R and TBR/R-items).

After studying all 24 pairs and solving further math problems for 30s, all participants performed a free recall test. As common in list-method DF (Sahakyan et al., 2014), all list-method participants were first asked to recall the words from the first (TBR/F) and then the words from the second (TBR/R) study list. Recall time was 60s for each list. Item-method participants were asked to freely recall as many words as possible from the learning phase for 120s, as typical in this paradigm (Basden et al., 1993; Sahakyan & Foster, 2009). Thus, total recall time was identical for both methods. Next, all participants performed the cued-recall test needed for the model application. For this test, participants were presented with the first word of each pair from the study phase and had to type the second word. In line with the free-recall procedure, list-method participants were first presented with the cues from the first (TBR/F) and then with those from the second (TBR/R) list (presentation order within lists was determined randomly). Item-method participants were also first presented with the cues (in random order) from the first half and then with cues from the second half of their study phase, both containing half TBR/F and half TBR/R-items. Before the final debriefing, participants who received forget instructions were asked whether they used a strategy to forget TBF-pairs and, if so, what kind of strategy they used.

Results and Discussion

An alpha level of .05 was adopted for all analyses. Most participants (i.e., 82.14% of the list-method–forget and 92.85% of the item-method–forget group) indicated that they used a forgetting strategy, when being asked after the experiment. Strategy-use frequencies did not differ between presentation methods, $\chi(1) = 1.47$, $p = .226$.

Behavioral Analyses. The total proportion of both cues and targets recalled in the free-recall test and the proportion of targets recalled in the cued-recall test were computed for TBF/R and TBR/R items separately and served as dependent variables in the behavioral analyses. Mean recall rates are displayed in Figure 2.

Free recall performance. Free-recall rates were submitted to a 2 × 2 × 2 mixed analysis of variance (ANOVA) with the between-subjects factors presentation method (list-method, item-method) and memory instruction (remember, forget) and the within-subjects factor item type (TBF/R, TBR/R). This analysis showed no main effects of method or instruction, $F < 1$, but a main effect of item type, $F(1, 108) = 32.43$, $p < .001$, $\eta^2_p = .231$. This effect was further qualified by significant 2-way interactions between item type and method, $F(1, 108) = 6.53$, $p = .012$, $\eta^2_p = .057$, and item type and instruction, $F(1, 108) = 43.71$, $p < .001$, $\eta^2_p = .288$. The interaction between method and instruction and the 3-way interaction were not significant, *F*s < 1.1.

In order to disentangle DF costs and benefits in the present design, we followed up on the significant 2-way interactions by conducting pairwise comparisons between remember and forget instruction groups, separately for each method and item type. For the list-method groups, the free-recall rate of TBF/R items (Figure 2, top panel, left side) was significantly lower in the forget than in the remember group, $F(1, 108) = 6.42$, $p = .013$, $\eta^2_p = .056$. The same was true when comparing the item-method–forget and the item-method–remember groups, $F(1, 108) = 16.723$, $p < .001$, $\eta^2_p = .134$. These findings indicate that DF costs occurred with both methods. Recall of TBR/R items (Figure 2, top panel, right side) was significantly better in the forget than in the remember group for list-method groups, $F(1, 108) = 7.68$, $p = .007$, $\eta^2_p = .066$, and for item-method groups, $F(1, 108) = 4.38$, $p = .039$, $\eta^2_p = .134$. This indicates that DF benefits were also present with both methods. Thus, results from our free-recall test replicate the typical DF costs and benefits.

An alternative approach, typically used in the item-method, to measuring DF effects, without differentiating between costs and benefits, is testing whether recall of TBF-items is reduced relative to TBR-items in the forget groups (see Sahakyan et al., 2014 for a critical discussion of this measure). For the present recall data, we found significant differences between TBF-item and TBR-item recall in the forget groups from both the list-method, $t(27) = 3.84$, $p = .001$, and the item-method, $t(27) = 8.38$, $p < .001$. Thus, this alternative measure also suggests that typical DF effects on free recall occurred for both methods.

Cued recall performance. The 2 × 2 × 2 mixed ANOVA with the between-subjects factors presentation method and memory instruction and the within-subjects factor item type for cued recall rates did not show a main effect of memory instruction, $F < 1$, but main effects of presentation method, $F(1, 108) = 4.87$, $p = .030$, $\eta^2_p = .043$, and item type, $F(1, 108) = 32.43$, $p = .007$, $\eta^2_p = .065$. These effects were further qualified by significant 2-way interactions between item type and method, $F(1, 108) = 24.09$, $p < .001$, $\eta^2_p = .182$, and between item type and instruction, $F(1, 108) = 37.64$, $p < .001$, $\eta^2_p = .258$. The interaction between method and instruction was not significant, $F < 1$, but the 3-way interaction was, $F(1, 108) = 5.67$, $p = .019$, $\eta^2_p = .050$.

We followed up on the significant 2-way and 3-way interactions by conducting the same pairwise comparisons as for the analysis of the free-recall rates. For list-method groups, cued-recall of TBF/R items—indicative of DF costs—did not vary with memory instructions, $F(1, 108) = 1.46$, $p = .230$, $\eta^2_p = .013$. For item-method groups, however, cued recall of TBF/R items was worse with forget than with the remember instructions, $F(1, 108) = 16.723$, $p < .001$, $\eta^2_p = .134$. This is evidence that DF costs in cued recall occurred with the item-method but not with the list-method. For the list-method groups, cued-recall of TBR/R items did not vary with memory instructions, $F(1, 108) = 1.16$, $p = .283$, $\eta^2_p = .011$, For the item-method groups, cued-recall of TBR/R items was marginally improved under forget compared to under remember instructions, $F(1, 108) = 3.56$, $p = .062$, $\eta^2_p = .032$. We also compared TBR- and TBF-item cued-recall performance in the two forget groups in the alternative DF measure that does not differentiate between costs and benefits. For the list-method forget group there was no indication of a DF effect on cued recall, $t < 1$. For the item method forget group, however, this comparison indicated a significant DF effect with higher cued recall of TBR- compared to TBF-items, $t(27) = 10.39$, $p < .001$. Thus, there was no evidence for DF costs and benefits with the list-method in the cued-recall test, whereas the item-method produced reliable DF effects (especially costs) on cued recall.

Summary of behavioral findings. The free-recall data replicate the typical patterns of DF costs and benefits with both the list and the item-method. With the present design, we were also able to differentiate between DF costs and benefits to both free and cued-recall performances in both methods. This differentiation is common for list-method but not for item-method studies. The interesting finding in this regard is that cost and benefits both seem to contribute to item-method DF effects on free recall. The cued-recall data provide preliminary support that the two methods of DF have different cognitive underpinnings as only the item-method produced a DF effect on cued-recall performance. This is in line with other studies reporting no list-method DF effects on retrieval-facilitated tests (i.e., recognition, implicit tests; Basden et al., 1993; Basden & Basden, 1998; E. L. Bjork & Bjork, 1996, 2003; Gottlob & Golding, 2007; Racsmany & Conway, 2006; Zellner & Bäuml, 2006) but reliable item-method DF effects on such tests (Basden et al., 1993; Basden & Basden, 1996; MacLeod, 1999). Nevertheless, our findings have to be interpreted with caution because the preceding free recall may have affected cued recall performance. We will discuss this issue further in the General Discussion section. The fixed order of the tests was necessary for the application of the storage–retrieval MPT model which will be presented next.

Model-based Analyses. The storage-retrieval MPT model for the free-then-cued-recall paradigm (Riefer & Rouder, 1992; Rouder & Batchelder, 1998) was applied to the present data to disentangle storage and retrieval contributions to the observed DF effects on free recall. For the model-based analysis, frequencies of the six critical events (E_1 –E_6; see Introduction) were aggregated separately for each of the eight experimental conditions (see Appendix A). The model applied to each condition contained five parameters (a, r, s, u, and l), resulting in 8 × 5 = 40 parameters in total. With six observable events per condition there were 8 × 5 = 40 free categories in the data. If the number of free parameters is equal to the number of free categories, parameters can be estimated but model fit cannot be tested (cf. Batchelder & Riefer, 1991). Because cued recall immediately followed free recall, which was timed to be 120s in all experimental conditions, loss from memory due to the time delay should be quite low and comparable across conditions. Thus, we set the l-parameter to be equal across all conditions to obtain a testable model with spare degrees of freedom. We used the software *multiTree* (Moshagen, 2010) to fit this restricted model to the data, which yielded a good fit, $G^2(7) =$

3.93, $p = .786$.[8] As expected, the probability of memory loss during the delay between the free and cued recall, was very low (parameter $l = .045$, $SE = .009$).

Directed-forgetting effects on storage (a) and retrieval (r). As we aimed to disentangle storage and retrieval processes of DF, the parameters *a* and *r* were of primary interest. Probability estimates of these two parameters are displayed in Figure 3.

For TBF/R items, *a*-estimates did not differ between the list-method–forget and list-method–remember groups, $\Delta G^2(1) = 2.46$, $p = .116$, but were significantly lower in the item-method–forget than in the item-method–remember group, $\Delta G^2(1) = 28.98$, $p < .001$ (Figure 3, top panel, left side). These results indicate that reduced storage of TBF-items underlies the item-method but not the list-method DF costs observed in recall performance.

For TBR/R items, *a*-estimates were significantly higher in the forget compared to the remember group with both the list-method, $\Delta G^2(1) = 14.05$, $p < .001$, and the item-method, $\Delta G^2(1) = 5.68$, $p = .017$ (Figure 3, top panel, right side). These results indicate that improved storage of TBR-items underlies the observed benefits of both DF methods.

For TBF/R items, *r*-estimates were significantly lower in the forget compared to the remember group with the list-method, $\Delta G^2(1) = 14.37$, $p < .001$, and with the item-method, $\Delta G^2(1) = 12.93$, $p < .001$ (Figure 3, bottom panel, left side). These results imply that retrieval processes underlie not only list-method but also item-method DF costs. This finding is *not* in line with current DF theorizing considering item-method DF costs a consequence of selective-rehearsal of TBR-items only (Basden & Basden, 1998; R. A. Bjork, 1970; MacLeod, 1999; Taylor, 2005).

For TBR/R items, there were no significant *r*-estimate differences between forget and remember groups, with either the list-method, $\Delta G^2(1) = 1.90$, $p = .167$, or the item-method, $\Delta G^2(1) = 1.55$, $p = .212$ (Figure 3, bottom panel, right side). Thus, as expected, retrieval processes did not significantly contribute to the observed DF benefits in either method.

Analogously to the behavior analyses, we compared *a* and *r*-estimate differences between TBF and TBR-items within the forget groups of both methods as an alternative measure of DF effects that does not differentiate costs and benefits. The *a*-estimates for list-method TBF and TBR-items did not differ, $\Delta G^2(1) = 1.64$, $p = .19$, but were significantly lower for item-method TBF-items than for item-method TBR-items, $\Delta G^2(1) = 55.46$, $p < .001$. The *r*-estimates for TBF-items, however, were lower than those for TBR-items, in both the list-method forget group, $\Delta G^2(1) = 24.91$, $p < .001$, and the item-method forget group, $\Delta G^2(1) = 24.81$, $p < .001$. Thus, analogous to the above analyses, there was evidence for storage-based DF in the item-method but not the list-method and evidence for retrieval-based forgetting with both methods with this alternative DF measure. We also formally tested whether the proportional changes in storage and retrieval with item type differed between methods using the

---

[8] A further common restriction for storage–retrieval models is setting the *s* and *u*-parameters to be equal (Riefer & Rouder, 1992). As this restriction would have resulted in a significantly worse fit, $\Delta G^2(8) = 16.51$, $p = .035$, this restriction was not applied here.

parametric-order-constraints function of *multiTree* (Moshagen, 2010).[9] We found significantly weaker forget-cue-induced changes (i.e., TBF vs. TBR) in storage (parameter *a*) in the list-method than in the item method, $\Delta G^2(1) = 31.57$, $p < .001$, but similar forget-cue-induced changes in retrieval (parameter *r*) for both methods, $\Delta G^2(1) = 0.01$, $p < .978$.

Singleton recall (parameters s and u). Parameters *s* and *u* measure singleton recall and reflect cognitive states that are psychologically plausible but assumed to occur rather infrequently. In line with this assumption, *s*-parameter and *u*-parameter estimates were generally rather low (i.e., estimates ranged from .00 to .10). Therefore, and because these parameters—that are crucial for model identification purposes—are considered a cover for nuisance effects (cf. Riefer & Rouder, 1992), we did not further interpret them. For transparency, parameter comparisons between remember and forget groups are displayed in Appendix B. These comparisons indicate that there were some differences in singleton recall, that are, however, not central to the present research question.

Summary of model-based analyses. The model-based results were largely in line with current DF theories. DF benefits were generally mostly reflected by changes in storage. Additionally, there were storage-based DF costs with the item-method but not with the list-method. As predicted by context accounts of DF (Pastötter & Bäuml, 2010; Sahakyan et al., 2014), list-method DF costs were reflected by retrieval changes. However, somewhat surprisingly, item-method DF costs were also partly driven by retrieval processes according to our modeling results. This finding is not in line with current theories that would only predict worse storage of TBF-items due to selective rehearsal of TBR-items (Basden & Basden, 1998; MacLeod, 1999; Taylor, 2005). To our knowledge this is the first behavioral demonstration that item-method DF is also a retrieval phenomenon (but see Nowicka, Jednoróg, Wypych, & Marchewka, 2009; Van Hooff, Whitaker, & Ford, 2009 for evidence from EEG-data that point in the same direction). However, the standard task settings and recall instructions for the item-method that we employed in Experiment 1 did not control for recall-output order: Whereas the typical list-method instructions employed here insured that participants first attempted to recall TBF items, participants in the item-method freely decided which items to recall first. It is thus possible that participants of the item-method forget group tended to start with recalling TBR-items—because they might have been more easily accessible in memory—and that the recall of these TBR-items proactively interfered with later attempts to retrieve TBF-items.[10] Although additional analyses provided no evidence that participants were more likely to recall TBR-items in the first than in the

---

[9] In doing so, we obtained new parameters that reflect proportional changes in storage and retrieval between TBF and TBR-items for both methods that we then compared with each other. This re-parametrization technique allows testing interactions by replacing some parameters with new parameters reflecting proportional changes between the original parameter and another model parameter (e.g., between storage *a* of TBF vs. TBR items) but do neither alter the overall number of parameters nor model fit. Details about this analysis can be obtained from the first author.

[10] We thank an anonymous reviewer for raising this point.

second half of their free-recall output,[11] the storage and retrieval effects of item-method DF may be different if participants deliberately attempt to recall TBF-items first, like in the list method. Therefore, we experimentally manipulated recall-output order of TBF- and TBR-items in Experiment 2. Thereby, this experiment gave us a chance to replicate the novel finding of retrieval effects in item-method DF and to evaluate its robustness.

Experiment 2

The central aim of Experiment 2 was to replicate the somewhat unexpected finding that item-method DF is not only storage- but additionally retrieval-based. Additionally, we wanted to rule out that the (method-typical) unrestricted recall-output order in the item-method free-recall test in Experiment 1 caused these retrieval effect on TBF-items due to output interference.

To this end, we realized three item-method forget groups, in which recall-output order was manipulated. In one forget group (TBF-first–forget), participants were first asked to recall TBF-items only and in a subsequent, second test to recall TBF-items. In a second forget group (TBR-first–forget), participants were first asked to recall TBR only and subsequently to also recall TBF-items. A third forget group was identical to the item-method forget group of Experiment 1, that is, recall was unrestricted (unrestricted–forget). We tested for differences between the three forget groups to examine whether the findings of Experiment 1 in the item-method conditions were specifically due to the unrestricted recall test. We also compared the forget groups with an item-method remember group, that was identical to the one in Experiment 1, to test whether Experiment 1's findings concerning the cognitive processes underlying item-method DF replicates across the different recall-order forget groups.

Method

Participants and Design. One hundred and twelve students from Heidelberg University (94 female; $M_{\text{age}} = 21.28$ years, range: 17 – 31 years) participated for course credit or monetary compensation. There were four experimental groups with $n = 28$ participants each (i.e., TBF-first–forget, TBR-first–forget, unrestricted–forget, and remember). One participant of the TBR-first–forget group did not follow task instructions and was removed from all analysis. All groups were presented with two different item types. TBF/R items received "forget" cues in the forget groups but "remember" cues in the remember group. TBR/R items received "remember" cues in all groups.

Materials and Procedure. Materials and procedure were identical to the ones used for the item-method conditions in Experiment 1. However, we realized three different forget groups and one

---

[11] An ANOVA for the proportion of TBR/R-items recalled with the within-subject factor first versus second half of total recall and the between-subjects factor instruction (remember, forget) showed a significant main effect of group, $F(1, 54) = 57.92$, $p < .001$, $\eta^2_p = .518$, but no main effect of total recall half and no interaction, both Fs < 1. That is, item-method forget participants were generally more likely to recall TBR/R ($M_{\text{first half}} = .86$, $SD = .20$; $M_{\text{second half}} = .83$, $SD = .23$) than item-method remember participants ($M_{\text{first half}} = .50$, $SD = .23$; $M_{\text{second half}} = .49$, $SD = .23$), but as evident from the comparison of TBR/R proportion between output halves, forget participants were not especially likely to start with recalling TBR items.

remember group. To control for recall-output-order effects, we presented participants of the TBF-first–forget and the TBR-first–forget groups with separate recall tests for the two item types. For the first test, we instructed the former group to recall the TBF and the latter to freely recall TBR-items only and then presented them with a cued-recall test for the corresponding items. For the second test, we asked them to only recall the items of the other type, respectively. The unrestricted–forget and the remember group were identical to the item-method groups in the first experiment.

Results and Discussion

All but the one excluded participant of the forget groups with recall-order restrictions followed the recall-order instructions. That is, they predominantly recalled items of the requested type (TBF, TBR) only (TBF-first: *M* = 88%; *SD* = 20; TBF-first: *M* = 87%, *SD* = 21). If participants recalled items of the currently not requested type, we still counted these as accurate for the overall recall performance.

Behavioral Analyses. Most participants of the forget groups reported that they used a forgetting strategy after the experiment (i.e., 97.61% of all participants). Strategy-use frequencies did not differ between forget groups, $\chi(2) = 1.02$, *p* = .599. Free and cued recall rates are displayed in Figure 4.

Free recall performance. A 4 × 2 mixed ANOVA for free recall rates with the between-subjects factor group (TBF-first–forget, TBR-first–forget, unrestricted–forget, remember) and the within-subjects factor item type (TBF, TBR) showed no main effect of group, *F*(1, 107) = 1.00, *p* = .395, η^2_p = .027, but a significant effect of item type, *F*(1, 107) = 61.00, *p* < .001, η^2_p = .363, that was qualified by a significant interaction, *F*(1, 107) = 7.36, *p* < .001, η^2_p = .171.

We followed up on the interaction by testing whether there were differences in the amount of forgetting between forget groups. The 3 × 2 mixed ANOVA for free recall rates with the between-subjects factor recall order (TBF-first, TBR-first, unrestricted) and the within-subjects factor item type (TBF, TBR) showed no main effect of recall order, *F*(1, 80) = 1.10, *p* = .338, η^2_p = .027, and no interaction, *F* < 1. Not surprisingly, there was a significant main effect of item type, *F*(1, 80) = 76.93, *p* < .001, η^2_p = .490, reflecting the DF effect of worse recall of TBF (*M* = .12; *SD* = .10) than TBR-items (*M* = .32; *SD* = .17).

Because the three forget groups did not differ in their recall performance, we combined them and compared them with the remember group in a next step. The 2 × 2 mixed ANOVA with the between-subjects factor memory instruction (forget, remember) and the within-subjects factor item type (TBF/R, TBR/R) showed no main effect of memory instruction, *F*(1, 109) = 1.17, *p* = .282, η^2_p = .011. However, replicating findings from Experiment 1, there was a significant main effect of item type, *F*(1, 109) = 20.10, *p* < .001, η^2_p = .156, that was further qualified by a significant interaction, *F*(1, 109) = 21.94, *p* < .001, η^2_p = .168. Simple-effect comparisons showed that forget participants (*M* = .12; *SD* = .10) recalled fewer of the TBF/R items than remember participants (*M* = .25; *SD* = .15), *F*(1, 109) = 24.10, *p* < .001, η^2_p = .181, indicating substantial DF costs. Further, forget participants (*M*

= .32; *SD* = .17) recalled more TBR/R items than remember participants (*M* = .25; *SD* = .15), indicating DF benefits, *F*(1, 109) = 4.65, *p* = .033, η^2_p = .041.

***Cued recall performance*.** We conducted the same 4 × 2 mixed ANOVA on cued-recall rates and again found no main effect of group, *F* < 1, but a significant effect of item type, *F*(1, 107) = 71.25, *p* < .001, η^2_p = .400, qualified by a significant interaction, *F*(1, 107) = 8.24, *p* < .001, η^2_p = .188.

The follow-up 3 × 2 mixed ANOVA for cued-recall rates in the forget groups with the between-subjects factor recall order (TBF-first, TBR-first, unrestricted) and the within-subjects factor item type (TBF, TBR) showed no main effect of recall order, *F* < 1, and no interaction, *F*(1, 80) = 1.47, *p* = .337, η^2_p = .035. There was a significant main effect of item type, *F*(1, 80) = 80.32, *p* < .001, η^2_p = .501, reflecting the DF effect of worse recall of TBF (*M* = .34; *SD* = .25) than TBR-items (*M* = .57; *SD* = .24).

The 2 × 2 mixed ANOVA with the between-subjects factor memory instruction (forget, remember) and the within-subjects factor item type (TBF/R, TBR/R) showed no main effect of memory instruction, *F*(1, 109) = 1.15, *p* = .286, η^2_p = .010. There was a significant main effect of item type, *F*(1, 109) = 24.72, *p* < .001, η^2_p = .185, that was further qualified by a significant interaction, *F*(1, 109) = 21.08, *p* < .001, η^2_p = .162. Simple-effect comparisons showed significant DF costs in terms of worse recall in forget (*M* = .34; *SD* = .25) compared to remember participants (*M* = .50; *SD* = .20), *F*(1, 109) = 8.73, *p* = .004, η^2_p = .074. There was no evidence for DF benefits in cued-recall, however, *F*(1, 109) = 1.26, *p* = .264, η^2_p = .011.

Summary of behavioral analyses. The behavioral analyses revealed that recall-output order did not affect item-method DF effects. Comparing the forget groups to the remember group, we replicated the item-method results observed in Experiment 1: There were DF costs and benefits in free-recall and DF costs in cued recall. In order to investigate the cognitive underpinnings of the DF effect we again applied the free-then-cued-recall storage–retrieval model to the present data.

Model-based Analyses. Event frequencies are displayed in Appendix A. As in Experiment 1, the *l*-parameter was set to be equal across all conditions. The restricted model had seven degrees of freedom and fit the data well, $G^2(7)$ = 9.65, *p* = .209. The probability of memory loss during the delay between the free and cued recall was very low (parameter *l* = .033, *SE* = .009).

Directed-forgetting effects on storage (a) and retrieval (r). Probability estimates are displayed in Figure 5. Analogously to the behavioral analyses, we first tested for differences between the three forget groups and then compared them with the remember group. For TBF/R-items, *a*-estimates in the two restricted recall groups did not differ, ΔG^2s < 1. However, *a*-estimates in the unrestricted forget group were significantly higher than in both the TBF-first forget group, $\Delta G^2(1)$ = 7.03, *p* = .008, and the TBR-first forget group, $\Delta G^2(1)$ = 5.19, *p* = .02. This finding was unexpected and—because storage processes should not be affected by the restriction of recall-output order—probably represents a sampling error. Importantly, *a*-estimates for TBF/R-items were still significantly lower in all three forget groups than in the remember group ($\Delta G^2_{\text{TBF-first}}(1)$ = 27.11, *p* < .001; $\Delta G^2_{\text{TBR-}}$

$_{\text{first}}(1) = 23.25, p < .001$; $\Delta G^2_{\text{unrestricted}}(1) = 6.68, p = .009$). Thus we replicated Experiment 1's finding of reduced storage of TBF-items underlying item-method DF costs.

For TBR/R-items, *a*-estimates did not differ between forget groups, $\Delta G^2(2) = 1.54, p = .462$, but storage of these items was significantly higher in the forget groups than in the remember group, $\Delta G^2(1) = 4.54, p = .033$. In line with Experiment 1, these results indicate that improved storage of TBR-items underlies the item-method DF benefits.

The *r*-estimates for TBF/R-items did not differ between forget groups, $\Delta G^2(2) = 1.60, p = .449$, but were significantly lower under forget compared to remember instructions, $\Delta G^2(1) = 8.00, p = .004$. Notably, we replicated Experiment 1's finding that item-method DF costs are also partly caused by impoverished retrieval of TBF-items.

There were no significant *r*-estimate differences between forget groups for TBR/R items, $\Delta G^2(2) = 2.18, p = .337$, but retrieval of these items was slightly worse in the remember compared to the forget group, $\Delta G^2(1) = 4.13, p = .042$. This finding may imply that some small parts of the DF benefits observed in Experiment 2 were due to a facilitated retrieval of TBR/R items in the forget groups. However, because this effect was very small and not present in Experiment 1 we refrained from further interpreting this finding.

We also compared *a* and *r*-estimate differences between TBF and TBR-items within each forget group as an alternative DF measure. *A*-estimates were significantly lower for TBF than for TBR-items, all $\Delta G^2(1) > 21.00, p < .001$. *R*-estimates were also lower for TBF than for TBR-items, all $\Delta G^2(1) > 9.81, p < .002$. Thus, both storage and retrieval processes reflected DF effects with this measure.

Singleton recall (parameters s and u). As in Experiment 1, *s*-parameter and *u*-parameter estimates were generally low (i.e., estimates ranged from .01 to .08) and parameters did not vary between forget groups for either item type, all $\Delta G^2(2) < 2.00, p > .377$. However, as in Experiment 1, there were some differences in singleton recall between forget and remember groups that are not central for our research question (see Appendix B).

Summary of behavioral analyses. We replicated the item-method findings of Experiment 1, including the novel finding that item-method DF is partly based on inabilities to retrieve to-be-forgotten information. As *r*-estimates of TBF-items were not affected by recall-order instructions, we also provided evidence that it is unlikely that recall-output order caused the retrieval effects observed in Experiment 1.

General Discussion

The present experiments were designed to re-evaluate differences in the cognitive underpinnings of DF effects between the list-method and the item-method and to further differentiate between storage and retrieval processes involved in cost and benefit effects obtained with the two methods.

