Disfluent difficulties are not desirable difficulties: the (lack of) effect of Sans Forgetica on memory

Andrea Taylor, Mevagh Sanson, Ryan Burnell, Kimberley A. Wade & Maryanne Garry

Published online: 04 May 2020.

Submit your article to this journal ⟋

View related articles ⟋

View Crossmark data ⟋

Disfluent difficulties are not desirable difficulties: the (lack of) effect of Sans Forgetica on memory

Andrea Taylor[a], Mevagh Sanson [a], Ryan Burnell[a], Kimberley A. Wade[b] and Maryanne Garry[a]

[a]Psychology, The University of Waikato, Hamilton, New Zealand; [b]Psychology, University of Warwick, Coventry, UK

ABSTRACT
Scientists working at the intersection of cognitive psychology and education have developed theoretically-grounded methods to help people learn. One important yet counterintuitive finding is that making information harder to learn – that is, creating *desirable difficulties* – benefits learners. Some studies suggest that simply presenting information in a difficult-to-read font could serve as a desirable difficulty and therefore promote learning. To address this possibility, we examined the extent to which Sans Forgetica, a newly developed font, improves memory performance – as the creators of the font claim. Across four experiments, we set out to replicate unpublished findings by the font's creators. Subjects read information in Sans Forgetica or Arial, and rated how difficult the information was to read (Experiment 1) or attempted to recall the information (Experiments 2–4). Although subjects rated Sans Forgetica as being more difficult to read than Arial, Sans Forgetica led to equivalent memory performance, and sometimes even impaired it. These findings suggest that although Sans Forgetica promotes a feeling of disfluency, it does not create a desirable difficulty or benefit memory.

One point of education is to help people learn new information. But sometimes people struggle to retain this information. It is not surprising, then, that research spanning the overlap between cognitive psychology and education has yielded several theoretically-grounded ways to help students learn. Many of these "ways" entail making that information harder for them to learn. This seemingly paradoxical suggestion is supported by a wealth of literature showing that making a task more difficult can be beneficial for learning (for reviews, see Bjork, 1994; Bjork & Bjork, 2011).

Such "desirable difficulty" can be created in many different ways. For instance, when studying information, having to generate some of that information oneself (given some guiding prompts), rather than simply reading it, leads to better memory for the generated information – even though generating it requires extra work at the time (deWinstanley, 1995; Slamecka & Graf, 1978). Likewise, people remember words they read aloud better than words they read silently – even though reading words aloud requires more effort (Hopkins & Edwards, 1972; MacLeod et al., 2010). Moreover, when studying information repeatedly, spacing out those repetitions, rather than having them back-to-back, leads to better memory for that information on a later test – even though restudying after a longer gap is more difficult in the moment (for a meta-analysis, see Cepeda et al., 2006).

Of course, not all difficulties are desirable, and desirable difficulties are notoriously fickle (see, for example, Morehead et al., 2019; c.f. Mueller & Oppenheimer, 2014). Manipulations that serve as desirable difficulties in one context, or for one person, might serve as undesirable difficulties in another context, or for another person (McDaniel & Butler, 2010). In designing desirable difficulties, then, it is important to consider both the background and ability of the learner, and the manner in which the information will later be recalled.

When successful, the idea is that desirable difficulties make the information harder to learn, but they also encourage people to process the information more deeply, slowly, and effortfully than they otherwise would have. This processing leads people to better integrate the information with what they already know and therefore have a better memory for it in the long run (cf. Craik & Tulving, 1975; McDaniel & Butler, 2010). This mechanism is reminiscent of a related literature on the effects of cognitive fluency. In simple terms, when information feels easy to process it is said to be fluent, whereas when information feels difficult to process it is said to be disfluent. People tend to take a feeling of fluency as an "all good" signal and therefore rely on faster, "shallower," heuristic processes. By contrast, they take a feeling of disfluency as a "danger" signal, and tend to switch to using slower, "deeper," systematic processes (Alter et al., 2007; see also, Lindsay, 2008; Oppenheimer, 2008). If we recast the

notion of desirable difficulties along fluency lines, then, we might say manipulations that create feelings of disfluency should push people towards the kind of effortful processing of information that results in better memory for that information. In short, the feeling of disfluency itself might be a desirable difficulty.

