Forgetting is a Feature, not a Bug: Intentionally Forgetting Some Things Helps Us Remember Others by Freeing up Working Memory Resources

Vencislav Popov[1,2,*], Ivan Marevic[3,*], Jan Rummel[3], & Lynne M. Reder[1,2]

[1] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA
[2] Center for the Neural Basis of Cognition, Pittsburgh, PA
[3] Department of Psychology, Heidelberg University, Heidelberg, Germany

* Equal contributions
Correspondence: vencislav.popov@gmail.com

Abstract

We used an item-method directed forgetting paradigm to test whether instructions to forget or to remember one item in a list affects memory for the subsequent item in that list. In two experiments, we found that free and cued recall were higher when a word-pair was preceded during study by a to-be-forgotten (TBF) word pair. This effect was cumulative – performance was higher when more of the preceding items during study were TBF. It also interacted with lag between study items – the effect decreased as the lag between the current and a prior item increased. Experiment 2 used a dual-task paradigm in which we suppressed either verbal rehearsal or attentional refreshing during encoding. We found that neither task removed the effect, thus the advantage from previous TBF items could not be due to rehearsal or attentional borrowing. We propose that storing items in long-term memory depletes a limited pool of resources that recovers over time, and that TBF items deplete fewer resources, leaving more available for storing subsequent items. A computational model implementing the theory provided excellent fits to the data.

Keywords: directed forgetting; item-method; directed-forgetting after-effects; computational modeling

I. Introduction

Associative memory formation is an effortful process that can be disrupted by many factors such as reduced study time (Malmberg & Nelson, 2003), divided attention (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996), aging (Castel & Craik, 2003), or instructions to forget (Bjork, 1972). The probability of forming novel associative memories also decreases with the difficulty of the material – for example, serial, free and cued recall, as well as associative recognition, are worse for low frequency words (e.g. Criss, Aue, & Smith, 2011; Hulme, Stuart, Brown, & Morin, 2003; Ward, Woodward, Stevens, & Stinson, 2003; for a review, see Popov & Reder, 2018), and the presence of low frequency words on a study list hurts memory for other items on the same list (Diana & Reder, 2006; Ozubko & Joordens, 2007). The ability to form long-term associative memories is also dependent on the individual's working memory capacity (Marevic, Arnold, & Rummel, 2017; Unsworth & Spillers, 2010).

To explain results like these, we have proposed a theory of episodic memory, according to which binding information together and storing it for long-term use depletes a limited working memory resource that recovers over time (Popov & Reder, 2018; Reder, Liu, Keinath, & Popov, 2016; Reder, Paynter, Diana, Ngiam, & Dickison, 2007; Shen, Popov, Delahay, & Reder, 2018). According to this model, storing weaker items such as low-frequency words requires more resources which leaves fewer resources for processing additional items. Since the resource does not recover immediately, the presence of weaker items on a list hurts memory for subsequent items processed during the same list.

In this article, we test a key prediction of the theory – memory performance for a studied item should be higher, if, during the study sequence, it was preceded by items that required less resources for their processing and storage. We tested this prediction in an item-method directed forgetting (DF) paradigm (Bjork, 1972; Golding & MacLeod, 1998). In this paradigm, items are studied for a later memory test. The items are presented sequentially, and each item is followed by either a forget instruction[1], asking participants to try to forget the previously studied item, or by a remember instruction, asking participants to remember the previously studied item, because it will be tested later. Following a short distractor, participants try to remember as many items as possible, regardless of the instructions that followed each item. Typically, recall of to-be-forgotten (TBF) items is worse than that of to-be-remembered (TBR) items (Bjork, 1972).

Different accounts suggest different cognitive mechanisms underlying the DF effect that are not necessarily mutually exclusive. According to the *selective rehearsal* account, DF results from rehearsing TBR-items during study while dropping TBF-items from rehearsal (Bjork, 1970; Davis & Okada, 1971). The *attention withdrawal* account, argues that TBF-items are not just dropped from rehearsal but that attention is actively withdrawn from TBF-items (Fawcett & Taylor, 2008; Taylor, 2005). Finally, the *executive inhibition* account posits that TBF-items are inhibited after they are stored (Geiselman & Bagheri, 1985; Zacks, Radvansky, & Hasher, 1996). Recent research further suggests that different cognitive mechanisms may underlie both the storage and retrieval of TBF and TBR items (Marevic et al., 2017).

Item-method DF studies have focused on the effects of TBF versus TBR instructions on recall of the preceding items and have found worse TBF than TBR recall (i.e., a DF effect). To our knowledge the effects of forget versus remember memory instructions on recall of subsequent items has not been investigated previously using the item-method (i.e., a DF after-effect). Investigating the effects and after-effects of memory instructions on recall in an item-method DF paradigm can shed new light on the role of working memory resources for remembering and forgetting information. According to the Resource Depletion Theory presented here, participants use more resources to store TBR compared to TBF items, which results in a stronger memory trace for TBR items. Specifically, we propose that before the instruction (TBR/TBF) appears, participants process each item at some base level, spending a proportion of the existing resource pool. Then, after the instruction appears, participants continue to process TBR items, but not TBF items, which results in more resources being depleted on trials with TBR items. As a result, more of the resources would be available for processing subsequent items, when they are preceded by one or more TBF items compared with one or more TBR items.

Early research with the list-method version of DF provides some support for this idea (for a review, see Epstein, 1972). For example, Bjork (1970) found that giving instructions in the middle of a study list to forget the items presented so far enhances memory for the second half of that list. While this result is consistent with our theory, it does not strongly constrain the range of possible explanations and, for example, Bjork argued that this memory advantage is due to participants not rehearsing the TBF part of the list while processing the second half. Furthermore, different mechanisms might underlie the list-method and item-method of DF (Basden, Basden, & Gargano, 1993; Rummel, Marevic, & Kuhlmann, 2016) and it is an open

[1] In the DF literature, these are often called forget cues and remember cues. In this paper we use the term forget instructions, remember instructions or instructions type, to avoid confusion when using the term cue to refer to the recall cue in the cued-recall tests used in the experiments.

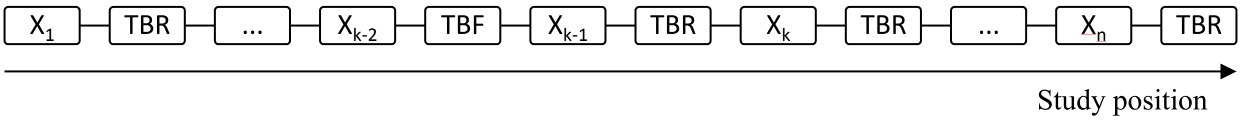

Figure *1*. *Order of items during study*

question whether DF after-effects will occur with the item-method. Finally, as we show below, investigating DF after-effects with the item-method DF allows us to characterize this phenomenon in much greater detail, and to discount alternative explanations for the results.