In their influential work, Basden et al. (1993) demonstrated that whether a DF effect occurs with a certain method or not strongly depends on the kind of memory test applied. Based on observations that only the item-method but not the list-method produces reliable DF-effects in memory tests that facilitate retrieval (recognition tests and implicit memory tests), they concluded that retrieval-inhibition processes underlie the list-method DF effect whereas selective-rehearsal processes underlie the item-method DF effect. Later research has challenged this assumption by showing that list-method benefits and costs can occur in recognition tests, at least under certain conditions (Benjamin, 2006; Lehman & Malmberg, 2009). The behavioral data of our first experiment are very much in line with the earlier work by Basden and colleagues. We found reliable DF effects with both methods in a free recall test but only an item-method DF effect in a subsequent cued-recall test (replicated in Experiment 2), where retrieval was facilitated through the provision of cue words. Extending Basden et al. (1993), we also included control groups always receiving a “remember” cue instead of the “forget” cue in our study design, which enabled us to differentiate between cost and benefit components of the DF effect. With the free-recall test, we found evidence for both DF costs and benefits with both methods. With the cued-recall test, however, only the item-method produced significant DF effects, mostly attributable to reliable DF costs. The list-method did not produce DF costs or benefits in the cued-recall test. It is important to note that in our experiments the cued-recall test always followed the free-recall test. This constant test order was necessary for applying the multinomial model, but the preceding free recall test might have influenced cued-recall performance. Whereas the model-based analyses take test order into account, the behavioral analyses do not. There is some empirical evidence that free-recall retrieval practice fosters subsequent cued-recall performance to a small but reliable extend compared to a control group with no retrieval practice (Cull, 2000, Experiment 2). Although we see no reason to assume that such retrieval-practice effects would interact with item type, our behavioral cued-recall results should therefore be interpreted with some caution. Nonetheless, the present cued-recall findings are largely in line with Basden et al.’s findings from an (immediate) recognition test: List-method but not item-method DF effects disappear on memory tests facilitating retrieval (Basden et al., 1993). Thus, despite the fact that the free-recall test always preceded the cued-recall test, the present behavioral findings on the cued-recall converge with previous findings and render further support to the assumption that list-method DF seems to be largely retrieval-based, whereas item-method DF seems to be based on encoding and/or maintenance processes.

This pattern of results may be of interest for DF researchers because a differentiation between costs and benefits has rarely been made in the item-method DF literature (R. A. Bjork, 1972; MacLeod, 1999; but see Foster & Sahakyan, 2012). In this regard, the present findings suggest that DF-costs and benefits are dissociable within both methods but that method comparisons that consider costs and benefits separately largely converge with earlier findings where costs and benefits were confounded (Basden et al., 1993).

Most importantly, adding an additional cued-recall test after the free-recall test allowed us to apply the storage-retrieval MPT model to obtain more process pure estimates of the storage and retrieval processes involved in DF effects with both methods. In line with current DF theorizing, the modeling results of Experiment 1 further bolstered the assumption that list-method DF costs are solely driven by hampered retrieval of TBF/R-items after a “forget” compared to after a “remember” cue. Thereby, this finding is in line with any DF account that postulates a retrieval-based forgetting mechanism for this method (i.e., retrieval-inhibition, context-inhibition, and context-change accounts). For the item-method, on the other hand, the model-based analyses revealed that both reduced storage and hampered retrieval of TBF-items after a “forget” compared to after a “remember” cue contribute to the observed DF costs. Whereas the former item-method finding can be explained in terms of classic selective-rehearsal accounts (R. A. Bjork, 1970; Johnson, 1994), the latter finding implies that parts of item-method DF costs are retrieval-mediated. Most importantly, in Experiment 2, we replicated this finding of a retrieval-based DF-cost effect independent of whether forget-group participants first recalled TBR or TBF-items, ruling out that this retrieval effect was caused by output interference due to participants possibly retrieving more TBR-items first when recall order is unrestricted (as in the standard item-method DF setting).

These results challenge the assumption that item-method DF effects occur solely due to selective rehearsal of TBR-items and a concurrent passive forgetting over time of TBF-items (R. A. Bjork, 1970; Eppstein, 1972) but support recent theorizing that inhibitory processing of TBF-items at the retrieval stage contribute to item-method DF (Nowicka et al., 2009; Van Hooff et al., 2009). The exact nature of this inhibition process, however, remains to be investigated. It may well be that the TBF-item itself is inhibited (Geiselman & Bagheri, 1985; Zacks, Radvansky, & Hasher, 1996). Alternatively, it may be that—analogously to context-based forgetting accounts of list-method DF—contextual information of TBF-items is blocked at the retrieval stage resulting in worse memory for these items (cf. Pastötter et al., 2008).

DF benefits in both methods were due to improved storage of TBR-items after receiving a “forget” cue for TBR-items (which became evident from the comparison with a “remember” control group), whereas retrieval processes played at most a minor role for DF benefits in both methods. Regarding the item-method, these findings are well in line with accounts proposing a selective rehearsal of TBR-items. Interestingly, DF benefits seem to be exclusively mediated by storage processes in the list-method as well, which may be due to a cue-initiated reset of encoding (Pastötter & Bäuml, 2010) or disengagement from rehearsal of TBF-items (Sahakyan & Delaney, 2010). Some researchers assume that the presentation of the list-method “forget” cue reduces proactive interference from the previous learned TBF-items onto the TBR-items. In as far as proactive interference is considered a purely retrieval-mediated phenomenon (Baddeley, 1990; Wixted & Rohrer, 1993), our results speak against a role of proactive interference for list-method DF benefits. Recent research, however, suggest that proactive interference also affects storage processes (Kliegl, Pastötter, & Bäuml, 2015; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011). Therefore, we would not conclude that proactive interference plays no role for DF benefits. However, if a reduction in proactive interference contributes to list-method DF benefits it will most likely do so by reducing item competition during storage (Pastötter et al., 2011; Sahakyan & Goodmon, 2007). Notably, we used a standard list-method paradigm in the present research where the TBF-list is tested prior to the TBF-list (Delaney & Sahakyan, 2007; Foster & Sahakyan, 2011; Pastötter & Bäuml, 2007; see also Sahakyan et al., 2014). Recent research showed, however, that DF benefits are more pronounced when forget participants are tested on the TBF-list only (Pastötter et al., 2012). Future applications of the storage–retrieval model should test whether such paradigm modifications will change the processes underlying DF benefits.

Importantly, as becomes evident in the preceding discussion, the model-based approach does not allow distinguishing between different processes that are engaged at the same processing (i.e., storage or retrieval) stage and one may argue that this limits the usefulness of this approach in the DF domain. Even though the question of whether DF is a storage or retrieval phenomenon is still not completely solved (Sahakyan et al., 2014; Sheard & MacLeod, 2005) there are also ongoing debates regarding the exact nature of retrieval-mediated DF. Specifically, an open question is whether retrieval processes of DF are manifestations of the inhibition of the TBF-items (R. A. Bjork, 1989), inhibition of the study context of TBF-items (Pastötter et al., 2008), or of an active context change initialized by the “forget” cue (Sahakyan & Kelley, 2002). Obviously, the storage–retrieval model does not allow measuring these retrieval-based processes separately. Therefore, thoughtful experimentation is necessary, to solve this question. However, we nonetheless suggest that the model-based approach can be useful even when investigating (and differentiating between) different processes engaged at the same memory stage because—by controlling for contributions of processes at stages other than the phenomenon of interest—the model provides a more precise and process-pure measure of the stage of the effect. We therefore suggest combining our model-based approach with other experimental approaches to processes dissociation in future DF research in order to achieve a more complete understanding of DF processes.

In sum, the present data replicate the basic findings of previous DF-method comparisons by Basden et al. (1993) and the model-based analyses further bolster some of the conclusions regarding the role of storage and retrieval processes in DF effects drawn by these authors. Further, the present results extend previous work by showing that the contributions of storage-mediated relative to retrieval-mediated forgetting to memory costs as well as benefits are also method-dependent. Finally, our results suggest that investigation regarding retrieval processes of DF should not only focus on the list-method but also on the item-method because (similar or different) retrieval processes contribute to DF costs in both methods.

References

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In C. M. MacLeod & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 159–190). Tokyo: Springer-Verlag.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn & Bacon.

Basden, B. H., & Basden, D. R. (1996). Directed forgetting : Further comparisons of the item and list methods. *Memory*, *4*(6), 633–653. doi:10.1080/096582196388825

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139–172). Mahwah, NJ: Erlbaum.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603–616. doi:10.1037/0278-7393.19.3.603

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. doi:10.3758/BF03210812

Bäuml, K.-H. (2008). Inhibitory processes. In H. L. Roediger III (Ed.), *Cognitive psychology of memory* (pp. 195–220). Oxford: Elsevier.

Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic Bulletin & Review*, *13*(5), 831–836. doi:10.3758/BF03194005

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition*, *5*(1), 176–196. doi:10.1006/ccog.1996.0011

Bjork, E. L., & Bjork, R. A. (2003). Intentional forgetting can increase, not decrease, residual influences of to-be-forgotten information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 524–531. doi:10.1037/0278-7393.29.4.524

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*(3), 255–268. doi:10.1016/S0022-5371(70)80059-7

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Washington, DC: Winston.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, *10*(2), 55–56.

Bjork, R. A., & Woodward, A. E. (1973). Directed forgetting of individual words in free recall. *Journal of Experimental Psychology*, *99*(1), 22–27. doi:10.1037/h0034757

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215–235. doi:10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Delaney, P. F., & Sahakyan, L. (2007). Unexpected costs of high working memory capacity following directed forgetting and contextual change manipulations. *Memory & Cognition*, *33*(5), 1074–1082.

Dempster, F. N., & Brainerd, C. J. (Eds.). (1995). *Interference and inhibition in cognition*. San Diego: Academic Press.

Eppstein, W. (1972). Mechanisms of directed forgetting. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 147–191). New York: Academic Press.

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*(3), 108–124. doi:10.1027/0044-3409.217.3.108

Foster, N. L., & Sahakyan, L. (2011). The role of forget-cue salience in list-method directed forgetting. *Memory*, *19*(1), 110–117. doi:10.1080/09658211.2010.537665

Foster, N. L., & Sahakyan, L. (2012). Metacognition influences item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1309–1324. doi:10.1037/a0027868

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition*, *13*(1), 57–62. doi:10.3758/BF03198444

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

Golding, J. M., & MacLeod, C. M. (1998). *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ US: Erlbaum.

Gottlob, L. R., & Golding, J. M. (2007). Directed forgetting in the list method affects recognition memory for source. *The Quarterly Journal of Experimental Psychology*, *60*(11), 1524–1539. doi:10.1080/17470210601100506

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB-eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20. doi:10.1026/0033-3042/a000029

Hourihan, K. L., & Taylor, T. L. (2006). Cease remembering: Control processes in directed forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1354–1365. doi:10.1037/0096-1523.32.6.1354

Johnson, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin*, *116*(2), 274–292. doi:10.1037/0033-2909.116.2.274

Kliegl, O., Pastötter, B., & Bäuml, K.-H. (2015). The contribution of encoding and retrieval processes to proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000096

Koppel, R. H., & Storm, B. C. (2012). Unblocking memory through directed forgetting. *Journal of Cognitive Psychology*, *24*(8), 901–907. doi:10.1080/20445911.2012.716822

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding , maintenance , and retrieval processes in the lag effect : A multinomial processing tree analysis. *Memory*, *20*(1), 37–48. doi:10.1080/09658211.2011.631550

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 970–88. doi:10.1037/a0015728

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*(1), 123–129. doi:10.3758/BF03210819

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. doi:10.3758/BRM.42.1.42

Nowicka, A., Jednoróg, K., Wypych, M., & Marchewka, A. (2009). Reversed old/new effect for intentionally forgotten words: An ERP study of directed forgetting. *International Journal of Psychophysiology*, *71*(2), 97–102. doi:10.1016/j.ijpsycho.2008.06.009

Pastötter, B., & Bäuml, K.-H. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 977–982. doi:10.1037/0278-7393.33.5.977

Pastötter, B., & Bäuml, K.-H. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 54–65. doi:10.1037/a0017406

Pastötter, B., Bäuml, K.-H., & Hanslmayer, S. (2008). Oscillatory brain activity before and after an internal context change-Evidence for a reset of encoding processes. *NeuroImage*, *43*(1), 173–181. doi:10.1016/j.neuroimage.2008.07.005

Pastötter, B., Kliegl, O., & Bäuml, K.-H. (2012). List-method directed forgetting: The forget cue improves both encoding and retrieval of postcue information. *Memory & Cognition*, *40*(6), 861–873. doi:10.3758/s13421-012-0206-4

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 287–297. doi:10.1037/a0021801

Racsmany, M., & Conway, M. A. (2006). Episodic inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 44–57. doi:10.1037/0278-7393.32.1.44

Reitman, W., Malin, J. T., Bjork, R. A., & Higman, B. (1973). Strategy control and directed forgetting. *Journal of Verbal Learning and Verbal Behavior*, *12*(2), 140–149. doi:10.1016/S0022-5371(73)80003-9

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339. doi:10.1037/0033-295X.95.3.318

Riefer, D. M., & Batchelder, W. H. (1991a). Age differences in storage and retrieval: A multonomial modeling analysis. *Bulletin of the Psychonomic Society*, *29*(5), 415–418.

Riefer, D. M., & Batchelder, W. H. (1991b). Statistical inference for multinomial tree models. In J.-C. Falmagne & J.-P. Doignon (Eds.), *Mathematical psychology: Current developments* (pp. 313–336). Berlin: Springer-Verlag.

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*(2), 184–201. doi:10.1037/1040-3590.14.2.184

Riefer, D. M., & LaMay, M. L. (1998). Memory for common and bizarre stimuli: A storage-retrieval analysis. *Psychonomic Bulletin & Review*, *5*(2), 312–317. doi:10.3758/BF03212957

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*(6), 601–611. doi:10.3758/BF03202710

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 195–225). New York: Psychology Press.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*(1), 63–77. doi:10.1037/h0031185

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, *9*(1), 99–105. doi:10.1016/S0022-5371(70)80015-9

Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language*, *48*(1), 195–206. doi:10.1016/S0749-596X(02)00524-7

Sahakyan, L., & Delaney, P. F. (2005). Directed forgetting in incidental learning and recognition testing: Support for a two-factor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 789–801. doi:10.1037/0278-7393.31.4.789

Sahakyan, L., & Delaney, P. F. (2010). Item-specific encoding produces an additional benefit of directed forgetting: Evidence from intrusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1346–1354. doi:10.1037/a0020121

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2014). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 59, pp. 131–190). Elsevier Inc.: Academic Press. doi:10.1016/b978-0-12-407187-2.00004-6

Sahakyan, L., & Foster, N. L. (2009). Intentional forgetting of actions: Comparison of list-method and item-method directed forgetting. *Journal of Memory and Language*, *61*(1), 134–152. doi:10.1016/j.jml.2009.02.006

Sahakyan, L., & Goodmon, L. B. (2007). The influence of directional associations on directed forgetting and interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1035–1049. doi:10.1037/0278-7393.33.6.1035

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1064–1072. doi:10.1037//0278-7393.28.6.1064

Sahakyan, L., & Smith, J. R. (2014). “A long time ago, in a context far, far away”: Retrospective time estimates and internal context change. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 86–93. doi:10.1037/a0034250

Sheard, E. D., & MacLeod, C. M. (2005). List method directed forgetting: Return of the selective rehearsal account. In N. Ohta, C. M. MacLeod, & B. Uttl (Eds.), *Dynamic cognitive processes*.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(5), 460–471. doi:10.1037/0278-7393.5.5.460

Taylor, T. L. (2005). Inhibition of return following instructions to remember and forget. *The Quarterly Journal of Experimental Psychology*, *58*(4), 613–629. doi:10.1080/02724980443000115

Van Hooff, J. C., Whitaker, T. A., & Ford, R. M. (2009). Directed forgetting in direct and indirect tests of memory: Seeking evidence of retrieval inhibition using electrophysiological measures. *Brain and Cognition*, *71*(2), 153–164. doi:10.1016/j.bandc.2009.05.001

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1024–1039. doi:10.1037/0278-7393.19.5.1024

Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology*, *89*(1), 109–116. doi:10.1037/h0031188

Zacks, R. T., Radvansky, G. A., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 143–156.

Zellner, M., & Bäuml, K.-H. (2006). Inhibitory deficits in older adults: List-method directed forgetting revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 290–300. doi:10.1037/0278-7393.32.3.290

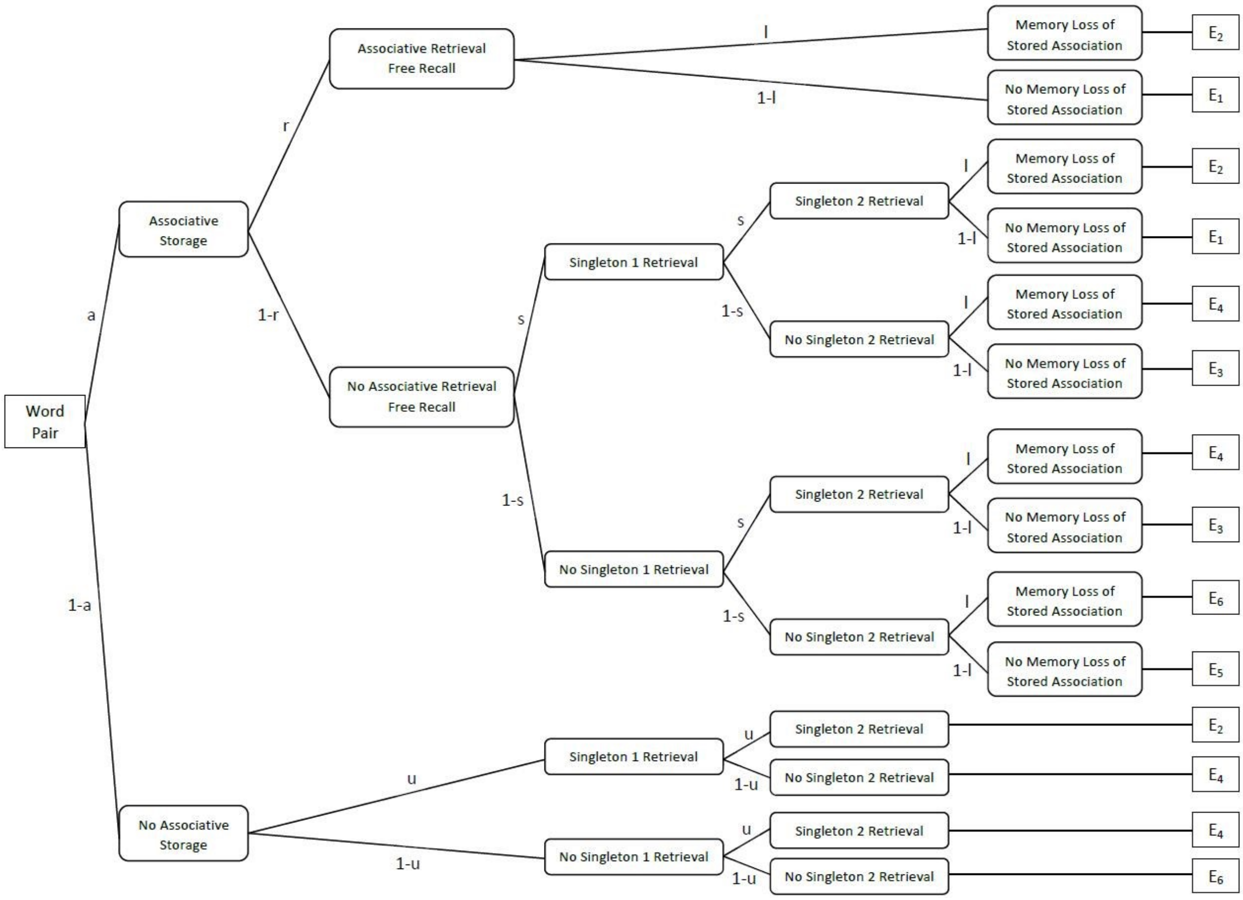


Figure 1. Multinomial processing-tree model (MPT) for a free-then-cued-recall paradigm, to separate storage and retrieval processes (adopted from Rouder & Batchelder, 1998). The processing tree represents the different latent cognitive states that lead to six observable recall events: E_1 = successful free recall of the complete pair, successful cued recall; E_2 = successful free recall of the complete pair, failed cued recall; E_3 = successful free recall of a single item from a pair (singleton), successful cued recall; E_4 = successful free recall of a singleton, failed cued recall; E_5 = failed free recall, successful cued recall; E_6 = failed free recall, failed cued recall. The rounded rectangles represent the latent cognitive states with the transition probabilities between the states being described by the model parameters: *a* = probability of associative storage during study; *r* = probability of associative retrieval during recall; *s* = probability of singleton retrieval of associatively stored items during recall; *u* = probability of singleton retrieval of individually stored items during recall; *l* = loss from memory due to the time delay between free and cued recall. Parameter *l* was restricted to be equal across all conditions to render the model testable.

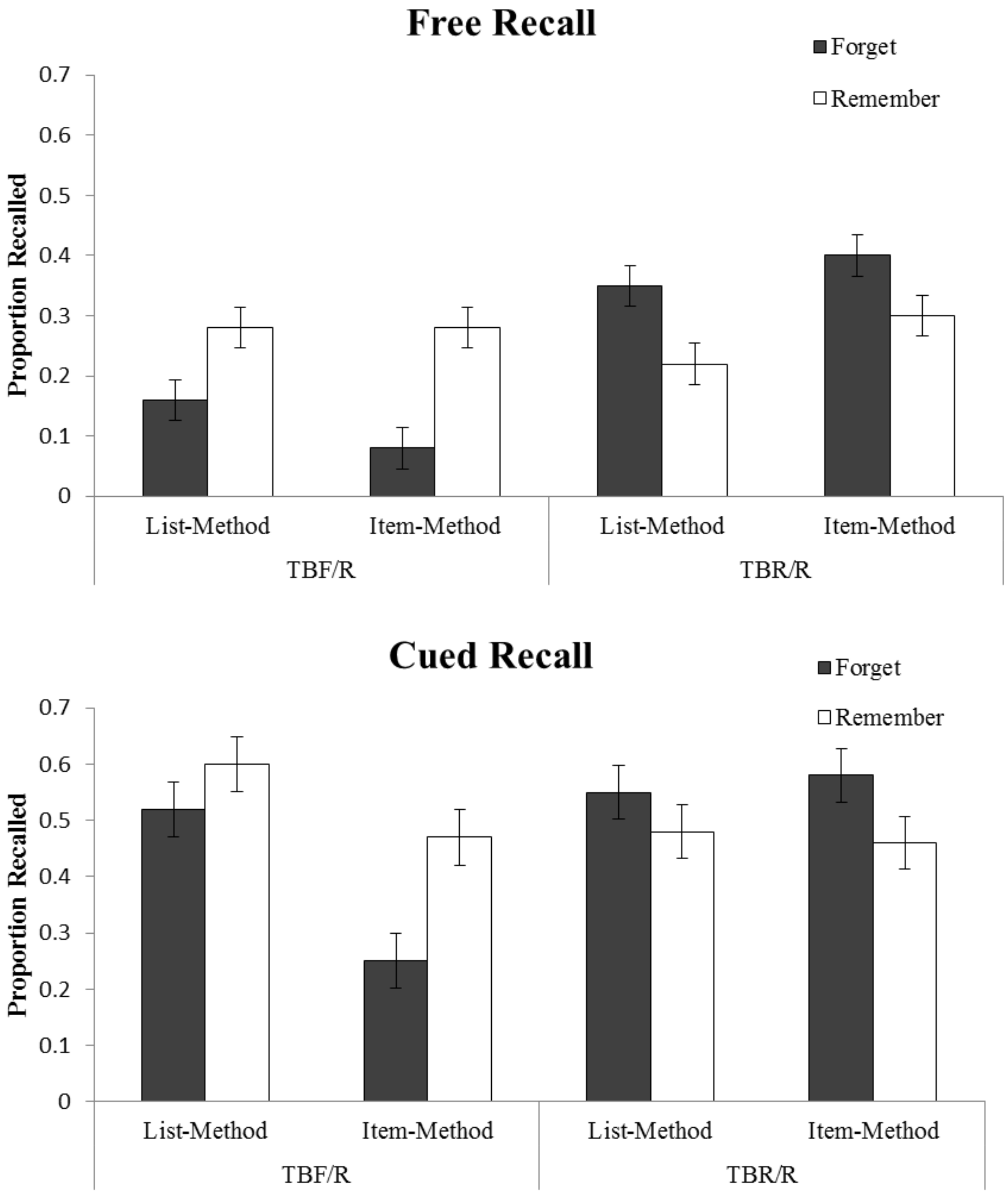


Figure 2. Mean proportion of TBF/R item and TBR/R item recall as a function of presentation method (list-method, item-method) and memory instruction (forget, remember) for free recall (top panel) and cued recall (bottom panel) in Experiment 1. TBF/R items were items for which participants in the forget groups received "forget" cues, whereas participants in the remember groups received "remember" cues. For TBR/R items, all participants received "remember" cues. Error bars represent standard errors of the mean.

