Memory

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pmem20

The role of prior familiarisation and meaningfulness of verbal and visual stimuli on directed forgetting

Yi-Pei Lo, Huiyu Ding, Jonathon Whitlock & Lili Sahakyan

MEMORY
https://doi.org/10.1080/09658211.2024.2358126

Routledge
Taylor & Francis Group

Check for updates

The role of prior familiarisation and meaningfulness of verbal and visual stimuli on directed forgetting

Yi-Pei Lo[a], Huiyu Ding[a], Jonathon Whitlock[a] and Lili Sahakyan[a,b]

[a]Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA; [b]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

ABSTRACT
Intentional forgetting of unwanted information is a crucial cognitive function that is often studied with directed forgetting (DF) procedure, whereby cuing some study materials with Forget (F) instruction impairs their memory compared to cuing with Remember (R) instruction. This study investigates how the nature of information (verbal or pictorial), its semantic significance (meaningful or meaningless), and the degree of prior episodic familiarity influence DF. Before the DF phase, stimuli were familiarised by pre-exposing them 0, 2, or 6 times in a prior preview phase. Finally, memory for all items was assessed with old/new recognition test. Experiment 1 employed words, Experiment 2 utilised fractal images, Experiment 3 featured both meaningful and meaningless object images, and Experiment 4 used words and nonwords. Our results indicate that materials that produced better memory performance are not always harder to intentionally forget. Previewed items showed reduced DF compared to non-previewed items regardless of the nature of information, and meaningless stimuli are challenging to intentionally forget regardless of their degrees of familiarisation unless they are meaningless verbal materials. Collectively, the results highlight the importance of joint consideration of the stimulus format, its meaningfulness, and its episodic familiarity in understanding conditions that interact with intentional forgetting.

ARTICLE HISTORY
Received 5 July 2023
Accepted 15 May 2024

KEYWORDS
directed forgetting; visual/verbal; meaningful/meaningless; unfamiliar/familiarised stimuli

We encounter enormous amounts of information in our daily lives, some of which is routine and familiar, and some that is novel. Not all information is worth maintaining in mind, and it may even be beneficial to forget some unimportant, irrelevant, or painful memories. The ability to intentionally forget unwanted information is a desired cognitive function because it benefits the efficiency of information processing (Bjork, 2011). One of the established laboratory paradigms for examining the control of unwanted memories is the directed forgetting (DF) procedure (Bjork et al., 1968). The current paper focuses on the item-method variant of DF, in which participants are presented with information to study for a subsequent memory test, and the presentation of each item is subsequently followed by a forget (F) or remember (R) cue. The typical findings demonstrate impaired memory for F-items compared to R-items, referred to as the DF effect, and the results generalise to various memory tests, and they emerge with a variety of stimuli (Basden & Basden, 1998; Bjork et al., 1998; Macleod, 1999; Sahakyan, 2024).

The theories proposed to explain the DF effect broadly construed focus on the upregulation of R-items or the downregulation of F-items. According to the *Selective Rehearsal Account* (Basden et al., 1993; Bjork, 1970; Macleod, 1999), R-items continue to receive rehearsal after the presentation of the cue, whereas rehearsal terminates for the F-items resulting in impoverished memory for F-items. Thus, the rehearsal account explains the emergence of the DF effect as the enhancement of R-items, whereas F-items passively decay in memory due to the lack of rehearsal. In contrast, other accounts propose that additional processes operate on F-items either by directly suppressing their representations in memory (*Active Inhibitory Account*) (Fawcett & Taylor, 2008, 2010, 2012; Fellner et al., 2020; Hauswald & Kissler, 2008; Lee, 2013; Ludowig et al., 2010; Nowicka et al., 2011; Oberauer, 2018; Oehrn et al., 2018; Paz-Caballero, 2004; Reber et al., 2002; Rizio & Dennis, 2013; Van Hooff & Ford, 2011), or by separating them from their encoding context (*Context Unbinding Account*) (Chiu et al., 2021; Whitlock et al., 2020). Multiple mechanisms have been shown to contribute to DF, as independently confirmed by behavioural data from participants' strategy reports indicating that different participants engage one or the other type of strategy to accomplish DF (Sahakyan, 2024), as well as from electrophysiological data (Fellner et al., 2020). A combination of these mechanisms may ultimately provide a more

CONTACT Yi-Pei Lo yipeilo2@illinois.edu Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, 61820, USA

comprehensive understanding of the various manifestations of DF in different contexts and populations.

The goal of this paper is to investigate whether certain items can be intentionally forgotten more easily than others. This investigation particularly delves into factors such as how the level of episodic familiarity, the nature of information (verbal or pictorial), and its semantic significance affect intentional forgetting. Assume that you had repeatedly used the same password for the last year and now have to change it. Forgetting the old password would help you remember the new one better by reducing proactive interference. Forgetting the old password may be optimal in this scenario, although potentially made more difficult due to the password being highly familiar and well entrenched in our memory. Furthermore, when we repeatedly park in the same parking garage, we need to remember today's parking location while "forgetting" our previous parking locations on other days. When considering these distinct cases of forgetting, we may ask whether it is more difficult to forget highly familiar passwords than parking spots that are unique to each day, when both differ in terms of familiarity and the nature of the information being either verbal or visual. In other words, what is the relationship between memory for information encountered repeatedly and the likelihood of successful DF, and how does the nature of information one is trying to forget affect DF?

Surprisingly, despite decades of research on DF, the answer to this question remains elusive. Various manipulations have been employed in trying to answer this question, though critically we note the inconsistency of these findings. The inconsistent findings in the literature may be partly driven by the slippery nature of *memory strength* as a construct, which is often defined circularly from memory performance levels, with better memory being attributed to stronger memories. Across different manipulations that affected memory performance levels, some studies revealed that conditions with better memory were associated with larger DF because the R-items benefitted more in the strong condition compared to F-items (Basden, 1996; Geiselman et al., 1985; Horton & Petruk, 1980), whereas others revealed smaller DF effect because the memory for F-items improved more than R-items in the strong condition (Bugelski, 1970; Dulaney et al., 2004; Geiselman et al., 1985; Hourihan & MacLeod, 2008; Lee, 2013; MacLeod & Daniels, 2000), and yet other studies found equivalent DF in strong and weak conditions (Geiselman et al., 1985; Hockley et al., 2016; Metzger, 2011).

The reasons for such inconsistencies are not fully understood and may be driven in part by the type of processing manipulations performed, and/or the nature of the stimuli. For example, in Geiselman et al. (1985), participants were asked to generate synonyms (i.e., deep encoding) or generate homonyms (i.e., shallow encoding) for each presented word. As expected, there was overall enhanced memory for the deep encoding condition, but importantly, the DF effect was reduced for deeply encoded items because they were harder to forget intentionally than shallowly processed items (see also Dulaney et al., 2004; Lee, 2013). However, other studies that used other depth of processing tasks, such as processing words depending on either their structural, phonemic, or semantic qualities, found that deeply encoded items showed *enhanced* DF rather than reduced DF compared to shallowly encoded items, because R-cued words benefited much more than F-cued words from deeper processing versus shallow processing (Horton & Petruk, 1980; Wetzel, 1975). Thus, findings on how depth of processing influences DF are inconsistent.

Likewise, study time manipulations influence memory performance with longer study duration improving overall memory, but the effect of study time on DF has also produced divergent results, with several studies finding no effect of study time on the DF magnitude (Geiselman et al., 1985; Hockley et al., 2016; Metzger, 2011), whereas other researchers found that longer study time reduced the DF magnitude (Gardiner et al., 1994; Wetzel & Hunt, 1977). Furthermore, studies that crossed DF with manipulations such as the generation effect (Bertsch et al., 2007; Slamecka & Graf, 1978), the enactment effect (Engelkamp & Krumnacker, 1980), and the production effect (Craik & Lockhart, 1972) found that making items more memorable through generating words, reading them aloud, or performing action phrases, reduced DF compared to the respective control conditions (Earles & Kersten, 2002; Hourihan et al., 2009; MacLeod & Daniels, 2000; Sahakyan & Foster, 2009). In all of these "encoding processing manipulation" studies, higher memory conditions were associated with reduced DF because F-items benefited more than the R-items. At the same time, another study found that visual imagery resulted in a larger DF effect, by improving memory only for R-items, whereas the F-items had comparable memory with and without mental imagery (Basden & Basden, 1996). Thus, higher memory conditions do not always interact with DF, and whenever they do, the direction of the effect is inconsistent.

The nature of materials can also influence memory and impact the magnitude of DF. Although prior studies have obtained DF using visual stimuli (Fawcett et al., 2016; Hauswald & Kissler, 2008; Hourihan et al., 2009; Metzger, 2011; Quinlan et al., 2010; Scotti & Maxcey, 2022), DF is usually more modest for concrete pictures (Hauswald & Kissler, 2008; Quinlan et al., 2010; Scotti & Maxcey, 2022), faces (Ding et al., 2022; Metzger, 2011), and abstract unnamable symbols (Hourihan et al., 2009) compared to verbal materials in general. Nonetheless, the factors contributing to the decreased magnitude of DF for pictorial stimuli remain unclear. Specifically, it is unclear whether the level of memory performance (i.e., high vs. low), the visual nature of stimuli, and/or the meaningfulness of stimuli contributes to the reduced DF effect. For example, concrete images may show reduced DF due to them being visual in nature and thus better remembered than verbal stimuli, as suggested by the dual coding hypothesis (Paivio, 1971). However, this explanation does not adequately account for the diminished DF observed with abstract symbols, which are also visual in nature but less meaningful than verbal stimuli (i.e., words, action phrases, or sentences), and in general less well remembered. Consequently, two potential explanations arise: either the visual nature of stimuli, regardless of their inherent meaning, is linked to reduced DF, or the meaningfulness of the stimuli itself impacts DF. In the latter case, abstract symbols could be harder to forget due to being semantically meaningless.

To sum up, previous literature on modulated magnitude of DF has yielded inconsistent results, which is likely due to the complex interactions of stimuli format, encoding strategies, and semantic meaningfulness. Therefore, the purpose of our investigation was to disentangle how the nature of study materials and episodic familiarity affect DF by employing visual and verbal stimuli that were either meaningful or meaningless while systematically manipulating the repetition of these stimuli before the DF phase. Instead of using different processing instructions to create stronger and weaker memories, we implemented the repeated presentation approach with consistent processing instruction. This method allowed us to enhance memory for items across different amount of repetition without altering the nature of the stimuli or requiring participants to engage in different levels of processing across items. This approach offers an advantage because it creates differences in memory performance for studied items by circumventing the complexity of processes that arise when some items are processed differently than others (i.e., level of processing or other processing manipulations at encoding) and their potential interactions with DF. Importantly, we repeated some stimuli in a preview phase 2 or 6 times prior to subjecting them to the DF manipulation, whereas other stimuli were not shown in the preview phase at all. Doing so allowed manipulating episodic familiarity of to-be-forgotten information, thereby creating stronger and weaker memories without employing different kinds of processing strategies. Across a series of experiments, repetition manipulations were applied to meaningful verbal and pictorial stimuli (i.e., words and pictures of everyday objects) and to meaningless stimuli (i.e., fractal images, abstract shapes, and non-words). This permitted investigating whether meaningful information is easier to intentionally forget compared to meaningless information, as well as whether the verbal/pictorial nature of information potentially interacts with DF. If the level of memory performance (i.e., high vs. low) interacts with DF in a consistent fashion regardless of stimulus type (verbal vs. pictorial; meaningful vs. meaningless), we should observe similar effects of repetition on DF across all types of stimuli. However, it could also be the case that the success of DF varies with the meaningfulness of studied information, and therefore the meaningfulness dimension may interact with DF on its own way regardless of the memory performance levels.

