Psychological Science
2019, Vol. 30(9) 1303 1317

DOI: 10.1177/0956797619859531
www.psychologicalscience.org/PS

recover over time, weaker items within a list especially hurt memory for subsequent items from the same list.

Here, we tested a key prediction of the theory: Memory should be higher for items that are, during study, preceded by items consuming fewer resources. We used an item-method directed-forgetting paradigm in which each study item was directly followed by either a to-be-forgotten (TBF) or a to-be-remembered (TBR) instruction, indicating whether it would be tested later (Bjork, 1972; Golding & MacLeod, 1998). Previous studies showed worse TBF than TBR item recall (i.e., a directed-forgetting effect), but it is unknown whether memory differs for items that follow a TBR or a TBF item (i.e., a

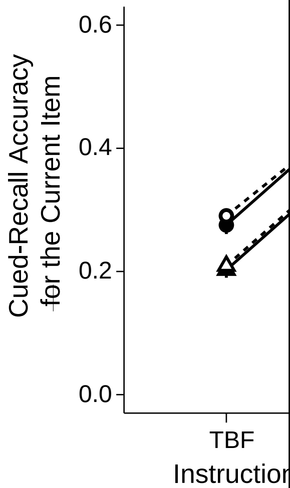
di 0 10 462sc[((ted1f)10(3r)12(er)(erng(her)1(y)10 462sc1(heing 4f)-5(f)(i 0 10 462s1998(I(erno)8st)462sig(hea462sin62sgten
- showvted-forgettbenefitu(ari0(obeca5(e)1ec)17( )]TJ34.035 T120.033 Tw

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 in the experimental phase. To control for item-specific effects, we counterbalanced the assignment of word-pair sets to phases. In each block, half of the word pairs were followed by instructions to forget them (TBF word pairs) and half by instructions to remember them (TBR word pairs).

Procedure. Experimental sessions started with a WM task (not analyzed here but reported by Marevic et al., 2018) and a practice phase in which participants studied 24 TBR and TBF word pairs. Participants were told to remember only the TBR word pairs for a later test and to forget the TBF word pairs. Each word pair was presented for 7 s in the center of the screen, followed by either a TBR or a TBF instruction for 2 s (i.e., the word remember or forget in German). Trials were separated by a 250-ms interstimulus interval. After all word pairs had been presented, participants solved math problems for 30 s before completing a free-recall test. In the practice phase, the free-recall test was followed by a cued-recall test for TBR items only. Recall cues were presented in random order for each participant. This practice phase was intended to familiarize participants with the paradigm and to increase their belief that the instruction to forget 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 WM task (not reported) and then were debriefed and compensated for their participation.

Data analysis. We employed Bayesian statistics for the new analyses of Marevic et al.s (2018) behavioral data. This approach had several advantages (Wagenmakers, Morey, & Lee, 2016), but most important to us was that Bayes factors (BFs) enabled us to quantify the evidence in favor of the null as well as the alternative hypotheses. We calculated BFs using bridge sampling for comparing models that included the effect of interest with models that did not. A BF close to 1 means that both models are equally likely, a BF greater than 3 is conventionally interpreted as moderate evidence, and a BF greater than 10 provides strong evidence in favor of the preferred model (Lee & Wagenmakers, 2013). We applied multilevel logistic Bayesian regressions as implemented in the brms package in the R programming environment (Bürkner, 2017; R Core Team, 2019), in which we included crossed random intercepts for participants and items as well as random participant slopes for directed-forgetting effects and aftereffects. The population-level regression coefficients had a weakly informative Students t distribution prior that was zero-centered with 3 degrees of freedom and a scale of 2.5 (Gelman, Jakulin, Pittau, & Su, 2008). For the free-recall analysis, words were coded as correctly recalled when both items of a pair were recalled. All models were run with 10,000 iterations, and half of those iterations were used as burn-in. Convergence was assessed using the potential scale-reduction factor $\hat{R}$. For all parameters, $\hat{R}$ was less than 1.01, indicating good convergence.

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. To measure the cumulative effect of successive cues, we also coded


Cued-Recall Accuracy
for the Current Item
0.6
0.4
0.2
0.0
TBF
Instruction

[redacted]

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 685 for cued recall and BF 977 for free recall). There was strong evidence that the directed-forgetting effect and the directed-forgetting aftereffect did not interact (BF_{cued}

Table 2. Parameter Estimates for the Bayesian Mixed-Effects Logistic Regression on Free Recall in Experiment 1