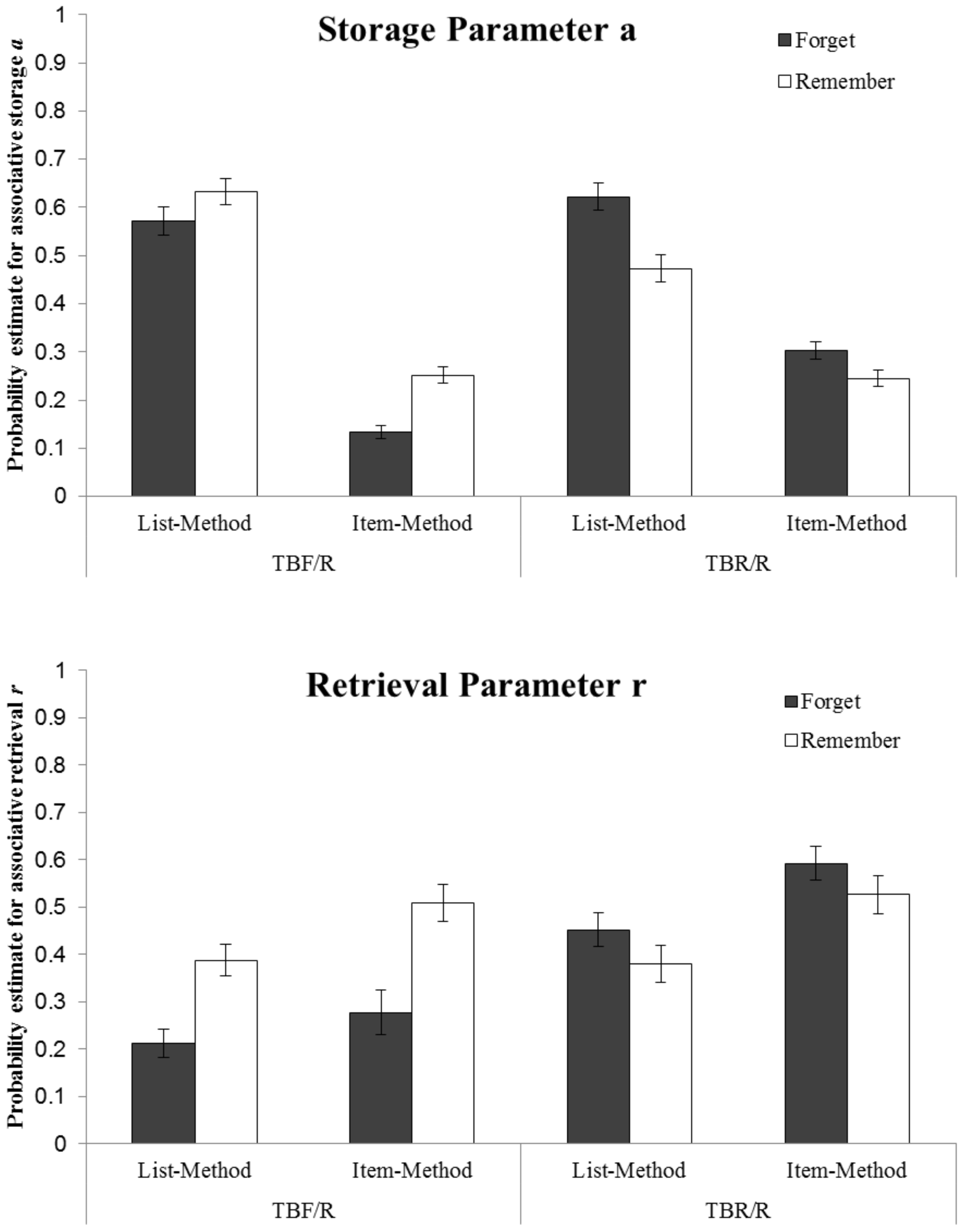
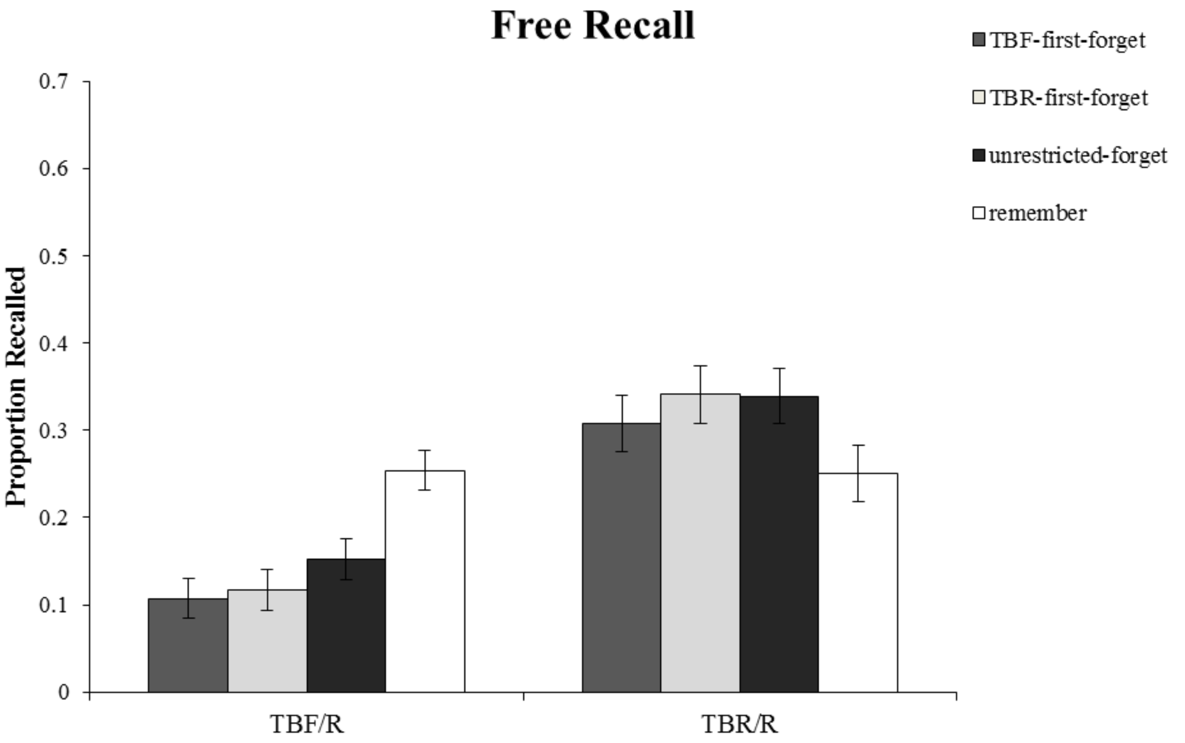


Figure 3. Parameter estimates for the probability of associative storage *a* (top panel), and for the probability of associative retrieval *r* (bottom panel) as a function of item type (TBF/R, TBR/R), presentation method (list-method, item-method), and memory instruction (forget, remember) in Experiment 1. TBF/R items were items for which participants in the forget groups received “forget” cues whereas participants in the remember groups received “remember” cues. For TBR/R items, all participants received “remember” cues. Error bars represent standard errrors.



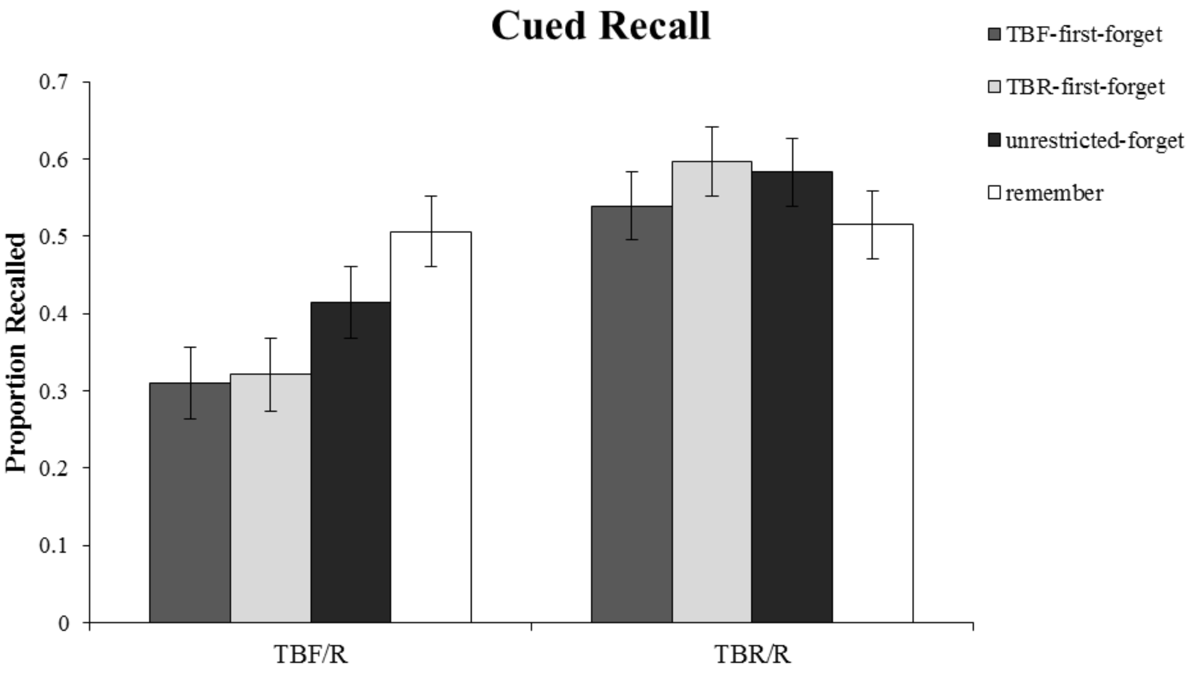
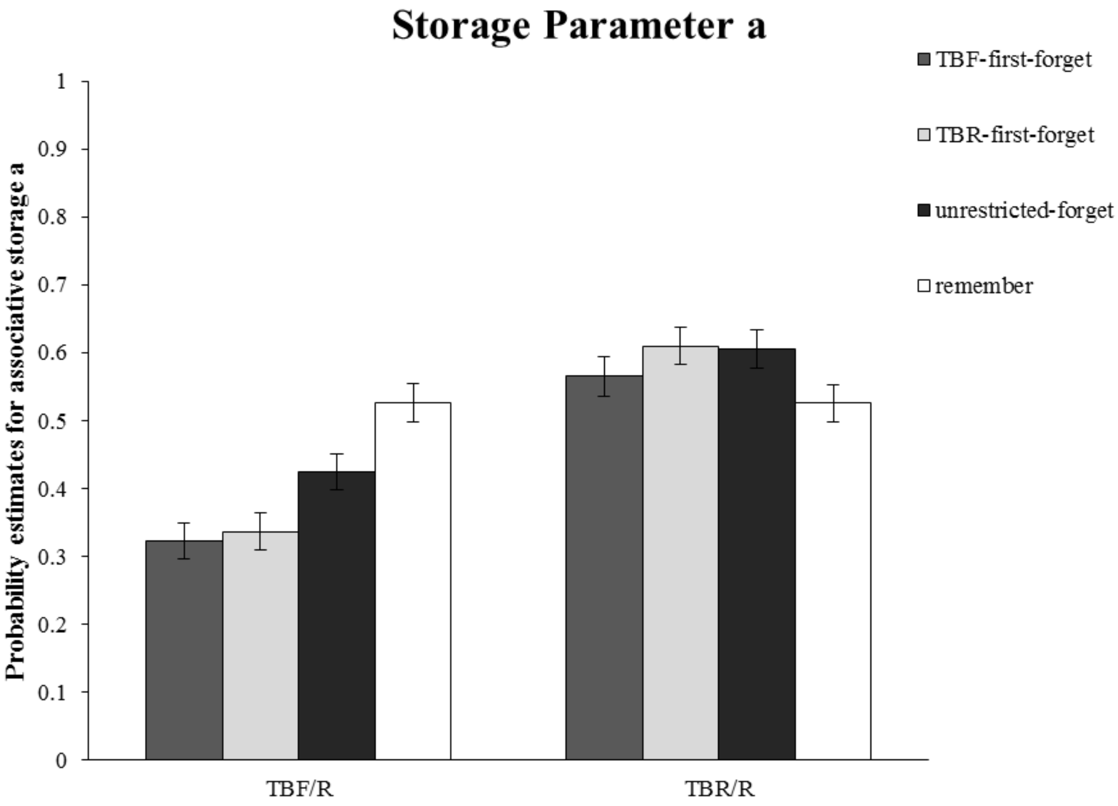


Figure 4. Mean proportion of TBF/R item and TBR/R item recall as a function of memory group (TBF-first–forget, TBR-first–forget, unrestricted–forget, remember) for free recall (top panel) and cued recall (bottom panel) in Experiment 2. TBF/R items were items for which participants in the forget groups received “forget” cues, whereas participants in the remember group received

“remember” cues. For TBR/R items, all participants received “remember” cues. Error bars represent standard errors of the mean.


Storage Parameter a
Probability estimates for associative storage a
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
TBF/R
TBR/R
TBF-first-forget
TBR-first-forget
unrestricted-forget
remember

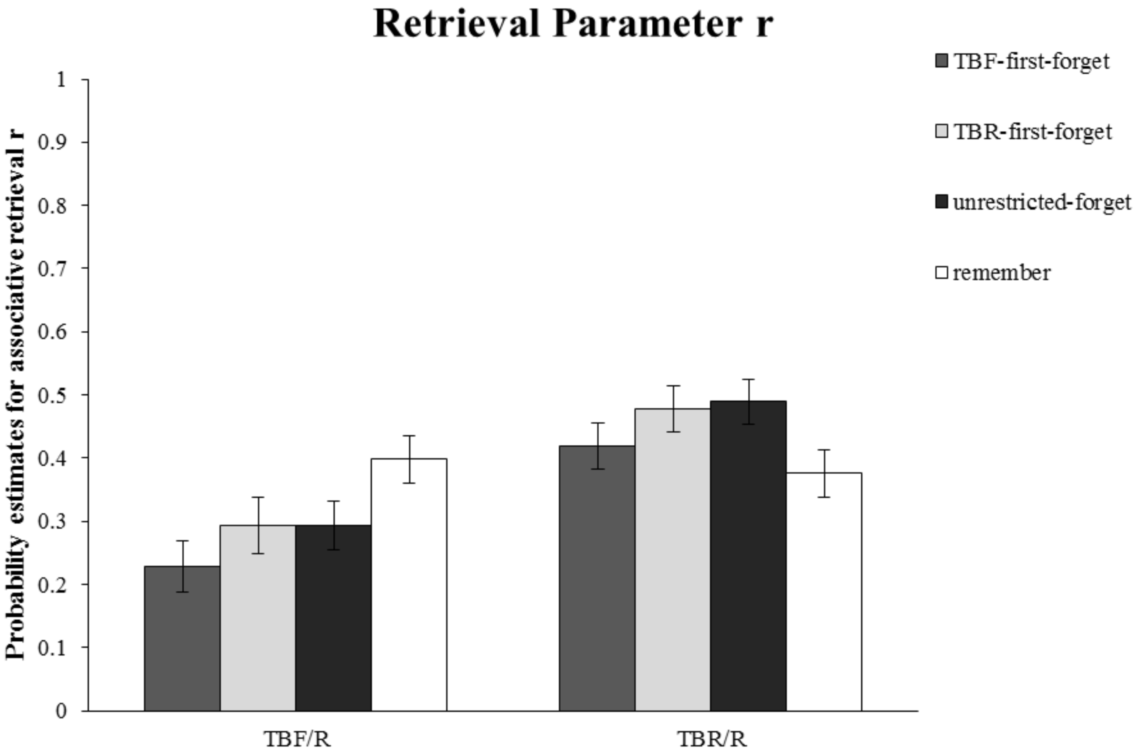

Retrieval Parameter r
Probability estimates for associative retrieval r
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
TBF/R
TBR/R
TBF-first-forget
TBR-first-forget
unrestricted-forget
remember

Figure 5. Parameter estimates for the probability of associative storage *a* (top panel), and for the probability of associative retrieval *r* (bottom panel) as a function of item type (TBF/R, TBR/R) and memory group (TBF-first–forget, TBR-first–forget, unrestricted–forget, remember) in Experiment 2. TBF/R items were items for which participants in the forget groups received “forget” cues whereas participants in the remember group received “remember” cues. For TBR/R items, all participants received “remember” cues. Error bars represent standard errrors.

Appendix A

Frequencies of all recall events by Experiment.

Condition	E_1	E_2	E_3	E_4	E_5	E_6
Experiment 1						
List-method forget instruction						
TBF/R items	38	4	8	136	17	133
TBR/R items	92	5	21	87	25	106
List-method remember instruction						
TBF/R items	79	4	11	113	10	119
TBR/R items	58	3	9	85	19	162
Item-method forget instruction						
TBF/R items	23	2	3	59	3	582
TBR/R items	118	3	8	71	24	448
Item-method remember instruction						
TBF/R items	82	4	4	75	13	494
TBR/R items	81	6	7	67	25	486
Experiment 2						
TBF-first forget instruction						
TBF/R items	23	2	2	78	13	214
TBR/R items	74	9	14	92	23	123
TBR-first forget instruction						
TBF/R items	30	2	2	72	8	208
TBR/R items	93	3	12	88	19	111
Unrestricted forget instruction						
TBF/R items	41	1	6	91	9	187
TBR/R items	97	2	9	90	17	117
Remember instruction						
TBF/R items	69	1	12	89	14	146
TBR/R items	66	2	14	93	21	143

Note. E_1 = both items freely recalled, correct cued recall; E_2 = both items freely recalled, incorrect cued recall; E_3 = one item freely recalled, correct cued recall; E_4 = neither item freely recalled, correct cued recall; E_5 = one item freely recalled, incorrect cued recall; E_6 = neither item freely recalled, incorrect cued recall. TBF/R items were items for which participants in the forget groups received “forget” cues whereas participants in the remember groups received “remember” cues. For TBR/R items, all participants received “remember” cues. Participants of the TBF-first group had to recall TBF-items first; participants of the TBR-first group had to recall TBR-items first.

Appendix B

Comparisons of parameters s *and* u *(singleton recall) between forget and remember groups in Experiments 1 and 2.*

	Forget		**Remember**			
	M	*SE*	*M*	*SE*	ΔG^2	*p*
Experiment 1						
***S*-parameter**						
List-Method						
TBF/R Items	.028	.010	.046	.013	1.118	.290
TBR/R Items	.107	.023	.050	.016	3.994	.045
Item-Method						
TBF/R Items	.024	.014	.025	.012	0.003	.951
TBR/R Items	.053	.018	.049	.018	0.019	.888
***U*-parameter**						
List-Method						
TBF/R Items	.064	.015	.040	.013	1.369	.241
TBR/R Items	.103	.020	.055	.012	4.226	.039
Item-Method						
TBF/R Items	.002	.001	.012	.003	8.243	.004
TBR/R Items	.025	.005	.025	.005	0.002	.959
Experiment 2						
***S*-parameter**						
TBF/R Items	.020	.006	.063	.018	6.912	.008
TBR/R Items	.060	.010	.070	.018	0.197	.656
***U*-parameter**						
TBF/R Items	.024	.004	.044	.011	2.944	.086
TBR/R Items	.079	.010	.066	.014	0.511	.474

Note. All comparisons with 1 df. TBF/R items were items for which participants in the forget groups received TBF cues whereas participants in the remember groups received TBR cues. For TBR/R items, all participants received TBR cues.

Appendix A2 – Article 2

Note:

This is a manuscript that is currently submitted to the journal *Psychological Research* [Publisher: Springer]

Retrieval-Reliant Mechanisms of Item-Method Directed Forgetting: The Effect of Semantic Cues

Ivan Marevic and Jan Rummel
Heidelberg University

Author Note

Ivan Marevic & Jan Rummel, Department of Psychology, Heidelberg University, Heidelberg, Germany.

Correspondence concerning this article should be addressed to Ivan Marevic, Department of Psychology, Heidelberg University, Hauptstrasse 47-51, D-69117 Heidelberg, Germany. E-mail: ivan.marevic@psychologie.uni-heidelberg.de

Abstract

Item-method directed forgetting is widely considered a storage phenomenon. However, by applying a multinomial model, which separates storage and retrieval processes, Rummel, Marevic, and Kuhlmann (2016) recently showed that item-method directed forgetting is reflected by changes in both storage and retrieval processes. The current investigation demonstrates that supposedly intentionally forgotten information in item-method forgetting can still be retrieved to some extent when semantic cuing facilitates retrieval of this information. Participants studied word-pairs, with some pairs being followed by a “forget” and others by a “remember” instruction. A subset of items was semantically related, that is, they shared the same superordinate category. In Experiment 1, a sub-portion of to-be-forgotten items was semantically related and less forgetting occurred selectively for these items when the category was reinstated during test. This finding was replicated and extended to reinstatement effects for to-be-remembered items in Experiment 2. The application of the storage–retrieval model confirmed that providing a category cue enhances retrieval processes for to-be-forgotten as well as to-be-remembered information. The results are discussed in the light of existing theories of DF.

Keywords: directed forgetting; item-method; context reinstatement, multinomial modeling; storage–retrieval model

Retrieval-Reliant Mechanisms of Item-Method Directed Forgetting: The Effect of Semantic Cues

Intentional forgetting is an adaptive memory process that prevents outdated or irrelevant information from interfering with relevant information (Golding & MacLeod, 1998). But what are the cognitive mechanisms that enable us to willingly forget previously learned information and is this information actually dismissed from memory or rather not readily accessible any longer? In the present research, we tackle these questions while focusing on intentional forgetting of information that was learned intermixed with to-be-remembered material.

In the laboratory, intentional forgetting can be studied within the list-method or the item-method directed forgetting (DF) paradigm. List-method participants study two lists of items and are instructed after the first list to forget the items they just learned (Bäuml, 2008; R. A. Bjork, 1989; Pastötter, Tempel, & Bäuml, 2017; Sahakyan, Delaney, Foster, & Abushanab, 2014). Item-method participants face a somewhat different situation, that is, during study every single item is followed by either a “forget” or a “remember” instruction varying from trial to trial (Basden & Basden, 1998; Fawcett & Taylor, 2010). Both methods produce reliable DF effects that are characterized by reduced recall of to-be-forgotten (TBF) compared to to-be-remembered (TBR) items (see Basden, Basden, & Gargano, 1993; Basden & Basden, 1996, 1998; Rummel, Marevic, & Kuhlmann, 2016 for a comparison of methods) but the present article focuses on the cognitive underpinnings of item-method DF.

Current theories largely consider item-method DF to occur due to a selective-rehearsal strategy (Basden et al., 1993; R. A. Bjork, 1972; MacLeod, 1999). That is, as a forget or a remember cue directly follows each item; TBF-items can be spared from rehearsal immediately, whereas TBR-items are maintained and rehearsed throughout the study episode. Such a strategy would result in the typical behavioral pattern of better memory for TBR than for TBF-items (aka a DF effect). Although the selective-rehearsal explanation can account for the item-method DF effect, others have argued that inhibitory and shifting processes are additionally recruited following the presentation of a forget cue (Geiselman & Bagheri, 1985; Zacks, Radvansky, & Hasher, 1996) or that a strategic withdrawal of attention from the TBF-items is initiated by the forget cue (Thompson, Hamm, & Taylor, 2014). That is, due to the presentation of a forget cue, executive attention is directed away from the previously encoded TBF-item. Such attention withdrawal is assumed to be an active process (Wylie, Foxe, & Taylor, 2008) in contrast to the selective rehearsal idea of passively dropping TBF-items from the rehearsal set (Lee, 2012). In a series of experiments, Taylor and colleagues found slower reaction times to secondary probes that were presented right after the memory cues as well as a more pronounced inhibition of return (IOR) effect following TBF compared to TBR-items (Fawcett & Taylor, 2008, 2010, 2012; Hourihan & Taylor, 2006; Taylor, 2005; Thompson et al., 2014; Wylie et al., 2008). These studies provide strong evidence that attention is actively withdrawn after the presentation of a forget cue but this does not necessarily imply that TBF-items are actively inhibited (Fawcett & Taylor, 2010).

Therefore, the question arises what the consequences of item-method DF on a process level are. Unlike list-method DF, item-method DF effects are present not only in free recall tests but also in other memory tests where retrieval is facilitated (e.g., cued recall, recognition, and implicit tests) (Basden et al., 1993; see also MacLeod, 1998). These findings suggest that item-method DF is a storage phenomenon because when a memory phenomenon is present on memory tests that rely on different levels of retrieval strength—such as free recall, cued recall, and recognition—the effect is assumed to be driven by storage rather than retrieval processes (Hanley & Morris, 1987; Hogan & Kintsch, 1971). However, there is recent model-based evidence that hampered retrieval of TBF-items contributes to the item-method DF effect (Rummel et al., 2016). In this study, we applied a multinomial model that measures storage and retrieval processes independently (Rouder & Batchelder, 1998) to item-method DF data and found that item-method DF was reflected by both storage and retrieval changes. The notion that storage changes cause item-method DF is in line with earlier findings that item-method DF occurs across memory test situations independent of their retrieval demands (Basden et al., 1993), but the involvement of additional retrieval processes in item-method DF challenges the assumption that storage processes are the only mechanism underlying this effect (see also Nowicka, Jednoróg, Wypych, & Marchewka, 2009; Van Hooff, Whitaker, & Ford, 2009). In a nutshell, the model-based findings suggest that whatever processing people mentally engage in to forget TBF-items in an item-method paradigm does not only lead to enhanced storage of TBR relative to the TBF-items but also to reduced retrieval of TBF-items relative to TBR-items.

The purpose of the present study is to identify and evaluate candidate retrieval-based mechanisms that contribute to item-method DF. Recent findings from item-method DF suggest that a loss of item-specific features could hamper the retrieval of TBF-items in the item-method. First evidence supporting this idea comes from studies in which item context during learning was reinstated at the final memory test (Burgess, Hockley, & Hourihan, 2017; Hourihan, Goldberg, & Taylor, 2007). In a study by Hourihan et al. (2007), participants studied items at varying screen locations that were followed by either forget or remember cues. At test, the studied items (intermixed with distractors) were presented at the same or at different locations as during study and participants made old/new recognition judgements. Results showed that the DF effect in recognition memory was eliminated when the location at test and at study matched, whereas a reliable DF effect was found when locations differed. Interestingly, only TBF but not TBR-items benefitted from the spatial-context reinstatement. Hourihan et al. (2007) argue that "elaborative encoding of TBR words makes location repetition a relatively weak cue compared to the more strongly encoded features that contribute to recollective memory (e.g., lexical and semantic information)" (*p*. 96). On the other hand, Burgess et al. (2017), found both TBF and TBR-items to profit from context reinstatement. These authors reinstated the study background for TBF and TBR-items during a recognition test and found enhanced hit and false alarm rates for TBF as well as TBR-items for which context was reinstated compared to items for which context was not reinstated. Notably, however, in both previous investigations of context-reinstatement effects on item-method DF, TBF-item recognition always profited from context reinstatement. Building on these findings, in the present study, we aimed to investigate whether the reinstatement of semantic item features can cause a similar release from directed-forgetting as context reinstatement and whether the expected beneficial effects of semantic reinstatement were limited to TBF-items or would also occur for TBR-items. Additionally, we intended to extend previous research on item-feature reinstatement effects on directed forgetting by using memory tests that more strongly rely on recollection (i.e., free and cued recall). To this end, in Experiment 1 we presented one group of participants with a category that was shared by some TBF-items to test whether this group would show reduced forgetting compared to a group that did not receive the category cue. In Experiment 2, we aimed to replicate findings from Experiment 1 and extend them to the investigation of context-reinstatement effects on TBR-items.

In both experiments we used word-pairs rather than single words in a free-then-cued-recall paradigm because we intended to disentangle storage and retrieval-mediated forgetting by applying the storage–retrieval multinomial model (Rouder & Batchelder, 1998; see also Marevic, Arnold, & Rummel, 2017; Rummel et al., 2016 for applications to DF). The storage–retrieval model enabled us to arrive at process-pure estimates of storage and retrieval and thus to investigate context reinstatement effects at the retrieval stage without confounds resulting from other cognitive processes.

The model is depicted in Figure 4 (see also Rummel et al., 2016) and further explained in Appendix A. Generally, this model can be applied to categorical data from a free and a subsequent cued recall test that provide six possible behavioral events, that is, (E_1) successful free recall of the complete pair and successful cued recall; (E_2) successful free recall of the complete pair but failed cued recall; (E_3) successful free recall of a single item from a pair (singleton) and successful cued recall; (E_4) successful free recall of a singleton but failed cued recall; (E_5) failed free recall but successful cued recall; (E_6) failed free recall and failed cued recall. The outcomes E_1-E_6 are modelled as a function of five latent cognitive states which are reflected by the five model parameters. As parameters reflect probabilities of cognitive states, parameter estimates lie between 1 and 0. The *storage parameter a* indicates the probability of associative storage of an item pair (i.e., of both items and their association) and its maintenance in memory until test. The *retrieval parameter r* indicates the probability of successful retrieval of a stored item pair during free recall. The other model parameters are nuisance parameters necessary for model identifiability that cover plausible but rare psychological states[12] that are of less interest for the present research (for a detailed description of the complete model and its application to DF see Rummel et al., 2016).

---

[12] *Parameters s* and *u* indicate singleton recall of items stored as pairs or singletons. *Parameter l* accounts for the possibility that some items are lost from memory during the delay between the first and the second memory test.

Pretest

As just described, the application of the storage–retrieval model requires participants to perform two successive memory tests, that is, a free and then a cued-recall test. In such a format one may assume that the preceding free recall test could influence subsequent cued recall performance. For example, it could be that retrieval during free recall strengthens certain items' memory trace, leading to enhanced cued recall rates. Although previous research has shown that such carry-over effects from free to cued recall do not occur in typical memory paradigms (Riefer & Rouder, 1992), we first intended to empirically rule out that a carry-over effect would occur in the item-method DF paradigm. To this end, we manipulated whether cued recall was preceded by free recall or not in an initial pretest.

Method

Participants. Eighty-one students from the University of Mannheim participated in the pretest in exchange for course credit or monetary compensation. Two participants were excluded because they did not follow the recall instructions (i.e., their recall performance was .00 and .13 for TBR-items), resulting in a final sample of seventy-nine participants (58 female, M_{age} = 21.76, age range: 18 – 34 years)[13].

Material. Hundred nouns of medium frequency were selected from the *dlex* database (Heister et al., 2011). A set of 50 word-pairs were formed by randomly pairing two words together. From this set, 20 word pairs were randomly selected to serve as items for the practice block, of which 10 pairs were always used as TBF and 10 as TBR-items. The remaining 30 word pairs were used for the real study phase; 15 served as TBF and 15 as TBR-items. The assignment of TBF or TBR instructions to word-pairs was counterbalanced across participants within groups.