In light of the mixed findings and the gaps identified in the literature, this study seeks to address several questions: (1) Does the nature of the information (verbal vs. pictorial) influence its susceptibility to DF? (2) Does the meaningfulness of the information impact the DF effect? (3) Does prior familiarity with to-be-forgotten information affect the success of DF, and if so, how does the degree of familiarity influence DF? By examining these dimensions, we aim to provide clarity on the conditions that facilitate or hinder DF, offering a more nuanced understanding of the interplay between memory performance levels, nature of stimuli, and their meaningfulness on the DF effect.

Experiment 1

The purpose of Experiment 1 was to examine how episodic familiarisation manipulated via objective repetitions affects the magnitude of item-method DF. The presented stimuli involved English words. There was an initial preview phase of the experiment in which half of the studied items were presented prior to the DF phase. Half of the items presented in the preview phase were presented twice whereas the other half were presented six times. All items in the preview phase were then presented in the DF phase along with an equal number of novel (i.e., non-studied) items.

Method

Participants

Participants were 160 undergraduates from UIUC, who received course credit for participation. All data were collected online. An a priori power analysis was conducted using G*Power version 3.1.9.6 (Faul et al., 2007) to determine the minimum sample size required. Results indicated the required sample size to achieve 80% power for detecting the smallest effect size ($f = 0.15$) that we were interested in, at an alpha level of .05, was $N = 90$ for the repeated-measures ANOVA analysis. All reported experiments exceeded the required sample size. No participants were excluded in any of the reported experiments.

Stimuli

The stimuli involved 160 medium frequency words selected from the MRC Psycholinguistic Database (Coltheart, 1981), with an average Kucera-Francis frequency score of $M = 50.03$ ($SD = 13.53$). The words were 5–7 letters long ($M = 6.06$, $SD = 0.80$). The study was programmed using PsychoPy (Peirce et al., 2019) and hosted by Pavlovia (Bridges et al., 2020).

Procedure

The experiment had three phases: a preview phase, a DF phase, and an old/new recognition test. The order of item presentations was randomised within all three phases. During the preview phase, participants were presented with forty unique words one at a time on a computer screen at a rate of 1.5 s per word, with a 1 s ISI between trials. Half of the words were presented in uppercase, and the other half were in lowercase. Participants were told that the preview phase was intended to help them familiarise with some words, and their task was to indicate if the word was in upper or lowercase using keyboard buttons. The purpose of the categorisation task in the preview phase was to ensure that participants paid attention to and processed the presented stimuli. By the end of the preview phase, half of the items were presented twice, and the other half were presented six times.

After the preview phase, participants underwent an item-method DF phase. They were presented with 80 words, each appearing on the screen for 1.5 s, followed by a Remember or a Forget memory cue for 3 s, with a 1 s ISI between trials. Participants were instructed that a Remember cue meant that the word would be tested, and they should keep that word in mind and memorise it, whereas a Forget cue meant that the word would not be on the test, so they should try to forget it. Out of 80 study words, 40 had appeared in the preview phase, and their letter case was preserved to be the same in the DF phase. The remaining 40 words were shown for the first time during the DF phase, half in upper and half in lowercase. Therefore, by the end of the DF phase, studied items included words that appeared once (non-previewed items), three times (twice previewed items), and seven times (six-time previewed items).

Immediately after the DF phase, participants completed a self-paced old/new recognition task in which they were presented with 160 words, containing 80 old words and 80 new words. Out of 80 old words that had appeared in the DF phase, half were non-previewed (40) and half were previewed either twice (20) or six times (20). All old words were presented in the same letter case as shown in the Preview and DF phases. Half of the new words (40) were in uppercase, and half (40) were presented in lowercase. Participants were instructed to endorse all words they remember seeing previously as "old", regardless of whether the word was followed by a remember or a forget cue, and to endorse the words they did not remember seeing in the experiment as "new" items. Responses were made by pressing keyboard buttons, either "z" for old or "m" for new. The assignment of words to Remember, Forget, and lure conditions, as well as assignments of times of repetitions in the preview phase, were fully counterbalanced.

Results & discussion

Recognition accuracy (*d′* scores) was calculated after hits and false alarms were transformed using a loglinear correction (Hautus, 1995; Stanislaw & Todorov, 1999). Hits and false alarm rates for this and the following experiments are reported in the Appendix. Recognition accuracy was first analysed using repeated-measures analysis of variance (ANOVA), with Familiarisation (previewed vs. non-previewed) x Cue (remember vs. forget) as factors to assess the effect of familiarisation on DF (visualised in Figure 1, left panel). The second analysis was aimed at assessing the effect of level of familiarisation on DF by performing Preview (twice vs. six times) x Cue (remember vs. forget) repeated-measures ANOVA (visualised in Figure 2, right panel). In both analyses, the number of items was therefore equated across previewed vs. non-previewed conditions, and across twice- vs. six-times preview conditions.

There was a significant main effect of Familiarisation, confirming that the overall accuracy was better for previewed items ($M = 1.22$, $SD = 0.88$) compared to non-previewed items ($M = 1.03$, $SD = 0.79$), $F(1, 159) = 26.22$, $MSE = 0.30$, $p < .001$. There was also a significant main effect of Cue, indicating that Remember-cued words were recognised better ($M = 1.33$, $SD = 0.92$) compared to Forget-cued words ($M = 0.98$, $SD = 0.74$), $F(1, 159) = 89.65$, $MSE = 0.34$, $p < .001$, indicating a DF effect. Importantly, there was a significant Familiarisation x Cue interaction, $F(1, 159) = 11.48$, $MSE = 0.11$, $p < .001$. Follow-up tests showed impaired

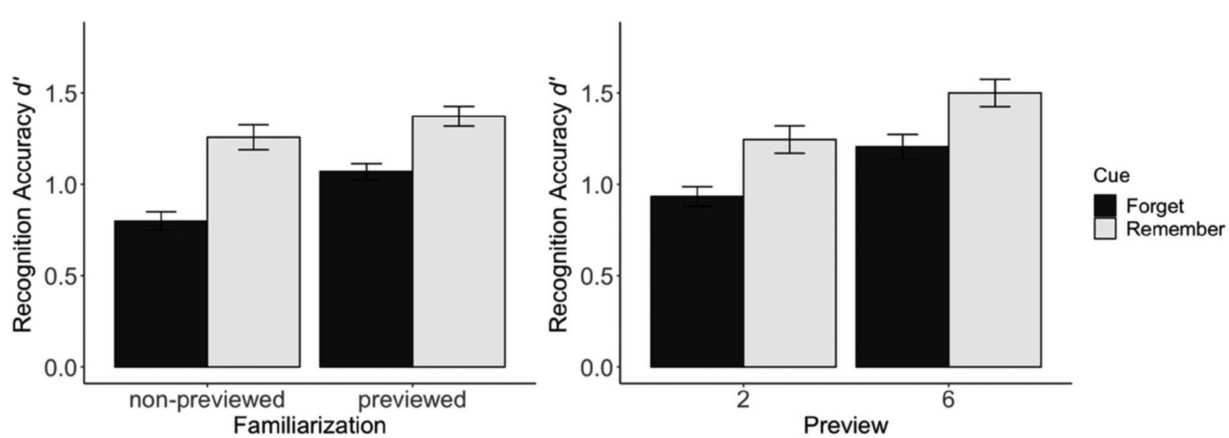

Figure 1. Recognition accuracy (*d′*) as a function of Familiarisation and Cue (left panel) and as a function of Preview and Cue (right panel) in Experiment 1. Error bars represent the standard error of the mean.

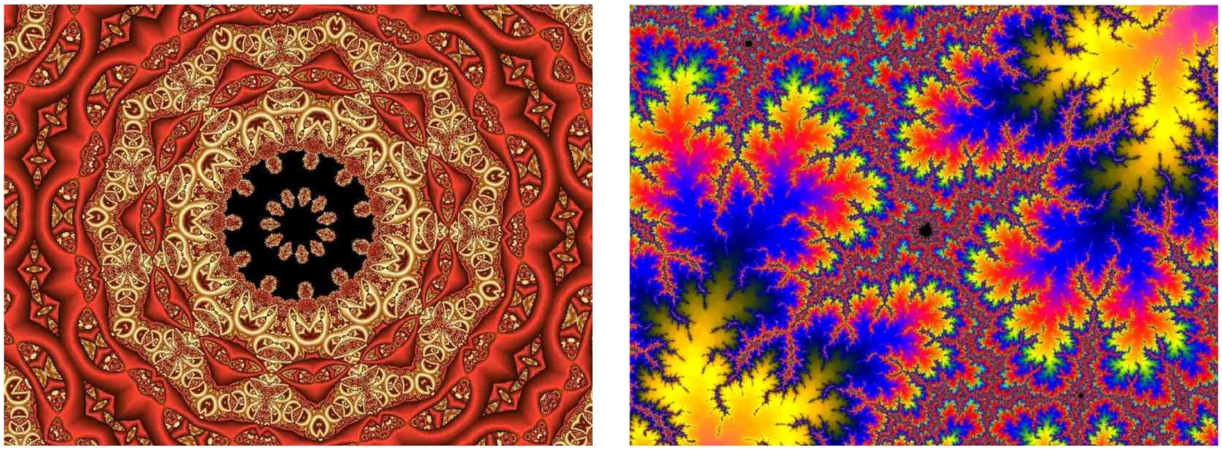

Figure 2. Examples of a circle shape (left) and a leaf shape (right) fractal images used in Experiment 2.

recognition for the Forget compared to the Remember condition for both non-previewed words, $t(159) = 10.81$, $p < .001$, $d_z = 0.85$, and previewed words, $t(159) = 7.66$, $p < .001$, $d_z = 0.42$. Although the DF effect was significant in both conditions, it was substantially reduced in the previewed condition because the F-items benefited more from familiarisation, $t(159) = 5.93$, $p < .001$, $d_z = 0.47$, compared to the R-items, $t(159) = 2.69$, $p < .05$, $d_z = 0.21$. In other words, it was more difficult to intentionally forget meaningful information that was familiar compared to unfamiliar.

To examine more specifically how the degree of familiarisation affects the magnitude of DF, Preview (twice vs. six times) x Cue (remember vs. forget) repeated ANOVA was conducted among previewed items. Results revealed a significant effect of Preview, confirming that overall recognition was higher for six times previewed items ($M = 1.35$, $SD = 0.91$) compared to twice previewed items ($M = 1.09$, $SD = 0.83$), $F(1, 159) = 63.34$, $MSE = 0.17$, $p < .001$. There was also a significant main effect of Cue, $F(1, 159) = 48.03$, $MSE = 0.31$, $p < .001$, indicating an overall DF effect. There was no interaction between Preview and Cue, suggesting that the DF magnitude did not differ between twice previewed words and six times previewed words, $F(1, 159) = .09$, $MSE = 0.11$, $p = .77$. Bayesian analysis ($BF_{10} = 0.074$) indicated the findings were 13.55 times more likely to be observed under a model that assumes there is no interaction between Preview and Cue than under a model that assumes there is a significant interaction between Preview and Cue.