Effect type and predictor	Parameter estimate	Odds ratio	Bayes factor
Fixed effect			
Intercept (TBF instructions)[a]	E 1.95	0.14 [0.10, 0.20]	
TBR instructions for the current item[a]	E 1.58	4.88 [6.82, 6.26]	BF^ 3.52 u 10^{82}
TBR instructions for the item at Lag 1	E 0.49	0.61 [0.48, 0.77]	BF^ 397
TBR instructions for the item at Lag 2	E 0.19	0.83 [0.67, 1.02]	BF^ 0.63
TBR instructions for the item at Lag 3	E 0.22	0.80 [0.65, 0.99]	BF^ 0.78
TBR instructions for the item at Lag 47 Td(BF)Tj5.85 0 0 5.85 440.8123 576.0721 Tm(^)Tj9 0 0 9 444.0883 572.9221 Tm( )Tj/			

Source-of-activation-confusion (SAC) computational model of results. Figure 2 also shows the fit of the SAC resource-depletion-and-recovery model. A full description of the model is available in the Supplemental Material and in Popov and Reders (in press) article; we describe it only briefly and note which of the model assumptions were specifically adapted for this study.

Our model posits that semantic, episodic, and contextual information is 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 increases with experience and decreases 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 and 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 list-context node and the cue-word-concept node is raised, which then spreads 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 simulated free recall by providing only the context node as a cue and evaluating the activation level of all items simultaneously. We also assumed that there would be output interference during free recall, which we simulated by exponentiating the activation values; this resulted in squashing the activation of weak items compared with 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 degree to 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 currently available resource pool is sufficient for storing an item, the memory trace is built or strengthened by the default learning rate. However, if there are currently fewer resources available than required, the memory trace is strengthened proportionally to the remaining resources. The resource pool recovers at a linear rate until it reaches the maximum WM resource capacity.

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 instruction type. Then, when the instruction appears, the episode node for TBR items is strengthened again, whereas the node for TBF items is not. We fitted the model by simulating data for each participant, given the specific trial sequence. Six parameters were optimized by minimizing the root-mean-square error (RMSE) of the cued-recall and free-recall data averaged over all participants, the current instruction type, and the number of consecutive preceding TBR or TBF items (24 data points; see Figs. 2b and 2e). In our initial modeling, we estimated separate learning rates for the strengthening during item and instruction presentation. These two estimates were roughly equal and the model did not fit the data significantly better than the simpler model with a single learning rate for the strengthening during both item and instruction presentation. The final model parameters consisted of a learning rate (G) of 0.553, which governs how much the base-level strength of nodes is increased with each exposure; a resource-recovery rate (w_r) of 0.526; retrieval thresholds (T) of 0.219 for cued recall and 0.167 for free recall; and standard deviation of the activation noise (V) of 0.831 for cued recall and 0.431 for free recall. All remaining parameters had the default values we have used in prior models. The model provided very good fits to the cued-recall data (RMSE 0.026, R^2 .963) and free-recall data (RMSE 0.034, R^2 .944). It is noteworthy that the model also captured the interaction between instruction type and lag (see Figs. 2c and 2f), although the parameters were not optimized to fit those data points.

Experiment 2

Despite good model fit, there were still alternative explanations for Experiment 1 s results. People may rehearse or reactivate the memory traces of preceding items while processing the current item (Camos, Lagner, & Barrouillet, 2009; McFarlane & Humphreys, 2012). Such rehearsal or attentional borrowing is more likely when the preceding item was TBR rather than TBF (Bjork, 1970), resulting in diminished processing for the current item. Similarly, the retrieving-effectively-from-memory (REM) model (Shiffrin & Steyvers, 1997; Lehmann & Malmberg, 2013) postulates that there is a limited rehearsal buffer and that memory-trace strength depends on how much of the buffer is currently available. The REM model would attribute the directed-forgetting aftereffect to the fact that TBF items are not rehearsed, which frees buffer space for the rehearsal of the current item.

In Experiment 2, we tested whether suppressing rehearsal during study would eliminate the directed-forgetting aftereffect to rule out that it is due to greater rehearsal of preceding TBR items (for a similar argument concerning the effect of articulatory suppression on rehearsal-based explanations for the regular directed-forgetting effect, see Hourihan, Ozubko, & MacLeod, 2009). We further tested whether the directed-forgetting aftereffect would be attenuated under divided attention to rule out that it is due to allocating attention to previous pairs (attentional refreshing) instead of the current pair (for illustrations, see Figs. S4 and S5 in the Supplemental Material). A stable directed-forgetting aftereffect under suppressed rehearsal or divided attention would support the resource-depletion-and-recovery explanation.

Method

The rationale, method, and original analysis plan for this experiment were preregistered on the Open Science Framework (https://osf.io/yugkt). We deviated from the preregistered analysis plan after realizing that a Bayesian analysis of variance (ANOVA) would not be appropriate for analyzing proportion data; instead, we conducted a Bayesian logistic regression. The parametric predictions were not included in the preregistration plan. This makes them exploratory for Experiment 1 but confirmatory for Experiment 2. The data, materials, and analysis code for this experiment are available at osf.io/5qd94.

proposes that different amounts of resources are depleted at Time t 1, the attention-borrowing explanation implies that the effect is retroactive; that is, during the current trial at Time t, participants redirected attention back to the item presented at Time t 1. The divided-attention task would remove the directed-forgetting aftereffect in the latter but not in the former case (for more information, see the Supplemental Material).

though the parameters were not optimized to fit those data points.