Procedure. After providing informed consent, participants studied word-pairs for a later memory test. In an initial practice phase, participants were presented with 20 word-pairs. Each word-pair was presented in the middle of the screen for five seconds and half of the pairs were followed by a forget cue (i.e. **VVV**; the first letters of the German word forget) whereas the other half were followed by a remember cue (i.e. **EEE**; the first letters of the German word remember). The memory cues were presented in the middle of the screen for 2 seconds and were followed by a 500 ms inter-stimulus-interval. After the study phase, participants solved math problems for 30 s. For a first memory test, all participants were asked to freely recall as many items as they could remember that were post-cued as TBR. If they were able to recall only one word of the pair, then they were asked to write only that word down. Next, they completed a cued recall test for TBR-items. That is, they were presented with the first words of the pairs (probes), one after the other, and had to type in the corresponding second ones (targets). The purpose of this practice phase was to acquaint participants with the task and to increase participants' trust in the forget cues (see Marevic et al., 2017; Rummel et

---

[13] The present results would not have differed when the full sample of 81 participants had been used.

al., 2016). Following the practice phase, the real DF task started. The procedure was identical to the practice block, with the exception that participants in the Free-Recall group were asked to recall as many items as they could remember in 2 min during the free recall test, regardless of the memory cue (forget vs. remember) and in any order. Then, they were presented with the first words of all pairs (randomly intermixed) for the cued recall test. Participants in the No-Free-Recall group did not receive a free recall test and continued to solve math problems for another 2 min (equal amount of time the Free-Recall group had to perform the free recall test), after which they also completed a cued recall test for all word pairs.

At the end of the experiment, all participants were asked to name the strategy they employed to forget the TBF-items (if they happened to use one), and to fill in a demographic questionnaire before they were debriefed and compensated.

Results and Discussion

An alpha level of .05 was adopted for all analyses. Partial eta-squared (η^2_p) was calculated to indicate effect size for analyses of variance (ANOVAs); Cohen's *d* to indicate effect size for *t*-tests. All pairwise comparisons were planned and theoretically driven. Most participants reported using a forget strategy (i.e., 33 of the Free-Recall group and 29 of the No-Free-Recall group). Strategy-use frequencies did not differ between the two experimental groups, $\chi^2(1) = .20$, $p = .652$.

The proportion of probe and target words recalled during the Free-Recall group's free recall test was calculated separately for TBF and TBR-items. The total proportion of target words recalled during both groups' cued recall test was also calculated separately for TBF and TBR-items. Then, the recall probability of TBF-items was subtracted from the recall probability of TBR-items for both free and cued recall data. The resulting R-F difference score was used as the dependent variable. This score reflects forgetting of TBF-items relative to the (baseline) memory for TBR-items for each participant. The R-F difference has been widely used in previous DF studies because it directly indicates the amount of forgetting that took place (Golding & MacLeod, 1998; MacLeod, 1999; Saletin, Goldstein, & Walker, 2011). Notably, an R-F difference score that differs significantly from zero indicates a DF effect.

Free recall performance. Mean free recall performance in the Free-Recall group was $M = .13$ ($SD = .10$) for TBF and $M = .41$ ($SD = .19$) for TBR-items. The R-F difference ($M = .29$; $SD = 24$) was significantly different from zero, $t(40) = 7.74$, $p < .001$, $d = 1.19$, indicating a significant DF effect.

Cued recall performance. Cued recall performance for TBF and TBR-items in both groups are displayed in Figure 1. We conducted two one-sample *t*-tests for the Free-Recall and No-Free-Recall group with the R-F difference measure serving as the dependent variable. As expected, a reliable DF effect was present in the Free-Recall group ($M = .29$, $SD = .24$), $t(40) = 7.60$, $p < .001$, $d = 1.18$, as well as the No-Free-Recall group, ($M = .25$, $SD = .15$), $t(37) = 10.12$, $p < .001$, $d = 1.64$. An independent-samples *t*-test further showed that the R-F difference score did not differ between the Free-Recall and the No-Free-Recall group, $t < 1$. Follow-up analyses confirmed that the cued recall probability of TBF, *t*(67) = .40, *p* = .683, *d* = .08, and TBR-items, *t*(67) = 1.15, *p* = .251, *d* = .26, did not differ between the Free-Recall and No-Free-Recall-groups. Thus, the preceding free recall test in our DF paradigm did not influence subsequent cued recall rates in any way.

Experiment 1

As outlined in the Introduction, the goal of Experiment 1 was to identify candidate mechanisms for retrieval-reliant forgetting in the item-method DF paradigm. One such candidate mechanism is the loss of item-specific information for TBF-items. This idea was tested in a recent study by Fawcett, Lawrence, and Taylor (2016). They presented participants with abstract and concrete images in different color that where post-cued as TBF or TBR. In a subsequent recognition test, participants made old-new and remember-know decisions and completed a color memory test. Not only were responses to TBF-items slower than to TBR-items, but color memory for successfully recognized TBR-items was more accurate than for successfully recognized TBF-items. This difference in accuracy of memory for item features provides compelling evidence that memory traces of TBF-items are impoverished compared to the memory traces of TBR-items. In the light of these findings it seems likely that a loss of item-specific information (or item fidelity, Fawcett et al., 2016), and an inability to reinstate these information at test, create a retrieval deficit responsible for retrieval-driven DF within the item-method.

Experiment 1 was designed to test whether semantic cuing of some TBF-items at test can lead to enhanced memory for the cued items. As the semantic cue was provided at the retrieval stage, any reinstatement effect should be retrieval-mediated and the application of the storage–retrieval model which allows separating storage and retrieval-mediated memory enhancement allows us to corroborate this theoretical assumption. We used an item-method DF paradigm in which participants were presented with three different types of items at study: unrelated word-pairs that were post-cued as TBF (TBF_{UR}-items), word-pairs that were post-cued as TBF but were semantically related (TBF_{SR}-items), that is, the first word of these pairs shared the same superordinate category, and unrelated word-pairs that were post-cued as TBR (TBR-items). A free recall and a cued recall test were administered. Building on associative memory models (Raaijmakers & Shiffrin, 1981), superordinate category cues can be used to reinstate semantic item features because category-to-item associations are assumed to be automatically co-activated when a category exemplar is processed. According to formal memory models, for items that share a superordinate category, additional category context features are simply appended to the item vector and are encoded the same way as any other item feature (Lehman & Malmberg, 2011). Thus, category reinstatement at retrieval provides an additional retrieval cue, resulting in a recall advantage for category-cued items. For items that share the same category but for which the category is not externally cued, retrieval is contingent on the internal generation of initially encoded cues (e.g., temporal cues) which does not necessarily include category features. Thus one can expect a higher retrieval probability in the presence versus absence of a category cue. The use of category cues also speaks to the generalizability of the finding that TBF-items profit from the reinstatement of item context-features (e.g., font color, background color; Burgess et al., 2017). More importantly, the category reinstatement manipulation has the additional advantage that it can be applied in recall paradigms whereas the previous reinstatement manipulations are only suitable for recognition paradigms and we are especially interested in the item-method DF processes involved in the more typical recall situation.

For the manipulation of category reinstatement, half of the participants of the present study were cued with the shared category of TBF_{SR}-items prior to free recall (Category-Cue group), whereas the other half were not cued with the shared category (No-Category-Cue group). According to Lehman and Malmberg (2011) reinstating a superordinate category that is shared by a subset of items functions as an effective cue in a free recall test when other cues are absent (Howard & Kahana, 2002; Malmberg & Shiffrin, 2005). Category cues seem to be better at guiding retrieval processes in free recall compared to other cues such as temporal context—a finding that has already been demonstrated in the list-method DF domain (Lehman & Malmberg, 2011). We hypothesize that, by cuing the superordinate category prior to free recall, item fidelity of the semantically-related TBF_{SR}-items should be restored, leading to a reduction of the DF effect. Even though there is some evidence that such a reduction in DF is item-specific (cf. Fawcett et al., 2016), it could also be that the category reinstatement benefits translate to other items from the same study episode whose features overlap with those of the reinstated items, namely semantically unrelated TBF-items. Similar observations have been made in the list-method DF paradigm where TBF-items are presented en bloc and thus form a more coherent study episode (Pastötter & Bäuml, 2010; Sahakyan & Kelley, 2002).

To disentangle storage-mediated and retrieval-mediated DF, we applied the multinomial storage–retrieval model to our data (Riefer & Rouder, 1992; Rouder & Batchelder, 1998; see Appendix A for a description of the model). In this model, we expect the release from DF due to the category cue to be reflected by the retrieval parameter only because only items that had been successfully stored can be brought back into consciousness via context reinstatement.

Method

Participants. Seventy-six participants from Heidelberg University (60 women, $M_{\text{age}} = 23.47$, age range: 19 – 34 years) participated in exchange for course credit or a monetary compensation. Recruitment of participants was organized with the software tool *hroot* (Bock, Baetge, & Nicklisch, 2014). Participants were randomly assigned to the two experimental groups (Category-Cue, No-Category Cue).

Material. A total of 120 nouns were used as stimuli, with 100 nouns of medium frequency being selected from the *dlex* database (Heister et al., 2011) and 20 nouns (10 clothes, 10 furniture) being selected from German category production norms (Mannhaupt, 1983). We did not use the four most typical exemplars of the category to keep their word frequency comparable to the other items. For the practice phase, we used the same material as in the pretest. For the experimental phase, two sets of 10 unrelated pairs each served as either TBF_{UR} or TBR-items. The assignment of sets to item types was counterbalanced within each group. Another two word-pair sets, with all first words of the pairs being members of the same superordinate category (i.e., either *clothes* or *furniture*) served as TBF_{SR}-items. For the semantically related pairs as well as for all other pairs, the first word was only weakly related to the second word of the pair. Half of the participants of each group received the clothes TBF_{SR} item set, the other half the furniture TBF_{SR} item set.

Procedure. The procedure was identical to the one used for the pretest, only the actual test phase differed. That is, all participants studied 30 word pairs for the actual test phase. Ten pairs were unrelated word pairs that were post-cued as TBF (TBF_{UR}-items). Another 10 pairs, which were semantically related because the first words of these pairs were exemplars of the same category, were also post-cued as TBF (TBF_{SR}-items). Yet another ten word pairs were unrelated pairs that were post-cued as TBR (TBR-items). With instructions for the free recall test (see above), participants of the Category-Cue group received a category cue (i.e. they were informed that some study items belong to the clothes or furniture category, respectively). Participants in the No-Category-Cue group did not receive this cue.

Results and Discussion

Most participants reported using a forget strategy (i.e., 35 of the Category-Cue group and 37 of the No-Category-Cue group). Strategy-use frequencies did not differ between the two experimental groups, $\chi^2(1) = 1.05$, $p = .304$.

Behavioral analyses of directed forgetting.

The proportion of probe and target words recalled during the free recall test and the proportion of target words recalled during the cued recall test were calculated separately for TBF_{UR}, TBF_{SR}, and TBR-items. The mean free and cued recall proportions are displayed in Figure 2. In order to assess the relative amount of forgetting, we calculated difference scores by subtracting the TBF_{UR}-item recall rates from the TBR-item recall rates (R-F_{UR} difference) as well as the TBF_{SR}-items recall rates from the TBR-item recall rates (R-F_{SR} difference) for both free and cued recall data. The resulting R-F difference scores were used as dependent variables.

Free recall performance. Separate one-sample *t*-tests were conducted for the Category-Cue and No-Category-Cue group on the free-recall R-F_{UR} and R-F_{SR} difference scores. R-F_{UR} difference scores differed significantly from zero in both the Category-Cue group ($M = .38$, $SD = .23$), $t(37) = 10.11$, $p < .001$, $d = 1.64$, and the No-Category-Cue group ($M = .42$, $SD = .26$), $t(37) = 10.04$, $p < .001$, $d = 1.62$. The same was true for the R-F_{SR} difference in both the Category-Cue group ($M = .27$, $SD = .24$), $t(37) = 7.12$, $p < .001$, $d = 1.15$ and the No-Category-Cue group ($M = .39$, $SD = .22$), $t(37) = 10.64$, $p < .001$, $d = 1.72$. These results indicate that there were reliable DF effects in all groups and conditions. Whereas the R-F_{UR} difference scores did not differ between the Category Cue and No-Category-Cue group, $t < 1$, the R-F_{SR} difference score in the Category-Cue group was significantly reduced compared to in the No-Category-Cue group, $t(74) = 2.11$, $p = .038$, $d = .48$. The smaller R-F_{SR}

difference in the Category-Cue group is indicative of the retrieval enhancing effects of context reinstatement on free recall.

Cued recall performance. One-sample *t*-tests for cued recall R-F_{UR} and the R-F_{SR} difference scores revealed that the R-F_{UR} difference scores differed significantly from zero, in both the Category-Cue group ($M = .36$, $SD = .25$), $t(37) = 8.77$, $p < .001$, $d = 1.42$ and the No-Category-Cue group ($M = .34$, $SD = .21$), $t(37) = 9.83$, $p < .001$, $d = 1.59$. The same applied to the R-F_{SR} difference in both the Category-Cue group ($M = .30$, $SD = .26$), $t(37) = 7.10$, $p < .001$, $d = 1.15$, and the No-Category-Cue group ($M = .35$, $SD = .20$), $t(37) = 10.45$, $p < .001$, $d = 1.69$. As expected for the cued recall test, in which the category information was carried by the probe word, neither the R-F_{UR} nor the R-F_{SR} difference scores varied between groups, both *t*s < 1.

Model-based analyses. The multinomial storage–retrieval model for free-then-cued-recall paradigms was applied to the present data (Riefer & Rouder, 1992; Rouder & Batchelder, 1998).

For the analysis, frequencies from the free and cued-recall tests of all six event categories (E_1-E_6) for each item type and experimental condition were aggregated, resulting in a total of six conditions. Event frequencies are provided in Appendix B. To achieve model identifiability, parameter restrictions had to be imposed. As the cued recall test immediately followed the free recall test, *l* parameter estimates that reflect memory loss from free recall to cued recall should be similar between conditions and generally close to zero. In accordance with our previous studies (Marevic et al., 2017; Rummel et al., 2016) we therefore set the *l* parameter to be equal across all conditions.

We used the software *multiTree* for the analysis (Moshagen, 2010). The restricted model had five degrees of freedom and fitted the data well, $G^2(5) = 4.53$, $p = .475$. As the main aim of the current investigation was to better understand the role of context memory for item-method DF and, especially, the cognitive routes that causes the release from forgetting after context reinstatement, we reparametrized the model in order to achieve proportional change parameters for the storage (*a*) and retrieval (*r*) parameters that can be interpreted analogously to the behavioral R-F difference scores. We used the parametric-order-constraints function of *multiTree* to impose the following constraints (see Klauer, Singmann, & Kellen, 2015; Knapp & Batchelder, 2004, for the mathematical background): The *a* parameters were replaced by a^* change parameters: $a(\text{TBF}_{UR}\text{-items}) < a(\text{TBR-items}) = a^*(\text{R-F}_{UR})$; $a(\text{TBF}_{SR}\text{-items}) < a(\text{TBR-items}) = a^*(\text{R-F}_{SR})$. Similarly, the *r* parameters were replaced by r^* change parameters: $r(\text{TBF}_{UR}\text{-items}) < r(\text{TBR-items}) = r^*(\text{R-F}_{UR})$; $r(\text{TBF}_{SR}\text{-items}) < r(\text{TBR-items}) = r^*(\text{R-F}_{SR})$. The new parameters reflect the engagement in storage and retrieval processing of TBF relative to TBR-items. After reparametrization, the probability of a cognitive event θ_1 is defined in terms of the probability of another cognitive event θ_2, with $\theta_1 = \alpha\ \theta_2$, $0 \leq \alpha, \theta_2 \leq 1$. The resulting change parameter α in the model thus reflects a proportion with $\alpha = \theta_1/\ \theta_2$ (see Method A from Knapp & Batchelder, 2004). High values of the change parameters a^* and r^* reflect low proportional forgetting (values of 1 would reflect no forgetting) and low values of the change parameters reflect high proportional forgetting in the model. As parameter estimates for TBF_{UR}-items and TBF_{SR}-items should not exceed those of TBR-items, the change parameters should always achieve positive values (see Horn, Pachur, & Mata, 2015, for a similar approach in the decision-making domain). Mean parameter estimates are displayed in Table 1.

Directed forgetting effects on storage (a). We first tested whether there was a DF effect of storing less TBF_{UR} and TBF_{SR} than TBR-items. In line with our expectations, in the Category-Cue group both the a*(R-F_{UR}) parameter, $\Delta G^2(1) = 119.02$, $p < .001$, and the a*(R-F_{SR}) parameter, $\Delta G^2(1) = 79.17$, $p < .001$, differed significantly from 1. Also in the No-Category-Cue group, both the a*(R-F_{UR}) parameter, $\Delta G^2(1) = 85.58$, $p < .001$, and the a*(R-F_{SR}) parameter, $\Delta G^2(1) = 93.46$, $p < .001$, differed significantly from 1. There were no differences between the Category-Cue group and the No-Category-Cue group regarding a* parameter estimates, neither for those reflecting R-F_{UR} changes, $\Delta G^2(1) = 2.35$, $p = .125$, nor for those reflecting R-F_{SR} changes, $\Delta G^2(1) = .88$, $p = .348$. According to these results, the DF effect was reflected by storage changes, but, as hypothesized, the context reinstatement manipulation did not affect storage processes.

Directed forgetting effects on retrieval (r). Again, we tested whether there were DF effects present on the retrieval change parameters r*(R-F_{UR}) and r*(R-F_{SR}). In the Category-Cue group, the r*(R-F_{UR}) parameter, $\Delta G^2(1) = 36.80$, $p < .001$, and the r*(R-F_{SR}) parameter, $\Delta G^2(1) = 12.04$, $p < .001$, differed from 1. The same was true in the No-Category-Cue group for both the r*(R-F_{UR}) parameter, $\Delta G^2(1) = 49.65$, $p < .001$, and the r*(R-F_{SR}) parameter, $\Delta G^2(1) = 35.40$, $p < .001$. The r*(R-F_{UR}) parameters did not differ between groups, $\Delta G^2(1) = .06$, $p = .811$, but r* (R-F_{SR}) estimates were significantly higher in the Category-Cue group compared to the No-Category-Cue group, $\Delta G^2(1) = 4.01$, $p = .045$. That is, the category context reinstatement resulted in a smaller TBR/TBF_{SR} retrieval change in the Category-Cue than in the No-Category-Cue group.

Taken together, the analyses indicate that the DF effect is driven by storage and retrieval processes and that reinstating semantic category context of TBF-items can significantly reduce their forgetting. Importantly, the reinstatement effect does not translate to other TBF-items which do not share the category cue. This speaks against the assumption that category reinstatement for semantically related TBF_{SR}-items generalizes to other TBF-items from the same study episode. The model-based results further corroborate the assumption that the category-cue-imposed forgetting reduction was retrieval-mediated. Finally, the present finding that the R-F_{SR} change score was still significantly different from 1 shows that, even though the DF effect can be reduced when category context is reinstated, the effect is not eliminated completely.

Experiment 2

Experiment 1 showed that the reinstatement of the superordinate category shared by a subset of TBF-items selectively improves TBF-item recall for the reinstated items. Experiment 2 was designed to replicate this finding and extend it to reinstatement effects on TBR-items. Different pattern of results could be expected for the TBR-items. It may be possible that features of TBR-items are as deeply encoded that the presentation of one particular feature does not improve TBR-item recall any further (cf. Hourihan et al., 2007). Alternatively, semantically related TBR-items could profit selectively from category reinstatement, just as TBF_{SR}-items did in Experiment 1 (cf. Burgess et al., 2017). A final alternative would be that category reinstatement effects translated to other, not semantically related TBR-items, because selective rehearsal of TBR-items may not only strengthen the memory traces of the individual items but also the associations between TBR-items.

In Experiment 2, we reinstated a shared category for TBF-items in one group and for TBR-items in an additional experimental group[14]. The benefit from category cuing in each reinstatement group was then compared to a corresponding control group in which the amount of TBF and TBR-items was equivalent to the reinstatement group but in which no category cue was provided. As in Experiment 1, we applied the storage–retrieval model to obtain purer estimates of the cognitive processes driving DF.

Method

Participants. Hundred and twenty participants from Heidelberg University participated in exchange for course credit or monetary compensation. Three participants were excluded because they told the experimenter that they knew the paradigm after participating. The final sample consisted of 117 participants (90 female, $M_{\text{age}} = 21.56$, age range: 17 – 34 years). Participants were randomly assigned to the four experimental groups.

Material. Materials were identical to the ones used in Experiment 1. In two conditions, TBF-items were semantically related and thus the *furniture* or *clothes* items were post-cued as TBF. In the other two conditions TBR-items were semantically related and thus the *furniture* or *clothes* items were post-cued as TBR. In both conditions the nature of the semantic relatedness was the same as in Experiment 1. That is, for the semantically related words, the first word of a pair was always related to another first word of another pair, with the first and second word of a pair remaining unrelated.

Procedure. The procedure was identical to the one used in Experiment 1 but the experiment consisted of four groups: A group in which semantically related study items were TBF-items and for which category context was reinstated prior to free recall (Category-Cue Forget group), a second group in which semantically related study items were TBF-items but for which category context was not reinstated prior to free recall (No-Category-Cue Forget group), a third group in which semantically related study items were TBR-items and for which category context was reinstated prior to free recall (Category-Cue Remember group), and a fourth group in which semantically related study items were also TBR-items but for which category context was not reinstated prior to free recall (No-Category-Cue Remember group).

Results and Discussion

Most participants reported using a forget strategy (i.e., 26 of the Category-Cue Forget group, 27 of the No-Category-Cue Forget group, 28 of the Category-Cue Remember group, and 24 of the No-Category-Cue Remember group). Strategy-use frequencies did not differ between the four experimental groups, $\chi^2(3) = 2.25$, $p = .521$.

Behavioral analyses of directed forgetting. Dependent variables were calculated as in Experiment 1. The mean free and cued recall proportions are displayed in Figure 3. To assess the relative amount of forgetting, we calculated difference scores by subtracting the $\mathrm{TBF_{UR}}$-item free and cued recall rates from the $\mathrm{TBR_{UR}}$-item free and cued recall rates ($\mathrm{R_{UR}}$-$\mathrm{F_{UR}}$ difference). To assess reinstatement effects on recall of semantically related TBF-items in the Category-Cue and No-Category-Cue Forget groups, we subtracted the $\mathrm{TBF_{SR}}$-item recall rates from the $\mathrm{TBR_{UR}}$-item recall rates for both free and cued recall data. This difference score was labeled $\mathrm{R_{UR}}$-$\mathrm{F_{SR}}$ difference. In the Category-Cue and No-Category-Cue Remember groups, where semantically related items were post-cued as TBR, we subtracted the semantically unrelated TBR-item recall rate from the semantically related TBR-item recall rate. This difference score was labeled $\mathrm{R_{SR}}$-$\mathrm{R_{UR}}$ difference.

Free recall performance. Separate one-sample *t*-tests were conducted for all groups on the free recall R-$\mathrm{F_{UR}}$, $\mathrm{R_{UR}}$-$\mathrm{F_{SR}}$, and $\mathrm{R_{SR}}$-$\mathrm{R_{UR}}$ difference scores, respectively. $\mathrm{R_{UR}}$-$\mathrm{F_{UR}}$ difference scores differed significantly from zero, in the Category-Cue Forget ($M = .34$, $SD = .17$), $t(27) = 10.58$, $p < .001$, $d = 2.00$, the No-Category-Cue Forget ($M = .38$, $SD = .23$), $t(29) = 8.92$, $p < .001$, $d = 1.65$, the Category-Cue Remember ($M = .13$, $SD = .22$), $t(29) = 3.23$, $p = .003$, $d = .59$ and the No-Category-Cue Remember groups ($M = .20$, $SD = .25$), $t(28) = 4.19$, $p < .001$, $d = .80$, groups. The same was true for the R-$\mathrm{F_{SR}}$ difference, in the Category-Cue ($M = .28$, $SD = .21$), $t(27) = 7.03$, $p < .001$, $d = 1.33$, and the No-Category-Cue Forget groups ($M = .39$, $SD = .25$), $t(29) = 8.53$, $p < .001$, $d = 1.56$, as well as for the $\mathrm{R_{SR}}$-$\mathrm{R_{UR}}$ difference in the Category-Cue Remember group, ($M = -.13$, $SD = .19$), $t(29) = 3.79$, $p = .001$, $d = .68$. For the No-Category-Cue Remember group, $\mathrm{R_{SR}}$-$\mathrm{R_{UR}}$ difference scores did not differ from zero ($M = .05$, $SD = .17$), $t(28) = 1.55$, $p = .131$, $d = .29$. These results indicate that there were reliable DF effects in all groups— independent of whether TBF-items were semantically related with each other or not and independent of whether a category cue was presented or not. Recall benefits for semantically related TBR-items relative to TBR-items, on the other hand, only occurred when the category context was reinstated.

Further, the $\mathrm{R_{UR}}$-$\mathrm{F_{UR}}$ difference did not differ between the Category-Cue and the No-Category-Cue Forget groups, $t < 1$ nor between the respective Remember groups, $t(113) = 1.17$, $p = .243$, $d = .07$. The $\mathrm{R_{UR}}$-$\mathrm{F_{SR}}$ difference was significantly smaller in the Category-Cue than in the No-Category-

---

[14] We deliberately choose to reinstate the shared category for TBF and TBR-items in separate experimental groups, as an initial pretest, in which both TBF and TBR-items shared two distinct superordinate categories for the same participants, revealed that participants notice the relatedness of

Cue Forget group, $t(113) = 2.01$, $p = .046$, $d = .48$. This finding is indicative of the retrieval enhancing effects of semantic reinstatement on free recall of TBF-items. However, there was no difference in the R_{SR}-R_{UR} difference between the Remember groups, $t(113) = 1.56$, $p = .121$, $d = .45$, suggesting that retrieval enhancing effects of context reinstatement on TBR-items were rather small as they were only present within the Category-Cue Remember group but were not reliable when comparing this group to the No-Category-Cue Remember group.

***Cued recall performance*.** Separate one-sample *t*-tests were conducted for the all groups on the cued recall R-F_{UR}, R_{UR}-F_{SR} and R_{SR}-R_{UR} difference scores. R_{UR}-F_{UR} difference scores differed significantly from zero, in the Category-Cue Forget ($M = .35$, $SD = .19$), $t(27) = 9.67$, $p < .001$, $d = .66$, the No-Category-Cue Forget ($M = .31$, $SD = .21$), $t(29) = 8.12$, $p < .001$, $d = 1.47$, the Category-Cue Remember ($M = .18$, $SD = .20$), $t(29) = 4.83$, $p < .001$, $d = .90$ and the No-Category-Cue Remember ($M = .22$, $SD = .22$), $t(28) = 5.22$, $p < .001$, $d = 1.00$, groups. The same was true for the R_{UR}-F_{SR} difference, in the Category-Cue Forget ($M = .35$, $SD = .23$), $t(27) = 8.21$, $p < .001$, $d = 1.52$, and the No-Category-Cue Forget groups ($M = .31$, $SD = .15$), $t(29) = 10.87$, $p < .001$, $d = 2.06$, whereas R_{SR}-R_{UR} difference scores did not differ from zero for the Category-Cue Remember ($M = -.03$, $SD = .24$), $t < 1$, and the No-Category-Cue Remember groups, ($M = -.06$, $SD = .22$), $t(28) = 1.48$, $p = .150$, $d = .27$. These results indicate that there were reliable DF effects in all groups, but no context reinstatement benefits for semantically related TBR-items.