Moreover, the fluency literature suggests that such difficulty can be created without making the information itself more difficult to learn. Instead, merely making the information look more difficult can make it feel more disfluent – and might promote this beneficial processing and therefore also improve memory (for a review, see Alter & Oppenheimer, 2009). The initial demonstration of this very effect asked subjects in one experiment to learn classroom-like material (Diemand-Yauman et al., 2011). Some of this material was presented in an easy-to-read, fluent font, and some was presented in one of several difficult-to-read, disfluent fonts (for instance, greyscale and italicised). Subjects remembered more of the information they saw in the disfluent fonts. In a second experiment, some high school students were given class materials in disfluent fonts across an entire semester, whereas others were given the same materials in fluent fonts. Students given materials in the disfluent fonts performed better on the assessments for that class. This study fits with the idea that indeed we can help students learn information just by making the font in which they see it more difficult to read.

The appeal of this simple intervention led to a flurry of work on the effects of disfluent fonts on memory. But the results were mixed: sometimes a disfluent font benefitted memory, sometimes it did little, and sometimes it even hurt memory (Eitel et al., 2014; Eitel & Kühl, 2016; French et al., 2013; Rhodes & Castel, 2008; Yue et al., 2013). These conflicting findings motivated a recent meta-analysis that aimed to more precisely estimate the effect of disfluent fonts on memory (Xie et al., 2018). Across 39 experiments, from 25 empirical articles, the overall estimate of the effect of disfluent fonts on memory was trivial, and plausibly zero. But the authors identified possible moderators they were unable to examine in the meta-analysis – most notably, for our purposes, the level of disfluency created by the fonts used in each study. A single study that manipulated degree of font legibility across several levels produced an inverted u-shape trend, whereby a moderately disfluent font appeared slightly more beneficial for learning than both more and less disfluent fonts (Seufert et al., 2017). What, then, should we conclude? Perhaps disfluent fonts do not produce a desirable difficulty that encourages deeper cognitive processing. Or perhaps researchers were yet to identify the font that produces the "optimal" level of disfluency.

Then, in 2018, a team comprised of academics with expertise in psychology, marketing, and graphic design created a new font purported to produce an optimal level of disfluency. The team chose this specific font – with an unusual back-slanted, broken appearance (see

Sans Forgetica

Figure 1. Text in Sans Forgetica. Sans Forgetica is licensed under the Creative Commons Attribution-NonCommercial License (CC BY-NC; https://creativecommons.org/licenses/by-nc/3.0/)

Figure 1) – based on the results of their first experiment. In that lab-based experiment, they showed university students ($N = 96$) several word pairs in three disfluent fonts – which respectively broke one, two, or three principles of good design – and a fluent font. Subjects remembered 69% of the pairs they saw in the moderately disfluent font, compared to 61% of the pairs they saw in the slightly disfluent and extremely disfluent fonts (and 68% of those in the fluent font; Earp, 2018). The team, therefore, identified the moderately-disfluent, back-slanted and broken font as the one best able to create a desirable difficulty that promotes learning. "When we want to learn something and remember it, it's good to have a little bit of an obstruction added to that learning process," one of the researchers, Dr. Janneke Blijlevens, explained to The Guardian (Martin, 2018). In their subsequent online experiment, students ($N = 303$) remembered 57% of information from text passages written in the new font, compared to 50% of information from passages written in Arial (Earp, 2018). The name of this new font? Sans Forgetica.

Since its release, Sans Forgetica has received accolades, including the "Best in class" GoodDesign award (Good Design, 2019) and a nomination for the "2018 Trade Name of the Year" (though it was trounced by the Philadelphia Flyers' new mascot, Gritty; Evans, 2019). What is more, the findings about San Forgetica's effects have enjoyed much media coverage, from The Guardian to the Smithsonian magazine to a BBC podcast (Curcic, 2019; Martin, 2018; Wu, 2018). Sans Forgetica was even the topic of a U.S. National Public Radio (NPR) interview, during which another of the researchers said that Sans Forgetica works because our automatic tendency to complete the broken font "slows down the process of reading inside your brain. And then it can actually trigger memory" (Simon, 2018). The problem with all this attention? The claims about Sans Forgetica's effects on memory have never been published in a peer-reviewed journal.