We can make several predictions concerning DF after-effects based on the resource depletion account. Consider Figure 1 which depicts a study-item sequence. We predict that memory for item X_k, $P(X_k)$, will be a function of for the instructions type given for the preceding items X_{k-1}, X_{k-2}, X_{k-i}, … where k denotes the position of the current item and i denotes the lag to the preceding item (e.g. lag of 2 indicates that the X_{k-2} item appeared two items ago). Specifically: 1) $P(X_k)$ will be higher when X_{k-1} is TBF; 2) these effects should be cumulative. That is, the more of the preceding items are TBR, the worse $P(X_k)$ will be, because more resources would have been spent; 3) this effect will interact with the lag i between study items – the effect of X_{k-1}'s cue should be stronger than the effect of X_{k-2}'s and in general the effect of the preceding items would decrease as the lag increases.

We tested these predictions in two experiments. The first involves a reanalysis of the data from Marevic et al. (2017) in which participants studied word pairs in an item-method DF paradigm and had to recall as many items as possible. The second experiment is a new report involving a dual-task procedure that allows us to discriminate among alternative explanations for the pattern of results we have predicted. Specifically, we examined whether suppressing rehearsal or dividing attention while concurrently performing the item-method DF task would negate DF after-effects. To show that the Resource Depletion Theory can capture the precise quantitative pattern of preceding item effects, we also fit a computational implementation of the account, the Source of Activation Confusion (SAC) model of memory, to the data. A full description of the model is available in Popov & Reder (2018). Finally, to further test whether DF after-effects can be traced back to differential item storage, we applied a multinomial storage-retrieval model that provides more precise storage and retrieval measures. Marevic et al. (2017) provide a comprehensive description of the model.

II. Experiment 1 – Reanalysis of Marevic, Arnold, & Rummel (2017)

A. Method

These methods were described in Marevic et al (2017) but are also included here to facilitate comprehension of the new information reported herein. The data, materials and analysis code for the current analysis are available at https://github.com/venpopov/directed-forgetting-after-effects.

1. Participants

There were 138 students recruited from Heidelberg University (110 female, M_{age} = 21.96, range: 19-34 years) and they received course credit or monetary compensation.

2. Materials.

A set of 96 nouns of medium frequency was drawn from the *dlex* database (Heister et al., 2011). Words were randomly paired and assigned to two sets with 24 word-pairs each. One set was used in an initial practice phase and the other was used for the real experimental phase. In order to control for item-specific effects the assignment of word-pair sets to phases was counter-balanced. In each block, half of the word pairs were succeeded by TBF and the other half by TBR instructions. For simplicity, we refer to items followed by TBR instructions as TBR items, and to items followed by TBF instructions as TBF items.

3. Procedure.

Experimental sessions started with basic demographic questions, a working-memory task (not analyzed here but reported in Marevic et al, 2017) and a practice phase. For practice, participants studied 24 TBR and TBF word pairs. Participants were told to only remember the TBR word pairs for a later test. Each word pair was presented for 7 seconds in the center of the screen, followed by either a TBR or TBF instruction for 2 seconds. Trials were separated by 250-ms inter-stimulus-intervals. After all word pairs had been presented, participants solved math problems for 30 seconds before completing a free recall test, that was followed by a cued recall test for TBR-items only. Recall order was randomized for the cued-recall task. This practice phase was intended to familiarize participants with the paradigm and to increase their reliance that the forget instruction was genuine. However, for the real task phase, the procedure was modified so that participants were, again, presented with TBF and TBR items but were asked to recall as many TBR *and* TBF items as possible in the subsequent free and cued-recall tests. Finally, participants performed another working-memory task (not reported), were debriefed, and received their compensation.

B. Data Analysis

For the behavioral and the multinomial analyses, we employed Bayesian statistics. This approach has several advantages (Wagenmakers, Morey, & Lee, 2016) but most important to us is that Bayes Factors (*BF*s) enabled us to quantify the evidence in favor of the null as well as the alternative hypotheses. *BF*s are reported in the direction of the favored model, such that BF_{21} denotes the evidence in favor of model two compared to model one. A $BF > 3$ is conventionally interpreted as moderate evidence and a $BF > 10$ as strong evidence in favor of the preferred model (Lee & Wagenmakers, 2013). We applied multilevel logistic Bayesian regressions as implemented in the *brms* R-package (Bürkner, 2017), in which we included crossed random intercepts for subjects and items, as well as random subject slopes for DF effect and after-effect. All models were run with 10,000 iterations and 5,000 iterations as burn-in. Convergence was assessed using the potential scale reduction factor $\hat{R}$. For all parameters, $\hat{R} < 1.1$, indicating good convergence. The statistical effects were identical across cued and free recall. For simplicity, we report only the analyses of cued-recall results (although the data from both types of test are presented in Figures 2 and 3).

For each item, we coded whether a TBR or TBF item preceded it. Given that the first item of a study sequence had no predecessor, it was not analyzed. In order to measure the cumulative effect of successive cues, we also coded how many consecutive TBR or TBF items preceded each item. We used a coding scheme that varied from -3 (3 or more consecutive TBF items preceded the current item) to +3 (3 or more consecutive TBR items preceded the current item). For example, if the current study item is preceded by a TBF and a TBR item, in that order, it was scored as -1, because there was only one immediately preceding TBF item. Finally, we also looked at the effect of the instructions at each lag individually, without considering other potential intervening items.

C. *Results*

1. Main effect of preceding item type

Figures 2a and *2d* plot the cued and free recall accuracy as a function of the instructions given for the current and the preceding item. There was a DF after-effect, such that both cued and free recall were higher for items that were preceded by TBF items than for those preceded by TBR items (BF_{10} = 2711 for the cued recall model with current and preceding instruction type vs. the null model with only current type). There was no evidence of an interaction between instructions for the preceding item and those for the current item (BF_{12} = 1.66 for the cued recall model with main effects only against the model with an interaction).

2. Cumulative effect of the number of consecutive preceding TBF or TBR items

Figures 2b and 2e show the cued and free recall accuracy as a function of the number of consecutive preceding TBF or TBR items. Both cued and free-recall performance for the current items were higher when it was preceded by a greater number of consecutive TBF items, and lower, when it was preceded by a greater number of consecutive TBR items**.** The model including the current item's instructions and the number of consecutive TBF or TBR preceding items fit the data better than the null model that included only the current item's instructions as a predictor (BF_{10} = 7103 for cued recall). Again, the DF effect and the DF after-effect did not interact (BF_{12} = 6.67 in favor of the cued recall model with main effects only versus the model with an interaction term).

3. Interaction between preceding item type and study position lag

Finally, Figures 2c and 2f plot the cued and free recall accuracy, respectively, as a function of the preceding item type and the lag between that preceding item and the current item on the study list. The plots clearly show that the DF after-effect interacted with the lag between the current item and the preceding item – the immediately preceding item had a stronger effect than the one two trials before, which in turn had a stronger effect than the one three trials before. We compared the full model, which included the instructions for items at lags 1, 2, 3 and 4, to identical models without the factor of interest. The posterior parameter estimates from the final model and the corresponding *BF*'s are reported in Table 1. The DF after-effect from lag 1 was greater than the DF after-effect from lag 2, and the after-effect from lag 3 was greater than the one from lag 4 (see Table 1 – parameter comparisons).