The R_{UR}-F_{UR} difference did not differ between the Category-Cue and the No-Category-Cue Forget groups, $t < 1$, nor between the Category-Cue and the No-Category-Cue Remember groups, $t < 1$. The R_{UR}-F_{SR} difference did not differ between the Category-Cue and the No-Category-Cue Forget groups, $t < 1$. The R_{SR}-R_{UR} difference did not differ between the respective Remember groups, $t < 1$. Thus, there was no evidence for context reinstatement effects on either forgetting of semantically related TBF-items or the benefits of semantically related TBR-items in cued recall.

Model-based analyses. As in Experiment 1, the multinomial storage–retrieval model for free-then-cued-recall paradigms was applied to the present data (Riefer & Rouder, 1992; Rouder & Batchelder, 1998). For the analysis, frequencies from the free and cued-recall tests of all six event categories (E_1-E_6) for each item type and experimental condition were aggregated, resulting in a total of six conditions (event frequencies are displayed in Appendix B). To achieve model identifiability, we set the *l* parameter to be equal across all conditions. Due to a cell value equal to zero for the semantically related item event category E_3 in the No-Category-Cue groups, the *s* parameter estimates for this item type in these groups lay outside the parameter space. We thus set the *s* parameter to be equal across groups for each item type—a common restriction that has been suggested by others

---

the items also in the absence of a category cue—probably because the proportion of items sharing a category was too high in the to-be-studied item set.

before (see Riefer & Batchelder, 1991b; Riefer & Rouder, 1992)[15]. The restricted model had twenty degrees of freedom and fitted the data well, $G^2(20) = 27.76$, $p = .114$. As in Experiment 1, we imposed parametric order constraints that reflect proportional changes of semantically unrelated TBF and semantically related TBF and TBR items relative to semantically unrelated TBR-items. As parameter estimates cannot exceed 1, estimates of the Category-Cue and the No-Category-Cue Forget groups for TBF_{SR}-items were set to be smaller than TBF_{UR} -item estimates. For the Remember groups, TBR_{SR}-items were set to be greater than TBR_{UR}-item estimates. Order constraints for TBF-items were imposed analogously to Experiment 1, resulting in the following constraints for the parameters of interest (a and r): The a parameters were replaced by a^* change parameters: $a(\text{TBF}_{\text{UR}}\text{-items}) < a(\text{TBR-items}) = a^*(\text{R}_{\text{UR}}\text{-F}_{\text{UR}})$; $a(\text{TBF}_{\text{SR}}\text{-items}) < a(\text{TBR-items}) = a^*(\text{R}_{\text{UR}}\text{-F}_{\text{SR}})$; $a(\text{TBR-items}) < a(\text{TBR}_{\text{SR}}\text{-items}) = a^*(\text{R}_{\text{SR}}\text{-R}_{\text{UR}})$. Similarly, the r parameters were replaced by r^* change parameters: $r(\text{TBF}_{\text{UR}}\text{-items}) < r(\text{TBR-items}) = r^*(\text{R}_{\text{UR}}\text{-F}_{\text{UR}})$; $r(\text{TBF}_{\text{SR}}\text{-items}) < r(\text{TBR-items}) = r^*(\text{R}_{\text{UR}}\text{-F}_{\text{SR}})$; $r(\text{TBR-items}) < r(\text{TBR}_{\text{SR}}\text{-items}) = r^*(\text{R}_{\text{SR}}\text{-R}_{\text{UR}})$. The new parameters reflect proportions in the engagement in storage and retrieval processing of TBF relative to TBR-items and processing of semantically related TBF and TBR-items relative to semantically unrelated TBR-items. Mean parameter estimates are presented in Table 2.

Directed forgetting effects on storage (a). We first tested whether there was a DF effect of storing less TBF_{UR} and TBF_{SR} than TBR-items. In line with our expectations, in the Category-Cue Forget group both the $a^*(\text{R}_{\text{UR}}\text{-F}_{\text{UR}})$, $\Delta G^2(1) = 61.63$, $p < .001$, and the $a^*(\text{R}_{\text{UR}}\text{-F}_{\text{SR}})$, $\Delta G^2(1) = 72.96$, $p < .001$, parameters differed significantly from 1. In the No-Category-Cue Forget group, both the $a^*(\text{R}_{\text{UR}}\text{-F}_{\text{UR}})$ parameter, $\Delta G^2(1) = 61.14$, $p < .001$, and the $a^*(\text{R}_{\text{UR}}\text{-F}_{\text{SR}})$ parameter, $\Delta G^2(1) = 59.32$, $p < .001$, differed significantly from 1. The $a^*(\text{R}_{\text{UR}}\text{-F}_{\text{UR}})$ parameter also differed significantly from 1, in the Category-Cue, $\Delta G^2(1) = 26.45$, $p < .001$, as well as the No-Category-Cue, $\Delta G^2(1) = 38.47$, $p < .001$, Remember groups.

Next, we tested whether there were any benefits of context reinstatement for the semantically related TBR-items. The $a^*(\text{R}_{\text{SR}}\text{-R}_{\text{UR}})$ parameter did not differ significantly from 1 in both the Category-Cue, $\Delta G^2(1) = 0.47$, $p = .490$, and No-Category-Cue, $\Delta G^2(1) = 1.44$, $p = .229$, Remember groups.

Further, there were no differences between the Category-Cue and the No-Category-Cue Forget groups regarding a^* parameter estimates, neither for those reflecting $\text{R}_{\text{UR}}\text{-F}_{\text{UR}}$ changes, $\Delta G^2(1) = 1.23$, $p = .268$, nor for those reflecting $\text{R}_{\text{UR}}\text{-F}_{\text{SR}}$ changes, $\Delta G^2(1) = 3.12$, $p = .077$. The same was true for the Category-Cue and the No-Category-Cue Remember groups regarding a^* parameter estimates for $\text{R}_{\text{UR}}\text{-F}_{\text{UR}}$ changes, $\Delta G^2(1) = 0.44$, $p = .506$, and for $\text{R}_{\text{SR}}\text{-R}_{\text{UR}}$ changes, $\Delta G^2(1) = .09$, $p = .759$.

---

[15] Setting the s and u parameters equal for each item type and condition resulted in a worse model fit, $\Delta G^2(3) = 38.33$, $p < .001$. The results of Experiment 1 did not change when we re-ran the analyses with the same restrictions as in Experiment 2.

In sum these results show that the DF effect was reflected by storage changes but, as expected, the context reinstatement manipulation did not affect storage processes in any way.

Directed forgetting effects on retrieval (r). Again, we first tested whether there was a retrieval-mediated DF effect. In line with our expectations, in the Category-Cue Forget group both the r*(R$_{UR}$-F$_{UR}$), $\Delta G^2(1) = 20.15$, $p < .001$, and the r*(R$_{UR}$-F$_{SR}$), $\Delta G^2(1) = 6.91$, $p = .008$, parameters differed significantly from 1. Also in the No-Category-Cue Forget group, both the r*(R$_{UR}$-F$_{UR}$), $\Delta G^2(1) = 33.06$, $p < .001$, and the r*(R$_{UR}$-F$_{SR}$), $\Delta G^2(1) = 41.84$, $p < .001$, parameters differed significantly from 1. The r*(R$_{UR}$-F$_{UR}$) parameter did, however, not differ significantly from 1, in both the Category-Cue, $\Delta G^2(1) = 1.37$, $p = .240$, and the No-Category-Cue, $\Delta G^2(1) = 2.47$, $p = .115$, Remember groups indicating the absence of a retrieval-mediated DF effect in these two groups. The r*(R$_{SR}$-R$_{UR}$) parameter differed significantly from 1 in the Category-Cue, $\Delta G^2(1) = 12.98$, $p < .001$, but not the No-Category-Cue, $\Delta G^2(1) = 0.01$, $p = .890$, Remember group indicating reliable retrieval benefits for TBR$_{SR}$ compared to TBR$_{UR}$-items after context reinstatement.

Furthermore, there was no difference between the Category-Cue and the No-Category-Cue Forget groups regarding r* parameter estimates reflecting R$_{UR}$-F$_{UR}$ changes, $\Delta G^2(1) < 0.01$, $p = .975$, but there was a difference between r* estimates reflecting R$_{UR}$-F$_{SR}$ changes, $\Delta G^2(1) = 4.51$, $p = .033$, indicating an isolated reduction in the DF effect on semantically related TBF-items as a result of context reinstatement. For the Category-Cue and the No-Category-Cue Remember group there was no difference regarding r* estimates for R$_{UR}$-F$_{UR}$ changes, $\Delta G^2(1) = 0.04$, $p = .836$, but r* estimates for R$_{SR}$-R$_{UR}$ changes were lower in the Category-Cue than in the No-Category-Cue Remember group, $\Delta G^2(1) = 5.51$, $p = .018$. This finding indicates reliable and isolated retrieval benefits of category reinstatement for TBR$_{SR}$-items.

Taken together, the behavioral and model-based results showed that the DF effect was driven by storage processes in all groups, but was only reflected by retrieval changes in the two groups where a sub-portion of the TBF-items shared a category. One possible reason for this finding could be that twice as many TBR-items had to be retrieved in the two groups in which a sub-portion of TBR-items shared a category. As a consequence of the increased number of to-be-retrieved items in these groups, TBR-items could have suffered more from retrieval competition between TBR-items (Raaijmakers & Shiffrin, 1981; Rundus, 1973). This phenomenon is referred to as the list-length effect in the memory literature (Murdock, 1962; Postman & Phillips, 1965; Watkins & Watkins, 1975) (see Aguirre, Gómez-Ariza, Andrés, Mazzoni, & Bajo, 2017 for a simmilar finding in selective directed forgetting). On the latent-process level, we observed a retrieval-mediated reduction of the DF effect for semantically related TBF-items as well as a retrieval-mediated memory benefit for semantically related TBR-items when context was reinstated. The finding of isolated reinstatement benefits in both TBF and TBR items is in line with Burgess et al.'s (2017) findings in a recognition memory test. Notably, the reinstatement effect on TBR-item memory was not evident from the behavior analyses alone. This aspect is further discussed in the General Discussion section.

General Discussion

In two experiments, we showed that the reinstatement of a superordinate category for semantically related TBF-items can reduce forgetting of these items, a finding that cannot be explained by accounts that consider item-method DF a mere storage phenomenon. In Experiment 1, the category shared by a subset of TBF-items was reinstated, leading to an isolated reduction of the DF effect for these items. That is, context reinstatement only fostered recall of those items that were reinstated and did not translate to other items of the same memory state. Additionally, the model-based results clearly show that the category-cue benefits are mediated by changes in item retrieval rendering further support to the models validity. Importantly, reinstatement effects in item-method DF seem to differ from those in list-method DF where reinstatement can help to recover the whole TBF study episode. For example, Sahakyan and Kelley (2002) found that the list-method DF effect is completely eliminated when study context is mentally reinstated prior to the recall test (e.g., thinking about the music that played directly prior to encoding the TBF items). Others have demonstrated similar reinstatement effects by providing some earlier studied TBF-items on recognition or stem-completion tests which then lead to the elimination of the DF effect (Bäuml & Samenieh, 2010; E. L. Bjork & Bjork, 1996; Goernert & Larson, 1993). One explanation for the divergent findings may be that different cuing mechanisms are at work within the two DF methods. That is, in the item-method, where reinstatement takes place for items that were instructed to be forgotten but were learned interleaved with to-be-remembered material, the TBF-items may not form a coherent study episode and thus, only item-feature cueing, but no item-item cueing, can cause a release from forgetting. Further research is necessary to test this idea.

The central goal of the present study was to identify candidate mechanisms for the retrieval-reliant effects in item-method DF. Because a selective rehearsal strategy (R. A. Bjork, 1972) that leads people to mentally repeat and elaborate on TBR but not TBF-items should predominantly affect storage of both item types differently, we find it unlikely that this strategy alone caused the retrieval differences between TBF and TBR-items that was consistently observed in our previous (Marevic et al., 2017; Rummel et al., 2016) and the present experiments. We find it more likely that an additional attention withdrawal strategy (Fawcett & Taylor, 2010; Taylor, 2005) is responsible for the DF effects at the retrieval stage. Fawcett et al. (2016) showed that TBF-item fidelity for incorrectly as well as correctly recognized TBF-items is greatly impoverished relative to TBR-item fidelity as a result of active attention withdrawal from TBF-item representations. Thus, it could be that the loss of item-specific context information, together with the inability to reinstate that context at test, is responsible for the aforementioned retrieval deficit. The results of the present investigation support this idea in that cuing a category shred by a subset of TBF-items reduced DF in Experiments 1 and 2 and the model-based results confirmed the retrieval-reliant nature of this process.

The observations of Experiments 1 and 2 that storage of TBF was worse than of TBR-items and that this storage deficit remained unaffected by the category-reinstatement manipulation suggests that a substantial part of the item-method DF effect is, in fact, due to generally reduced storage of TBF-items. In the light of current theorizing, these storage differences may likely be caused by selective rehearsal of TBR-items at the cost of storing TBF-items (Lee, 2012). Thus, our conclusion would not be that the selective-rehearsal explanation of item-method DF has to be dismissed. Rather, our results support the assumption that selective rehearsal is the strongest factor driving DF in the item-method but that smaller parts of this effect may be caused by other mechanisms such as attention withdrawal that hampers memory for item features.

Another goal of the present investigation was to shed light on the question of whether semantic item-feature reinstatement, that reduces forgetting of TBF information, would also enhance memory for TBR information. The (model-based) analyses of Experiment 2 showed not only retrieval-mediated DF reductions after semantic reinstatement for semantically related items that were post-cued as TBF but also retrieval-mediated benefits for reinstated semantically related items that were post-cued as TBR. This finding is well in line with early studies on context memory showing that context reinstatement benefits recall of previously learned material (Godden & Baddeley, 1975; Smith, 1979). Notably, the behavioral data did not reveal significant reinstatement benefits for TBR-items—likely because storage processes had a larger impact on the observable memory performance than retrieval processes and thus retrieval-mediated benefits became only evident after being isolated.

The present interpretation of results relies on the assumption that the storage–retrieval model is able to separate the storage and retrieval contributions of an observed effect and thus creates added value relative to mere behavioral analyses. The storage-retrieval model has been validated in its ability to separate storage and retrieval processes with regard to many memory effects, such as the bizarreness effect (Riefer & LaMay, 1998; Riefer & Rouder, 1992), memory deficits in clinical populations (Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002), or lag effects (Küpper-Tetzel & Erdfelder, 2012), and several simulation studies, in which some parameter estimates were varied while leaving others constant, demonstrate that the model provides reliable storage and retrieval estimates (see Riefer & Batchelder, 1991b; Rouder & Batchelder, 1998 for an overview). Nevertheless, some researchers may argue that using ANOVAs on the free and cued recall data would similarly allow for making inferences about the contribution of storage and retrieval processes rendering the application of the storage–retrieval model obsolete. However, there are good arguments that this is not the case. Rouder & Batchelder (1998) and Riefer & Rouder (1992) report a simple simulation that illustrates this point: They set equal parameter values for all parameters of two hypothetical conditions, but specified values for the a and r parameter that reflect an equal difference between the two conditions for both a and r. From these values, the authors then computed free and cued recall probabilities based on the model equations for each recall event. The resulting probabilities reflect equal differences across conditions for both free and cued recall. Such a pattern would imply that a given effect is an isolated storage phenomenon as it is present in both free and cued recall, but as both storage and retrieval were set to differ between conditions in this simulation the behavioral results did not correctly recover the data pattern. This divergence between multinomial modeling and ANOVA approaches, that is not only present in simulation studies but also in several empirical data sets, can be intuitively understood by considering the simple fact that mean free and cued recall rates do not take into account whether the same or different items are recalled in the two tests. The storage–retrieval model, however, considers this information and is thus able to separate storage and retrieval processes more precisely (Batchelder & Riefer, 1986).

One possible limitation of the present investigation is that the category cues may have influenced free recall output strategies. For instance, participants could have started recalling the semantically related items first, leading to an output order bias (Anderson, 2005) that caused a recall benefit for semantically related items independent of the reinstatement itself. To rule out this alternative explanation, we conducted output order analyses on the free recall data from Experiments 1 and 2. The results showed that participants were similarly likely to recall semantically related items in the first quarter of the recall episode as in all remaining quarters of the recall episode (see Appendix C for details). Additionally, Rummel et al. (2016) demonstrated empirically that output order is not responsible for the retrieval deficit in item-method DF. Taken together, an output order bias is unlikely to be responsible for the present reinstatement effects.

The separate manipulation of semantic relatedness of TBF and TBR-items imposes another possible limitation. As a consequence, only one-third of all items was semantically related. If the relative proportion of related to unrelated items facilitated recall for the related items, then we should have observed better recall of semantically related items across all experimental groups. The absence of such an effect in the groups without a category cue speak against this potential confound.[16]

Finally, reinstating only one semantic feature of TBF-items (i.e., their shared superordinate category) could have limited the retrieval-reliant reduction of DF: A single item feature can be considered a rather impoverished cue compared to the variety of item-specific information that is usually stored with an item (e.g., semantic, spatiotemporal, emotional and physical information; Johnson, Hashtroudi, & Lindsay, 1993), or idiosyncratically generated context features (Johnson & Raye, 1981). Thus, with the reinstatement of just one item feature that may have been stored or not for one particular item in the present research, we probably underestimate the true impact that feature reinstatement can have on releasing to-be-forgotten material from forgetting. However, it seems impressive that this rather weak manipulation already substantially reduced the DF effect.

To conclude, the present findings further bolster recent evidence that item-method DF is caused by both storage and retrieval mechanisms (Rummel et al., 2016) and extends the existing body

---

[16] In Experiment 2, there was neither a significant difference between the TBF_{SR} and TBF_{UR}-items for free recall, $F(1, 29) = .35$, $p = .556$, $\eta^2_p = .01$, and cued recall, $F(1, 29) = .01$, $p = .931$, $\eta^2_p < .01$, in the No-Category-Cue Forget group nor was there a significant difference between the TBR_{SR} and TBR_{UR}-items for free recall, $F(1, 28) = 2.42$, $p = .131$, $\eta^2_p = .08$, and cued recall, $F(1, 28) = 2.19$, $p = .150$, $\eta^2_p = .07$, in the No-Category-Cue Remember group.

of research by identifying item-specific context as one factor that can selectively influence retrieval-reliant mechanisms of item-method DF. The retrieval-reliant recall recovery is not limited to TBF information, but also aids TBR information. The latter, however, could only be identified when process-pure measures of retrieval were obtained. Importantly, the retrieval-reliant recovery does not generalize to other item types. The exact nature of the retrieval-mediated processes underlying item-method DF is still an issue of debate and the current investigation does certainly not solve it completely. However, we believe that the present research is a step towards better understanding retrieval-reliant mechanisms of item-feature reinstatement in relation to intentional forgetting that also demonstrates the usefulness of model-based analysis in this research domain.

Compliance with Ethical Standards

The first author (Ivan Marevic) and second author (Jan Rummel) both declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Aguirre, C., Gómez-Ariza, C. J., Andrés, P., Mazzoni, G., & Bajo, M. T. (2017). Exploring Mechanisms of Selective Directed Forgetting. *Frontiers in Psychology*, *8*, 1–15. http://doi.org/10.3389/fpsyg.2017.00316

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In C. M. MacLeod & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 159–190). Tokyo: Springer-Verlag.

Basden, B. H., & Basden, D. R. (1996). Directed forgetting : Further comparisons of the item and list methods. *Memory*, *4*(6), 633–653. http://doi.org/10.1080/096582196388825

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139–172). Mahwah, NJ: Erlbaum.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603–616. http://doi.org/10.1037/0278-7393.19.3.603

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology*, *39*, 129–149.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. http://doi.org/10.3758/BF03210812

Bäuml, K.-H. (2008). Inhibitory processes. In H. L. Roediger III (Ed.), *Cognitive psychology of memory* (pp. 195–220). Oxford: Elsevier.

Bäuml, K.-H., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science*, *21*(6), 793–795. http://doi.org/10.1177/0956797610370162

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition*, *5*(1), 176–196. http://doi.org/10.1006/ccog.1996.0011

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Washington, DC: Winston.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot - Hamburg registration and organization online tool. *European Economic Review*, *71*, 117–120. http://doi.org/10.1016/j.euroecorev.2014.07.003

Burgess, N., Hockley, W. E., & Hourihan, K. L. (2017). The effects of context in item-based directed forgetting: Evidence for “one-shot” context storage. *Memory & Cognition*. http://doi.org/10.3758/s13421-017-0692-5

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*(3), 108–124. http://doi.org/10.1027/0044-3409.217.3.108

Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology: General*, *145*(1), 56–81. http://doi.org/10.1037/xge0000128

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition*, *36*(6), 1168–1181. http://doi.org/10.3758/MC.36.6.1168

Fawcett, J. M., & Taylor, T. L. (2010). Directed forgetting shares mechanisms with attentional withdrawal but not with stop-signal inhibition. *Memory & Cognition*, *38*(6), 797–808. http://doi.org/10.3758/MC.38.6.797

Fawcett, J. M., & Taylor, T. L. (2012). The control of working memory resources in intentional forgetting: Evidence from incidental probe word recognition. *Acta Psychologica*, *139*(1), 84–90. http://doi.org/10.1016/j.actpsy.2011.10.001

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition*, *13*(1), 57–62. http://doi.org/10.3758/BF03198444

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331. http://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Goernert, P. N., & Larson, M. E. (1993). The initiation and release of retrieval inhibition. *The Journal of General Psychology*, *121*(1), 61–66.

Golding, J. M., & MacLeod, C. M. (1998). *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ US: Erlbaum.

Hanley, J., & Morris, P. (1987). The effects of the amount of processing on recall and regocnition. *Quarterly Journal of Experimental Psychology*, *39*(A), 431–449. http://doi.org/10.1080/14640748708401797

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB-eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20. http://doi.org/10.1026/0033-3042/a000029

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 562–567. http://doi.org/10.1016/S0022-5371(71)80029-4

Horn, S. S., Pachur, T., & Mata, R. (2015). How does aging affect recognition-based inference? A hierarchical Bayesian modeling approach. *Acta Psychologica*, *154*, 77–85. http://doi.org/10.1016/j.actpsy.2014.11.001

Hourihan, K. L., Goldberg, S., & Taylor, T. L. (2007). The role of spatial location in remembering and forgetting peripheral words. *Canadian Journal of Experimental Psychology*, *61*(2), 91–101. http://doi.org/10.1037/cjep2007010

Hourihan, K. L., & Taylor, T. L. (2006). Cease remembering: Control processes in directed forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1354–1365. http://doi.org/10.1037/0096-1523.32.6.1354

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. http://doi.org/10.1006/jmps.2001.1388

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. http://doi.org/10.1007/BF02294263

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. http://doi.org/10.1037/0033-2909.114.1.3

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. http://doi.org/10.1037/0033-295X.88.1.67

Klauer, K. C., Singmann, H., & Kellen, D. (2015). Parametric order constraints in multinomial processing tree models: An extension of Knapp and Batchelder (2004). *Journal of Mathematical Psychology*, *64–65*, 1–7. http://doi.org/10.1016/j.jmp.2014.11.001

Knapp, B. R., & Batchelder, W. H. (2004). Representing parametric order constraints in multi-trial applications of multinomial processing tree models. *Journal of Mathematical Psychology*, *48*, 215–229. http://doi.org/10.1016/j.jmp.2004.03.002

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding , maintenance , and retrieval processes in the lag effect : A multinomial processing tree analysis. *Memory*, *20*(1), 37–48. http://doi.org/10.1080/09658211.2011.631550

Lee, Y.-S. (2012). Cognitive load hypothesis of item-method directed forgetting. *Quarterly Journal of Experimental Psychology*, *65*(6), 1110–1122. http://doi.org/10.1080/17470218.2011.644303

Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition*, *39*(2), 335–47. http://doi.org/10.3758/s13421-010-0025-4

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

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*(1), 123–129. http://doi.org/10.3758/BF03210819

Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 322–36. http://doi.org/10.1037/0278-7393.31.2.322

Mannhaupt, H.-R. (1983). Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien. *Sprache & Kognition*, *2*(4), 264–278.

Marevic, I., Arnold, N. R., & Rummel, J. (2017). Item-method directed forgetting and working memory capacity: A hierarchical multinomial modeling approach. *The Quarterly Journal of Experimental Psychology*, advanced online publication.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. http://doi.org/10.3758/BRM.42.1.42

Murdock, B. B. (1962). The serial position curve of free recall. *Journal of Experimental Psychology*, *64*, 482–488.

Nowicka, A., Jednoróg, K., Wypych, M., & Marchewka, A. (2009). Reversed old/new effect for intentionally forgotten words: An ERP study of directed forgetting. *International Journal of Psychophysiology*, *71*(2), 97–102. http://doi.org/10.1016/j.ijpsycho.2008.06.009

Pastötter, B., & Bäuml, K.-H. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 54–65. http://doi.org/10.1037/a0017406

Pastötter, B., Tempel, T., & Bäuml, K.-H. (2017). Long-term memory updating: The reset-of-encoding hypothesis in list-method directed forgetting. *Frontiers in Psychology*, *8*(2076), 1–6. http://doi.org/10.3389/fpsyg.2017.02076

Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *The Quarterly Journal of Experimental Psychology*, *17*(2), 132–138. http://doi.org/10.1080/17470216508416422

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. http://doi.org/10.1037/0033-295X.88.2.93

Riefer, D. M., & Batchelder, W. H. (1991a). Age differences in storage and retrieval: A multonomial modeling analysis. *Bulletin of the Psychonomic Society*, *29*(5), 415–418.

Riefer, D. M., & Batchelder, W. H. (1991b). Statistical inference for multinomial tree models. In J.-C. Falmagne & J.-P. Doignon (Eds.), *Mathematical psychology: Current developments* (pp. 313–336). Berlin: Springer-Verlag.

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*(2), 184–201. http://doi.org/10.1037/1040-3590.14.2.184

Riefer, D. M., & LaMay, M. L. (1998). Memory for common and bizarre stimuli: A storage-retrieval analysis. *Psychonomic Bulletin & Review*, *5*(2), 312–317. http://doi.org/10.3758/BF03212957

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*(6), 601–611. http://doi.org/10.3758/BF03202710

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 195–225). New York: Psychology Press.

Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1526–1543. http://doi.org/10.1037/xlm0000266

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 43–50. http://doi.org/10.1016/S0022-5371(73)80059-3

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2014). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 59, pp. 131–190). Elsevier Inc.: Academic Press. http://doi.org/10.1016/b978-0-12-407187-2.00004-6

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1064–1072. http://doi.org/10.1037//0278-7393.28.6.1064

Saletin, J. M., Goldstein, A. N., & Walker, M. P. (2011). The role of sleep in directed forgetting and remembering of human memories. *Cerebral Cortex*, *21*(11), 2534–2541. http://doi.org/10.1093/cercor/bhr034

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(5), 460–471. http://doi.org/10.1037/0278-7393.5.5.460

Taylor, T. L. (2005). Inhibition of return following instructions to remember and forget. *The Quarterly Journal of Experimental Psychology*, *58*(4), 613–629. http://doi.org/10.1080/02724980443000115

Thompson, K. M., Hamm, J. P., & Taylor, T. L. (2014). Effects of memory instruction on attention and information processing: Further investigation of inhibition of return in item-method directed forgetting. *Attention, Perception & Psychophysics*, *76*(2), 41–6. http://doi.org/10.3758/s13414-013-0584-0

Van Hooff, J. C., Whitaker, T. A., & Ford, R. M. (2009). Directed forgetting in direct and indirect tests of memory: Seeking evidence of retrieval inhibition using electrophysiological measures. *Brain and Cognition*, *71*(2), 153–164. http://doi.org/10.1016/j.bandc.2009.05.001

Watkins, O. C., & Watkins, M. J. (1975). Build up of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology Human Learning and Memory*, *1*, 442–452.