In summary, previously familiarised verbal stimuli (achieved through objective repetitions) were overall better remembered than non-familiarised stimuli, confirming that repetition improved memory performance. Furthermore, previously familiarised stimuli became harder (but not impossible) to intentionally forget compared to less familiarised meaningful verbal materials. Interestingly, the degree of episodic familiarity did not appear to interact with DF effect, with the same magnitude of DF being observed for twice as six times previewed items. Thus, familiarisation hindered the ability to intentionally forget that item to some extent, but once an item had been familiarised, additional degree of familiarisation did not make that item completely immune to DF in the current study.

Experiment 2

Previous literature suggests the DF effect is more modest for visual stimuli than verbal stimuli (Hauswald & Kissler, 2008; Hourihan et al., 2009; Metzger, 2011; Quinlan et al., 2010). For example, Quinlan et al. (2010) compared verbal and visual stimuli directly within the same experiment, where they equated the content of line drawings and words (i.e., the word APPLE and a line drawing of an APPLE), and the results showed reduced DF with line drawings compared to DF with words, along with overall better memory for visual stimuli (see also Hauswald & Kissler, 2008). However, it is unclear whether the reduced DF effect may be driven by the visual nature of stimuli or higher memory performance level for pictures compared to words.

By using images that were completely novel, such as fractal images, and improving memory for these images through objective repetitions, Experiment 2 aimed to investigate the gradations of memory quality (using recognition accuracy as a proxy), and its effect on DF using visual stimuli. If the level of memory performance for studied items is unrelated to the size of DF for visual stimuli, we would anticipate obtaining a significant DF effect for fractal images and the same magnitude of DF effect across repetition conditions. In contrast, if the reduced DF observed in previous literature resulted from better memory associated with visual stimuli, we would expect a larger DF for non-previewed images relative to previewed images (conceptually replicating Experiment 1 findings with visual images). In summary, Experiment 2 pursued two objectives to disentangle the quality of memory and the nature of stimuli on the effect of modulated DF effect: first, to ascertain whether the DF effects observed in prior research with images could be replicated using fractal images, and second, to examine the impact of varying levels of memory performance on DF by manipulating through repetitions familiarity of the fractal images.

Method

Participants

Participants were 161 undergraduates from UIUC, who received course credit for participation. All data were collected online and all participants were recruited through the university SONA system. None of these participants had taken part in the Experiment 1.

Stimuli

Eighty-two colour images of fractals used in previous studies (Kinnell & Dennis, 2012; Osth et al., 2014) were used as stimuli in Experiment 2, each 600 × 400 pixels in size. Half of the fractals were circle-like shape, and the other half were leaf-like shape (see Figure 2, for examples). One circle fractal image and one leaf fractal image were used as practice examples, and the remaining 80 images were used as experimental stimuli.

Procedure

Procedures were very similar to Experiment 1, with a few exceptions. During the preview phase, participants were presented with a series of fractal images one at a time, at a rate of 3 s per image. Twenty unique fractal images were presented, out of which 10 were previewed twice, and 10 were previewed six times. Also, half of the fractals were circle-like, and the other half were leaf-like. To ensure that participants were paying attention to stimuli, they were told to classify each image into either a circle shape or a leaf shape category using keyboard buttons.

During the DF phase, participants were presented with 40 images, half of which were followed by a Remember memory cue and the other half by a Forget memory cue. A pilot study we conducted revealed that fractals were more difficult to recognise than verbal stimuli, and therefore we shortened the study list while also extending the presentation rate to 3 s per image to facilitate encoding. The remaining timing parameters were similar to Experiment 1. Twenty images had appeared in the preview phase, and the remaining 20 images were shown for the first time during the DF phase.

The recognition test contained 80 images, 40 of which had been presented in the DF phase and the other 40 were new items. Half of the items from the DF phase had been previewed either twice or six times prior to the DF phase, whereas the other half were only presented once in the DF phase only.

Results & discussion

Recognition accuracy (*d′* scores) was first analysed with Familiarisation x Cue repeated-measures ANOVA (visualised in Figure 3, left panel), to assess the effects of familiarisation on DF of meaningless visual stimuli. The second analysis used Preview x Cue repeated-measures ANOVA (visualised in Figure 3, right panel), to assess the effect of the level of familiarisations on DF for meaningless visual stimuli.

There was a significant main effect of Familiarisation, confirming that overall discrimination was better for previewed fractal images ($M = 1.09$, $SD = 0.75$) compared to non-previewed fractal images ($M = 0.70$, $SD = 0.61$), $F(1, 160) = 140.27$, $MSE = 0.23$, $p < .001$, confirming that repetition manipulation improved memory. There was neither a main effect of Cue, $F(1, 160) = 0.18$, $MSE = 0.22$, $p = .67$, $BF_{10} = 0.076$, nor a Familiarisation x Cue interaction, $F(1, 160) = 0.68$, $MSE = 0.19$, $p = .41$, $BF_{10} = 0.084$. Thus, there was no DF effect with fractal images, regardless of whether they were familiarised in the preview phase or seen for the first time in the DF phase.

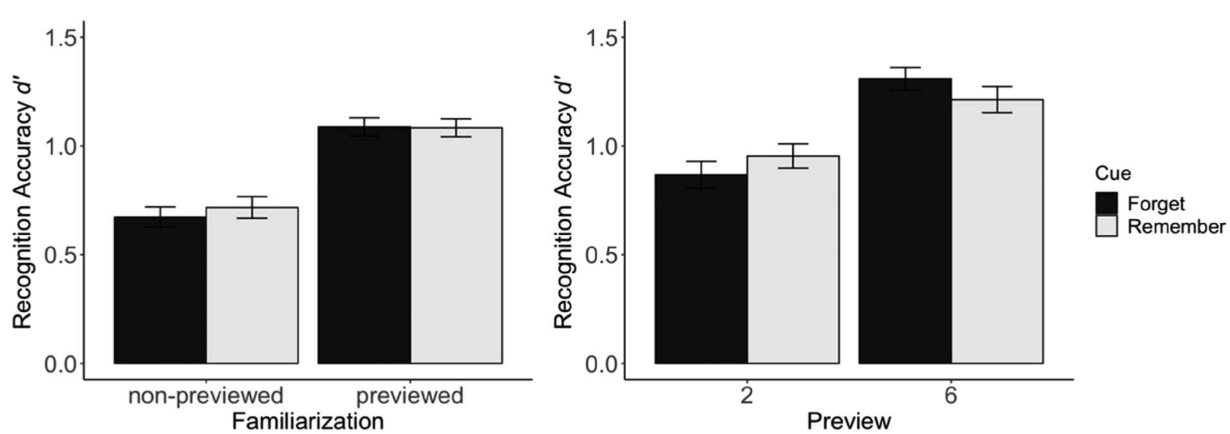

Figure 3. Recognition accuracy (*d′*) as a function of Familiarisation and Cue (left panel) and as a function of Preview and Cue (right panel) in Experiment 2. Error bars represent the standard error of the mean.

To examine more specifically how familiarity affects the magnitude of DF, Preview (twice vs. six times) x Cue (remember vs. forget) repeated ANOVA was conducted among previewed images. Results showed a significant effect of Preview, confirming that overall recognition accuracy was better for six times previewed items ($M = 1.26$, $SD = 0.71$), compared to twice previewed items ($M = 0.91$, $SD = 0.75$), $F(1, 160) = 75.59$, $MSE = 0.26$, $p < .001$. There was no main effect of Cue, $F(1, 160) = 0.01$, $MSE = .003$, $p = .91$, $BF_{10} = 0.176$, confirming the previous findings. However, there was a significant interaction between Preview and Cue, $F(1, 160) = 6.09$, $MSE = 0.30$, $p = .01$. Follow-up tests indicated that there was no DF for twice previewed fractal images $t(160) = 1.54$, $p = .12$, $d_z = 0.12$, $BF_{10} = 0.197$, but there was a marginally significant reversed DF effect for six times previewed fractals $t(160) = 1.99$, $p = .05$, $d_z = 0.16$, $BF_{10} = 0.329$.

In summary, the results demonstrated that unlike the words in Experiment 1, fractal images were not susceptible to the DF manipulation irrespective of overall memory performance level. The lack of a DF effect across all levels of familiarity suggests that familiarity is unlikely to be a contributing factor to the absence of the DF effect. This conclusion is also supported by the fact that memory performance level of six times previewed fractal images in the current study was comparable to twice previewed verbal materials in Experiment 1, and yet meaningful verbal materials showed significant DF effect in the previous experiment, whereas meaningless fractal images in the current experiment did not show DF (if anything at all, DF was in the reverse direction) despite approximately similar memory performance levels across experiments.

Previous research with simple line drawings of everyday objects showed that DF can be obtained with visual stimuli, albeit the effect was quite modest compared to DF obtained with verbal labels denoting those objects (Quinlan et al., 2010). In Quinlan et al. (2010), the line drawings were meaningful and could benefit from dual coding (Paivio, 1971), and such stimuli produced significant DF despite being better remembered than corresponding verbal label condition in the same experiment. Overall, the results of the current experiment as well as those obtained by Quinlan et al. (2010) suggest that the meaningfulness of the stimuli may play an additional and previously unappreciated role in DF effect. Interestingly, this experiment observed a reversed DF effect for fractals previewed six times during the preview phase. We are not aware of any previous "reverse-DF effect" in the literature and are cautious to overinterpret this finding, given its marginal significance.

Experiment 3

The results from Experiment 2 revealed that fractal images were not susceptible to DF, even when overall memory increased with repetitions, with the memory performance in the six times repeated condition reaching the levels of twice-repeated verbal materials observed in Experiment 1, where significant DF was observed. These findings suggest that degree of familiarity and performance level are unlikely to be responsible for the lack of DF with fractal images. The critical aspect that is different between words and fractal images may be their semantic meaningfulness, which could be a contributing factor to an emergence of DF with words but not fractal images.

Experiment 3 aimed to investigate the effects of semantic meaning on DF by using both meaningful and meaningless visual materials. To our knowledge, this is the first experiment to directly contrast the effects of meaningfulness as well as episodic familiarity on the magnitude of DF. Specifically, some visual stimuli involved line drawings of everyday objects that were meaningful to participants, whereas other stimuli consisted of line drawings of abstract shapes that were meaningless and novel. Again, we manipulated episodic familiarity of items by showing some of the to-be-studied materials in the preview phase, with the number of repetitions identical to those used in earlier experiments. Notably, similar to the previous two experiments, we used the term "episodic familiarisation" to denote the improvement in memory due to the effect of repetitions in the preview phase, which is distinct from the concept of "familiarity rating" that assesses how common an item is in daily life. Based on the first two experiments wherein we obtained a DF effect for meaningful stimuli but not for meaningless stimuli, we hypothesised that we would observe a significant DF effect with meaningful line drawings of everyday objects, and a reduced, if not negligible, DF effect for meaningless line drawings of abstract objects. Additionally, we anticipated a reduced DF effect for meaningful items after previewing, with the magnitude of reduction either remaining consistent for items previewed twice and those previewed six times (analogous to what was observed with meaningful words in Experiment 1), or the degree of familiarisation may interact with meaningful visual stimuli in unique ways, producing different magnitude of DF across repetition conditions. Given we obtained no DF in Experiment 2 for fractal images, it is likely that we would again obtain no DF with meaningless visual shapes, and this absence of DF would persist across different levels of familiarisation.