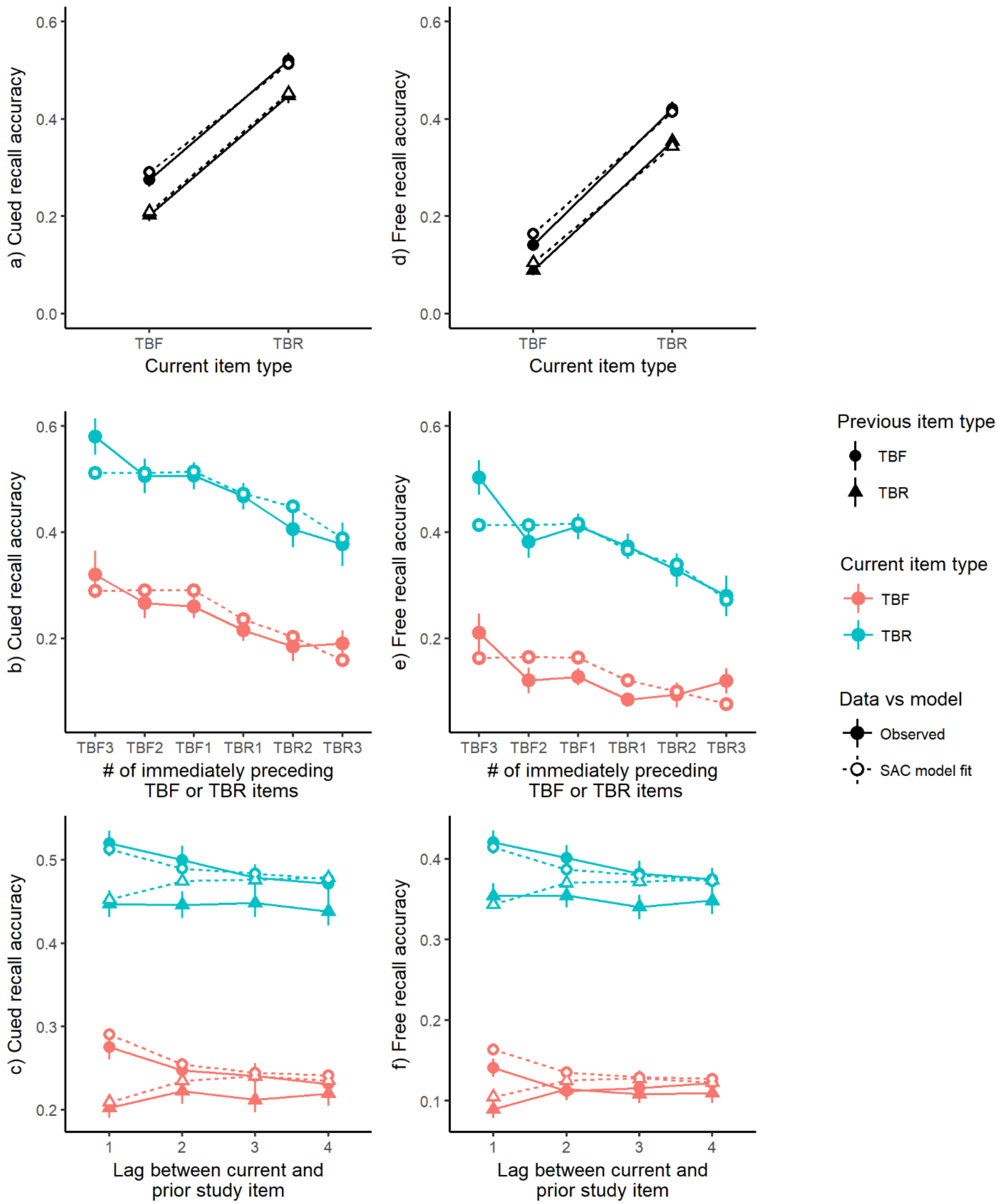

Figure 2. *Results of Marevic et al (2017) reanalysis and fit of the SAC model – cued recall (a,b,c) and free recall (d,e,f) for the current to-be-remembered (TBR) or to-be-forgotten (TBF) item, depending on: a, d) whether it was preceded during study by a TBR or a TBF item; b, e) how many of the immediately preceding items during study were TBR or TBF; c, f) what was the study position lag between the current and the prior item (e.g., how many trials ago did the previous item occur). Error bars represent ±1 SE. Solid points and lines represent the data, the empty points and dashed lines represent the predictions of the SAC model.*

Table 1 Parameter estimates for the Bayesian mixed-effects logistic regression

Fixed-effects	**β**	**Odds Ratio**	**Odds ratio 95% BCI**	**BF^**
Intercept (TBF instructions) *	0.88	0.41	0.30 – 0.58	
TBR instructions for the current item*	1.17	3.22	2.61 – 4.01	3.52×10^{82}
TBR instructions for the item at lag1	-0.41	0.66	0.54 – 0.81	2711
TBR instructions for the item at lag2	-0.26	0.77	0.64 – 0.92	8.84
TBR instructions for the item at lag3	-0.23	0.79	0.66 – 0.95	5.61
TBR instructions for the item at lag4	-0.13	0.88	0.73 – 1.05	0.53
Subject random-effects	**σ**	**95% BCI**		
Intercept	0.79	0.63 – 0.97		
TBR instructions for the current item*	0.51	0.15 – 0.79		
TBR instructions for the item at lag1	0.34	0.02 – 0.69		
Item random-effect	**σ**	**95% BCI**		
Intercept	0.48	0.32-0.69		
Parameter comparisons	**BF+**			
Lag1 < Lag2	7.28			
Lag2 < Lag3	1.41			
Lag3 < Lag4	3.48			

Note: Instructions = whether the current item or the items at lag *i* had to be remembered (TBR) or forgotten (TBF). BCI = Bayesian Credible Interval. * indicates models for which the reference category was TBF instruction, so the parameter estimates of the memory instruction effects reflect the odds for correct recall with TBR instructions; ^ Bayes Factor (*BF*) for the model that includes the parameter vs a model that does not. + the Bayes Factor (*BF*) evidence for the difference between the directed forgetting after-effect at different lags.

4. SAC computational model of results.

Figure 2 also shows the fit of the SAC resource depletion model. A full description of the model is available in Popov & Reder (2018); we will describe it only briefly and note the model assumptions that were specifically adapted for this study.