Wylie, G. R., Foxe, J. J., & Taylor, T. L. (2008). Forgetting as an active process: An fMRI investigation of item-method-directed forgetting. *Cerebral Cortex*, *18*(6), 670–682. http://doi.org/10.1093/cercor/bhm101

Zacks, R. T., Radvansky, G. A., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 143–156. http://doi.org/10.1037/0278-7393.22.1.143

Table 1

Proportional changes in storage and retrieval due to forgetting instructions as a function of group and context-reinstatement in Experiment 1.

	Category-Cue Group		**No-Category-Cue Group**			
	M	*SE*	*M*	*SE*	ΔG^2	*p*
***a** parameter**						
R-F_{UR} change	0.385	0.038	0.472	0.042	2.349	.125
R-F_{SR} change	0.506	0.042	0.451	0.041	0.880	.348
***r** parameter**						
R-F_{UR} change	0.452	0.073	0.429	0.064	0.056	.811
R-F_{SR} change	0.714	0.075	0.509	0.069	4.007	.045

Note. Storage (a^*) and retrieval (r^*) change parameters reflect the proportional change in storage and retrieval of unrelated TBF-items (R-F_{UR}) and semantically related TBF-items (R-F_{SR}) relative to TBR-items. Thus, $a^* = a(\text{TBF})/\ a(\text{TBR})$ and $r^* = r(\text{TBF})/\ r(\text{TBR})$, for both unrelated and semantically related TBF-items. Notably, the higher (lower) a^* and r^* parameter estimates are the less (more) relative forgetting they indicate. The Category-Cue Group received a category cue prior to the free-recall test; the No-Category-Cue Group did not receive a category cue. All comparisons with 1 *df*.

Table 2

Proportional changes in storage and retrieval due to forgetting instructions and semantic relatedness as a function of group and context-reinstatement in Experiment 2.

	Category-Cue Forget Group		**No-Category-Cue Forget Group**			
	M	*SE*	*M*	*SE*	ΔG^2	*p*
***a** parameter**						
R_{UR}-F_{UR} change	0.444	0.050	0.521	0.046	1.225	.268
R_{UR}-F_{SR} change	0.408	0.047	0.528	0.047	3.124	.077
***r** parameter**						
R_{UR}-F_{UR} change	0.495	0.092	0.499	0.071	0.000	.975
R_{UR}-F_{SR} change	0.699	0.105	0.439	0.068	4.506	.033

	Category-Cue Remember Group		**No-Category-Cue Remember Group**			
	M	*SE*	*M*	*SE*	ΔG^2	*p*
***a** parameter**						
R_{UR}-F_{UR} change	0.553	0.066	0.494	0.059	0.441	.506
R_{SR}-R_{UR} change	0.940	0.083	0.906	0.073	0.093	.759
***r** parameter**						
R_{UR}-F_{UR} change	0.834	0.132	0.797	0.119	0.042	.836
R_{SR}-R_{UR} change	0.708	0.070	0.986	0.099	5.510	.018

Note. Storage (a^*) and retrieval (r^*) change parameters reflect the proportional change in storage and retrieval of unrelated TBF-items (R_{UR}-F_{UR}) and semantically related items (R_{UR}-F_{SR} for the Forget groups; R_{SR}-R_{UR} for the Remember groups) relative to TBR-items. Thus, $a^* = a(\text{TBF})/\ a(\text{TBR})$ and $r^* = r(\text{TBF})/\ r(\text{TBR})$, for unrelated TBF-items of all groups, $a^* = a(\text{TBF}_{SR})/\ a(\text{TBR})$ and $r^* = r(\text{TBF}_{SR})/\ r(\text{TBR})$ for semantically related TBF-items of the Category-Cue Forget and No-Category-Cue Forget group, and $a^* = a(\text{TBR})/\ a(\text{TBR}_{SR})$ and $r^* = r(\text{TBR})/\ r(\text{TBR}_{SR})$ for semantically related TBR-items of the Category-Cue Remember and No-Category-Cue Remember group. Notably, the higher (lower) a^* and r^* parameter estimates are the less (more) relative forgetting or benefits from semantic relatedness they indicate. The Category-Cue Forget group and Category-Cue Remember group received a category cue prior to the free-recall test; the No-Category-Cue Forget group and No-Category-Cue Remember group did not receive a category cue. All comparisons with 1 df.

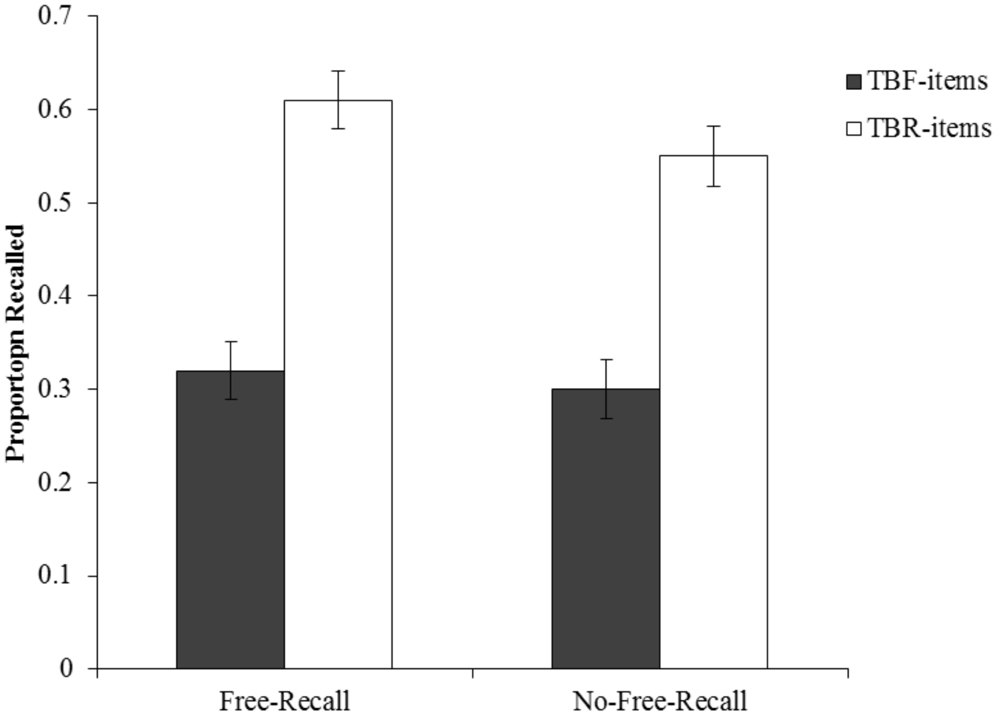


Figure 1. Cued recall probabilities of the Pretest as a function of group (Free-Recall, No-Free-Recall) and item type (to-be-forgotten items (TBF-items), to-be-remembered items (TBR-items)). Error bars represent standard errors.

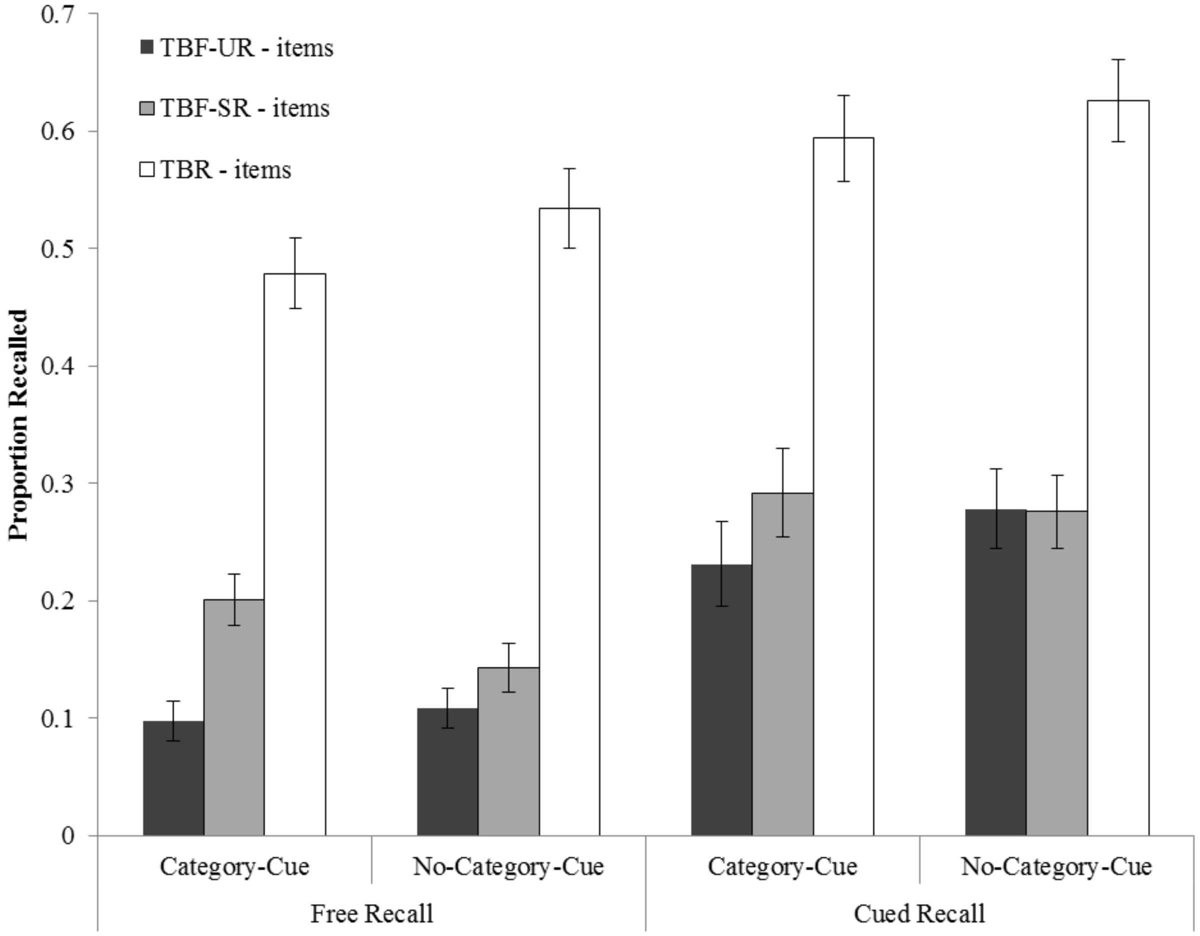


Figure 2. Recall probabilities of Experiment 1 for free recall (left side) and cued recall (right side) as a function of item type (to-be-forgotten unrelated items (TBF_{UR}-items), to-be-forgotten semantically related items (TBF_{SR}-items), to-be-remembered items (TBR-items)) and group. The Category-Cue groups were pre-cued with the semantic category shared by a subset of TBF-items whereas the No-Category Cue groups did not receive this cue. Error bars represent standard errors.

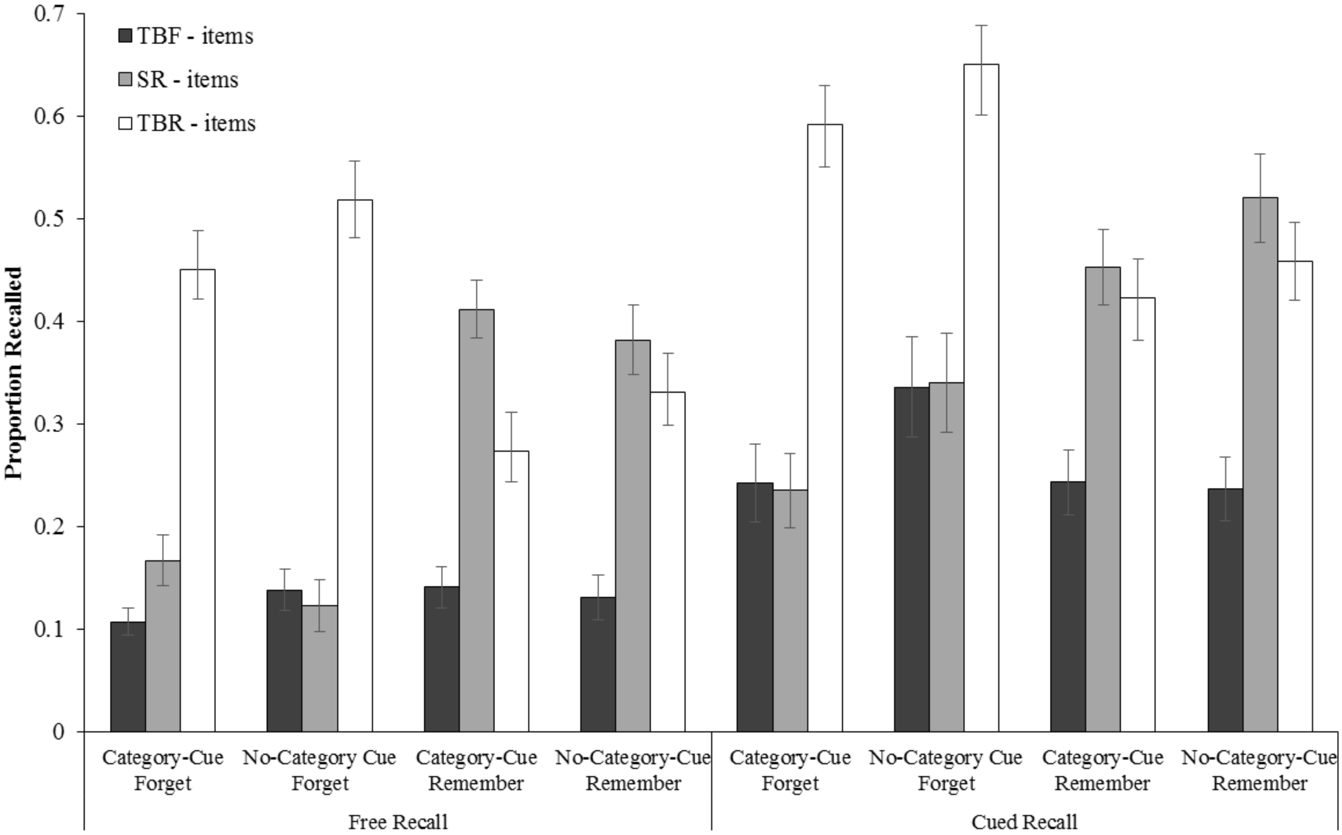


Figure 3. Recall probabilities of Experiment 2 for free recall (left side) and cued recall (right side) as a function of group (Category-Cue Forget, No-Category-Cue Forget, Category-Cue Remember, No-Category-Cue Remember) and item type (to-be-forgotten items (TBF-items), semantically related items (SR-items), to-be-remembered items (TBR-items)). SR-items were post-cued as TBF in the Category-Cue Forget group and No-Category-Cue Forget group and post-cued as TBR in the Category-Cue Remember group and No-Category-Cue Remember group. Error bars represent standard errors.

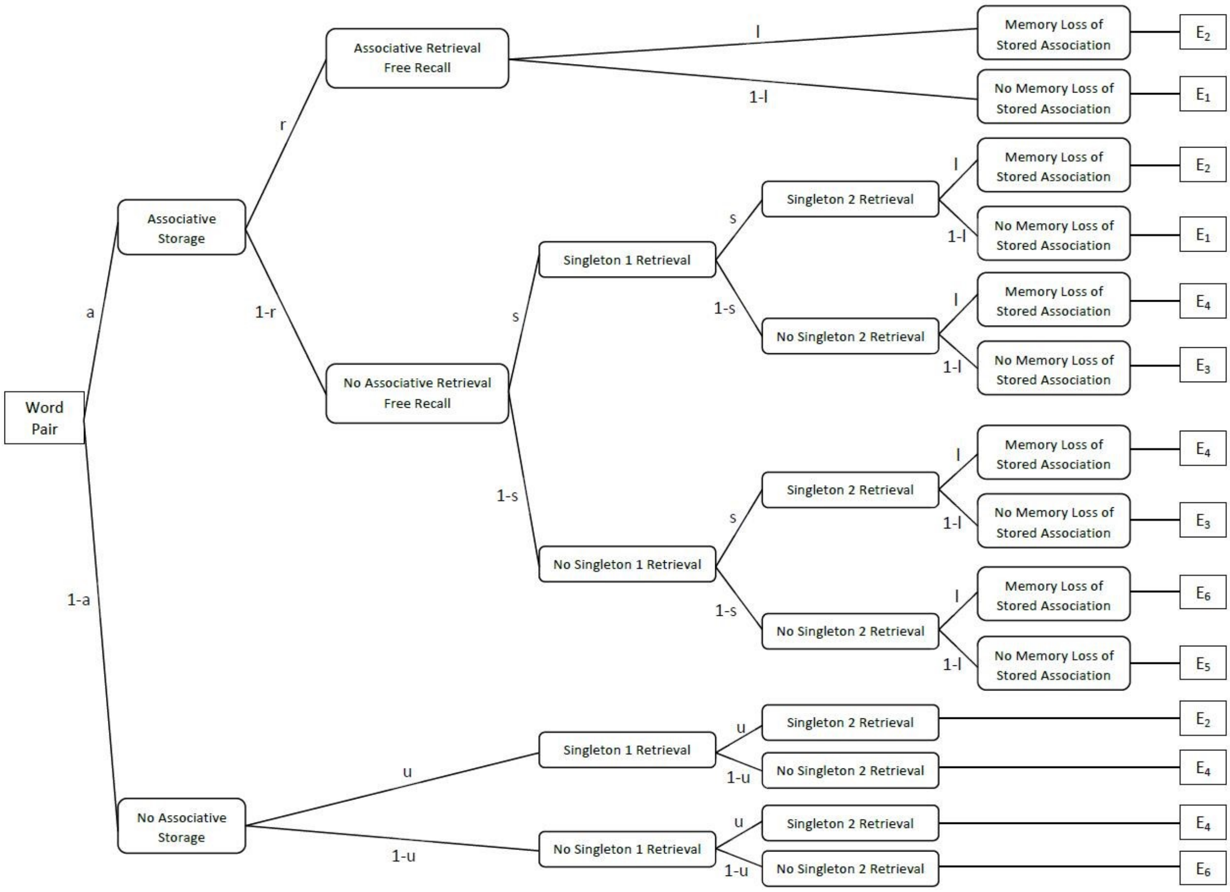


Figure 4. Multinomial processing tree model (MPT) for a free-then-cued-recall paradigm, to separate storage and retrieval processes (based on Rouder & Batchelder, 1998; adopted from Rummel et al., 2016). The processing tree represents the different latent cognitive processes that lead to six observable recall events (E_1-E_6). Rounded rectangles represent latent cognitive states with transition probabilities being described by the model parameters: *a* = probability of associative storage; *r* = probability of associative retrieval; *s* = probability of singleton retrieval given association was stored; *u* = probability of singleton retrieval given association not stored; *l* = memory loss due to time delay between free and cued recall. Parameter *l* was restricted to be equal across all conditions to render the model testable.

Appendices

Appendix A. *Description of the Storage–Retrieval Model and its Application*

The storage–retrieval model used in the present investigation is a multinomial processing tree (MPT) model first proposed by Riefer and Rouder (1992) that can be applied to memory paradigms employing a free-then-cued recall memory test. The model has been used to investigate storage and retrieval processes in many different memory domains, including the bizarreness effect (Riefer & LaMay, 1998; Riefer & Rouder, 1992), lag effects (Küpper-Tetzel & Erdfelder, 2012), age-related memory differences (Riefer & Batchelder, 1991a), clinically related memory deficits (Riefer et al., 2002), and recently directed-forgetting effects (Marevic et al., 2017; Rummel et al., 2016).

To apply the model to the current data, the free-then-cued recall test results of Experiments 1 and 2 were scored as either correct or incorrect paired free recall, singleton free recall, or cued recall. The combination of the resulting recall possibilities yields six possible recall events (E_1-E_6) for each studied item-pair: E_1, successful free recall of the complete pair and successful cued recall; E_2, successful free recall of the complete pair but failed cued recall; E_3, successful free recall of a single item from a pair (singleton) and successful cued recall; E_4, successful free recall of a singleton but failed cued recall; E_5, failed free recall but successful cued recall; E_6, failed free recall and failed cued recall. From these outcome frequencies the model parameters (see Figure 4) were estimated:

Associative storage: Storing and maintaining an item-pair association until the free recall memory test. These processes occur with probability a, $0 \leq a \leq 1$.

Associative retrieval: Retrieval of both items of a pair, given the pair was stored in the first place. The pair does not necessarily need to be retrieved associatively, as singleton-linked retrieval is also possible. The model does not differentiate between these two types of associative retrieval. These retrieval processes occur with probability r, $0 \leq r \leq 1$.

Stored singleton retrieval: Retrieval of only one item of a previously stored pair. These processes occur with probability s, $0 \leq s \leq 1$.

Memory loss of stored association: Even though a successfully free recalled association is assumed to be associatively stored, memory loss from free to cued recall can nevertheless occur. The probability l, $0 \leq l \leq 1$ accounts for such memory loss.

Non-stored singleton retrieval: If an item-pair is not stored associatively, singletons from the pair can still be stored and retrieved independently. These processes occur with probability u, $0 \leq u \leq 1$.

The model has five parameters (a, r, s, u, and l), resulting in 6 x 5 = 30 parameters in total. There were six event categories per condition and therefore 5 degrees of freedom per condition, totaling 6 x 5 = 30 degrees of freedom. If the number of parameters and the number of degrees of freedom are equal, model parameters can be estimated but model fit cannot be assessed (Riefer & Batchelder, 1991b). As in Experiments 1 and 2 the cued recall test immediately followed the free recall test, l parameter estimates that reflect memory loss from free recall to cued recall should be similar between conditions and generally close to zero. Therefore, in the current studies, the l

parameter was set to be equal across all item types and groups. In Experiment 2, the additional restriction of setting the *s* parameter to be equal across groups for each item type was applied.

In fitting our data to the storage–retrieval model by using the software *multiTree* (Moshagen, 2010), the following steps were performed (see Erdfelder et al., 2009 for a detailed description of hypothesis testing with MPT models): First, parametric order constraints were imposed that allow to evaluate proportional changes from one item type to the other (e.g., $a(\text{TBF}_{\text{UR}}\text{-items}) < a(\text{TBR-items})$ for the $a^*(\text{R-F}_{\text{UR}}$ difference) for each experimental group. Next, the parameters of the model were estimated through minimization of the log-likelihood ratio statistic $G^2(df)$ using the expectation-maximization (EM) algorithm (Hu & Batchelder, 1994). In a third step, we assessed the model's fit to our data by comparing the previously minimized G^2 statistic against the $\chi^2(df)$ statistic. If the fit statistic falls below the $(1 - \alpha)$ percentage of the distribution, the model is retained. In our case the model fitted the data well in Experiment 1, $G^2(5) = 4.53$, $p = .475$, as well as in Experiment 2, $G^2(20) = 27.76$, $p = .114$. The last step involves hypothesis testing through parameter comparison. In doing so, we imposed restrictions on the parameters of interest and compared the resulting restricted version of the model with the superordinate model by assessing the ΔG^2 difference statistic (Batchelder & Riefer, 1999). Testing each of our hypotheses this way allowed us to draw interpretations from the results about the underlying cognitive processes of interest (e.g., storage and retrieval).

Appendix B. *Recall frequencies of all recall events by Experiment*

Condition	E_1	E_2	E_3	E_4	E_5	E_6
Experiment 1						
Category-Cue Group						
TBF_{UR}-items	27	3	1	60	13	276
TBF_{SR}-items	51	12	2	58	22	235
TBR-items	151	21	4	71	16	117
No-Category-Cue Group						
TBF_{UR}-items	31	5	2	73	8	261
TBF_{SR}-items	37	4	1	65	18	255
TBR-items	162	16	5	64	21	112
Experiment 2						
Category-Cue Forget Group						
TBF_{UR}-items	21	2	1	46	13	183
TBF_{SR} -items	29	3	2	35	28	183
TBR_{UR}-items	106	3	11	49	23	88
No-Category-Cue Forget Group						
TBF_{UR}-items	35	2	3	63	6	191
TBF_{SR} -items	31	2	1	70	7	189
TBR_{UR}-items	135	8	3	57	22	75
Category-Cue Remember Group						
TBF_{UR}-items	32	1	4	37	15	211
TBR_{SR} -items	98	11	1	37	28	125
TBR_{UR}-items	64	8	5	58	15	150
No-Category-Cue Remember Group						
TBF_{UR}-items	32	1	1	36	9	211
TBR_{SR} -items	87	9	0	64	30	100
TBR_{UR}-items	74	11	4	55	18	128

Note. E_1 = both items freely recalled, correct cued recall; E_2 = both items freely recalled, incorrect cued recall; E_3 = one item freely recalled, correct cued recall; E_4 = neither item freely recalled, correct cued recall; E_5 = one item freely recalled, incorrect cued recall; E_6 = neither item freely recalled, incorrect cued recall. TBF_{UR}-items were items that did not share a superordinate category cue and were post-cued as TBF. TBF_{SR}-items were semantically related items in that they shared a superordinate category cue and were post-cued as TBF. SR-items were semantically related items in that they shared a superordinate category cue; in the Category-Cue Forget and No-Category-Cue Forget group these items were post-cued as TBF and in the Category-Cue Remember and No-Category-Cue Remember group they were post-cued as TBR. TBR-items were unrelated items that were post-cued as TBR.

Appendix C. *Output Order Analysis for Semantically Related Items*

Output order analyses were conducted to test whether semantically related items were recalled earlier in the recall test by the Category-Cue groups than by the No-Category-Cue groups. For these analyses, we counted the number of items that were recalled by each participant, including intrusions (i.e., items that were not studied but incorrectly recalled). Then, for each participant we divided the total number of items recalled by four to obtain four output order bins. Finally, the proportion of semantically related items was determined separately for each of the four bins and submitted to an ANOVA.

In Experiment 1 the proportion of semantically related items recalled did not vary with bins, *F*(3, 222) = .40, $p = .748$, $\eta^2_p = .005$,or groups, *F*(1, 74) = 1.47, $p = .228$, $\eta^2_p = .020$. There was also no interaction between group and bin, *F*(3, 222) = 2.07, $p = .104$, $\eta^2_p = .027$. Thus, participants were generally not more likely to recall TBF_{SR}-items early compared to later during recall (see Figure C-1).

In Experiment 2, the proportion of semantically related items recalled varied with bins, *F*(3, 339) = 6.46, $p < .001$, $\eta^2_p = .05$, and groups, *F*(3, 113) = 24.82, $p < .001$, $\eta^2_p = .40$. There was also a significant interaction between bin and group, *F*(3, 339) = 2.55, $p = .007$, $\eta^2_p = .06$. Post-hoc Tukey tests, however, revealed that the proportion of semantically related items recalled did not differ between bins for the Category-Cue Forget and the No-Category-Cue Forget group, all *p*s > .05. For the Category-Cue Remember group, recall of semantically related items was lower in bin 1 compared to bin 2, *t*(339) = 3.14, $p = .009$, but did not differ compared to bin 3 and bin 4, all *p*s > .05. For the No-Category-Cue Remember group recall of semantically related items was lower in bin 1 compared to bin 2, *t*(339) = 2.51, $p = .05$, and bin 3, *t*(339) = 3.23, $p = .007$, but did not differ compared to bin 4, *t*(339) = 1.80, $p = .273$. Again, these results indicate that participants in the Category-Cue Forget and No-Category-Cue Forget group were not more likely to start recalling SR-items at the beginning of the recall episode and if anything, in the Category-Cue Remember and No-Category-Cue Remember groups recall of semantically related items increased from bin 1 to bin 3 (see Figure C-2).