Considered together, this collection of factors – the high stakes for educators, the seemingly small effect, the conflict with prior work and open theoretical question, the apparent "one-size-fits-all" benefit for memory, the accolades and media coverage, and the lack of peer-review – mean it is important to replicate the Sans Forgetica team's findings. After all, if Sans Forgetica promotes learning, then there is, indeed, a font that produces a level of disfluency that is beneficial. Furthermore, these findings would demonstrate that it is possible to improve learning with a straightforward manipulation and without making the information more difficult to learn. In other words, this font would have the "right" level of perceptual

fluency to prompt the deeper, more beneficial encoding processes engaged by other desirable difficulties. But if Sans Forgetica has no benefit or even hurts learning, then such a finding would suggest, at the very least, that the media has failed in its duty as society's "reality monitors" (Johnson, 2007).

Accordingly, we set out to answer this question: to what extent does Sans Forgetica boost people's memory for information? To answer this question, we first gathered evidence that Sans Forgetica creates a feeling of disfluency in readers. We then carried out three experiments examining how well people remember information in Sans Forgetica, the first two of which we did not preregister, but the third we did. The numeric data for all experiments, and the corresponding R code, are available on the Open Science Framework: https://osf.io/b6wd9/.

Experiment 1

The purpose of this experiment was to establish the extent to which material written in Sans Forgetica feels disfluent.

Method

Subjects
A total of 151 undergraduate psychology students at the University of Waikato participated in partial fulfilment of course requirements (M_{age} = 21.14, SD_{age} = 6.53; 28% identified as men, 72% as women, and <1% as gender diverse). Of these subjects, 87% reported English as their first language.

Design
We manipulated font (Sans Forgetica, Arial) between subjects.

Procedure
All experiments reported here were approved by the University of Waikato's School of Psychology Research and Ethics Committee under delegated authority of the University's Human Research Ethics Committee. The experiment consisted of two phases, administered online in Qualtrics (https://www.qualtrics.com/). Subjects were told they would "read passages about a topic and then be asked some questions about those passages".

Phase one. We randomly presented subjects with one of two prose passages designed to mimic educational materials (Butler et al., 2007) – one about the creation of basketball and one about the artist Georgia O'Keefe. For some subjects, the passage was presented in Sans Forgetica, and for others the passage was presented in Arial. The passages were both 298 words long, and split into four paragraphs, each of which was presented on a separate page. Subjects were presented with the first paragraph and were instructed to progress to the next page once they had finished reading the paragraph on the screen.

Phase two. Next, to measure how difficult subjects found it to read the passages, we asked them "How difficult was it to read the font the passages were presented in?" (1 = *Very easy*, 7 = *Very difficult*).[1]

Results

According to the creators of Sans Forgetica, the font improves memory performance via its disfluency, which promotes deeper processing. Therefore, before testing the effects of Sans Forgetica on memory performance, we first sought to establish that people find Sans Forgetica more difficult to read than a commonly-used font (Arial). As expected, subjects rated Sans Forgetica as more difficult to read (M = 3.04, SD = 1.74) than Arial (M = 2.32, SD = 1.51; M_{diff} = 0.71, 95% CI [0.19, 1.24]; for each experiment we also conducted Null Hypothesis Significance Test analyses, which appear in the Supplemental Materials and lead to the same conclusions). This finding suggests that material written in Sans Forgetica feels more disfluent than its more ordinary counterpart, Arial.

The next two experiments we present comprised an attempt to replicate the experiments as reported by the Sans Forgetica team in developing the font. Using a published interview with Teacher Magazine and information provided by the Sans Forgetica team, we were able to reconstruct most of the methods originally used, including information about the type of materials, exposure time, and test format (J. Blijlevens, personal communication, 12 December, 2018; Earp, 2018). The exact length of the delay between exposure and test was never specified, so we tested a range of delays in Experiment 2, and in Experiment 3 used the delay that Butler et al. (2007) used in their learning experiments. The Sans Forgetica team used Albion as the comparison font in one experiment and Arial in the other, but we used Arial in both.

Experiment 2

Method

Subjects
We recruited subjects on Amazon's Mechanical Turk (MTurk) using the TurkPrime platform (Litman et al., 2017). A total of 156 subjects completed the experiment in return for U.S.$0.30 ($M_{age}$ = 44.53, SD_{age} = 14.00; 37% identified as men, 73% as women, and 0% as gender diverse; 97% reported English as their first language).