Our model posits that semantic, episodic and contextual information are represented as a network of interconnected nodes that vary in strength. Each node has a current activation value that increases when a node is perceived or when it receives activation from other nodes. This activation decays with time according to an exponential law to a base-level strength of the node. The base-level strength also gets strengthened with experience and decays with time according to a power law. When new information is studied, two processes occur. First, the current and the base level activation values of the preexisting concept nodes are increased. Second, if this is the first occurrence of the study event, a new event node is created, and it gets associated with the corresponding concept nodes, as well as with the general and specific context nodes. If, however, the study event has occurred previously, the existing event node and its links associated with the concept and context nodes are strengthened instead.

During cued-recall, the activation of the general context node, the specific list node and the cue word concept node are raised, which then spread activation to all nodes to which they are connected. The amount of activation that is spread from a node to any given association is multiplied by the strength of its association and divided by the sum total strength of all associated links that emanate from that node. If the current activation of an event node that is connected to the cue concept node surpasses a retrieval threshold, then the correct target word is recalled. The model was not designed to model free recall; however, we simulate free recall by providing only the context node as a cue and evaluating the activation level of all items simultaneously. We also assume that there is output interference during free recall, which we simulate by exponentiating the activation values – this results in squashing the activation of weak items compared to stronger items.

The model also includes a resource pool that is used every time a node is retrieved, created or strengthened. The resource cost of strengthening a node is equal to the amount by which a node is strengthened. Similarly, the resource cost of retrieving a node is equal to the amount of activation necessary to reach the retrieval threshold. During study, if the resource pool is greater than the required resource for storing an item, the memory trace is strengthened by the default learning rate. However, if the resource pool is less than what is required, the memory trace is strengthened proportionally to the remaining resources. The resource pool recovers at a linear rate.

For the current experiment, we assumed that when an item appears, an episode node is created with a default base-level strength, regardless of the cue. Then, when the cue appears, the episode node for TBR items is strengthened again, while the node for TBF items is not. We fit the model by simulating data for each subject, given their specific trial sequence. Six parameters were optimized by minimizing the root mean squared error of the cued recall and free recall data averaged over all subjects, the current cue type and the number of consecutive preceding TBR or TBF items (24 data points; Figure 2b/e). The parameters were the learning rate $\delta = 0.553$, which governs how much the base-level strength of nodes is increased with each exposure, the resource recovery rate $\gamma = 0.526$, the retrieval thresholds for cued-recall $\theta_{cued} = 0.219$ and for free-recall $\theta_{free} = 0.167$, and the standard deviation of the activation noise $\sigma_{cued} = 0.831$ and $\sigma_{free} = 0.431$. All remaining parameters had the default values we have used in prior models. The model provided very good fits to the cued recall (*RMSE* = 0.026, $R^2 = 0.963$) and free recall data (*RMSE* = 0.034, $R^2 = 0.944$). It is noteworthy that the model also captured the interaction between cue type and lag (Figure 2c/f), although the parameters were not optimized to fit those data points.

5. Multinomial modeling.

A hierarchical version of the multinomial storage–retrieval model (Riefer & Rouder, 1992; Rouder & Batchelder, 1998) was fit to the data, which allows for separately estimating storage (*a parameter*) and retrieval (*r parameter*) from six discrete combinations of free and cued recall events for each participant and memory instruction (current item type, preceding type) (see Appendix A for a detailed description of the model). We only considered discrete recall combinations from the current item's instructions and the first preceding item's instructions for modeling, because considering more than one preceding item (as was done in the above parametric analyses) would have resulted in sparse frequency counts for each cue type. Model parameters were estimated using the *R*-package *TreeBUGS* (Heck, Arnold, & Arnold, 2017), which uses the Markov chain Monte Carlo (MCMC) sampling routine implemented in JAGS for model parameter estimation (Plummer, 2003). The algorithm was run with 1,000,000 iterations retaining every 300th sample and 2,000 iterations as burn-in. Convergence was assessed using the potential scale reduction factor $\hat{R}$. For all parameters, $\hat{R} < 1.2$, indicating good convergence. The model fit the data well, as indicated by individual posterior predictive *p* values (*PPP*), all *PPP*s > .05.

Within- comparisons with respect to the posterior group-level mean parameters measuring storage and retrieval (*a* parameter and *r* parameter) were conducted for current and preceding items' memory instructions. Mean parameter estimates for the *a* and *r* parameters of both item types are reported in Table 2. The results revealed lower *a* parameter estimates for TBF compared to TBR items, as the 95% Bayesian Credible Interval (*BCI*) of the posterior difference did not include zero, *BCI* [.25; .33]. The same was true for *r*-parameter difference estimates, *BCI* [.25; .41]. This replicates the finding that the item-method DF effect seems to be driven by both storage and retrieval processes (Marevic et al., 2017; Marevic & Rummel, submitted; Rummel et al., 2016). Regarding the effect of the preceding item's instructions, there were higher *a* parameter estimates for items preceded by TBF compared to TBR items, as the *BCI* of the posterior difference did not include zero, *BCI* [.06; .14]. This was not the case, however, for *r* parameter difference estimates, *BCI [*-.001; .16]. This finding supports our view that the beneficial encoding effect for items that followed TBF items seems to be driven by storage processes only.

Table 2 Means (*M*) and standard deviations (*SD*) of storage (*a*) and retrieval (*r*) parameter estimates for the dependent measure of Marevic, Arnold, and Rummel (2017).

	Parameters as a function of current items' TBF/TBR instructions		Parameters as a function of preceding items' TBF/TBR instructions	
	M (SD)	95% BCI	M (SD)	95% BCI
a parameter				
TBF	.21 (.01)	[.18; .25]	.42 (.01)	[.38; .45]
TBR	.51 (.02)	[.47; .55]	.31 (.02)	[.27; .35]
r parameter				
TBF	.33 (.03)	[.26; .39]	.60 (.02)	[.54; .65]
TBR	.66 (.02)	[.61; .71]	.52 (.03)	[.45; .59]

Note: TBR: to-be-remembered; TBF: to-be-forgotten; *a* parameter: storage estimate for TBF and TBR current and preceding item cue types; *r* parameter: retrieval estimate for TBF and TBR current and preceding item cue types; *BCI*: Bayesian Credible Interval

III. Experiment 2

Although the resource depletion model fit the data well, alternative explanations for Experiment 1's results should be considered. When people study an item, they may also rehearse or reactivate the memory traces of the preceding items (Camos, Lagner, & Barrouillet, 2009; Loaiza, Duperreault, Rhodes, & McCabe, 2014; McFarlane & Humphreys, 2012). Such rehearsal or attentional borrowing will be more likely when the preceding item was TBR rather than TBF (Bjork, 1970), causing interference with the processing of the current item.

Experiment 2 tested these alternative explanations. In a dual-task learning scenario, we investigated whether suppressing rehearsal or dividing attention during study would erase or at least minimize the DF after-effect. If the DF after-effect was due to greater rehearsal of the preceding TBR items, then preventing rehearsal should effectively remove it. Similarly, if the DF after-effect was due to allocating attention towards previously presented pairs instead of processing the current pair, then such an effect should be strongly attenuated when attention is already occupied by a secondary task. Finally, if the DF after-effects remain stable after suppressing rehearsal or dividing attention, our favored resource depletion explanation will receive strong support.