General Discussion

We demonstrated a novel directed-forgetting aftereffect: When an item is TBF rather than TBR, memory for the subsequent item benefits. This effect occurs in both cued and free recall and is cumulative: The more that preceding items are TBF, the higher the memory benefits; the effect decreases when memory is conditioned on instructions for items appearing further back in the study list. The directed-forgetting aftereffect was replicable and remarkably consistent across the two experiments; the cued-recall odds ratios associated with items preceded by TBR items relative to TBF items were 0.66 and 0.67, respectively.

Previous research has also shown improved memory for whole lists when a preceding list was TBF rather than TBR (Bjork, 1970; Epstein, 1972). This is, however,

Table 5. Parameter Estimates for the Bayesian Mixed-Effects Logistic Regression on Free Recall in Experiment 2

Effect type and predictor	Parameter estimate	Odds ratio	Bayes factor
Fixed effect			
Intercept (TBF instructions; control)[a]	E 0.65	0.52 [0.34, 0.78]	
Effect of dual-task condition			
Divided-attention condition	E 0.77	0.46 [0.31, 0.69]	BF^ !15 u 10^{3}
Rehearsal-suppression condition	E 0.65	0.52 [0.34, 0.79]	BF^ 651.17
Divided-attention condition rehearsal-suppression condition	E 1.15	0.32 [0.19, 0.51]	BF^ !15 u 10^{3}
Effect of instructions			
TBR instructions for the item at Lag 1	E 0.48	0.62 [0.47, 0.81]	BF^ 30.53
TBR instructions for the item at Lag 2	E 0.12	0.89 [0.72, 1.10]	BF^ 0.15
TBR instructions for the item at Lag 3	E 0.09	0.92 [0.75, 1.13]	BF^ 0.05
TBR instructions for the item at Lag 4	E 0.08	0.92 [0.74, 1.14]	BF^ 0.06
Participant random effect			
Intercept (control)	V 0.63 [0.42, 0.90]		
Divided-attention condition	V 0.21 [0.01, 0.58]		
Rehearsal-suppression condition	V 0.44 [0.04, 0.86]		
Divided-attention condition rehearsal-suppression condition	V 0.67 [0.19, 1.18]		
TBR instructions for the item at Lag 1	V 0.38 [0.03, 0.77]		
Item random effect			
Intercept	V 0.70 [0.54, 0.87]		
Parameter comparison			
Lag 1 Lag 2			BF 69.42
Lag 2 Lag 3			BF 1.37
Lag 3 Lag 4			BF 1.04

Note: The parameter estimates reflect the means of the posterior distribution. Values in brackets are 95% Bayesian credible intervals. BF^ refers to the Bayes factor for the model that includes the parameter versus a model that does not; BF refers to Bayes factor evidence for the difference between the directed-forgetting aftereffect at different lags. Participants were instructed whether the current item or the items at lag i were to be remembered (TBR) or to be forgotten (TBF).

[a]The reference category for this analysis was TBF instruction in the control condition, so the parameter estimates of the memory-instruction effects reflect the odds for correct recall with TBR instructions.

the first study to demonstrate directed-forgetting aftereffects on an item level and to characterize in detail how the precise order of TBR and TBF items affects memory for subsequent items. The present findings indicate similarities between the two directed-forgetting methods but also provide new theoretical insight, because the item method allows for a more fine-grained investigation of directed-forgetting aftereffects. For example, researchers have argued that the list-method directed-forgetting aftereffect is due to less rehearsal borrowing (Bjork, 1970; Sahakyan & Kelley, 2002). This explanation is unlikely to hold for the item method because the directed-forgetting aftereffects in our experiments were not attenuated when rehearsal was prevented.

What causes item-method directed-forgetting aftereffects? We propose that memory formation and storage deplete a limited resource that recovers over time

Marevic, I., Arnold, N. R., & Rummel, J. (2018). Item-method directed forgetting and working memory capacity: A hierarchical multinomial modeling approach. The Quarterly Journal of Experimental Psychology, 71, 1070 1080.
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, 1001 1018.
Pastötter, B., Tempel, T., & Bäuml, K.-H. T. (2017). Long-term memory updating: The reset-of-encoding hypothesis in list-method directed forgetting. Frontiers in Psychology, 8, Article 2076. doi:10.3389/fpsyg.2017.02076
Popov, V., & Reder, L. (in press). Frequency effects on memory: A resource-limited theory. Psychological Review. doi:10.1037/rev0000161
Popov, V., So, M., & Reder, L. (2019). Word frequency affects binding probability not memory precision. Retrieved from