Design
We used a mixed design with font (Sans Forgetica, Arial) as a within-subject factor and delay (10s, 20s, 30s) as a between-subjects factor.

Procedure
The experiment consisted of two phases, administered online in Qualtrics.

Phase one. As in the Sans Forgetica team's lab-based experiment, subjects saw a series of 20 word pairs, each of which was presented on the computer screen for 100 ms. These pairs were highly associated words (for example, accident – crash, chip – potato) taken from the University of Florida word association norms (Nelson et al., 2004).[2] Half of the word pairs were presented in Sans Forgetica and half in Arial. We created four versions of the word pair list, varying the font and order of pairs across these counterbalances to ensure: [1] pairs were presented equally often in each font across subjects; and [2] there were no large clusters of pairs all presented in the same font. All counterbalances contained the same word pairs. Prior to viewing the word pairs, subjects were shown a black fixation cross in the centre of the screen for two seconds. There was an inter-stimulus interval of 1.5 s between word pairs, in which subjects viewed a visual mask.

Phase two. Because we did not have access to the length of delay between exposure and test used in the original experiment, we randomly assigned subjects to one of three short delays (10 s, 20 s, or 30 s). During this delay, subjects were presented with a series of multiplication problems and asked to solve each one. After this short delay, we tested subjects' memory for the word pairs by presenting the first word of the pair and asking them to type the second word of that pair into a box. We presented the memory test in Times New Roman font, so as not to reinstate the context of one particular font. The cuing words were presented one-by-one, in a random order.

Results

Recall our main research question was: to what extent does Sans Forgetica boost people's memory for information? To answer this question, we calculated the percentage of word pairs subjects correctly recalled that they had seen

in Sans Forgetica, and the percentage of word pairs subjects correctly recalled that they had seen in Arial. We display these data in Panel A of Figure 2. As the figure shows, subjects recalled fewer word pairs that they had seen in Sans Forgetica ($M = 40.26\%$, $SD = 24.07$) than those that they had seen in Arial ($M = 50.51\%$, $SD = 24.54$; $M_{diff} = 10.26\%$, 95% CI [6.36, 14.15]). Furthermore, this pattern held true regardless of whether the delay was 10 seconds ($M_{diff} = 6.36\%$ [−0.74, 13.47]), 20 seconds ($M_{diff} = 11.06\%$ [3.92, 18.21]), or 30 seconds ($M_{diff} = 13.52\%$ [7.24, 19.80]). Contrary to the claim that Sans Forgetica boosts memory, these results suggest that presenting word pairs in Sans Forgetica actually impairs people's ability to recall them.

It is possible, however, that this finding is due to the fact that subjects were presented with each of the word pairs for only a short time (100 ms). As a result, subjects might have struggled to read the word pairs properly, or to process the word pairs deeply enough, for any benefits of Sans Forgetica to take effect. Furthermore, although these findings suggest Sans Forgetica might impair people's ability to recall word pairs, students often want to remember information other than word pairs, such as passages of prose. The results of this experiment cannot speak to the extent to which Sans Forgetica benefits people's memories for prose passages. Therefore, we address this issue in Experiment 3.