A. Method

The rationale, method and parts of the analyses for this experiment were pre-registered at the Open Science Framework. The pre-registration is available at https://osf.io/b45tn/. The analysis was changed from the pre-registration from a Bayesian ANOVA to a Bayesian logistic regression, because ANOVA is not appropriate for analyzing proportion data (Jaeger, 2008). The parametric predictions were not included in the pre-registration report. This makes them exploratory for Experiment 1, but confirmatory for Experiment 2. The data, materials and analysis code are available at https://github.com/venpopov/directed-forgetting-after-effects.

1. Participants

The 33 student participants from Heidelberg University (22 female, M_{age} = 22.36, range: 18-31 years) received course credit or monetary compensation.

2. Materials

Words of medium frequency were selected from the *dlex* database (Heister et al., 2011), 448 in all so that they could be randomly paired to form 224 word pairs. The task was divided into eight task blocks. Each block consisted of 12 TBF and 12 TBR word pairs. The memory instructions for individual item pairs were randomized for each participant. The first four items (two TBF, two TBR) of each block served as primacy buffers and were not included in the analyses.

3. Procedure

Participants first received general instructions for the DF task asking them to only remember items that were followed by TBR instructions. Participants were informed that they were about to complete eight study-test blocks of this task while performing a different secondary task in each block. At the beginning of each block, the respective secondary task was explained (see below). Then, each block featured a study phase, in which 12 TBF and 12 TBR items were presented sequentially with a random permutation of the pair type order. During study, participants performed different secondary tasks, which changed every two blocks. The order of secondary tasks was systematically varied across participants using a Latin Square (see Table 3).

Table 3 Counterbalancing orders for the four experimental conditions according to a balanced Latin Square Design

	Block 1 & 2	Block 3 & 4	Block 5 & 6	Block 7 & 8
Order 1	Reh	Att	Reh + Att	Control
Order 2	Att	Control	Reh	Reh + Att
Order 3	Control	Reh + Att	Att	Reh
Order 4	Reh + Att	Reh	Control	Att

Note: Each row represents a unique order, ensuring that each secondary task was followed and preceded by each other condition at least once. Secondary tasks of the same type were always grouped in two consecutive blocks. Reh = rehearsal suppression task, Att = divided attention task, Reh + Att = combined rehearsal suppression and divided attention task, Control = control condition with no secondary task.

In the control blocks, no secondary task was added to the study phase. For the rehearsal suppression blocks, participants were continuously presented via headphones with 60-beats-per-minute (BPM) metronome sounds and were asked to say the German word "der" [the equivalent word to "the" in English] aloud every time they heard the metronome. Additionally, they had to press the j-key or f-key (counterbalanced) whenever saying "der," to keep the motor component equal across blocks. For the divided attention blocks, participants were continuously presented via headphones with even and odd two-digit numbers. They had to press the j-key (f-key) for even and the f-key (j-key) for odd numbers. The assignment of keys was counterbalanced across participants. A new number was presented every 2000 ms on average but inter-stimulus-intervals varied between 1250 and 2750 ms to avoid habituation. For the combined rehearsal suppression and divided attention task, participants were also presented with even and odd two-digit numbers but made verbal odd/even judgements. Additionally, they had to press the j or f-key (counterbalanced) with each judgment to align motor demands to the other secondary tasks.

Following each block's study phase, participants always solved math problems for 30 seconds before they performed a free recall test. For these tests, they were always asked to recall as many TBR items as possible in two minutes. We did not ask participants to recall TBF items because there were multiple study-test blocks and thus a TBF recall instruction would not have come at a surprise any more from the second block on. Participants were specifically encouraged to recall both words of the pairs if possible, but if they could recall only one word of the pair, they should report it as well. Then, participants performed a cued-recall test for which they were presented with the first words of all TBR item pairs they had studied (in random order) and were asked to recall the second word. After four blocks, participants were given a three-minute break in which they received water but had to stay in the laboratory. After completing all eight blocks, participants were asked whether they used a certain forgetting strategy and some demographic questions.

B. Results

1. Main effect of preceding item type and dual task condition.

Figures 3a and 3d plot the cued and free recall accuracy as a function of the instructions given for the current and the preceding item. Both cued and free recall were higher for items that were preceded by TBF items rather than TBR items (BF_{10} = 117 for the cued recall model with preceding items' instructions vs the null model). Overall, memory performance was lower in all dual-task conditions compared to the control condition (BF_{21} = 168570 for the model with preceding items' instructions and condition as main factors, against the model with only preceding item's instructions type). Nevertheless, the DF after-effect was present in all conditions, and preceding items' instructions did not interact with dual-task condition (BF_{23} =

1.38 for the cued recall model with main effects only against the model with an interaction). Because the main effect of preceding cue type did not differ between conditions, we report all remaining analyses collapsed over conditions.

2. Cumulative effect of the number of consecutive preceding TBF or TBR items.

Figures 3b and 3e show the cued and free recall accuracy as a function of the number of consecutive preceding TBF or TBR items. Both cued and free recall performance for the current item were higher when it was preceded by a greater number of consecutive TBF items, and lower, when it was preceded by a greater number of consecutive TBR items. The model including the number of consecutive TBF or TBR items fit the data better than the null model (BF_{10} = 7610 for cued recall).

3. Interaction between preceding cue and study position lag.

Finally, the DF after-effect interacted with the study lag between the current item and the preceding item – the immediately preceding item had a stronger effect than the one two trials before, which in turn had a stronger effect than the one that occurred three trials before (*Figure 3c/f*). We compared the full model, which included the instructions for items at lags 1, 2, 3 and 4, to identical models without the factor of interest. The posterior parameter estimates from the final model and the corresponding *BF*'s are reported in Table 4.

4. SAC computational modeling.

Similar to Experiment 1, we fit the SAC model by simulating data for each subject, given their specific trial sequence. There is no rehearsal mechanism in the model and, for that reason, we ignored the dual-task conditions and focused only on modeling the effect of the prior cue. The same six parameters were optimized by minimizing the root mean squared error of the cued recall and free recall data averaged over the number of consecutive preceding TBR or TBF items (12 data points; Figure 3b/e). In addition, we had to increase the free recall output interference exponent parameter, to account for the different performance in free and cued-recall. The estimated parameters were very similar to those of Experiment 1 – learning rate $\delta = 0.639$, resource recovery rate $\gamma = 0.551$, the retrieval thresholds for cued-recall $\theta_{cued} = 0.279$ and for free-recall $\theta_{free} = 0.457$, and the standard deviation of the activation noise $\sigma_{cued} = 0.451$ and $\sigma_{free} = 0.868$. All remaining parameters had the default values we used in prior models. The model provided excellent fits to the cued recall (*RMSE* = 0.008, $R^2 = 0.991$) and free recall data (*RMSE* = 0.005, $R^2 = 0.984$). It is noteworthy that the model also captured the fact that the DF after-effect decreases with lag (Figure 3c/f), even though the parameters were not optimized to fit those data points.