Method

Participants

There were 154 (88 men) participants recruited using Prolific (www.prolific.co) ($M_{age} = 26.04$, $SD = 5.40$). Participants were compensated with $4.40 upon completion. All data were collected online.

Stimuli

A total of 160 images were used in Experiment 3. Eighty images were line drawings of daily objects that were

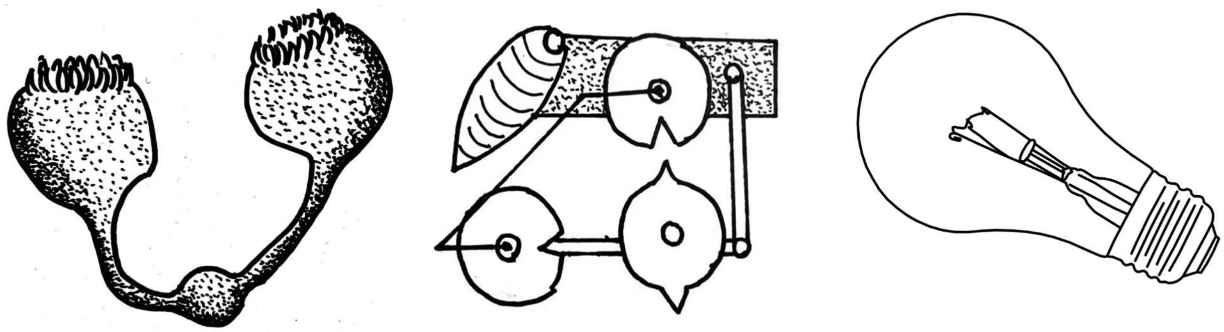

Figure 4. Examples of meaningless (left and middle figure) and meaningful (right figure) visual stimuli used in Experiment 3. The left figure is an image of a "germ", the middle figure is an image of a "machine", and the right figure is an image of a familiar everyday object.

meaningful to participants, which were selected from the Bank of Standardised stimuli (BOSS) (Brodeur et al., 2010, 2014). The other 80 stimuli were line drawing of objects designed to resemble biological organisms ("germs") or mechanical devices ("machines") that were meaningless to participants prior to the experiment. Half of the meaningless images were selected from published literature (Smith & Federmeier, 2020), and the other half were kindly created by a professional artist, who is a friend of the first author, Kefan Zhang (see Figure 4 for examples of meaningful and meaningless images).

Procedure

Procedures were very similar to Experiment 1, with a few exceptions. Forty unique images were presented, out of which 20 images were shown twice, and 20 were shown six times. Also, twenty were images of meaningful daily objects, and the other twenty were meaningless object images. Half of the meaningless stimuli were images of "germs", and the other half were images of "machines". Participants were told to classify the presented item according to whether the image contained any straight line. During the DF phase, forty images had presented in the preview phase, and the remaining 40 were shown for the first time during the DF phase. The recognition test contained 160 images, with 80 old items that had appeared in the DF phase, and 80 new items. All remaining parameters were similar to Experiment 1.

Results & discussion

The interplay between the effect of repetition and meaningfulness on DF effect was investigated by using Familiarisation (previewed vs. non-previewed) x Cue (remember vs. forget) x Stimulus Type (meaningful vs. meaningless) repeated-measures ANOVA on recognition accuracy (d' scores). Different false alarm rates were used to calculate d' scores for meaningful and meaningless items respectively. To further examine the influence of degree of familiarity on DF, recognition accuracy was also analysed using Preview (twice vs. six times) x Cue (forget vs. remember) x Stimulus Type (meaningful vs. meaningless) repeated-measures ANOVA.

There was a significant main effect of Familiarisation, indicating that overall accuracy was higher for previewed images ($M = 1.86$, $SD = 0.97$) compared to non-previewed images ($M = 1.20$, $SD = 0.95$), $F(1, 153) = 396.38$, $MSE = 0.45$, $p < .001$, confirming that repetition improved overall memory. There was also a significant main effect of Stimulus Type, $F(1, 153) = 556.62$, $MSE = 459.81$, $p < .001$, indicating that meaningful object images were better recognised ($M = 2.13$, $SD = 0.93$) than meaningless object images ($M = 1.13$, $SD = 0.82$). No main effect of Cue was observed, $F(1, 153) = 0.86$, $MSE = 0.15$, $p = .36$, $BF_{10} = 0.450$. However, there was a significant Cue x Familiarisation interaction, $F(1, 153) = 16.02$, $MSE = 0.16$, $p < .001$ (see Figure 5, left panel). Follow-up tests showed a significant DF effect for non-previewed images, $t(153) = 3.69$, $p < .001$, $d_z = 0.30$, but no significant DF for images that were presented in the preview phase, $t(153) = 1.56$, $p = .12$, $d_z = 0.13$, $BF_{10} = 0.061$. These findings confirm the previous experiments, demonstrating that familiarisation reduced the DF effect. Also, there was a significant Cue x Stimulus Type interaction, $F(1, 153) = 3.94$, $MSE = 0.17$, $p = .049$ (see Figure 5, right panel). Follow-up tests showed a significant DF effect for meaningful daily objects, $t(153) = 2.25$, $p = .03$, $d_z = 0.18$, but no DF for meaningless visual stimuli, $t(153) = 0.70$, $p = .49$, $d_z = 0.06$, $BF_{10} = 0.041$. These results suggest that meaningfulness of the study materials also affects DF in that meaningful visual stimuli were susceptible to DF, whereas meaningless stimuli were not. Finally, there was a significant Familiarisation x Stimulus Type interaction, $F(1, 153) = 9.27$, $MSE = 0.25$, $p = .003$ (see Figure 6). Follow-up tests showed that previewing improved memory for both meaningful daily object images, $t(153) = 16.30$, $p < .001$, $d_z = 0.63$, and for meaningless object images, $t(153) = 15.6$, $p < .001$, $d_z = 1.26$, but memory for meaningless object images benefited more from episodic familiarisation compared to meaningful everyday objects. There was no 3-way interaction, $F(1, 153) = 0.04$, $MSE = 0.19$, $p = .85$, $BF_{10} = 0.095$.

The results revealed a complex interplay between episodic familiarisation and stimulus type in the DF effect. The findings showed that overall recognition was better

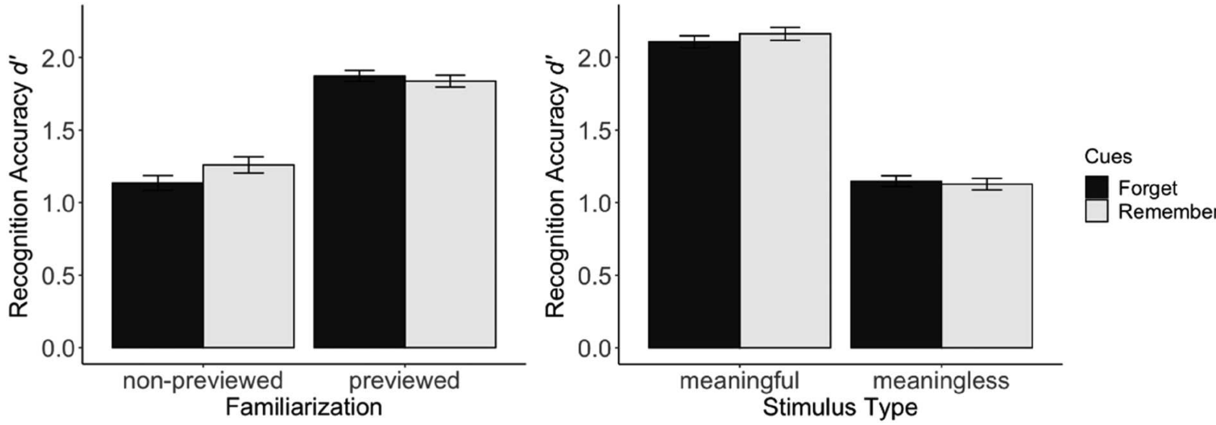

Figure 5. Recognition accuracy (*d′*) as a function of Familiarisation and Cue (left panel) and as a function of Stimulus and Cue (right panel) in Experiment 3. Error bars represent the standard error of the mean.

for previewed compared to non-previewed images, and it was also better for meaningful daily object images than meaningless object images. Moreover, familiarisation reduced the DF effect, with a significant DF effect for non-previewed images and no DF effect for images that were presented in the preview phase. This finding confirms the results of Experiment 1 and highlights the role of familiarisation in modulating the DF effect. Importantly, meaningfulness dimension also modulated the DF effect, with meaningless items being immune to DF compared to meaningful stimuli. The absence of the 3-way interaction suggests that the effects of familiarisation and meaningfulness on DF were independent of each other, with each of these dimensions reducing DF regardless of the other dimension.

Note, that contrary to the reduced yet significant DF effect for previewed words observed in Experiment 1, visual stimuli in the current experiment (combined across the meaning dimension) did not demonstrate any DF in the previewed condition. This could be driven by the fact that DF with words produced a large size DF effect in the non-previewed condition (consistent with established literature in DF), and previewing manipulation reduced the large size effect to a medium size effect. In contrast, visual stimuli produced small-to-medium effect size DF in the non-previewed condition (again, consistent with prior literature), making it more susceptible to elimination through previewing manipulation. Thus, the difference in the magnitude of the DF effect between verbal and visual items could explain why previewing did not eliminate DF with words, but eliminated it with visual materials.

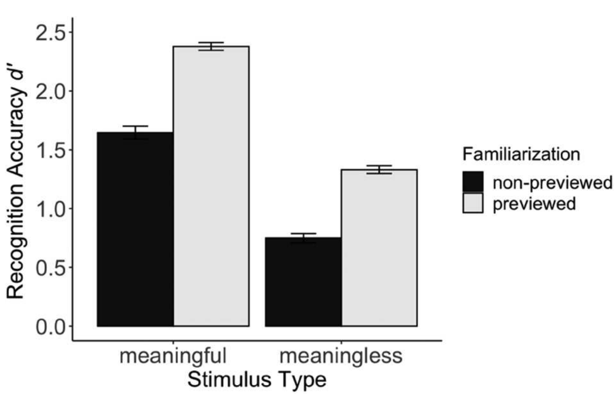

Figure 6. Recognition accuracy (*d′*) as a function of Familiarisation and Stimulus Type in Experiment 3. Error bars represent the standard error of the mean.

It is also worth noting that although meaningless images were overall recognised worse than meaningful images in the non-previewed condition, repeating meaningless objects six times in the preview phase brought their memory up to the levels comparable to non-previewed meaningful stimuli, and yet meaningless stimuli did not show DF, suggesting that episodic familiarisation and meaningfulness of stimuli independently influence DF. Nevertheless, while previewing led to a reduction in DF, the level of recognition performance did not have a differential impact on the efficacy of the DF manipulation between meaningful and meaningless images, given that there was no 3-way interaction.