Experiment 3

Method

Subjects

A total of 300 subjects took part in the experiment ($M_{age} = 31.04$, $SD_{age} = 13.94$; 30% identified as men, 69% as women, and 2% as gender diverse; 91% reported English as their first language). Of these subjects, 155 were
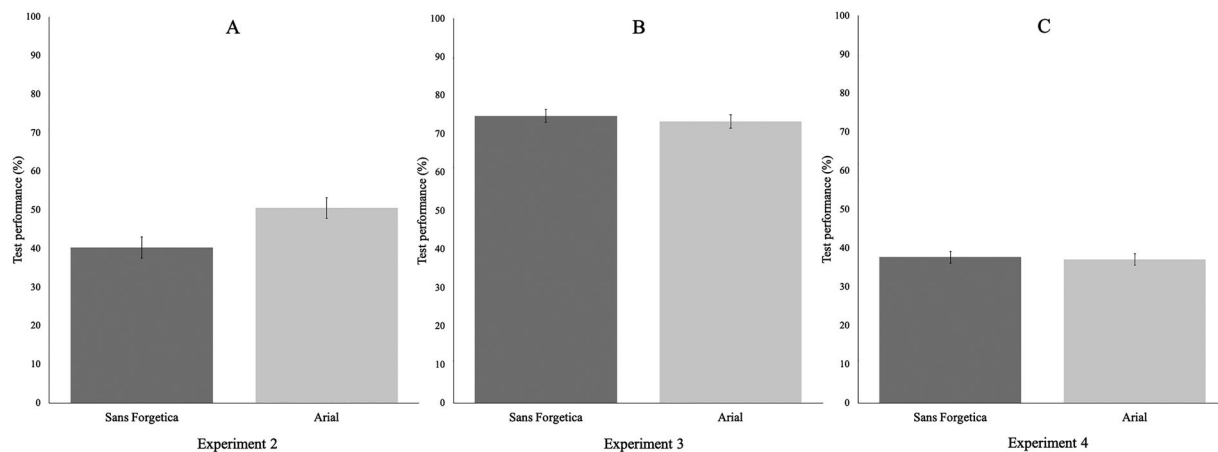
Figure 2. Panel A: Percentage of word pairs correctly recalled when presented in Sans Forgetica and Arial in Experiment 2. Panel B: Percentage of multichoice questions answered correctly about information presented in Sans Forgetica and Arial in Experiment 3. Panel C: Percentage scores on cued-recall questions about information presented in Sans Forgetica and Arial in Experiment 4. Error bars represent 95% Cousineau-Morey within-subject confidence intervals.

MTurk workers living in the United States and Canada, who received U.S.$0.50 on completion of the experiment (M_{age} = 40.10, SD_{age} = 12.79). The remaining 145 subjects were first year psychology students at the University of Waikato who took place in partial fulfilment of course requirements (M_{age} = 21.35, SD_{age} = 6.76).

Design
We manipulated font (Sans Forgetica, Arial) within subjects.

Procedure
The experiment consisted of two phases and was again administered online in Qualtrics.

Phase one. First, we asked subjects to read, at their own pace, several prose passages containing some information in Sans Forgetica and some in Arial. More specifically, subjects read, in a random order, five prose passages about different topics, all designed to mimic educational materials (Butler et al., 2007). Each of these passages was between 297 and 299 words long, and split into four paragraphs. Subjects were instructed to continue to the next page once they had finished reading the paragraph on the screen. In total, each subject read 20 separate paragraphs of information – five of these (one per passage) were randomly selected to be presented in Sans Forgetica and the remaining 15 were presented in Arial.

Phase two. After subjects had played a card-matching game for five minutes, we tested their memory for the information presented in the prose passages. We asked subjects to answer 20 multi-choice questions, designed to test their knowledge for facts contained in the 20 paragraphs they had read (Butler et al., 2007). This test contained one question about each paragraph, meaning subjects answered five questions about information they read in Sans Forgetica and 15 questions about information they read in Arial. As for Experiment 2, these questions were presented in Times New Roman font, one-by-one, in a random order.

Results

To what extent did Sans Forgetica boost people's memory for information presented in prose form? To answer this question, we calculated the percentage of questions subjects correctly answered about paragraphs they read in Sans Forgetica and the percentage of questions subjects correctly answered about paragraphs they read in Arial, and then compared these percentages. We display these data in Panel B of Figure 2.

As the figure shows, performance on questions about information read in Sans Forgetica was only trivially different from performance on questions about information read in Arial ($M_{Sans\ Forgetica}$ = 74.73%, SD = 23.37; M_{Arial} = 73.24%, SD = 17.30; M_{diff} = 1.49%, 95% CI [−0.94, 3.92]). Using an equivalence-testing approach (Lakens, 2017), we tested the hypothesis that subjects' performance was equivalent across both font conditions. We set the

smallest effect size of interest as a difference of five percentage points (the equivalent of one letter grade), and found the two fonts produced equivalent performance (p = .002). Consistent with the findings of Experiment 2, these results provide evidence against the idea that reading information in Sans Forgetica boosts recall of that information.

But if it were true that Sans Forgetica creates desirable difficulty, which promotes deeper processing that leads people to better conceptually integrate information, we might expect this font would produce the greatest benefit for questions that require such integration of information. Therefore, in Experiment 4 we tested the effect of presenting information in Sans Forgetica on people's ability to answer higher-level conceptual questions about that information.