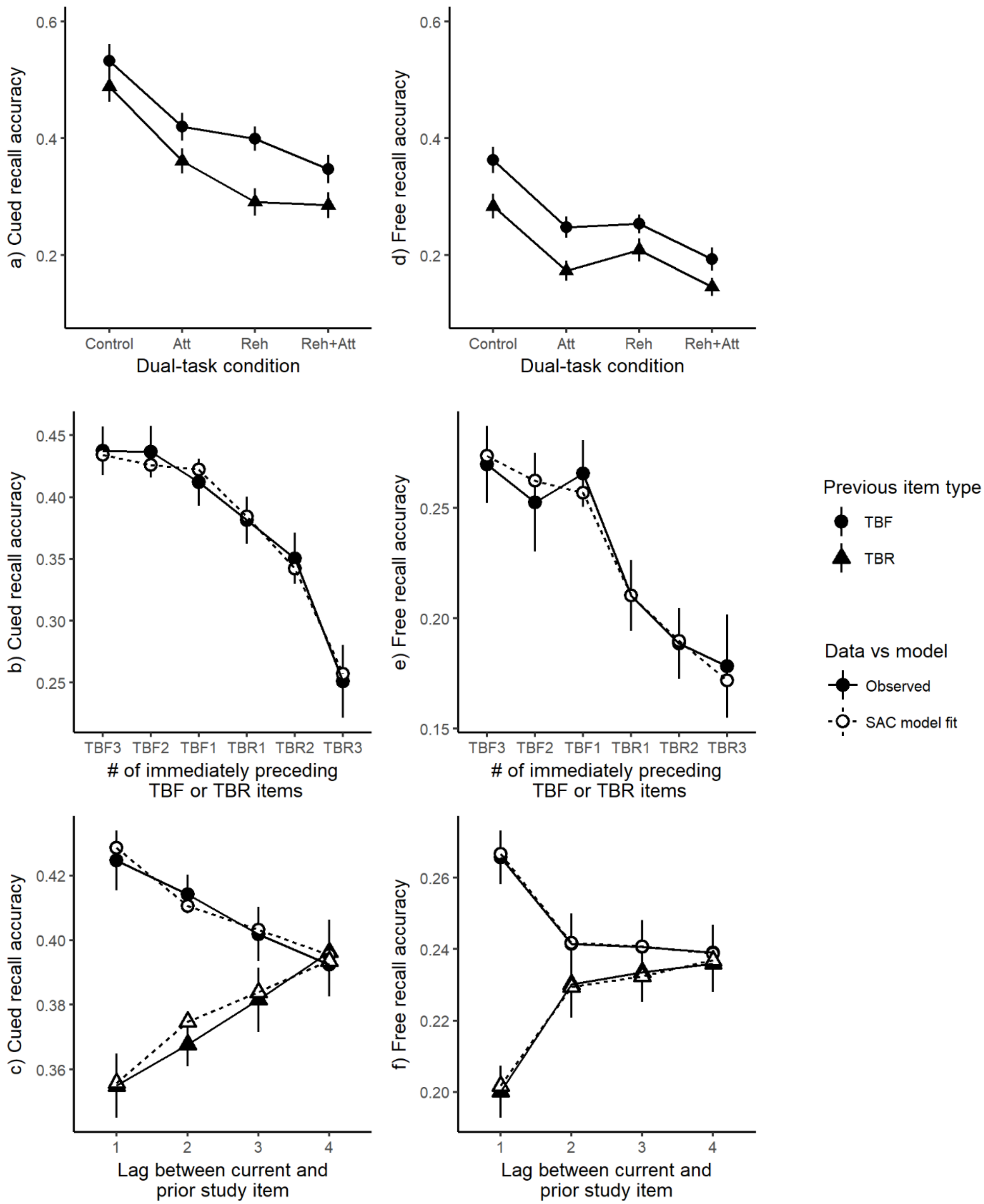

Figure 3 *Results of Experiment 2 and SAC model fits – cued recall (a,b,c) and free recall (d,e,f) for the current item depending on a, d) whether it was preceded during study by a TBR or a TBF item and the dual task condition (Control = No dual task, Att = Divided attention, Reh = suppressed rehearsal, Reh+Att = simultaneous divided attention and suppressed rehearsal; b, e) how many of the immediately preceding items during study were TBR or TBF; c, f) what was the study position lag between the current and the prior item (e.g., how many trials ago did the previous item occur). Error bars represent ±1 SE.*

Table 4 Parameter estimates for the Bayesian mixed-effects logistic regression

Fixed-effects	**β**	**Odds Ratio**	**Odds ratio 95% BCI**	**BF^**
Intercept (TBF instructions) *	-0.15	0.86	0.53 – 1.40	
TBR instructions for the item at lag1	-0.40	0.67	0.53 – 0.84	51.21
TBR instructions for the item at lag2	-0.28	0.76	0.62 – 0.92	10.21
TBR instructions for the item at lag3	-0.15	0.86	0.71 – 1.05	0.72
TBR instructions for the item at lag4	-0.01	0.99	0.81 – 1.21	0.20
Subject random-effects	**σ**	**95% BCI**		
Intercept (control)	1.14	0.85 – 1.52		
Divided attention	0.65	0.19 – 1.12		
Rehearsal suppression	0.56	0.11 – 1.00		
DA + RS	0.69	0.28 – 1.13		
TBR instructions for the item at lag1	0.28	0.02 – 0.69		
Item random-effect	**σ**	**95% BCI**		
Intercept	0.98	0.82 – 1.15		
Parameter comparisons	**BF+**			
Lag1 < Lag2	3.67			
Lag2 < Lag3	4.98			
Lag3 < Lag4	5.27			

Note: Instructions = whether the current item or the items at lag *i* had to be remembered (TBR) or forgotten (TBF). * the reference category was TBF instructions, so the parameter estimates of the instruction effects reflect the odds for correct recall with TBR instructions; ^ Bayes Factor (BF) for the model that includes the parameter vs a model that does not. + the Bayes Factor (BF) evidence for the difference between the cue effect at different lags. BCI = Bayesian Credible Interval

5. *Multinomial modeling.*

The multinomial storage–retrieval model was fit hierarchically to the data, using the same settings as the model fit in Experiment 1. The algorithm was run with 1,000,000 iterations retaining every 500th sample and 20,000 iterations as burn-in. Convergence was assessed using the potential scale reduction factor $\hat{R}$. For all parameters convergence was good, $\hat{R} < 1.01$. The model fit the data well, as indicated by individual posterior predictive *p* values (*PPP*), all *PPP*s > .05. The posterior group-level mean parameters measuring storage and retrieval (*a parameter* and *r parameter*) were analyzed as a function of preceding item instructions across all four experimental conditions. Mean parameter estimates for the *a* and *r* parameters are reported in Table 5. Estimates of the *a* parameters for items preceded by TBF instruction were higher than for those preceded by a TBR instruction, *BCI* of the difference [.03; .13]. For *r* parameter estimates, the lower bound of the *BCI* for the respective difference was barely above 0 [.003; .159]. Thus, in this experiment there was some evidence for effects of preceding items' instructions on retrieval as well. However, given the marginal bounds of the *BCI*, given that we did not find evidence for such an effect in

Experiment 1, and given the excellent fits of the SAC model, which implements only a storage mechanism to account for these effects, we believe that this experiment lends further support to the notion that this effect is best reflected by changes in item storage.