In the succeeding analysis, we focused exclusively on previewed items, with the objective of examining how different levels of episodic familiarisation affected the magnitude of DF of meaningful and meaningless stimuli. Preview x Cue x Stimulus Type repeated-measures ANOVA analysis was conducted. The results showed a significant main effect of Preview, $F(1, 153) = 83.89$, $MSE = 0.26$, $p < .001$, confirming that overall recognition was better for six times previewed images ($M = 1.99$, $SD = 0.93$) compared to twice previewed images ($M = 1.72$, $SD = 1.00$). Again, there was a significant main effect of Stimulus Type, $F(1, 153) = 494.82$, $MSE = 0.68$, $p < .001$, indicating better recognition for meaningful stimuli ($M = 2.38$, $SD = 0.81$) than meaningless stimuli ($M = 1.33$, $SD = 0.93$). There was no main effect of Cue, $F(1, 153) = 2.44$, $MSE = 0.38$, $p = .12$, $BF_{10} = 0.060$. None of the two-way

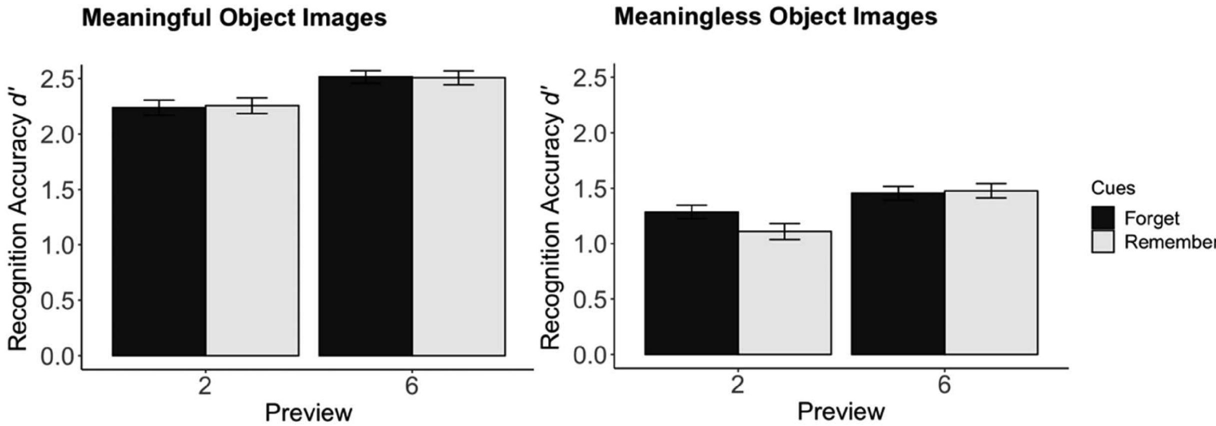

Figure 7. Recognition accuracy (*d′*) of as a function of Preview and Cue for previewed items (left panel: meaningful stimuli; right panel: meaningless stimuli) in Experiment 3. Error bars represent the standard error of the mean.

interactions were significant (Preview x Cue interaction, $F(1, 153) = 3.55$, $MSE = 0.58$, $p = .06$, $BF_{10} = 0.141$; Stimulus Type x Cue interaction, $F(1, 153) = 2.97$, $MSE = 0.51$, $p = .09$, $BF_{10} = 0.151$; Preview x Stimulus Type interaction, $F(1, 153) = 0.007$, $MSE = 0.001$, $p = .93$, $BF_{10} = 0.058$). However, there was a significant 3-way interaction, $F(1, 153) = 5.45$, $MSE = 0.18$, $p = .02$.

To follow-up the interaction, we performed repeated-measures ANOVAs for meaningful stimuli and meaningless stimuli separately, using Cue and Preview as variables. The results showed that there was a significant Cue x Preview interaction for meaningless object stimuli, $F(1, 153) = 7.10$, $MSE = 0.21$, $p < .009$ (Figure 7, right panel), but not for meaningful stimuli, $F(1, 153) = 0.20$, $MSE = 0.13$, $p = .65$, $BF_{10} = 0.070$. (Figure 7, left panel). Follow-up tests indicated that, for meaningless visual stimuli, there was a significant reversed DF for twice previewed items, $t(153) = 3.20$, $p = .002$, $d_z = 0.26$, meaning that participants were more likely to recognise those meaningless objects that were followed by a Forget cue compared to those that were followed by a Remember cue. No DF was observed for six times previewed meaningless visual stimuli, $t(153) = 0.48$, $p = .63$, $d_z = 0.04$, $BF_{10} = 0.053$.

To summarise, among previewed items, the level of familiarisation and the meaningfulness of the stimuli influenced the effectiveness of DF manipulation. Memory for previewed meaningful visual stimuli is not intentionally forgettable, at least for an immediate recognition test. When the stimuli were meaningless images, we again observed a reversed DF effect. Notably, this reversed effect was only apparent with objects that had been previewed twice, but not for six times previewed objects. This suggests that a certain degree of familiarisation with meaningless visual stimuli might result in a reversed DF effect.

In the current experiment, our aim was to clarify the influence of two factors on the DF effect, specifically the meaningfulness of visual stimuli and the degree of episodic familiarisation. Replicating findings from previous experiments, we observed DF for meaningful stimuli but not for meaningless stimuli. Furthermore, increasing the degree of episodic familiarisation prior to the DF phase negated the DF effect for all previewed line drawings. Therefore, our results indicate that both the meaningfulness of stimuli and episodic familiarisation independently influence the magnitude of the DF effect.

Experiment 4

The outcomes of the previous experiments suggest that both the meaningfulness of stimuli and episodic familiarity independently influence the magnitude of DF. This was evident as DF effects were only observed with meaningful stimuli, such as daily object images (Experiment 3) and words (Experiment 1), whereas they were absent with meaningless stimuli such as fractals (Experiment 2) and abstract object images (Experiment 3).

In the case of meaningful stimuli, episodic familiarisation prior to the DF phase reduced the size of DF. More specifically, DF was less effective for previewed items compared to those not previewed. However, further boosting memory performance via additional repetitions (previewed six times) only increased the overall memory but did not further diminish the DF effect in comparison to items previewed twice. The meaningless visual stimuli such as fractal images (Experiment 2) and abstract object images (Experiment 3) did not show any DF effect, and in some cases, a reversed DF effect was observed.

In Experiment 3, the DF effect was compared between meaningful and meaningless visual stimuli. The outcomes suggested that participants were only able to intentionally forget unfamiliar (i.e., not previewed) meaningful visual items but not unfamiliar meaningless items. Notably that even when meaningless items were made familiar through multiple repetitions and achieved recognition accuracy comparable to non-previewed meaningful images, there was still no DF effect for those previewed meaningless images, indicating that for visual stimuli, DF is contingent on the meaningfulness of the stimuli, and that meaningless stimuli of comparable memory performance levels are nevertheless immune to DF.

Experiment 4 aims to extend these findings into the verbal domain, because the impact of semantic meaning could differ between verbal and visual information, and given that majority of DF research involves verbal information, it is important to examine how meaningfulness and episodic familiarity influence DF in verbally meaningful and meaningless stimuli.

Participants

Participants were 112 undergraduates from UIUC, who received course credit for participation. All data was collected in person, in Fall of 2022.

Stimuli

There were 160 verbal stimuli used in Experiment 4, including 80 medium frequency words and 80 nonwords. The words were 5–7 letters long ($M = 6.15$, $SD = .80$), selected from the MRC Psycholinguistic Database (Coltheart, 1981), with an average Kucera-Francis frequency score of $M = 50.16$ ($SD = 13.48$).

Nonwords were selected from the ARC Nonword Database (Rastle et al., 2002). The number of letters has the same criteria as for the word selection criteria ($M = 6.39$, $SD = .70$). To select nonword stimuli that looked more word-like (as opposed to a random string of letters), we selected only nonwords with orthographically existing onsets and orthographically existing bodies, and restricted the number of orthographic neighbours and phonological neighbours to zero to minimise the probability of associating the nonwords with existing words. These restrictions produced a set of nonword stimuli that lacked meaning while being verbal in nature (e.g., *cralph, juths, splauve*).

Procedure

Procedures were similar to Experiment 3, with an exception that in lieu of meaningful and meaningless object images, we used words and nonwords as verbal stimuli. Given the verbal nature of stimuli, the orienting instruction in the preview task was the same as Experiment 1 (judging uper or lower case).

Results & discussion

To investigate how episodic familiarisation and meaningfulness of verbal stimuli influence the DF effect, recognition accuracy (d' scores) was analysed using Familiarisation (previewed vs. non-previewed) x Cue (remember vs. forget) x Stimulus Type (meaningful vs. meaningless) repeated-measures ANOVA. Different false alarm rates were used to calculate d' scores for meaningful and meaningless items respectively. Similar to previous experiments, we further explored how varying degrees of familiarity and meaningfulness impacted DF by analysing recognition accuracy using a Preview (twice vs. six times) x Cue (remember x forget) x Stimulus Type (meaningful x meaningless) repeated-measures ANOVA.

A Familiarisation x Cue x Stimulus Type repeated-measures ANOVA showed a significant main effect of Familiarisation, confirming that overall accuracy was better for previewed items ($M = 1.37$, $SD = 0.82$) compared to non-previewed items ($M = 0.99$, $SD = 0.74$), $F(1, 111) = 167.33$, $MSE = 0.27$, $p < .001$. There was also a significant main effect of Stimulus Type, $F(1, 111) = 91.94$, $MSE = 0.32$, $p < .001$, indicating that words were recognised better ($M = 1.49$, $SD = 0.83$) than nonwords ($M = 1.00$, $SD = 0.72$). In addition, there was a significant effect of Cue, $F(1, 111) = 30.71$, $MSE = 0.32$, $p < .001$, indicating that R-items ($M = 1.34$, $SD = 0.85$) were recognised better compared to F-items ($M = 1.16$, $SD = 0.77$). Moreover, results showed a significant Cue x Familiarisation interaction, $F(1, 111) = 4.04$, $MSE = 0.17$, $p = .047$ (see Figure 8, left panel). This interaction indicated that although there was a significant DF effect for both non-previewed stimuli, $t(111) = 4.91$, $p < .001$, $d_z = 0.46$, and previewed stimuli, $t(111) = 4.35$, $p < .001$, $d_z = 0.41$, the DF effect was smaller for previewed items than non-previewed items, because F-items benefited more from previewing, $t(111) = 11.1$, $p < .001$, $d_z = 1.05$, than R-items did, $t(111) = 9.11$, $p < .001$, $d_z = 0.86$, consistent with previous experiments. Another important significant interaction emerged between Stimulus Type and Cue, $F(1, 111) = 2.54$, $MSE = 0.20$, $p = .01$ (see Figure 8, right panel). Follow-up analyses showed that although there was a significant DF effect for both words, $t(111) = 5.34$, $p < .001$, $d_z = 0.51$, and nonwords, $t(111) = 3.08$, $p = .003$, $d_z = 0.29$, the DF effect was smaller for nonwords compared to words because R-items were better recognised in the word than nonword condition, $t(111) = 9.60$, $p < .001$, $d_z = 0.91$, compared to F-items, $t(111) = 7.69$, $p < .001$, $d_z = 0.73$. Overall, these results confirm that meaningfulness of materials affects DF such that meaningful verbal materials are more susceptible to DF compared to meaningless nonwords. There was neither a Familiarisation x Stimulus Type interaction, $F(1, 111) = 6.19$, $MSE = 0.21$, $p = .11$, $BF_{10} = 0.130$, nor a 3-way interaction, $F(1, 111) = 0.001$, $MSE = 0.19$, $p = .99$, $BF_{10} = 0.126$.