Experiment 4

Method

We preregistered this experiment and posted our preregistration here: https://osf.io/b6wd9/

Subjects
We set our target sample size at 271 subjects, allowing us 90% power to detect a difference of d = 0.2 with an equivalence test for dependent means (Lakens, 2017). We recruited 319 MTurk workers living in the United States and Canada, who received U.S.$0.75 upon completion of the experiment. After exclusions (see Results) we retained 275 subjects for analysis (M_{age} = 41.58, SD_{age} = 12.28; 36% identified as men, 64% as women, and 0% as gender diverse; 94% reported English as their first language).

Design
We manipulated font (Sans Forgetica, Arial) within subjects.

Procedure
The procedure for Experiment 4 was identical to that of Experiment 3, with the following two exceptions. First, for this experiment, subjects read six prose passages about different topics, designed to mimic educational materials (Butler, 2010). Each of these passages was between 502 and 516 words long, and split into four paragraphs. For each passage, there were two paragraphs that contained concepts subjects would later be asked questions about. Each subject was randomly assigned to see one of these paragraphs presented in Sans Forgetica and the other presented in Arial. Subjects were also randomly assigned to see one of the remaining two paragraphs for each passage in Sans Forgetica and the other in Arial. In total, each subject read 24 paragraphs of prose – 12 presented in Sans Forgetica, and 12 in Arial.

Second, we asked subjects to answer 12 cued-recall questions designed to test their understanding for concepts contained in the passages they read (Butler, 2010). These conceptual questions required subjects to integrate

information from different sentences within a paragraph (Bloom, 1956). For each passage, subjects were asked one question about a concept contained in a Sans Forgetica paragraph and one question about a concept contained in an Arial paragraph. Therefore, subjects answered six conceptual questions about information they read in Sans Forgetica and six conceptual questions about information they read in Arial. Subjects typed their answers into open text-entry boxes.

Results

To address our primary research question, we first scored subjects' performance on each question. We used a coding scheme that identified the important pieces of information that subjects' answers needed to contain to be considered correct (similar to Butler, 2010) and scored responses as either incorrect (0 points), partially correct (1 point), or correct (2 points). For example, subjects' responses to the question "Gas exchange occurs in a part of the human respiratory system called the alveoli. How does the process of gas exchange work?" were awarded 1 point for mentioning diffusion, and 1 point for mentioning that oxygen moves in one direction with carbon-dioxide moving in the other. A research assistant coded subjects' answers, and one of the authors (AT) coded 20% of the answers to examine inter-rater reliability. The two coders agreed on 86% of answers, and all disagreements were resolved through discussion.

To what extent did Sans Forgetica boost performance on conceptual questions? We calculated (as a percentage) the number of points subjects gained on the memory test for questions that related to paragraphs they read in Sans Forgetica. We then compared this percentage to subjects' percentage score for questions that related to paragraphs they read in Arial. We display these data in Panel C of Figure 2.

As the figure shows, we again found no evidence that Sans Forgetica provides any performance boost. Subjects' performance on conceptual questions about information presented in Sans Forgetica was almost identical to their performance on conceptual questions about information presented in Arial ($M_{Sans\ Forgetica} = 37.64\%$, $SD = 20.94$; $M_{Arial} = 37.06\%$, $SD = 19.93$; $M_{diff} = 0.58\%$, 95% CI [−1.53, 2.69]). Using an equivalence-testing approach with the smallest effect size of interest set at five percentage points, we again found that the two fonts produced equivalent performance ($p < .001$). Once again, these results suggest Sans Forgetica does not meaningfully benefit memory performance.

General discussion

Across four experiments with 882 people, we set out to answer the question: to what extent does Sans Forgetica boost people's memory for information? In Experiment 1, we showed Sans Forgetica feels less fluent than its more

ordinary counterpart, Arial. In Experiment 2, when we showed subjects word pairs in Sans Forgetica or Arial, they recalled fewer of the "Sans Forgetica" pairs than the Arial ones – a finding in line with the idea that Sans Forgetica did not improve their memory, but instead hurt it. In Experiment 3, we showed subjects prose passages developed to mimic educational materials and within each passage presented some information in Sans Forgetica and some in Arial. When we tested people's memory for facts in the passages, we found no evidence that Sans Forgetica improved their performance. This pattern continued in Experiment 4, when we instead tested people's understanding of concepts presented in either Sans Forgetica or Arial. Taken together, our findings converge on a failure to replicate, and furthermore suggest that although Sans Forgetica promotes a feeling of disfluency, it does not create desirable difficulty.