Table 5 Means (*M*) and standard deviations (*SD*) of storage (*a*) and retrieval (*r*) parameter estimates for the dependent measure of Experiment 2.

Parameters depending on preceding item's instructions		
	M (SD)	95% BCI
a parameter		
TBF	.43 (.03)	[.35; .50]
TBR	.34 (.03)	[.26; .42]
r parameter		
TBF	.56 (.03)	[.51; .62]
TBR	.48 (.03)	[.41; .56]

Note. BCI = Bayesian Credible Interval; TBF = to-be-forgotten items; TBR = to-be-remembered items.

IV. General Discussion

We demonstrated a previously unknown DF after-effect of remember and forget instructions in an item-method DF paradigm on memory for the items that *follow* a pair that was to be remembered versus forgotten: cued and free recall for word pairs was higher when people were instructed to forget the preceding word pair. This effect was cumulative, such that performance was even better when more of the preceding pairs had to be forgotten. The size of the DF after-effect depended on how many pairs ago the DF instruction appeared during study. Specifically, the immediately preceding word-pair provided a stronger DF after-effect than when the DF instruction appeared several word-pairs ago. Finally, neither increased rehearsal nor attentional borrowing of TBR items could explain why memory for the subsequent item was worse in those cases – the DF after-effects remained stable, even when rehearsal was suppressed or attention divided in a dual-task paradigm.

The DF after-effects are replicable and are remarkably consistent across the two experiments – the odds ratio associated with items preceded by TBR items rather than TBF items at lag one was 0.66 in the prior study and 0.67 in the new experiment. Similarly, the odds ratio for the effect of cues at lag two were 0.77 and 0.76 in the two studies. Thus, this represents a robust and replicable phenomenon. Additionally, the multinomial storage–retrieval model confirmed that DF after-effects are clearly a storage phenomenon.

Previous work with the list-method DF paradigm has also shown improved memory for the part of a list that follows the TBF items (Bjork, 1970; Epstein, 1972); however, this is the first study to demonstrate DF after-effects in an item-method paradigm, and to characterize in detail how the precise order of TBR and TBF items affects memory for subsequent items. The fact that DF after-effects also appear in the item-method is important, because different processes are assumed to be involved in the two methods (Rummel et al., 2016), and because the item-method allows for a more fine-grained investigation of these effects. Furthermore, researchers have argued that the list-method DF after-effect is due to less rehearsal borrowing

(Bjork, 1970). The current study provides strong evidence that this explanation is unlikely to hold for the item-method, because the DF after-effect was not attenuated when rehearsal was prevented.

The specific pattern of DF after-effects observed here was predicted by a theory of episodic memory, which proposes that memory formation and storage operations deplete a limited pool of resources that recover over time (Reder et al, 2007; Popov & Reder, 2018). Within this framework, TBR items deplete more resources, and they leave fewer resources for processing the subsequent item. A computational model implementing the theory provided excellent fits to the data from both experiments. The model fits, combined with the discounting of alternative explanations, provide support for the resource depletion account of the preceding item cue effects. This account suggests that forgetting is a feature, not a bug – it is an adaptive process because it prevents cognitive resources from being wasted on maintaining irrelevant information.

V. Author Contributions

L. M. Reder developed the theory. V. Popov developed the study concept. Data for Experiment 1 came from Marevic et al (2017). V. Popov, I. Marevic, and J. Rummel designed Experiment 2. Testing and data collection were performed by I. Marevic. I. Marevic and V. Popov performed the data analyses. V. Popov implemented the SAC model. I. Marevic implemented the multinomial model. V. Popov and I. Marevic drafted the initial manuscript. J. Rummel and L. M. Reder provided critical feedback and revisions. All authors approved the final version of the manuscript for submission.

VI. Acknowledgements

We are grateful to Matt So for proofing the final manuscript.

VII. References

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.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*(3), 255–268.

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.

Bürkner, P.-C. (2017). **brms** : An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1).

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469.

Castel, A. D., & Craik, F. I. M. (2003). The Effects of Aging and Divided Attention on Memory for Item and Associative Information. *Psychology and Aging*, *18*(4), 873–885.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology. General*, *125*(2), 159–180.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*(2), 119–132.

Davis, J. C., & Okada, R. (1971). Recognition and recall of positively forgotten items. *Journal of Experimental Psychology*, *89*(1), 181–186.

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(4), 805–815.

Epstein, W. (1972). Mechanisms of directed forgetting. Psychology of Learning and Motivation - Advances in Research and Theory, 6(C), 147–191.

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.

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

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.

Heck, D., Arnold, N. R., & Arnold, D. (2017). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, 1–21.

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.

Hulme, C., Stuart, G., Brown, G. D. ., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500–518.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.

Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., & McCabe, D. P. (2014). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychonomic Bulletin and Review*, *22*(1), 274–280.

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*(1), 35–43.

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.

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

McFarlane, K. A., & Humphreys, M. S. (2012). Maintenance rehearsal: The key to the role attention plays in storage and forgetting. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(4), 1001–1018.

Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review*, *14*(5), 871–876.

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.

Popov, V., & Reder, L. (2018, June 5). Frequency Effects on Memory: A Resource-Limited Theory. *PsyArXiv*. Retrieved from http://doi.org/10.17605/OSF.IO/DSX6Y

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin and Review*, *23*(1), 271–277.

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a Double-Edged Sword: A Computational Model of The Encoding/Retrieval Trade-Off With Familiarity. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 48, pp. 271–312). Elsevier.

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

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.

Shen, Z., Popov, V., Delahay, A. B., & Reder, L. M. (2018). Item strength affects working memory capacity. *Memory and Cognition*, *46*(2), 204–215.

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

Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review*, *17*(2), 200–205.

Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian Benefits for the Pragmatic Researcher. *Current Directions in Psychological Science*, *25*(3), 169–176.

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 186–210.

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.