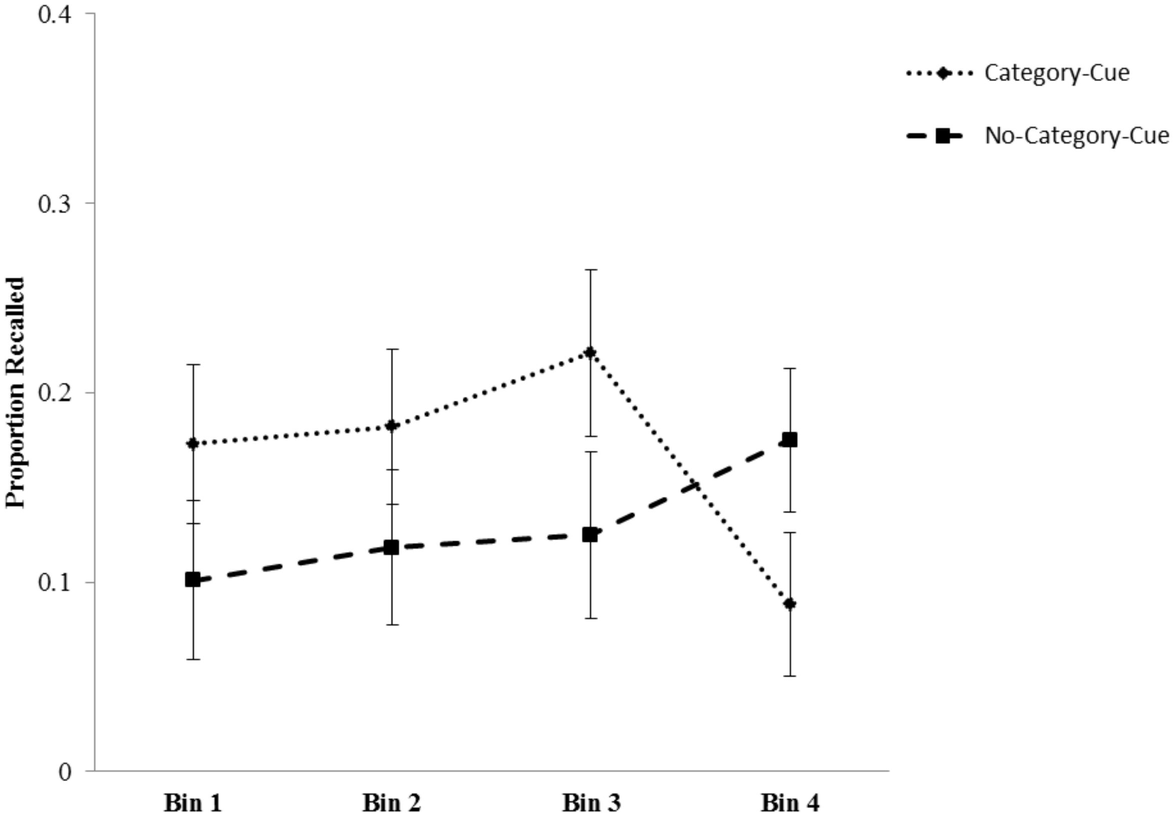


Figure C-1. Recall probabilities of semantically related items (TBF$_{SR}$-items) from Experiment 1 as a function of output order (Bin 1 – Bin 4) and group (Category-Cue, No-Category-Cue). Error bars represent standard errors.

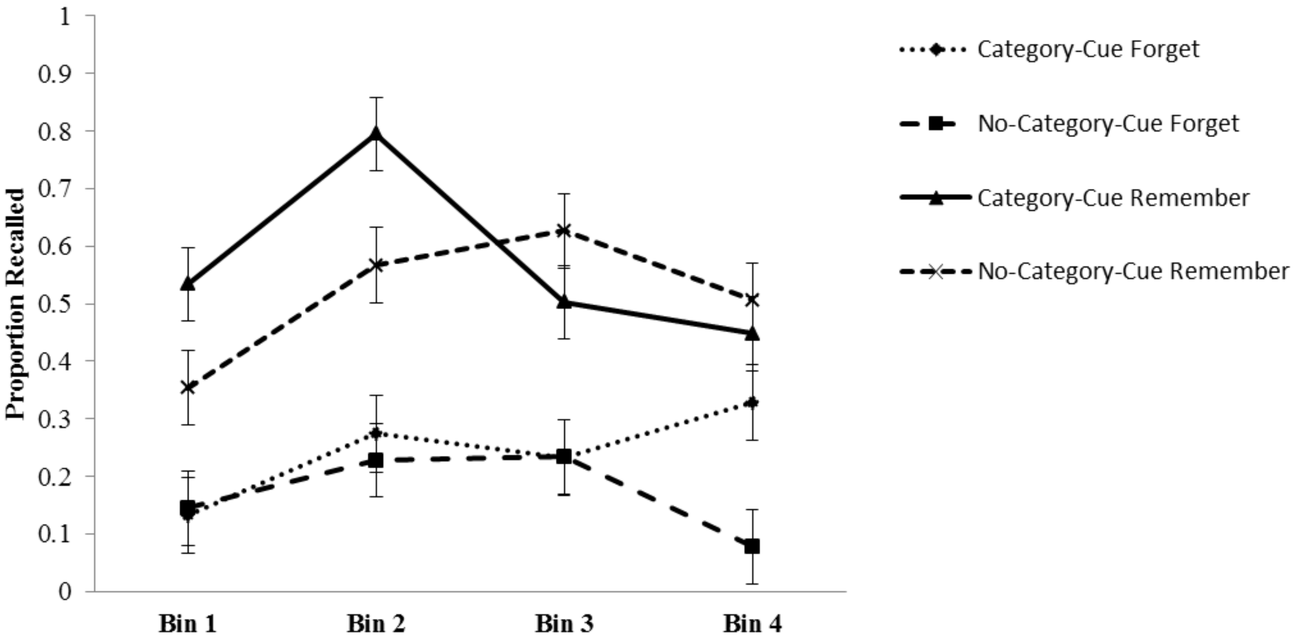


Figure C-2. Recall probabilities of semantically related items (SR-items) from Experiment 2 as a function of output order (Bin 1 – Bin 4) and group (Category-Cue Forget, No-Category-Cue Forget, Category-Cue Remember, No-Category-Cue Remember). Error bars represent standard errors.

Appendix A3 – Article 3

Note:
This is a previous version of an article that has been published in its final version in the *Quarterly Journal of Experimental Psychology* [copyright SAGE Publishing], doi: 10.1080/17470218.2017.1310270

Item-Method Directed Forgetting and Working Memory Capacity: A Hierarchical Multinomial Modeling Approach

Ivan Marevic
Heidelberg University

Nina R. Arnold
University of Mannheim

Jan Rummel
Heidelberg University

Author Note

Ivan Marevic & Jan Rummel, Department of Psychology, Heidelberg University, Heidelberg, Germany. Nina R. Arnold, Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany.

Correspondence concerning this article should be addressed to Ivan Marevic, Department of Psychology, Heidelberg University, Hauptstrasse 47-51, D-69117 Heidelberg, Germany. E-mail: ivan.marevic@psychologie.uni-heidelberg.de

Abstract

Intentional forgetting of information that has recently been encoded is regarded an active and adaptive process and is widely studied using the item-method or the list-method directed forgetting (DF) paradigm. In the present research, we tested whether inter-individual differences in working-memory capacity (WMC), that have been identified as a relevant predictor of DF within the list-method, are also related to stronger DF effects within the item-method. Furthermore, we investigated relationships between WMC and item-method DF at different processing stages by applying the multinomial storage–retrieval model (Riefer & Rouder, 1992) hierarchically to our data. Results showed that individuals with high WMC are better able to store to-be-remembered information than individuals with low WMC; whereas WMC was not related to retrieval of to-be-remembered information or to either storage or retrieval of to-be-forgotten information. Implications for theoretical accounts of item-method DF are discussed.

Keywords: directed forgetting; item-method; working-memory capacity; hierarchical multinomial modeling; storage–retrieval model

Item-Method Directed Forgetting and Working Memory Capacity: A Hierarchical Multinomial Modeling Approach

Forgetting is mostly viewed a passive process with memories decaying over time. However, sometimes people intentionally try to forget information that is outdated, no longer relevant, or that comes from an unreliable source. This type of forgetting is an adaptive process that prevents irrelevant information from interfering with relevant information (Golding & MacLeod, 1998). Recent evidence suggests that intentional forgetting of information that has just been encoded is an active process that is caused by selective rehearsal of to-be-remembered (TBR) information (Bjork, 1972), as well as attention withdrawal (Taylor, 2005; Wylie, Foxe, & Taylor, 2008) or inhibition (Zacks, Radvansky, & Hasher, 1996). In the present research we investigate this issue from an individual difference standpoint, by exploring whether individual's working-memory capacity (WMC), is related to the ability of storing or retrieving information that is intended to-be-remembered (TBR) versus information that is attempted to-be-forgotten (TBF).

WMC has been identified as a very important predictor of cognitive abilities across a variety of cognitive tasks (see Unsworth & Engle, 2007). For example, high-WMC individuals show better performance in higher-order tasks of fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Unsworth, 2010b; Unsworth, Brewer, & Spillers, 2009; Unsworth, Fukuda, Awh, & Vogel, 2014), but also in rather basic memory tasks, like free recall (Spillers & Unsworth, 2011), cued recall (Unsworth, 2009), or prospective memory tasks (Arnold, Bayen, & Smith, 2015; Brewer, Knight, Marsh, & Unsworth, 2010). Interestingly, high-WMC individuals seem not only to be better able to remember information but also more efficient in willingly forgetting information. For instance, high-WMC individuals are better able to suppress thoughts when instructed to do so (Brewin & Beaton, 2002; Wessel, Overwijk, Verwoerd, & de Vrieze, 2008; but see Waldhauser, Johansson, Bäckström, & Mecklinger, 2011) In such forgetting tasks, irrelevant information competes with relevant information that is attempted to be held active in memory and the typical pattern of results involves lower recall of suppressed or inhibited information compared to the actively maintained information. In order to achieve the goal of maintaining relevant information in memory and keeping irrelevant information from interfering, an adaptive mechanism has been proposed (Hasher & Zacks, 1988; Zacks et al., 1996). The maintenance of relevant information is achieved via a general attention component that controls memory contents through monitoring of relevant information and inhibition of irrelevant information. In the literature such adaptive control mechanisms have been conceptualized in different ways (e.g., Central Executive, Baddeley, 2000; attentional control, Unsworth & Engle, 2007; executive functions, Miyake et al., 2000; inhibitory functions, Friedman & Miyake, 2004).

Independent of the exact conceptualization of these cognitive control mechanisms, it seems to be the case that WMC is highly correlated with different facets of cognitive control, such as short-term memory capacity, attention control, retrieval from short-term memory, shifting, updating, or inhibition

(Miyake & Friedman, 2012; Unsworth, 2016). Consequently, it has been argued that high-WMC tend to have better cognitive control abilities than low-WMC individuals (Unsworth, 2016). For this reason, high-WMC individuals should be better able than low-WMC individuals to remember relevant information from long-term memory (Unsworth, 2017) and maybe even to get rid of irrelevant information.

One special situation in which people have to maintain relevant and to adaptively forget irrelevant information is mimicked by the directed-forgetting (DF) paradigm. In this paradigm, participants are presented with two types of items, that is, items that are to be remembered (TBR) and items that are to be forgotten (TBF). Critically, in this paradigm, TBF-information is irrelevant to the task goal of remembering relevant TBR-information and efficient task performance requires keeping as many TBR-items in mind as possible for a later recall test as well as preventing the TBF-items from interfering. Thus, the question arises whether high-WMC individuals perform better than low-WMC individuals in DF tasks and, if so, why that is the case. It is possible that high-WMC individuals are better able to focus on the rehearsal of TBR-items and that the TBF-items are therefore passively dropped from rehearsal (Bjork, 1972) but it is also possible that they are especially efficient in actively withdrawing attention from the TBF-items (attention withdrawal account; Taylor, 2005) or actively inhibiting retrieval of the TBF-items (item inhibition account; Zacks et al., 1996). If the former were the case, high-WMC participants should remember more TBR-items than low-WMC participants in a DF task. If the latter were the case, high WMC participants should also forget more TBF-items than low-WMC participants.

Up to date there are only few studies that addressed this question and all of these studies have focused on the role of WMC in relation to list-method DF. In the list-method paradigm, participants study two item lists and are told after studying the first list, to forget this list (Bäuml, 2008; Bjork, 1989; Sahakyan, Delaney, Foster, & Abushanab, 2014) and high-WMC individuals show more TBF-item forgetting as well as better TBR-item recall than low-WMC individuals in this paradigm (Aslan, Zellner, & Bäuml, 2010; Delaney & Sahakyan, 2007; Soriano & Bajo, 2007). Another widely used DF paradigm is the item-method paradigm. Participants of this paradigm study items and receive instructions to either forget or remember the just encoded item after each item presentation (Basden & Basden, 1998; Fawcett & Taylor, 2010). Because the item status (TBF, TBR) changes item-by-item rather than list-wise, the processes involved in item-method DF seem to differ from those of list-method DF (see Basden, Basden, & Gargano, 1993; Rummel et al., 2016 for a comparison of methods). List-method DF is assumed to be mostly driven by processes that are active during retrieval, that is, inhibition of TBF-items, inhibition of the study context of TBF-items or a mental context change (Bjork, 1989; Pastötter & Bäuml, 2010; Pastötter, Bäuml, & Hanslmayer, 2008; Sahakyan & Kelley, 2002) whereas item-method DF is assumed to be primarily driven by storage processes (such as selective rehearsal of TBR-items). Some researchers assume, however, that retrieval-based forgetting processes also contribute to forgetting in the item-method (Nowicka, Jednoróg, Wypych, &

Marchewka, 2009; Rummel et al., 2016). However, to our knowledge, there is no published study that investigated the role of WMC for item-method DF.

In examining the relationship between WMC and item-method DF we had two aims: First, we were interested in whether WMC generally moderates the item-method DF effect. That is, we wanted to test whether, analogous to the list-method DF and WMC studies, high-WMC individuals exhibit a greater item-method DF effect than low-WMC individuals. This endeavor is also of theoretical importance, because some researchers assume that list-method and item-method DF effects are driven by different cognitive mechanisms (Basden & Basden, 1998; Basden et al., 1993; Sahakyan et al., 2014) and a correlational pattern between item-method DF and WMC that diverges or converges with those found within the list-method paradigm would speak to this issue. As recent evidence from our lab suggests that both storage and retrieval processes are involved in item-method DF (Rummel et al., 2016), we were especially interested in whether the possibly behaviorally observed relation between WMC and TBF and/or TBR-item recall is driven by storage and/or retrieval processes. To shed light on these questions, we applied behavioral analyses and a hierarchical version of the multinomial storage–retrieval model (Riefer & Batchelder, 1991) that we used in previous studies of DF (Marevic & Rummel, submitted.; Rummel et al., 2016) to data from an item-method DF paradigm and correlated estimates of storage (*a*) and retrieval (*r*) parameters for TBR and TBF information with a general estimate of WMC (for an application of the model to DF as well as a detailed description of the model see Rummel et al., 2016). Unlike the standard version, the hierarchical model version takes parameter heterogeneity across individuals into account by drawing each individual's parameter estimates from an overarching distribution. Using this approach parameter estimates are obtained for each individual and can be correlated with WMC.

Second, we were interested in why WMC may moderate the item-method DF effect. It could be that high-WMC individuals display better recall of TBR-items or reduced recall of TBF-items, or both. This question is particularly interesting, as one can derive different predictions regarding these correlations from the selective rehearsal account (Bjork, 1972), the attention withdrawal account (Taylor, 2005), and the inhibition account (Zacks et al., 1996) of item-method DF. According to the selective rehearsal view, WMC should be positively related to the storage of TBR-items, as these items are actively rehearsed while TBF-items are passively dropped from the rehearsal set. Proponents of the attention withdrawal account, on the other hand, argue that following the forget instruction attention is withdrawn from the TBF-item's representation (Taylor, 2005; Thompson, Hamm, & Taylor, 2014) and that this mechanism is mediated by frontal areas of the brain that involve cognitive control (Wylie et al., 2008). Relatedly, according to the inhibitory view of item-method DF, an additional controlled process is engaged to inhibit the retrieval of TBF-items. Thus, if attention is withdrawn form TBF-item representations during the storage process, we would expect an additional negative relation between WMC and storage of TBF-items. If inhibition prevents TBF-items from being retrieved, then we would expect a negative relation between WMC and retrieval of TBF-items.

Method

Participants

One-hundred and thirty-eight students from Heidelberg University (110 female; $M_{age} = 21.96$, range: 18-34 years) participated in the study in exchange for course credit or a monetary compensation. All participants were recruited via the recruitment management tool *hroot* (Bock, Baetge, & Nicklisch, 2014).

Material

Two automated span tasks, that is, the operation span (Ospan; Unsworth, Heitz, Schrock, & Engle, 2005) and the running span (RunSpan; Harrison et al., 2013) tasks were used to assess WMC. In the Ospan task, participants study series of letters of varying length (3 to 7 letters) over 15 blocks. Letters are presented individually on the screen and after each letter presentation participants solve a simple math problem. Following each letter series, participants have to select the previously presented letters in the order they were studied from a set of letters presented on the screen. In the RunSpan task, participants also study series of letters of varying length over five blocks. At the beginning of each letter series, participants are told that they should remember the last x letters of the upcoming series for a subsequent memory test but they are not informed how many letters the series will contain. The number of last letters of a series that have to be recalled (i.e., $x = [3; 7]$) varies for each block. At the end of each letter series, participants are to select from a set of letters the correct amount of last studied letters in the order they were studied in.

For the item-method DF study material, a set of 96 nouns of medium frequency was selected from the *dlex* database (Heister et al., 2011) and 48 word-pairs were formed by randomly pairing two words together. Twenty-four word-pairs were used in a practice block (half as TBF and the other half as TBR-items). For the real task the remaining 24 pairs were randomly divided into two task sets of twelve word-pairs each. Each set served as TBF-items for half of the participants and as TBR-items for the other half. This counterbalancing was applied to control for item specific effects.

Procedure

After providing informed consent, participants were asked to answer some demographic questions. Next, they completed the Ospan task and then started the DF experiment, in which they were asked to study word-pairs for a later memory test. In an initial practice phase, participants were presented with 24 word-pairs for study. Each word-pair was presented in the middle of the screen for seven seconds and half of the pairs were followed by a forget cue (VVV – Initial letters of the German word “forget” [vergessen]), whereas the other half were followed by a remember cue (EEE – initial letters of the German word “remember” [erinnern]). The forget or remember cues were presented in the middle of the screen for 2 seconds and were followed by a 500 ms inter-stimulus-interval. Following the study phase, participants solved math problems for 30 s. For a first memory test, they were asked to freely recall as many items as they could remember that were post-cued as TBR. Next, they completed a cued recall test for TBR-items only. That is, they were presented with the first words of the pairs, one after the other, and had to type in the corresponding ones. The purpose of this practice phase was to acquaint participants with the task and to increase participants' trust in the forget cues. Following this initial practice block, the real DF task started. Here, the procedure was identical to the practice block, with the exception that participants were asked to recall as many items as they could remember during the free recall test, regardless of memory cue (forget vs. remember) and in any order. Then, they were presented with the first words of all pairs (randomly intermixed) for the cued recall test.

Next, participants completed the RunSpan task and participated in an unrelated prospective memory experiment (not reported). Finally, they were debriefed and received their participation compensation.

Results

Partial eta-squared (η^2_p) indicates effect size for analyses of variance (ANOVAs) Cohen's *d* indicates effect size for *t* tests, and Bayes Factors (*BF*) are reported in addition to *p*-values for each analysis. An alpha level of .05 was adopted for all analysis. Pearson's *r* was used as the correlation coefficient for the correlational analyses. The raw data can be obtained from the data repository Zenodo (https://doi.org/10.5281/zenodo.259352).

Behavioral analyses

Behavioral analyses were conducted with JASP, using the default settings (The JASP Team, 2016). The proportion of items recalled during free recall and during cued recall were calculated for each item type separately. To achieve a more reliable WMC measure we z-standardized the OSpan and RunSpan scores and calculated their mean. As both measures have been shown to have good validity and reliability and to be highly correlated, the resulting WM composite score (WM_{comp}) provides a more reliable WMC indicator than the separate measures (Broadway & Engle, 2010; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Unsworth et al., 2005).

Free recall. To test whether a DF effect was present in free recall, the free recall proportions of TBF and TBR-items were submitted to a repeated-measures ANOVA with item type (TBF, TBR) as the within-subjects factor. This analysis showed a significant effect of item type, $F(1, 136) = 291.23$, $\eta^2_p = .682$, $p < .001$ ($BF_{10} = 3.69 \times 10^{41}$), reflecting a robust DF effect. Next, we included the WM_{comp} score as a covariate into the model, to test whether individuals' variance in WM can account for parts of the observed DF effect. In doing so, we compared the full model that included WM_{comp} to the model containing the factor item type only. The analysis revealed a significant interaction between item type and the WM_{comp} score, $F(1, 136) = 4.75$, $\eta^2_p = .034$, $p = .031$ ($BF_{10} = 10.74$) indicating that WMC moderated the size of the DF effect. An additional correlational analysis revealed that WM_{comp}

was positively correlated with TBR-item recall, $r = .26$, $p = .002$ ($BF_{10} = 11.74$), but not correlated with TBF-item recall, $r = .10$, $p = .229$ ($BF_{10} = 0.21$) (see Table 1).[17]

Cued recall. Analogously to the free recall data analysis, the cued recall proportions of TBF and TBR-items were submitted to a repeated-measures ANOVA with item type (TBF, TBR) as the within-subjects factor. The analysis revealed a significant effect of item type, $F(1, 136) = 189.29$, $\eta^2_p = .582$, $p < .001$ ($BF_{10} = 9.58 \times 10^{24}$), reflecting a robust DF effect. Next, we included WM_{comp} as a covariate to test whether variation in WM can account for parts of the DF effect. Again, we compared the full model that included WM_{comp} to the model containing the factor item type only. The analysis revealed a significant interaction between item type and WM_{comp}, $F(1, 136) = 3.92$, $\eta^2_p = .028$, $p = .05$ ($BF_{10} = 20.24$). As in the free recall data, TBR-item recall, $r = .30$, $p < .001$ ($BF_{10} = 60.97$), but not TBF-item recall, $r = .14$, $p = .094$ ($BF_{10} = 0.42$), correlated positively with the WM_{comp} score (see Table 1 for all correlations).[18]

Model-based analyses

For the model-based analyses the storage–retrieval model (Riefer & Rouder, 1992; Rouder & Batchelder, 1998) was applied to the present data (see Figure 1 for a depiction of the multinomial processing tree (MPT) of the model). Six possible categories of events can be derived from the free-then-cued-recall paradigm: successful free and cued recall of the complete pair (E_1); successful free recall of the complete pair but failed cued recall (E_2); successful free recall of a single item from a pair (singleton) and successful cued recall (E_3); successful free recall of a singleton but failed cued recall (E_4); failed free recall but successful cued recall (E_5); failed free recall and failed cued recall (E_6). These outcomes (E_1-E_6) are then modelled as a function of five latent cognitive states which reflect the five model parameters. The *storage parameter a* indicates the probability of associatively storing an item pair and maintaining it in memory until test. The *retrieval parameter r* indicates the probability

---

[17] Additional between-subjects analyses for the free recall rates, where participants with WM_{comp} scores of one standard deviation (SD) below and above the group mean where classified as low and high WMC participants, revealed significantly higher recall rates in high compared to low WMC participants for TBR-items, $t(47) = 2.64$, $d = .75$, $p = .011$ ($BF_{10} = 4.44$), but not for TBF-items, $t < 1$.

[18] Analogously to the free recall analysis, a between-subjects analyses with cued recall rates as dependent measures, where participants with WM_{comp} scores of one SD below and above the group mean where classified as low and high WMC participants, revealed significantly higher recall rates in high compared to low WMC participants for TBR-items, $t(47) = 2.52$, $d = .72$, $p = .015$ ($BF_{10} = 3.54$), but not for TBF-items, $t < 1$.

of successful retrieval of a stored item pair during free recall (for a detailed description of the model and its application to DF see Marevic & Rummel, submitted; Rummel et al., 2016)[19].

In traditional MPT analyses data are aggregated over items and participants. This implies that there is one set of parameters for all participants and thus homogeneity for items and participants is assumed. The homogeneity assumption implies that data from different item types and participants are considered to be independent and identically distributed. However, homogeneity assumptions may often lead to biased parameter estimates (Klauer, 2006, 2010, Smith & Batchelder, 2008, 2010). As the traditional approach only yields group level parameter estimates, and often violates parameter homogeneity, recent approaches have addressed this issue by drawing each participant's parameters from an overarching distribution. In the latent-trait approach individual parameters are drawn from a multivariate normal distribution of probit transformed parameters (Klauer, 2010). By using Bayesian modeling techniques that treat parameters as random variables and update a prior distribution for each parameter to a posterior distribution given the data (Bayes' theorem), individual parameter distributions can be obtained. The Bayesian Credibility Interval (BCI) of these parameter distributions indicates the range of values where the true estimates are found with a 95% confidence. By using Markov chain Monte Carlo (MCMC) algorithms a large number of draws can be obtained from the posterior distribution. These draws allow us to arrive at a posterior distribution for each parameter that is informative of the individual model parameter's properties.

For the model-based analyses we conducted traditional MPT analyses on aggregated data with the software *multiTree* (Moshagen, 2010) and hierarchical modeling with the latent-trait approach (Klauer, 2010) using the software *TreeBUGS* (Heck, Arnold, & Arnold, submitted).

Traditional MPT analysis. First, we aggregated frequencies of the six possible recall events (E_1-E_6) for each item type (see Table 2). As cued recall always directly followed free recall, estimates of the probability of memory loss (l) from free to cued recall should be similarly high and close to zero for both item types. Thus, we set the l-parameter to be equal for all conditions in order to achieve model identifiability (cf. Marevic & Rummel, submitted; Rummel et al., 2016). This restricted model had one degree of freedom and fitted the data well $G^2(1) = .57$, $p = .448$. Further, we obtained a reliable DF effect in terms of higher storage (a), $\Delta G^2(1) = 255.83$, $p < .001$ and retrieval (r), $\Delta G^2(1) = 110.10$, $p < .001$, estimates for TBR than for TBF-items (group parameter estimates are presented in Table 3). Thus, we successfully replicated earlier findings from our lab showing that the item-method DF effect is driven by both storage and retrieval processes (Marevic & Rummel, submitted; Rummel et al., 2016).

---

[19] The model parameters s, u, and l can be regarded as nuisance parameters as they describe possible but very rare psychological states that are not of interest for the present investigation. Parameter s and u indicate the probability of singleton recall of items stored as pairs or singletons. Parameter l indicates the probability of memory loss form free to cued recall.

Hierarchical MPT analysis. We calculated model parameters for the storage–retrieval model hierarchically using *TreeBUGS* (Heck et al., submitted). *TreeBUGS* uses the basic Markov chain Monte Carlo (MCMC) algorithm implemented in JAGS (Plummer, 2003). The algorithm was run with 1,000,000 iterations retaining every 300th sample and 2,000 iterations as burn-in period. We assessed convergence using the potential scale reduction factor $\hat{R}$. For all parameters, $\hat{R} < 1.05$, indicating good convergence.