These findings have theoretical implications. For one, they add to our understanding of the role of perceptual disfluency in fostering desirable difficulties. We have long known that interventions that encourage deeper, slower, more effortful processing can help people remember information (Bjork, 1994). But our findings fit with the idea that disfluent fonts do not encourage this kind of processing (see also, Xie et al., 2018). As to the question of whether Sans Forgetica produces the "optimal" level of disfluency – our findings suggest it does not. In fact, we found no evidence Sans Forgetica reliably leads to better recall. Although Sans Forgetica is novel and hard to read, its effects might well end there.

Of course, it is possible that Sans Forgetica has beneficial effects on learning that we were not able to detect here. For instance, it is possible that there are individual differences in the extent to which Sans Forgetica boosts people's memory for information (see preliminary evidence from Eskenazi & Nix, 2020). Perhaps Sans Forgetica is only beneficial to those people with the cognitive capacity to accommodate the challenge of reading the font at the same time as comprehending and encoding the information presented (McDaniel & Butler, 2010). Future research should address this issue.

In addition, we cannot rule out that benefits of Sans Forgetica might emerge after a longer interval between study and test, as often happens with desirable difficulties. Because disfluent fonts are thought to promote deeper processing resulting in better integration of information into memory, benefits of disfluent fonts are most likely to be detected in situations that test how strongly information is stored in memory, rather than simply how accessible information is during an immediate test (for a review, see Bjork & Bjork, 2011). But this issue of delay was examined in the recent meta-analysis, which found that the effects of disfluent fonts on recall were trivial, regardless of whether studies had delays of more or less than 10 minutes (Xie et al., 2018). This finding is relatively uninformative about the effects after more ecologically-valid delays. After all, one study suggests that the benefits of

disfluent fonts on memory are particularly pronounced after a delay of two weeks (Weissgerber & Reinhard, 2017). Given students typically have to retain information for longer than a few minutes, this issue is one that could be examined in future work.

Considered more broadly, our findings can also be understood in light of the media's role as our institutional "reality monitors" (Johnson, 2007). When the media fails to help readers evaluate what is true or even what is plausible, it abdicates its responsibility. The problem becomes especially thorny when the typical consumer cannot easily remedy the media's failure. The typical reader of The Guardian or NPR listener could not reasonably be expected to evaluate the Sans Forgetica coverage as we have here, let alone carry out attempts to replicate the basic claims about Sans Forgetica's effects. Therefore, just as peer-review can be understood as the essential mechanism by which the scientific community does its own reality monitoring, work such as ours can be understood as the essential mechanism by which the scientific community redresses the media's failures to do reality monitoring.

In addition to these theoretical implications, our work also has practical implications for all learners. Our findings suggest that disfluent fonts are ineffective at improving people's memory. Furthermore, adoption of these fonts could have more serious ramifications. After all, if students put their study materials into Sans Forgetica in the mistaken belief that the feeling of difficulty created is benefitting them, they might forego other, effective study techniques. Instead, we should encourage learners to rely on the robust, theoretically-grounded techniques developed by scientists working at the forefront of cognitive psychology and education, that really do enhance learning.

Notes

1. As exploratory measures, we also asked subjects three questions to measure the extent to which they believed they would retain the information and would be willing to use Sans Forgetica to study. Because these findings are not central to our research question, we report these data in the Supplemental Materials.
2. We thank Andrew Butler for providing us with his composed list of highly associated word pairs.