VIII. Appendix A. Description of the storage–retrieval model

The storage–retrieval model (Riefer & Rouder, 1992) allows for measuring the relative contribution of storage and retrieval to associative memory from data derived from a free-then-cued-recall paradigm and has been applied to DF paradigms in previous studies (Marevic, Arnold, & Rummel, 2017; Marevic & Rummel, submitted; Rummel, Marevic, & Kuhlmann, 2016). To apply the model to the present data, recall frequencies were calculated differentiating between free recall of a complete study pair, free recall of a singleton from a pair, and cued recall of the second word of a pair, leading to the following combinations of recall events (E_1-E_6): 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 storage–retrieval model accounts for all of these recall event combinations by specifying the probability *p* of an event falling in one of the categories (E_1-E_6) through the following 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 states that lead to each of the six event categories. Consequently, parameter estimates must always fall in the range [0; 1]. The above equations can also be visualized in the form of a multinomial processing tree (MPT) with each branch representing a series of latent states leading to one of the six recall events (see Figure 1-A). The model parameters are as follows:

- *a parameter (associative storage)*: Probability of storing and maintaining an item-pair association until the free recall memory test ($0 \leq a \leq 1$).
- *r parameter (associative retrieval)*: Probability of retrieving both items of a pair, given that the pair was stored ($0 \leq r \leq 1$). The pair does not necessarily need to be retrieved associatively, as ***as subsequent singleton retrieval at a time*** is also possible and the model does not differentiate between these two types of associative retrieval.
- *s parameter (stored singleton retrieval)*: Probability of retrieving only one item of an associativelystored pair ($0 \leq s \leq 1$).
- *l parameter (memory loss of stored association)*: Probability of memory loss from successfully free to cued recall ($0 \leq l \leq 1$).
- *u parameter (non-stored singleton retrieval)*: Probability of retrieving a singleton that was individually stored ($0 \leq u \leq 1$).

In the present experiments free recall was immediately followed by cued recall and thus *l* parameter estimates that reflect memory loss from free recall to cued recall should be similar between item types and should generally be close to zero. Therefore, the *l* parameter was set to be equal across conditions (see Marevic et al., 2017; Marevic & Rummel, submitted; Rummel et al., 2016).

In order to obtain reliable parameter estimates for each individual as well as on the group level, the hierarchical Bayesian latent-trait approach was applied (Klauer, 2010). In the latent-trait approach, parameters are drawn from a multivariate normal distribution and are monotonically mapped from [0; 1] to real numbers. The estimation of summary statistics of the model parameters (e.g. posterior mean) is then achieved through Markov Chain Monte Carlo (MCMC) sampling, where a large amount of draws from the posterior of each parameter is obtained. The convergence of the model is then assessed by inspecting the $\hat{R}$ statistic, which compares the variance within to the variance between these MCMC chains (Gelman & Rubin, 1992). An $\hat{R}$ close to 1 is indicative of good convergence. Model fit can be assessed using posterior predictive *p* values (*PPP*) that quantify the discrepancy of the observed and from the model predicted data by computing the proportion of samples for which the observed data is smaller than the predicted data. A *PPP* > .05 is generally indicative of a good model fit. Inferences can then be drawn in the Bayesian framework by assessing whether the Bayesian Credible Interval of the posterior samples of the difference between two parameters of interest (e.g. TBF, TBR) excludes zero, as this resulting posterior distribution summarizes the knowledge about the effect of interest (Lee & Wagenmakers, 2013).

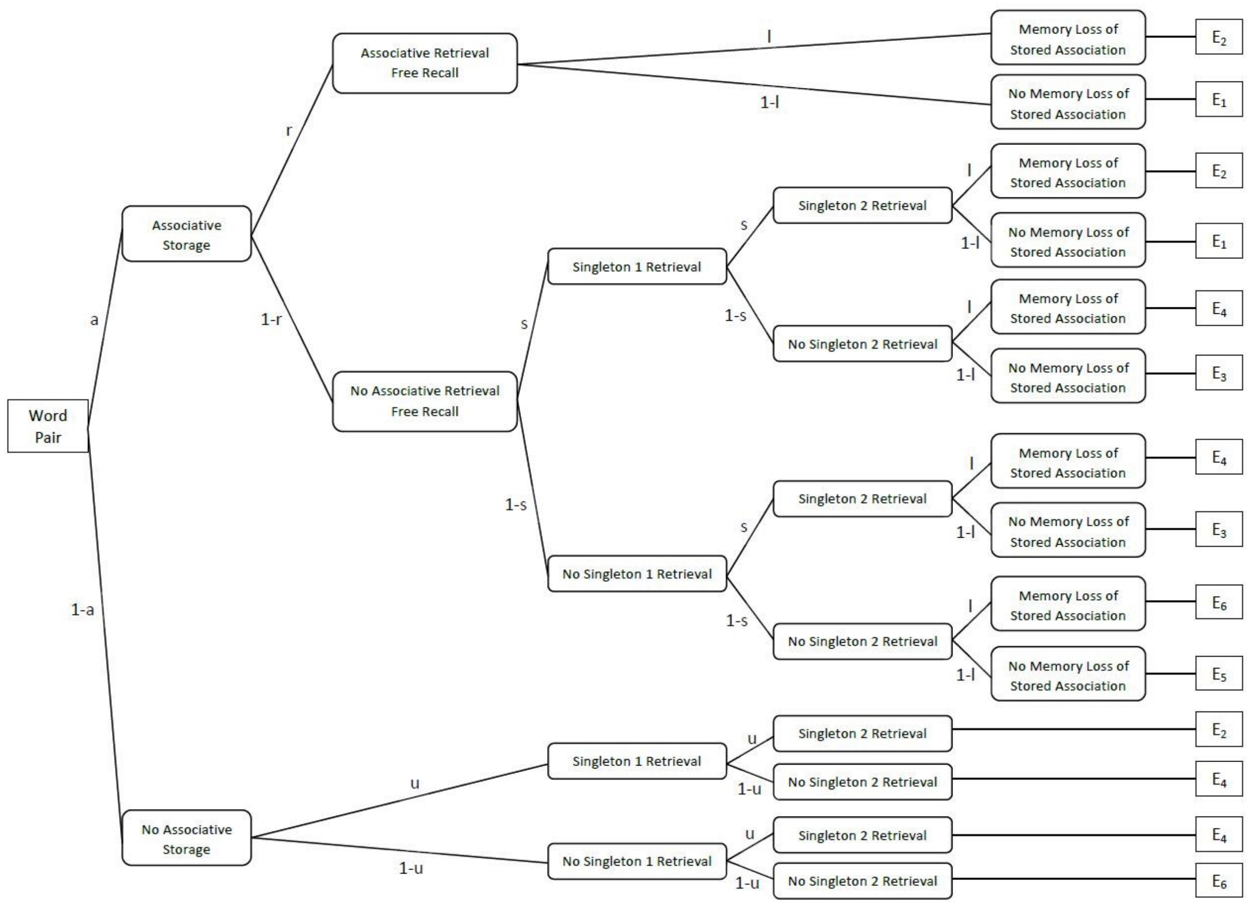

Figure 1-A. Multinomial processing tree model (MPT) for a free-then-cued-recall paradigm, to separate storage and retrieval (based on Rouder & Batchelder, 1998, adopted from Rummel et al. 2016). The processing tree represents the different latent cognitive states that lead to the six observable recall events (E_1-E_6). Rounded rectangles represent latent cognitive states with transition probabilities 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.