Overall, the first set of analyses revealed that participants had higher recognition accuracy for previewed than non-previewed items, as well as higher recognition accuracy for words than nonwords. Results also confirmed an overall DF effect, with R-items being better recognised than F-items. All of these effects were expected based on previous experiments. Importantly, DF was less effective for nonwords than words, confirming that the meaningfulness of stimuli affects DF, with meaningless stimuli showing reduced DF. Also, DF was smaller for previewed than non-previewed items, indicating that verbal stimuli with greater familiarity due to repetitions are

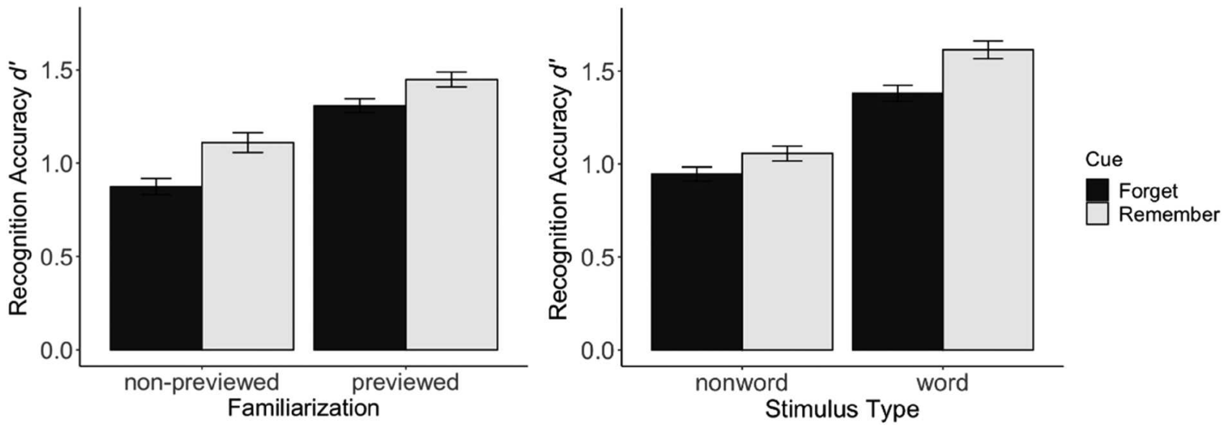

Figure 8. Recognition accuracy (*d*′) as a function of Familiarisation and Cue (left panel) and as a function of Stimulus Type and Cue (right panel) in Experiment 4. Error bars represent the standard error of the mean.

more immune to DF compared to less familiar non-repeated verbal stimuli. These results confirm the influence of material meaningfulness and familiarity on DF and are consistent with prior experiment. Importantly, episodic familiarity and semantic meaningfulness independently affected DF, as there was no 3-way interaction, replicating Experiment 3 findings but using verbal stimuli.

In the following set of analysis, we focused exclusively on previewed items and aimed to examine whether different extent of prior familiarity affected DF for meaningful and meaningless verbal stimuli. A Preview x Cue x Stimulus Type repeated-measures ANOVA analysis was conducted on recognition accuracy of previewed items. The results showed a significant main effect of Preview, $F(1, 111) = 24.92$, $MSE = 0.27$, $p < .001$, confirming that overall recognition was better for six times previewed items ($M = 1.46$, $SD = 0.84$) compared to twice previewed items ($M = 1.29$, $SD = 0.80$). A significant main effect of Stimulus Type was found, $F(1, 111) = 90.08$, $MSE = 0.68$, $p < .001$, showing that words were recognised better ($M = 1.64$, $SD = 0.83$) than nonwords ($M = 1.12$, $SD = 0.73$). Also, there was a significant main effect of Cue, $F(1, 111) = 18.95$, $MSE = 0.23$, $p < .001$, with R-items being recognised better ($M = 1.45$, $SD = 0.85$) than F-items ($M = 1.31$, $SD = 0.79$), confirming the DF effect. These findings are fully consistent with previous experiments. There was a significant Cue x Preview interaction, $F(1, 111) = 4.54$, $MSE = 0.23$, $p = .03$ (see Figure 9, left panel). Follow-up tests showed that there was a significant DF effect for twice previewed items, $t(111) = 4.22$, $p < .001$, $d_z = 0.40$, but not for six times previewed items, $t(111) = 1.77$, $p = .08$, $d_z = 0.17$, $BF_{10} = 0.111$. In addition, six times previewed items were better recognised than twice previewed items for both Remember items, $t(111) = 2.32$, $p = .02$, $d_z = 0.21$, and Forget items, $t(111) = 4.91$, $p < .001$, $d_z = 0.46$, but the F-items benefitted more from additional repetitions than R-items. There was also a Cue x Stimulus Type interaction, $F(1, 111) = 3.87$, $MSE = 0.22$, $p = .050$ (see Figure 9, right panel). Follow-up tests showed significant DF effects for words, $t(111) = 3.92$, $p < .001$, $d_z = 0.37$, and also a significant DF effect for nonwords, $t(111) = 2.15$, $p = .03$, $d_z = 0.20$, with a larger DF effect for words compared to nonwords. In addition, words were better recognised than nonwords for both Remember items, $t(111) = 9.60$,

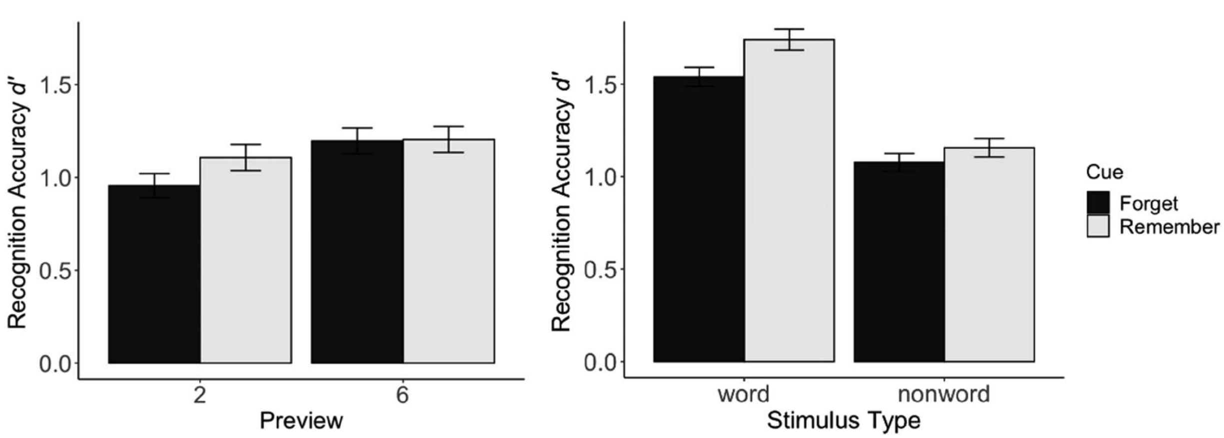

Figure 9. Recognition accuracy (*d*′) as a function of Preview and Cue (left panel) and as a function of Stimulus Type and Cue (right panel) for previewed items in Experiment 4. Error bars represent the standard error of the mean.

$p < .001$, $d_z = 0.91$, and Forget items, $t(111) = 7.69$, $p < .001$, $d_z = 0.73$, but the improvement in memory was larger for Remember items. No Stimulus Type x Preview interaction was found, $F(1, 111) = 0.02$, $MSE = 0.19$, $p = .90$, $BF_{10} = 0.068$. Finally, there was no 3-way interaction, $F(1, 111) = 0.02$, $MSE = 0.19$, $p = .88$, $BF_{10} = 0.136$, indicating that the observed effects of semantic meaningfulness and the degree of episodic familiarisation influenced the magnitude of DF independently and were not contingent on each other.

In summary, the results affirm the effectiveness of DF, rate of repetition, and item meaningfulness on recognition accuracy amongst previewed items. Our findings extended previous patterns observed in visual and verbal materials and demonstrated that both words and non-words with greater familiarity through repetition were more resistant to DF. For the first time, we observed DF in the context of meaningless stimuli.

General discussion

This investigation was prompted by the broad question of whether conditions that produce better memory impact DF in a consistent manner (i.e., by diminishing or enhancing it). The review of the literature did not yield a clear answer to this question because many conditions could affect memory performance, ranging from different processing instructions at encoding to different format/nature of the study stimuli. Therefore, the aim of our investigation was to uncover how format and semantic content of information, coupled with the level of episodic familiarisation, interact to influence the efficacy of DF. By enhancing memory performance through objective repetitions without invoking different processing instructions at encoding, our studies aimed to provide a more unambiguous assessment of how prior exposure of an item impacts its later susceptibility to DF. We examined these questions with visual and verbal stimuli that varied in the amount of semantic meaning they convey.

With visual materials used in Experiments 2 and 3, our results showed that DF was absent when stimuli are meaningless. Conversely, DF was present for meaningful non-previewed items, but it was eliminated after being previously previewed. This suggests that DF is less effective with more familiarised visual stimuli compared to unfamiliar visual stimuli. Additionally, there was a Cue by Stimulus Type interaction. DF was only effective with meaningful visual stimuli, but ineffective with meaningless visual stimuli. Neither the fractals, nor the abstract meaningless shapes showed any DF even in the novel, non-previewed condition. Among previewed stimuli, no DF was observed for meaningful visual images. In the case of meaningless previewed images, DF was either absent or occasionally even reversed. Note, that even after the meaningless stimuli had been previewed multiple times, which in some conditions raised their recognition accuracy levels comparable to the meaningful stimuli, meaningless stimuli nevertheless were not susceptible to DF. Collectively, these results suggest that the effectiveness of DF manipulation is independently influenced by both the item's semantic meaningfulness and the extent of prior episodic familiarisation. Curiously, a reversed DF effect was observed for twice previewed meaningless images in Experiment 3 and a marginally significant reversed DF effect for six times repeated fractal images in Experiment 2. The reasons for such reversed DF effects remain unclear, and to our knowledge there are no other reports of reversed DF in published studies. Additional research is needed to understand the nature of these findings.

When considering verbal materials, which were used in Experiments 1 and 4, a pattern similar to visual stimuli emerged: non-previewed items were more susceptible to DF manipulation compared to previewed items. However, unlike visual stimuli, verbal stimuli retained their susceptibility to the DF manipulation after previewing, albeit the effect size was substantially reduced compared to non-previewed verbal items. There was once again a Cue by Stimulus Type interaction similar to visual materials. Meaningful verbal materials resulted in greater DF than meaningless ones. However, unlike with visual stimuli, we obtained a DF effect with meaningless verbal items. With previewed stimuli, both twice and six times repeated verbal materials remained susceptible to DF.