Acknowledgements

Disclosure statement

ORCID

Mevagh Sanson http://orcid.org/0000-0003-4520-224X

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13, 219–235. https://doi.org/10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth Publishers.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Essex, England: Harlow.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, And Cognition*, 36, 1118–1133. https://doi.org/10.1037/a0019902

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13, 273–281. https://doi.org/10.1037/1076-898x.13.4.273

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Curcic, S. (2019). *BBC World Service – Digital Planet, Sans Forgetica – a font to remember*. Retrieved 11 December 2019, from https://www.bbc.co.uk/programmes/p06y5n60.

deWinstanley, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin & Review*, 2, 538–541. https://doi.org/10.3758/BF03210990

Diemand-Yauman, C., Oppenheimer, D., & Vaughan, E. (2011). Fortune favors the Bold (and the Italicized): effects of disfluency on educational outcomes. *Cognition*, 118, 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Earp, J. (2018). *Q&A: Designing a font to help students remember key information*. Retrieved 12 November 2019, from https://www.teachermagazine.com.au/articles/qa-designing-a-font-to-help-students-remember-key-information.

Eitel, A., & Kühl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, 11, 107–121. https://doi.org/10.1007/s11409-015-9145-3

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder-to-read mean better-to-understand? *Applied Cognitive Psychology*, 28, 488–501. https://doi.org/10.1002/acp.3004

Eskenazi, M. A., & Nix, B. (2020). Individual differences in the desirable difficulty effect during Lexical Acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, https://doi.org/10.1037/xlm0000809. Advance Online Publication.

Evans, C. (2019). Name of the year 2018. *Names*, *67*, 113–117. https://doi.org/10.1080/00277738.2019.1589678

French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., Poland, R. M., Reece, M., & Tabor, J. (2013). Changing fonts in education: How the benefits vary with ability and dyslexia. *The Journal of Educational Research*, *106*, 301–304. https://doi.org/10.1080/00220671.2012.736430

Good Design. (2019). *Sans Forgetica – good design*. Retrieved 12 November 2019, from https://good-design.org/projects/sans-forgetica/.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior*, *11*, 534–537. https://doi.org/10.1016/S0022-5371(72)80036-7

Johnson, M. K. (2007). Reality monitoring and the media. *Applied Cognitive Psychology*, *21*, 981–993. https://doi.org/10.1002/acp.1393

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*, 355–362. https://doi.org/10.1177/1948550617697177

Lindsay, D. S. (2008). Source monitoring. In J. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory* (pp. 325-348). Oxford: Elsevier.

Litman, L., Robinson, J., & Abberbock, T. (2017). Turkprime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*, 433–442. https://doi.org/10.3758/s13428-016-0727-z

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671–685. https://doi.org/10.1037/a0018785.

Martin, L. (2018). *Font of all knowledge? Researchers develop typeface they say can boost memory*. Retrieved 12 November 2019, from https://www.theguardian.com/artanddesign/2018/oct/04/font-of-all-knowledge-researchers-develop-typeface-they-say-can-boost-memory.

McDaniel, M. A., & Butler, A. C. (2010). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). New York, NY: Psychology Press.

Morehead, K., Dunlosky, J., & Rawson, K. A. (2019). How much mightier is the pen than the keyboard for note-taking? A replication and extension of Mueller and Oppenheimer (2014). *Educational Psychology Review*, *31*, 753–780. https://doi.org/10.1007/s10648-019-09468-2.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*, *25*, 1159–1168. https://doi.org/10.1177/0956797614524581

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402–407. https://doi.org/10.3758/bf03195588

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*, 237–241. https://doi.org/10.1016/j.tics.2008.02.014

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625. https://doi.org/10.1037/a0013684

Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science*, *45*, 221–238. https://doi.org/10.1007/s11251-016-9387-8

Simon, S. (2018). *Sans Forgetica: a font to remember*. Retrieved 12 November 2019, from https://www.npr.org/2018/10/06/655121384/sans-forgetica-a-font-to-remember.

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

Weissgerber, S. C., & Reinhard, M. (2017). Is disfluency desirable for learning? *Learning and Instruction*, *49*, 199–217. https://doi.org/10.1016/j.learninstruc.2017.02.004

Wu, K. J. (2018). *Sans Forgetica is the typeface you won't forget*. Retrieved 12 November 2019, from https://www.smithsonianmag.com/smart-news/sans-forgetica-trippy-new-typeface-could-help-readers-remember-what-they-read-180970485/.

Xie, H., Zhou, Z., & Liu, Q. (2018). Null effects of perceptual disfluency on learning outcomes in a text-based educational context: A meta-analysis. *Educational Psychology Review*, *30*, 745–771. https://doi.org/10.1007/s10648-018-9442-x

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*, 229–241. https://doi.org/10.3758/s13421-012-0255-8