WM_{comp} was entered as predictor for individual parameter estimates of interest (a and r). The BCI for the regression coefficient of WM_{comp} and a estimates for TBR items did not include zero indicating that the WM_{comp} score predicted storage of TBR items. WM_{comp} did not predict a estimates of TBF-items, and also not r estimates of either TBR or TBF-items (see Table 3 for posterior estimates and BCIs).

Discussion

In the present study, we were interested in the role of individual differences in WMC for item-method DF and especially in their relation to storage and retrieval-based DF in the item-method. We found reliable forgetting effects in both free and cued recall. Additionally, the ANCOVA analyses showed that individuals' WMC interacted with the DF effect. This finding is evidence that differences in WMC are related to the amount of forgetting that took place. Additional correlational analyses further showed that WMC was correlated with TBR but not TBF-item recall, suggesting that high-WMC individuals recalled more TBR-items than low-WMC individuals but did not forget more of the TBF-items. The application of the storage–retrieval model to the data lend further support to recent findings from our lab that item-method DF is driven by both storage and retrieval processes. The correlations between the parameter estimates and WMC paralleled and extended the behavioral correlational results by showing a positive correlation of WMC with TBR-item recall but not with TBF-item recall at the storage stage (parameter a). No correlations were found between WMC and retrieval parameters. Thus, the moderating role of WMC for item-method DF seems to result from better storage of TBR information in high-WMC individuals compared to low-WMC individuals.

As common practice in the item-method DF paradigm (Fawcett & Taylor, 2010; Taylor, 2005; see Golding & MacLeod, 1998 for a review), we did not restrict output order in the free recall test, to not interfere with participants' individual retrieval strategies. Additionally, we determined output order for the cued recall randomly. Some memory researchers (Anderson, 2003, 2005; Golding & Gottlob, 2005) that investigate other forms of forgetting (e.g., list-method DF or retrieval-induced forgetting) recently argued that output order of different item types can affect memory performance, such that the recall of one item type produces certain levels of output interference on subsequent recall of the other item type that differs from the output interference level for the reversed test order (e.g., likelihood of recalling TBF-items due to the initial recall of TBR-items is stronger than vice versa in the list-method DF paradigm; Golding & Gottlob, 2005; but see Pastötter, Kliegl, & Bäuml, 2012). Recent evidence from our lab showed, however, that neither storage nor retrieval parameter estimates of TBF or TBR-items are affected by the order in which the items are recalled (see Rummel et al., 2016, Experiment 2). Therefore, we did not expect a systematic output order effect to emerge in the present study, especially because we avoided any systematic output order effect by randomizing cue presentation in the cued recall test.

With regard to the application of the storage–retrieval model one could speculate that the preceding free recall test may have influenced the following cued recall test. Notably, such carry-over effects from free to cued recall have been investigated by others that used the present storage–retrieval model. Riefer and Rouder (1992) for example showed, in a study that investigated storage and retrieval processes in bizarre imagery, that cued recall rates remain the same, regardless of whether free recall preceded cued recall or not. We recently showed that the same applies to the item-method DF paradigm. In our study, participants studied word-pairs in a standard item-method DF paradigm. At test, one group of participants performed a free recall prior to a cued recall test, while another group only performed a cued recall test. Results showed that cued recall rates for both TBF and TBR-items did not differ depending on whether free recall preceded cued recall or not (Marevic & Rummel, submitted). Therefore, we are confident that test type order effects did not influence the current results in any major way.

The application of the storage–retrieval MPT model allowed us to gain more process pure estimates of the cognitive processes involved in item-method DF (see also Rummel et al., 2016). Extending our own previous work, we introduce a latent-trait approach (Klauer, 2010) in the present article. Besides methodological advantages, this approach comes with the great benefit of allowing for investigating the relationship between different DF processes and individual difference factors such as WMC. Thus, the MPT approach presented in this paper may be a useful tool for researchers in the area of DF who want to gain more process pure estimates of the DF effect and relate these to external variables.

Beyond these methodological advances, the present study is also of interest for current theorizing in the DF area. Based on the selective rehearsal account of DF (Basden et al., 1993; Bjork, 1972; MacLeod, 1999), which attributes the DF effect to better storage of TBR than TBF-items, as a result of selectively rehearsing the TBR-items and (passively) dropping TBF-items from the rehearsal set, we predicted a *positive* relationship between WMC and TBR-item storage. The present findings are well in line with this notion. Based on the attention withdrawal account (Taylor, 2005; Thompson et al., 2014) and the item inhibition account (Zacks et al., 1996), that argue that forgetting of TBF-items is a controlled attention-mediated process, we further expected TBF-item storage (due to attention withdrawal) and retrieval (due to TBF-item inhibition) to be *negatively* associated with WMC. But we did not find any evidence for such relations.

Taken together, these results imply that high-WMC individuals store TBR-information in the item-method situation better than low-WMC individuals but they are not more efficient in withdrawing their attention from or inhibiting the retrieval of TBF-items. Even though our results do not support the idea that high-WMC individuals are better at withdrawing attention from TBF-items (Fawcett & Taylor, 2008, 2010, 2012; Quinlan, Taylor, & Fawcett, 2010; Taylor, 2005) or inhibiting the retrieval of TBF-items (Zacks et al., 1996), there is strong empirical evidence for the occurrence of attention withdrawal after the presentation of TBF-items in the item-method DF paradigm. In several studies, Taylor and colleagues investigated whether attention control differed for items that were followed by a forget compared to a remember instruction by interleaving a secondary reaction time (RT) task into the item-method paradigm. Following each memory cue (forget vs. remember), participants had to respond as quickly as possible to a variety of secondary probes. They found slowed RTs as well as reduced inhibition of return (IOR) to secondary probes following TBF compared to TBR-items (Fawcett, Lawrence, & Taylor, 2016; Fawcett, Taylor, & Nadel, 2013; Fawcett & Taylor, 2010; Hourihan, Goldberg, & Taylor, 2007; Taylor, 2005; Taylor & Fawcett, 2012; Thompson et al., 2014; Wylie et al., 2008). The slowed responses on the secondary RT task are evidence for active attention withdrawal from TBF-item representations in memory that should rely on cognitive control mechanisms. But why then did WMC not correlate with TBF-item processing?

One (statistical) concern could be that the variance of TBF-item recall was artificially reduced due to floor performance. This would render it generally difficult to find a meaningful correlation between TBF-item recall and any other measure. However, both free, $t(137) = 12.92$, $d = 1.10$ $p < .001$ ($BF_{10} = 2.19 \times 10^{22}$), and cued recall rates, $t(137) = 13.99$, $d = 1.19$, $p < .001$ ($BF_{10} = 9.81 \times 10^{24}$), were significantly different from zero and the variances of TBF and TBR-item cued recall were furthermore quite similar in size (TBF-item 95% CI [.040; .071]; TBR-item 95% CI [.050 - .087]). TBF-item free recall variance was indeed smaller than TBR-item free recall variance (TBF-item 95% CI [.011; .019]; TBR-item 95% CI [.034; .059]) (see Table 1, diagonal) but—as free and cued recall rates for TBF-items were highly correlated—it seems that there was still enough variance left to potentially achieve substantial correlations in both TBF-item recall rates. For these reasons, we find it unlikely that a near-floor effect was responsible for the null correlation between TBF-item recall and WMC.

Another methodological concern may be that the WMC measure used in the current study does not tap into the specific executive functions (inhibition, shifting; Miyake & Friedman, 2012) that are recruited during attention withdrawal from, or inhibition of TBF-items. Even though the span tasks measures we used have been shown to be strongly related to these executive functions (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Unsworth, 2010a) recent findings suggest that the speed of removal of outdated information, when updating information in memory, is not related to WMC as measured by most span tasks (Ecker, Lewandowsky, & Oberauer, 2014; Ecker, Oberauer, & Lewandowsky, 2014)[20]. Similarly, one could assume that the withdrawal of attention from TBF-items may be mediated by such active removal processes that are not captured by the current standard WMC

---

[20] We thank an anonymous reviewer for raising this point.

tests. Alternatively, it may also be that low-WMC and high-WMC individuals are similarly capable of withdrawing their attention from the TBF-items but differ with regard to how efficiently they use the freed attentional capacity. For example, for high-WMC individuals the forget cue may serve as a prompt to engage in further elaboration on TBR information whereas for low-WMC individuals the forget cue may just signal that they can now lower their attention, allowing their mind to wander. Findings from the mind wandering literature that high-WMC individuals' minds mind-wander generally less than low-WMC individuals (McVay & Kane, 2009; Unsworth & McMillan, 2014) and are, additionally, better able to adjust their task-focus to current task demands (Rummel & Boywitt, 2014) are generally in line with this idea. However, this potential explanation for the present findings remains speculative at this point and further research is needed to test it. Nevertheless, such inter-individual differences in the attention withdrawal mechanism would also be well in line with the present finding that high-WMC store more TBR information than low-WMC individuals.

Finally, the absent relation between WMC and recall of TBF-items in item-method DF lends further support to the notion that the processes engaged in the item-method differ from those in list-method DF. As outlined in the Introduction section, previous research showed that high-WMC individuals are better to inhibit or change context from TBF-items in the list-method of DF compared to low-WMC individuals (Aslan et al., 2010; Delaney & Sahakyan, 2007). As such a pattern was not observed in the present item-method DF study, it seems plausible to assume that the active forgetting processes underlying list-method DF differ from those underlying item-method DF. Cognitive control processes that correlate highly with WMC are likely responsible for forgetting in the list-method; as well as for related phenomena such as retrieval-induced forgetting (Anderson, 2003; Schilling, Storm, & Anderson, 2014). Other processes, however, into which WMC does not tap to the same extent (e.g., removal processes of outdated information in memory; Ecker, Lewandowsky, et al., 2014; Ecker, Oberauer, et al., 2014), may be responsible for forgetting in the item-method. Further research is needed to test this idea.

To conclude, the current findings further bolster recent evidence that item-method DF seems to be driven by both storage and retrieval processes (Rummel et al., 2016) and extends research on DF in showing that the size of the item-method DF effect is greatly influenced by variations of WMC. Higher WMC is associated with an enhanced ability to store TBR information but does not seem to play a major role for the storage or retrieval of TBF-information maybe because different attentional processes are at work for high-WMC and low-WMC individuals when it comes to active forgetting

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. http://doi.org/10.1016/j.jml.2003.08.006

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In C. M. MacLeod & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 159–190). Tokyo: Springer-Verlag.

Arnold, N. R., Bayen, U. J., & Smith, R. E. (2015). Hierarchical multinomial modeling approaches: An application to prospective memory and working memory. *Experimental Psychology*, *62*(3), 143–152. http://doi.org/10.1027/1618-3169/a000287

Aslan, A., Zellner, M., & Bäuml, K.-H. (2010). Working memory capacity predicts listwise directed forgetting in adults and children. *Memory*, *18*(4), 442–450. http://doi.org/10.1080/09658211003742698

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. http://doi.org/10.1016/S1364-6613(00)01538-2

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139–172). Mahwah, NJ: Erlbaum.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603–616. http://doi.org/10.1037/0278-7393.19.3.603

Bäuml, K.-H. (2008). Inhibitory processes. In H. L. Roediger III (Ed.), *Cognitive psychology of memory* (pp. 195–220). Oxford: Elsevier.

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Washington, DC: Winston.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot - Hamburg registration and organization online tool. *European Economic Review*, *71*, 117–120. http://doi.org/10.1016/j.euroecorev.2014.07.003

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition*, *38*(3), 304–311. http://doi.org/10.3758/MC.38.3.304

Brewin, C. R., & Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behaviour Research and Therapy*, *40*(8), 923–930. http://doi.org/10.1016/S0005-7967(01)00127-9

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*(2), 563–70. http://doi.org/10.3758/BRM.42.2.563

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. http://doi.org/10.1016/S0160-2896(01)00096-4

Delaney, P. F., & Sahakyan, L. (2007). Unexpected costs of high working memory capacity following directed forgetting and contextual change manipulations. *Memory & Cognition*, *33*(5), 1074–1082.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language*, *74*, 77–90. http://doi.org/10.1016/j.jml.2013.09.003

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language*, *74*, 1–15. http://doi.org/10.1016/j.jml.2014.03.006

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. http://doi.org/10.1037/0096-3445.128.3.309

Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology: General*, *145*(1), 56–81. http://doi.org/10.1037/xge0000128

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition*, *36*(6), 1168–1181. http://doi.org/10.3758/MC.36.6.1168

Fawcett, J. M., & Taylor, T. L. (2010). Directed forgetting shares mechanisms with attentional withdrawal but not with stop-signal inhibition. *Memory & Cognition*, *38*(6), 797–808. http://doi.org/10.3758/MC.38.6.797

Fawcett, J. M., & Taylor, T. L. (2012). The control of working memory resources in intentional forgetting: Evidence from incidental probe word recognition. *Acta Psychologica*, *139*(1), 84–90. http://doi.org/10.1016/j.actpsy.2011.10.001

Fawcett, J. M., Taylor, T. L., & Nadel, L. (2013). Intentional forgetting diminishes memory for continuous events. *Memory*, *21*(6), 675–94. http://doi.org/10.1080/09658211.2012.748078

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. http://doi.org/10.1037/0096-3445.133.1.101

Golding, J. M., & Gottlob, L. R. (2005). Recall order determines the magnitude of directed forgetting in the within-participants list method. *Memory & Cognition*, *33*(4), 588–594. http://doi.org/10.3758/BF03195326

Golding, J. M., & MacLeod, C. M. (1998). *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ US: Erlbaum.

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*(12), 1–11. http://doi.org/10.1177/0956797613492984

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Heck, D., Arnold, N. R., & Arnold, D. (submitted). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB-eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20. http://doi.org/10.1026/0033-3042/a000029

Hourihan, K. L., Goldberg, S., & Taylor, T. L. (2007). The role of spatial location in remembering and forgetting peripheral words. *Canadian Journal of Experimental Psychology*, *61*(2), 91–101. http://doi.org/10.1037/cjep2007010

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. http://doi.org/10.1037/0096-3445.133.2.189

Klauer, K. C. (2006). Hierarchical multinomial processing tree models - A latent-class approach. *Psychometrika*, *71*(1), 7–31. http://doi.org/10.1007/S11336-004-1188-3

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. http://doi.org/10.1007/S11336-009-9141-0

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*(1), 123–129. http://doi.org/10.3758/BF03210819

Marevic, I., & Rummel, J. (submitted). Retrieval-reliant mechanisms of item-method directed forgetting: New evidence from behavioral and model-based approaches.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The Relationship Between Working Memory Capacity and Executive Functioning: Evidence for a Common Executive Attention Construct. *Neuropsychology*, *24*(2), 222–243. http://doi.org/10.1037/a0017619

Mcvay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *35*(1), 196–204. http://doi.org/10.1037/a0014104

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. http://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychologysychology*, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. http://doi.org/10.3758/BRM.42.1.42

Nowicka, A., Jednoróg, K., Wypych, M., & Marchewka, A. (2009). Reversed old/new effect for intentionally forgotten words: An ERP study of directed forgetting. *International Journal of Psychophysiology*, *71*(2), 97–102. http://doi.org/10.1016/j.ijpsycho.2008.06.009

Pastötter, B., & Bäuml, K.-H. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 54–65. http://doi.org/10.1037/a0017406

Pastötter, B., Bäuml, K.-H., & Hanslmayer, S. (2008). Oscillatory brain activity before and after an internal context change: Evidence for a reset of encoding processes. *NeuroImage*, *43*(1), 173–181. http://doi.org/10.1016/j.neuroimage.2008.07.005

Pastötter, B., Kliegl, O., & Bäuml, K.-H. (2012). List-method directed forgetting: The forget cue improves both encoding and retrieval of postcue information. *Memory & Cognition*, *40*(6), 861–873. http://doi.org/10.3758/s13421-012-0206-4

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, (Dsc), 1–10.

Quinlan, C. K., Taylor, T. L., & Fawcett, J. M. (2010). Directed forgetting: Comparing pictures and words. *Canadian Journal of Experimental Psychology*, *64*(1), 41–6. http://doi.org/10.1037/a0016569

Riefer, D. M., & Batchelder, W. H. (1991). Statistical inference for multinomial tree models. In J.-C. Falmagne & J.-P. Doignon (Eds.), *Mathematical psychology: Current developments* (pp. 313–336). Berlin: Springer-Verlag.

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition*, *20*(6), 601–611. http://doi.org/10.3758/BF03202710

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 195–225). New York: Psychology Press.

Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review*, *21*(5), 1309–1315. http://doi.org/10.3758/s13423-013-0580-3

Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1526–1543. http://doi.org/10.1037/xlm0000266

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2014). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 59, pp. 131–190). Elsevier Inc.: Academic Press. http://doi.org/10.1016/b978-0-12-407187-2.00004-6

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1064–1072. http://doi.org/10.1037//0278-7393.28.6.1064

Schilling, C. J., Storm, B. C., & Anderson, M. C. (2014). Examining the costs and benefits of inhibition in memory retrieval. *Cognition*, *133*(2), 358–70. http://doi.org/10.1016/j.cognition.2014.07.003

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review*, *15*(4), 713–731. http://doi.org/10.3758/PBR.15.4.713

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167–183. http://doi.org/10.1016/j.jmp.2009.06.007

Soriano, M. F., & Bajo, M. T. (2007). Working memory resources and interference in directed forgetting. *Psicológica*, *28*, 63–85.

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1532–1539. http://doi.org/10.1037/a0024852

Taylor, T. L. (2005). Inhibition of return following instructions to remember and forget. *The Quarterly Journal of Experimental Psychology*, *58*(4), 613–629. http://doi.org/10.1080/02724980443000115

Taylor, T. L., & Fawcett, J. M. (2012). Does an instruction to forget enhance memory for other presented items? *Consciousness and Cognition*, *21*(3), 41–6. http://doi.org/10.1016/j.concog.2012.05.002

The JASP Team (2016). JASP (Version 0.7.5.5) [Computer software]. Retrieved from http://jasp-stats.org

Thompson, K. M., Hamm, J. P., & Taylor, T. L. (2014). Effects of memory instruction on attention and information processing: Further investigation of inhibition of return in item-method directed forgetting. *Attention, Perception & Psychophysics*, *76*(2), 41–6. http://doi.org/10.3758/s13414-

013-0584-0

Unsworth, N. (2009). Examining variation in working memory capacity and retrieval in cued recall. *Memory*, *17*(4), 386–96. http://doi.org/10.1080/09658210902802959

Unsworth, N. (2010a). Interference control, working memory capacity, and cognitive abilities: A latent variable analysis. *Intelligence*, *38*(2), 255–267. http://doi.org/10.1016/j.intell.2009.12.003

Unsworth, N. (2010b). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, *134*(1), 16–28. http://doi.org/10.1016/j.actpsy.2009.11.010

Unsworth, N. (2016). The Many Facets of Individual Differences in Working Memory Capacity. In *Psychology of Learning and Motivation* (Vol. 65, pp. 1–46). Elsevier Ltd. http://doi.org/10.1016/bs.plm.2016.03.001

Unsworth, N. (2017). Examining the dynamics of strategic search from long-term memory. *Journal of Memory and Language*, *93*, 135–153. http://doi.org/10.1016/j.jml.2016.09.005

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity-fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin & Review*, *16*(5), 931–937. http://doi.org/10.3758/PBR.16.5.931

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. http://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. http://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. http://doi.org/10.3758/BF03192720

Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, *150*, 14–25. http://doi.org/10.1016/j.actpsy.2014.04.001

Waldhauser, G. T., Johansson, M., Bäckström, M., & Mecklinger, A. (2011). Trait anxiety, working memory capacity, and the effectiveness of memory suppression. *Scandinavian Journal of Psychology*, *52*(1), 21–27. http://doi.org/10.1111/j.1467-9450.2010.00845.x

Wessel, I., Overwijk, S., Verwoerd, J., & de Vrieze, N. (2008). Pre-stressor cognitive control is related to intrusive cognition of a stressful film. *Behaviour Research and Therapy*, *46*(4), 496–513. http://doi.org/10.1016/j.brat.2008.01.016

Wylie, G. R., Foxe, J. J., & Taylor, T. L. (2008). Forgetting as an active process: An fMRI investigation of item-method-directed forgetting. *Cerebral Cortex*, *18*(6), 670–682. http://doi.org/10.1093/cercor/bhm101

Zacks, R. T., Radvansky, G. A., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 143–156. http://doi.org/10.1037/0278-7393.22.1.143

Table 1

Means (M), standard deviations (diagonal), and correlations of the working memory score and dependent measures of the present study

	M	WM_{comp}	TBF_{FR}	TBR_{FR}	TBF_{CR}	TBR_{CR}
WM_{comp}	**.0079**	**.86**				
TBF_{FR}	**.11**	.103	**.10**			
TBR_{FR}	**.39**	$.260^{**}$	.103	**.18**		
TBF_{CR}	**.23**	.143	$.640^{***}$	.100	**.19**	
TBR_{CR}	**.49**	$.301^{***}$	.165	$.653^{***}$	$.424^{***}$	**.21**

Note. WM_{comp} = working memory composite score from z-standardized Ospan and RunSpan scores (sum of correct serial letter recall); TBF =to-be-forgotten items; TBR = to-be-remembered items; R-F = amount of forgetting as indexed by TBR-TBF recall probability; FR = free recall; CR = cued recall.

** $p < .01$

*** $p < .001$

Table 2

Aggregated recall frequencies of to-be-forgotten (TBF) and to-be-remembered (TBR) items for all recall events

Item Type	E_1	E_2	E_3	E_4	E_5	E_6
TBF-items	135	7	15	237	74	1188
TBR-items	541	39	32	242	121	681

Note. E_1 = both items freely recalled, correct cued recall; E_2 = both items freely recalled, incorrect cued recall; E_3 = one item freely recalled, correct cued recall; E_4 = neither item freely recalled, correct cued recall; E_5 = one item freely recalled, incorrect cued recall; E_6 = neither item freely recalled, incorrect cued recall.

Table 3

Means (M) and standard deviations (SD) of storage (a) and retrieval (r) parameter estimates for TBR-items, and TBF-items, and correlations with WM composite score (WM_{comp})

	Estimates		Correlations with WM_{comp}	
	M (SD)	*95% BCI*	*Posterior* (*SD*)	*95% BCI*
a parameter				
TBF-items	.21 (.01)	[.18 ; .25]	.11 (.07)	[-.02; .25]
TBR-items	.52 (.02)	[.48; .56]	.19 (.05)	[.07; .30]
r parameter				
TBF-items	.34 (.02)	[.29; .40]	-.03 (.07)	[-.19; .11]
TBR-items	.68 (.02)	[.63; .72]	.05 (.07)	[-.09; .19]

Note. WM_{comp} = working memory composite score from z-standardized Ospan and RunSpan scores; *a* parameter = storage estimate for TBF-items, and TBR-items; *r* parameter = retrieval estimate for TBF-items, and TBR-items; Posterior = posterior distribution of the correlation with mean and 95% Bayesian Confidence Interval (*BCI*).

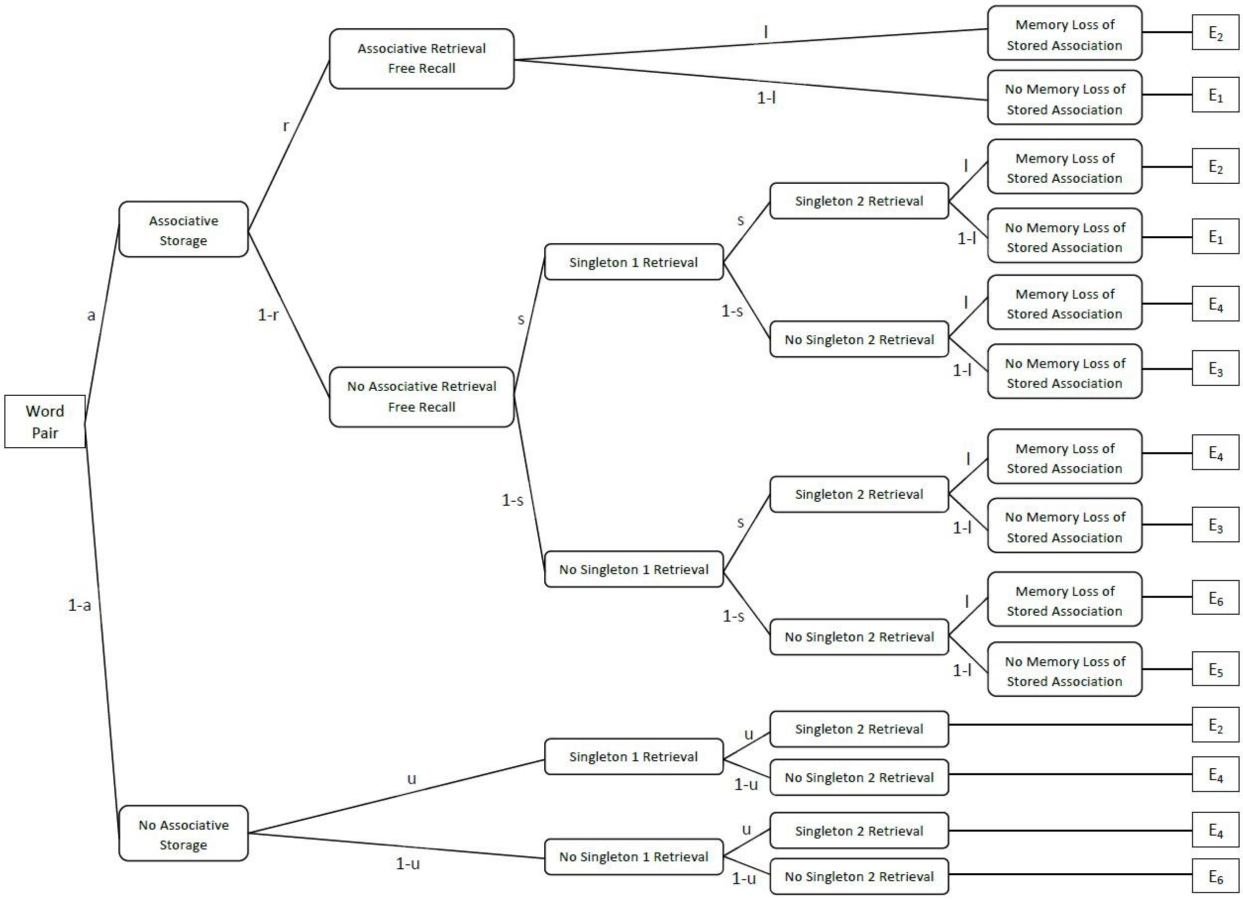


Figure 1. Multinomial processing tree model (MPT) for a free-then-cued-recall paradigm, to separate storage and retrieval processes (based on Rouder & Batchelder, 1998; adopted from Rummel et al., 2016). The processing tree represents the different latent cognitive processes that lead to six observable recall events (E_1-E_6). Rounded rectangles represent latent cognitive states with transition probabilities being described by the model parameters: *a* = probability of associative storage; *r* = probability of associative retrieval; *s* = probability of singleton retrieval given association was stored; *u* = probability of singleton retrieval given association not stored; *l* = memory loss due to time delay between free and cued recall. Parameter *l* was restricted to be equal across both item-types to render the model testable.

Erklärung gemäß § 8 Abs. 1 Buchst. c) und d) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften

---

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

FAKULTÄT FÜR VERHALTENS- UND EMPIRISCHE KULTURWISSENSCHAFTEN

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg
Doctoral Committee of the Faculty of Behavioural and Cultural Studies, of Heidelberg University

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Declaration in accordance to § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.
I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Declaration in accordance to § 8 (1) d) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.
I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname
First name Family name ________ Ivan Marevic ________

Datum, Unterschrift
Date, Signature ____________________