There is a tentative indication that the study design might contribute to the effects of familiarisation on DF – namely, when the study list contained only meaningful verbal stimuli (Experiment 1), we observed a similar magnitude of DF for twice and six times previewed words, suggesting that the degree of familiarisation does not interact with DF. In contrast, when the study list contained a mixture of meaningful and meaningless verbal materials (Experiment 4), we observed a significant DF for twice-repeated verbal stimuli, but not with six times previewed stimuli, where DF was eliminated. Likewise, when the list contained only meaningless visual stimuli (Experiment 2), there was no DF effect with previewed materials. However, when the list contained a mixture of meaningless and meaningful visual stimuli (Experiment 3), DF was not only absent with twice previewed stimuli, but it was even reversed for six times previewed materials. Published findings in other memory phenomena indicate that the study list composition interacts with encoding effects like generation, production, enactment, and many others in consistent ways (for a review, see McDaniel & Bugg, 2008). Given that DF effect is also a phenomenon that emerges from processes operating during the encoding stage, it is possible that the list composition (mixed/pure lists) mediates the effects of familiarisation on DF. Collectively, these findings suggest the need for additional research on this topic.

Several DF mechanisms could be used to explain the reduced DF effects for meaningless visual and verbal materials, compared to meaningful materials, including the context unbinding account, inhibitory account, and selective rehearsal account. The context unbinding theory proposes that episodic context is crucial to DF effectiveness (Chiu et al., 2021; for review, see Sahakyan, 2024). Prior studies have shown that meaningful items bound to source information such as episodic context better than meaningless ones (Charness, 1976; Chase & Simon, 1973; Reder et al., 2006; Simon, 1974). The difficulty in intentionally forgetting meaningless visual and verbal materials may be due to a lack of encoded contextual information, vital for DF. While the context unbinding account could be one mechanism explaining the reduced DF for meaningless stimuli, other processes are likely to be involved as well. Proponents of the selective rehearsal account could posit that the pronounced DF in meaningful items stems from their amenability to rehearsal; hence, they are more easily rehearsed with the R instruction compared to meaningless stimuli. This logic could extend to explain the difference in DF between visual and verbal stimuli too: even though visual items can be rehearsed, verbal ones might be more susceptible to rehearsal because of its ease of verbal labelling, leading to a greater DF effect. The inhibition theory, conversely, would suggest that our cognitive systems are better adapted to suppress meaningful items. Items devoid of semantic content, like fractals or nonwords, might resist DF instructions, since they are less exposed to inhibitory processes, and therefore less natural or less routinely encountered by our cognitive system. Indeed, forgetting meaningless stimuli may not serve an adaptive function in the same way as forgetting outdated, incorrect, or distressing meaningful information does. As such, it may not be a commonly encountered situation in daily life, thus it may not be an activity that participants would typically engage in. Furthermore, there may be some benefit to rendering seemingly meaningless information less susceptible to forgetting processes, as it may be adaptive to remember information until we find meaning in it. Collectively, all three theoretical frameworks – selective rehearsal, inhibition, and context unbinding – can account for our findings on the effect meaningfulness on DF across the four experiments. Determining which theory better accounts for these findings is beyond the scope of the current investigation and requires additional investigation.

It should be acknowledged that the present study treated meaningfulness in a binary fashion – either an item is meaningful, or it is completely meaningless. For a more nuanced understanding, future studies could consider adopting a spectrum of meaningfulness. For instance, a more authentic reflection of real-life circumstances might be achieved by using stimuli that participants are vaguely aware of, but not necessarily familiar with. This could include music they have not listened to, faces they have not seen before, or locations they are aware of but have not visited. Another limitation of our current design is the inclusion of all previewed items in the subsequent DF phase, and therefore participants could use the prior exposure as a memory cue in the later recognition test. Specifically, if they remember seeing the item in the preview, they might infer it to be an "old" item in the recognition test. It is plausible that this source memory for the preview phase, rather than the degrees of episodic familiarisation of an item, played a role in attenuating the DF magnitude. To address this, future research using a similar design could incorporate items in the preview phase that do not appear in the later DF phase, and ask participants to only endorse items they remember seeing in the DF phase as old to eliminate potential confounding effects from source memory for the preview phase.

We note that some of our findings are different from with those reported by Hourihan et al. (2009) and Fawcett et al. (2016), where DF effects were observed with abstract visual stimuli, whereas we did not obtain such effects in Experiments 2 and 3. A number of factors might explain the reasons for these discrepancies. In Hourihan et al. (2009), the authors used abstract visual symbols and administered a delayed recognition test after 24 h instead of the immediate test because performance was at ceiling on the immediate test. The authors thereby obtained a significant DF with abstract visual stimuli on the delayed test (in Experiment 1). Recent findings indicate that DF is associated with accelerated rate of forgetting – over time, the Forget condition forgets at a faster rate compared to the Remember condition (Nickl & Bäuml, 2023; Whitlock et al., 2024). This is particularly noteworthy in light of the findings of Whitlock et al. (2024) study, who intentionally set up an experiment where the DF effect was absent on an immediate test, but DF started to emerge after the delayed tests. Results from fitting of forgetting functions to the data confirmed that the results were driven by the accelerated forgetting in the Forget condition, akin to those reported by Nickl and Bäuml (2023). Therefore, significant DF with abstract stimuli after a 24-hour period obtained by Hourihan et al. (2009) may have been due to the delayed testing in that study. In their second experiment, Hourihan et al. (2009) used a concurrent task during the DF phase to prevent verbal rehearsal (a concurrent task was also implemented by Fawcett et al. (2016) after Forget/Remember memory cues). Previous findings demonstrate that DF is more robust with increased cognitive load from concurrent tasks (Lee & Lee, 2011). Hence, the use of the concurrent task could have also contributed to these studies obtaining significant DF with abstract stimuli compared to our studies, where no concurrent tasks were used.

Another potential factor that may explain the differences in findings between our studies and those in prior literature might be the nature of abstract stimuli used. Specifically, compared to the abstract stimuli used in our experiments (fractals and novel object stimuli), those used in previous research contained fewer diagnostic features to differentiate one shape from another. In contrast, our abstract stimuli had to be categorised during the encoding phase based on their features (i.e., circles vs. leaf shapes for fractals in Experiment 2, or germs vs. tools for novel images in Experiment 3). Our images contained more differentiating features to permit better discrimination (colours, shapes, ad-hoc categories). The similarity of shapes used in previous published studies likely leads to the recognition process relying more on retrieval of more specific details because there are fewer alternative bases for recognition, and previous research indicates that recognition driven by recollection (vs. pure familiarity) contributes to exaggerated DF effect (Sahakyan et al., 2009). Therefore, it is plausible that the stimuli used in previous research contributed to more recollection driven recognition, thereby enabling DF effect to emerge with those types of stimuli.

Finally, differences in list length used between the studies could also be a factor that needs additional scrutiny. In general, comparing the number of stimuli across our studies with previously published studies suggests that the number of stimuli we employed is unlikely to explain the difference in these findings. With meaningful verbal stimuli, prior studies use much longer lists than what we had employed, using up to 240 stimuli (Scholz et al., 2021), and nevertheless obtaining significant DF. Similarly, our Experiment 1, using words, employed a larger number of stimuli than Experiment 2 with fractals, and nevertheless yielded a robust DF. Admittedly, these are all studies involving meaningful verbal stimuli, whereas the lack of DF in our studies was obtained with meaningless stimuli, and it is possible that shorter lists might lead to DF effect with abstract stimuli. Although this could be the case, as the study by Fawcett et al. (2016) using fewer meaningless stimuli (15 R and 15 F) observed a significant DF, the recognition accuracy (calculated from the reported hit rates and false alarm rates) was *lower* compared to what we obtained using *longer* lists, a pattern not consistent with list length effect. Additionally, this explanation appears to be less parsimonious because it hinges on the assumption that list length interacts with the meaningfulness of materials, such that it does not interact with meaningful stimuli (such as words) to yield DF, but interacts with meaningless materials (such as fractals) to eliminate DF. If longer list lengths in general prompt participants to ignore the DF instruction, then it is unclear a priori why the number of study stimuli would differentially affect the effectiveness of DF depending on the nature of stimuli. This might indeed be the case, but it needs to be demonstrated and examined further. Overall, a number of critical factors could account for the differences between previous findings and our findings indicating no DF with abstract meaningless stimuli. We note, that our experiments recruited substantially larger samples, and were more powered compared to previous research, where critical tests often hinged on samples of fewer than 20 participants.

In conclusion, the results of our experiments highlight the crucial roles of the level of familiarity and stimulus type in DF. Our findings show that recognition performance levels do not solely determine the magnitude of DF. Although episodic familiarity tends to reduce the size of DF, the format and the semantic meaningfulness are equally important when we evaluate the effectiveness of the DF manipulation. We observed that visual items, despite their higher recognition performance, yield reduced DF. Similarly, items familiarised through repetitions, which also showed higher recognition performance compared to unfamiliarised items, showed reduced DF sizes. Conversely, words, despite their higher recognition performance, showed larger DF size compared to nonwords. The current studies highlight the importance of considering the format, meaningfulness, and episodic familiarity of stimuli to enrich our understanding of factors contributing to the manifestation of the DF effect.

Data availability statement

The data that support the findings of this study are openly available in *Episodic memory, stimulus types and DF* [https://doi.org/10.17605/OSF.IO/28NTW].

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Basden, B., & Basden, D. (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory*, *4*(6), 633–654. https://doi.org/10.1080/741941000

Basden, B. H. (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory*, *4*(6), 633–654. https://doi.org/10.1080/741941000

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). Lawrence Erlbaum Associates Publishers.

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. https://doi.org/10.1037/0278-7393.19.3.603

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

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). Lawrence Erlbaum Associates Publishers.

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

Bjork, R. A. (2011). On the symbiosis of remembering, forgetting, and learning. In A. S. Benjamin (Ed.), *Successful remembering and* *successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 1–22). Psychology Press.

Bjork, R. A., Laberge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, *10*, 55–56. https://doi.org/10.3758/BF03331404

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, Article e9414. https://doi.org/10.7717/peerj.9414

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a New Set of 480 normative photos of objects to Be used as visual stimuli in cognitive research. *PLoS One*, *5*(5), Article e10773. https://doi.org/10.1371/journal.pone.0010773

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PLoS One*, *9*(9), Article e106953. https://doi.org/10.1371/journal.pone.0106953

Bugelski, B. R. (1970). Words and things and images. *American Psychologist*, *25*(11), 1002–1012. https://doi.org/10.1037/h0030150

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(6), 641–653. https://doi.org/10.1037/0278-7393.2.6.641

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chiu, Y.-C., Wang, T. H., Beck, D. M., Lewis-Peacock, J. A., & Sahakyan, L. (2021). Separation of item and context in item-method directed forgetting. *NeuroImage*, *235*, Article 117983. https://doi.org/10.1016/j.neuroimage.2021.117983

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Ding, H., Whitlock, J., & Sahakyan, L. (2022). Can intentional forgetting reduce the cross-race effect in memory? *Psychonomic Bulletin & Review*, *29*, 1387–1396. https://doi.org/10.3758/s13423-022-02080-6

Dulaney, C. L., Marks, W., & Link, K. E. (2004). Aging and directed forgetting: Pre-cue encoding and post-cue rehearsal effects. *Experimental Aging Research*, *30*(1), 95–112. https://doi.org/10.1080/03610730490251504

Earles, J. L., & Kersten, A. W. (2002). Directed forgetting of actions by younger and older adults. *Psychonomic Bulletin & Review*, *9*(2), 383–388. https://doi.org/10.3758/BF03196297

Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. *Zeitschrift Für Experimentelle Und Angewandte Psychologie*, *27*(4), 511–533.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

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. https://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. https://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*, 797–808. https://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. https://doi.org/10.1016/j.actpsy.2011.10.001

Fellner, M.-C., Waldhauser, G. T., & Axmacher, N. (2020). Tracking selective rehearsal and active inhibition of memory traces in directed forgetting. *Current Biology*, *30*(13), 2638–2644.e4. https://doi.org/10.1016/j.cub.2020.04.091

Gardiner, J. M., Gawlik, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review*, *1*(1), 107–110. https://doi.org/10.3758/BF03200764

Geiselman, R. E., Rabow, V. E., Wachtel, S. L., & Mackinnon, D. P. (1985). Strategy control in intentional forgetting. *Human Learning: Journal of Practical Research & Applications*, *4*(3), 169–178.

Hauswald, A., & Kissler, J. (2008). Directed forgetting of complex pictures in an item method paradigm. *Memory*, *16*(8), 797–809. https://doi.org/10.1080/09658210802169087

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Hockley, W. E., Ahmad, F. N., & Nicholson, R. (2016). Intentional and incidental encoding of item and associative information in the directed forgetting procedure. *Memory & Cognition*, *44*(2), 220–228. https://doi.org/10.3758/s13421-015-0557-8

Horton, K. D., & Petruk, R. (1980). Set differentiation and depth of processing in the directed forgetting paradigm. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 599–610. https://doi.org/10.1037/0278-7393.6.5.599

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *62*(4), 242–246. https://doi.org/10.1037/1196-1961.62.4.242

Hourihan, K. L., Ozubko, J. D., & MacLeod, C. M. (2009). Directed forgetting of visual symbols: Evidence for nonverbal selective rehearsal. *Memory & Cognition*, *37*(8), 1059–1068. https://doi.org/10.3758/MC.37.8.1059

Kinnell, A., & Dennis, S. (2012). The role of stimulus type in list length effects in recognition memory. *Memory & Cognition*, *40*(3), 311–325. https://doi.org/10.3758/s13421-011-0164-2.

Lee, Y.-S. (2013). Costs and benefits in item-method directed forgetting: Differential effects of encoding and retrieval. *The Journal of General Psychology*, *140*(3), 159–173. https://doi.org/10.1080/00221309.2012.750591

Lee, Y.-S., & Lee, H.-M. (2011). Divided attention facilitates intentional forgetting: Evidence from item-method directed forgetting. *Consciousness and Cognition*, *20*(3), 618–626. https://doi.org/10.1016/j.concog.2010.09.008

Ludowig, E., Möller, J., Bien, C. G., Münte, T. F., Elger, C. E., & Rosburg, T. (2010). Active suppression in the mediotemporal lobe during directed forgetting. *Neurobiology of Learning and Memory*, *93*(3), 352–361. https://doi.org/10.1016/j.nlm.2009.12.001

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. https://doi.org/10.3758/BF03210819

MacLeod, C. M., & Daniels, K. A. (2000). Direct versus indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review*, *7*(2), 354–359. https://doi.org/10.3758/BF03212993.

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*(2), 237–255. https://doi.org/10.3758/PBR.15.2.237

Metzger, M. M. (2011). Directed forgetting: Differential effects on typical and distinctive faces. *The Journal of General Psychology*, *138*(2), 155–168. https://doi.org/10.1080/00221309.2011.557407

Nickl, A. T., & Bäuml, K.-H. T. (2023). To-be-forgotten information shows more relative forgetting over time than to-be-remembered information. *Psychonomic Bulletin & Review*, *31*, 156–165. https://doi.org/10.3758/s13423-023-02330-1

Nowicka, A., Marchewka, A., Jednoróg, K., Tacikowski, P., & Brechmann, A. (2011). Forgetting of emotional information is hard: An fMRI study of directed forgetting. *Cerebral Cortex*, *21*(3), 539–549. https://doi.org/10.1093/cercor/bhq117

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences*, *1424*(1), 239–255. https://doi.org/10.1111/nyas.13603

Oehrn, C. R., Fell, J., Baumann, C., Rosburg, T., Ludowig, E., Kessler, H., Hanslmayr, S., & Axmacher, N. (2018). Direct electrophysiological evidence for prefrontal control of hippocampal processing during voluntary forgetting. *Current Biology*, *28*(18), 3016–3022.e4. https://doi.org/10.1016/j.cub.2018.07.042

Osth, A. F., Dennis, S., & Kinnell, A. (2014). Stimulus type and the list strength paradigm. *Quarterly Journal of Experimental Psychology*, *67*(9), 1826–1841. https://doi.org/10.1080/17470218.2013.872824

Paivio, A. (1971). *Imagery and verbal processes* (pp. xi, 596). Holt, Rinehart & Winston.

Paz-Caballero, M. (2004). Predictive validity of event-related potentials (ERPs) in relation to the directed forgetting effects. *Clinical Neurophysiology*, *115*(2), 369–377. https://doi.org/10.1016/j.clinph.2003.09.011

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Quinlan, C. K., Taylor, T. L., & Fawcett, J. M. (2010). Directed forgetting: Comparing pictures and words. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *64*(1), 41–46. https://doi.org/10.1037/a0016569

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword database. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1339–1362. https://doi.org/10.1080/02724980244000099

Reber, P. J., Siwiec, R. M., Gitleman, D. R., Parrish, T. B., Marsel Mesulam, M., & Paller, K. A. (2002). Neural correlates of successful encoding identified using functional magnetic resonance imaging. *The Journal of Neuroscience*, *22*(21), 9541–9548. https://doi.org/10.1523/JNEUROSCI.22-21-09541.2002

Reder, L. M., Oates, J. M., Thornton, E. R., Quinlan, J. J., Kaufer, A., & Sauer, J. (2006). Drug-induced amnesia hurts recognition, but only for memories that can be unitized. *Psychological Science*, *17*(7), 562–567. https://doi.org/10.1111/j.1467-9280.2006.01744.x

Rizio, A. A., & Dennis, N. A. (2013). The neural correlates of cognitive control: Successful remembering and intentional forgetting. *Journal of Cognitive Neuroscience*, *25*(2), 297–312. https://doi.org/10.1162/jocn_a_00310

Sahakyan, L. (2024). Current perspectives on directed forgetting. In A. Wagner, & M. Kahana (Eds.), *Oxford handbook of human memory* (pp. 1257–1277). Oxford University Press.

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. https://doi.org/10.1016/j.jml.2009.02.006

Sahakyan, L., Waldum, E. R., Benjamin, A. S., & Bickett, S. (2009). Where is the forgetting with list-method directed forgetting in recognition? *Memory & Cognition*, *37*(4), 464–476.

Scholz, S., Dutke, S., & Busch, N. A. (2021). Oscillatory correlates of intentional forgetting: The role of theta and alpha power in item-method directed forgetting. *Eneuro*, *8*(5), ENEURO.0022-21.2021. https://doi.org/10.1523/ENEURO.0022-21.2021

Scotti, P. S., & Maxcey, A. M. (2022). Directed forgetting of pictures of everyday objects. *Journal of Vision*, *22*(10), 8–8. https://doi.org/10.1167/jov.22.10.8

Simon, H. A. (1974). How big is a chunk? *Science*, *183*(4124), 482–488. https://doi.org/10.1126/science.183.4124.482

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592.

Smith, C. M., & Federmeier, K. D. (2020). Neural signatures of learning novel object–scene associations. *Journal of Cognitive Neuroscience*, *32*(5), 783–803. https://doi.org/10.1162/jocn_a_01530

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Van Hooff, J. C., & Ford, R. M. (2011). Remember to forget: ERP evidence for inhibition in an item-method directed forgetting paradigm. *Brain Research*, *1392*, 80–92. https://doi.org/10.1016/j.brainres.2011.04.004

Wetzel, C. D. (1975). Effect of orienting tasks and cue timing on the free recall of remember- and forget-cued words. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(5), 556–566. https://doi.org/10.1037/0278-7393.1.5.556.

Wetzel, C. D., & Hunt, R. E. (1977). Cue delay and the role of rehearsal in directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(2), 233–245. https://doi.org/10.1037/0278-7393.3.2.233

Whitlock, J., Ding, H., Hubbard, R., & Sahakyan, L. (2024). *Directed forgetting and thought substitution in delayed testing*. Manuscript submitted for publication; revision submitted.

Whitlock, J., Lo, Y.-P., Chiu, Y.-C., & Sahakyan, L. (2020). Eye movement analyses of strong and weak memories and goal-driven forgetting. *Cognition*, *204*, Article 104391. https://doi.org/10.1016/j.cognition.2020.104391

Appendix

Table A1. Hit (HT) and false alarm rates (FAR) in Experiment 1 and Experiment 2.

		HT			FAR
			M (SD)		*M* (SD)
		Non-previewed	Previewed, twice	Previewed, six-times	
Exp. 1 (words)	Remember	0.689 (0.18)	0.672 (0.20)	0.655 (0.19)	0.270 (0.17)
	Forget	0.530 (0.19)	0.577 (0.20)	0.743 (0.22)	
Exp. 2 (fractals)	Remember	0.599 (0.18)	0.674 (0.19)	0.748 (0.20)	0.345 (0.15)
	Forget	0.584 (0.17)	0.648 (0.20)	0.781 (0.16)	

Table A2. Hit (HT) and false alarm rates (FAR) in Experiment 3.

		HT			FAR
			M (SD)		*M* (SD)
		Non-previewed	Previewed, twice	Previewed, six-times	
Meaningful images	Remember	0.622 (0.24)	0.788 (0.18)	0.860 (0.10)	0.123 (0.12)
	Forget	0.579 (0.22)	0.782 (0.19)	0.866 (0.12)	
Meaningless images	Remember	0.589 (0.23)	0.685 (0.22)	0.796 (0.17)	0.331 (0.20)
	Forget	0.561 (0.21)	0.741 (0.19)	0.796 (0.15)	

Table A3. Hit (HT) and false alarm rates (FAR) in Experiment 4.

		HT			FAR
			M (SD)		*M* (SD)
		Non-previewed	Previewed, twice	Previewed, six-times	
Words	Remember	0.640 (0.20)	0.730 (0.21)	0.767 (0.18)	0.202 (0.14)
	Forget	0.537 (0.22)	0.655 (0.21)	0.734 (0.22)	
Nonwords	Remember	0.658 (0.19)	0.722 (0.19)	0.754 (0.18)	0.360 (0.18)
	Forget	0.593 (0.20)	0.678 (0.20)	0.750 (0.18)	