

1

Sans Forgetica is Really Forgettable

2

Jason Geller¹, Sara D. Davis², & Daniel Peterson²

3

¹ University of Iowa

4

² Skidmore College

5

Abstract

6

Keywords: fluency

Word count: X

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

Students want to remember more and forget less. Decades of research have put forth the paradoxical idea that making learning harder (not easier) should have the desirable effect of improving long-term retention of material—called the desirable difficulty principle (Bjork, 1994). Notable examples of desirable difficulties include having participants generate information from word fragments instead of passively reading intact words (e.g., Slamecka & Graf, 1978), spacing out study sessions instead of massing them (e.g., Carpenter, 2017), and having participants engage in retrieval practice after studying instead of simply restudying the information (Kornell & Vaughn, 2016). Another simple strategy that has gained some attention is to make material more perceptually disfluent. This can be done by changing the material’s perceptual characteristics (Diemand-Yauman, Oppenheimer, & Vaughan, 2011; French et al., 2013). Visual material that is masked (Mulligan, 1996), inverted (Sungkhasette, Friedman, & Castel, 2011), presented in an atypical font (Diemand-Yauman et al., 2011), blurred (Rosner, Davis, & Milliken, 2015), or even in handwritten cursive (Geller, Still, Dark, Carpenter, 2018) have all been shown to produce memory benefits. The desirable effect of perceptual disfluency on memory is called the disfluency effect (Bjork, 2016)

23

24

25

26

27

Although appealing as a pedagogical strategy, there have been several experiments that failed to find memorial benefits for perceptually disfluent materials (e.g., Magreehan, Serra, Schwartz & Narciss, 2016; Rhodes & Castel, 2008, 2009; Rummer, Scheweppe, & Schewede, 2016; Yue, Castel, & Bjork, 2013), casting doubt upon the veracity of the disfluency effect. A recent meta-analysis (), Recent studies by Geller et al.(2018) and Geller

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Jason Geller, Postal address. E-mail: jason-geller@uiowa.edu

28 & Still (2018) found that perceptual disfluency can have a beneficial effect on memory, but
 29 seems to be rather fickle, thus delimiting its educational usefulness.

30 Given the weak evidence, it came as a surprise to me when a little over a year ago, a
 31 font by the name of Sans Forgetica (SF) started getting a ton of press coverage. The mnn-
 32 menomic benefits of this font, *based on cognitive psychology*, were being touted in reputable
 33 news sources like Washington Post ([https://www.washingtonpost.com/business/2018/](https://www.washingtonpost.com/business/2018/10/05/introducing-sans-forgetica-font-designed-boost-your-memory/)
 34 [10/05/introducing-sans-forgetica-font-designed-boost-your-memory/](https://www.washingtonpost.com/business/2018/10/05/introducing-sans-forgetica-font-designed-boost-your-memory/)) and NPR ([https://](https://www.npr.org/2018/10/06/655121384/sans-forgetica-a-font-to-remember)
 35 www.npr.org/2018/10/06/655121384/sans-forgetica-a-font-to-remember, amongst others.
 36 The creators even made the SF font available for mac and pc operating systems—all you
 37 have to do is downlaod the font file and you to can remember everything you read :). There
 38 is even a Chrome browser extension and cellphone application that allows users to place
 39 material in Sans Forgetica. With this much attention and marketing, there has to be solid
 40 empirical evidence backing it up, right? Not quite.

41 Despite the weak evidence for perceptual disfluency, it came as a surprise
 42 to me when a little over a year ago, I saw a font by the name of Sans For-
 43 getica (SF) getting a ton of press coverage. The mnnmenomic benefits of this
 44 font, *based on cognitive psychology*, were being touted in reputable news sources like
 45 Washington Post ([https://www.washingtonpost.com/business/2018/10/05/introducing-](https://www.washingtonpost.com/business/2018/10/05/introducing-sans-forgetica-font-designed-boost-your-memory/)
 46 [sans-forgetica-font-designed-boost-your-memory/](https://www.washingtonpost.com/business/2018/10/05/introducing-sans-forgetica-font-designed-boost-your-memory/)) and NPR ([https://www.npr.org/2018/](https://www.npr.org/2018/10/06/655121384/sans-forgetica-a-font-to-remember)
 47 [10/06/655121384/sans-forgetica-a-font-to-remember](https://www.npr.org/2018/10/06/655121384/sans-forgetica-a-font-to-remember), amongst others. The creators even
 48 made the SF font available for mac and pc operating systems—all you have to do is down-
 49 laod the font file and you to can remember everything you read :). There is even a Chrome
 50 browser extension and cellphone application that allows users to place material in Sans For-
 51 getica. With this much attention and marketing, there has to be solid empirical evidence
 52 backing it up, right? Not quite.

53 What do we know about SF?

54 There is not information about SF. The typyface itself is a variation of a sans-serif
 55 typeface. It is a typeface that consists of intermitten gaps in letters that are back slanted
 56 (see below picture). The design features of this typeface require readers of it to “fill-in”
 57 the missing pieces like a puzzle. As it pertains to the empirical validation of the claims
 58 made, the website does offer some information about SF and how the original results were
 59 obtained, but not enough information to replicate the studies.

60 Earp (2018) conducted an interview with the creators of SF and I was able to glean
 61 some details about how SF ws validated. Apparently two studies were conducted. In a
 62 lab experiment ($N=96$), they had participants read 20 word pairs (e.g., girl - guy; called a
 63 paried associates task in cognitive parlance) in three new fonts (one of them being SF) and a
 64 typical or common font. The font pairs were presented in was counterbalanced participants.
 65 What this means is that all fonts were showns, but the same pairs were never presneted in
 66 more than one type of font. Each word pair was presnted on the screen for 100 ms (that is
 67 super fast. . .). For a final test, they were given the cue (e.g., *girl*) and had to respond with
 68 the target (*guy*). What did they find? According to the interview, targets were recalled

69 68% of time when presented in a common font. For cue-target pairs in SF, targets were
70 recalled 69% of the time—a negligible difference.

71 In an online experiment, participants were presented passages (250 words in total)
72 where one of the paragraphs was presented in SF. Each participant saw five different texts in
73 total. For each text they were asked one question about the part written in SF and another
74 question about the part written in standard Arial. Participants remembered 57% of the
75 text when a section was written in Sans Forgetica, compared to 50% of the surrounding
76 text that was written in a plain Arial font.

77 At the time of this writing, these studies have not been published nor is there a
78 preprint available. I reached out to the creators of SF, but they refused to share the
79 materials with me. Instead of waiting, I elicited the help of Sara Davis and Daniel Peterson
80 at Skidmore university to test the mnemonic benefits of Sans Forgetica.

81 Experiment 1

82 In the first study we compared the mnemonic benefits of SF against a robust tech-
83 nique known to enhance memory—generation. The generation effect is a phenomenon where
84 information is better remembered when retrieved than if it is simply read. In a prototypical
85 experiment, participants are asked to generate words from word fragments DOLL - DR___
86 or read intact cue-target pairs (*DOLL-DRESS*). Compared to the intact condition, indi-
87 viduals recall the generated target words at a higher rate. The nature of generation is
88 where the supposed mnemonic benefit of SF comes from. We examined this in the current
89 experiment.

90 Participants

91 We recruited 230 people from Amazon’s Mechanical Turk Service. Sample size was
92 calculated based on the smallest effect of interest (SEOI; Lakens & Evers, 2014). In this
93 case, we were interested in powering our study to detect a medium-sized effect size ($d =$
94 $.35$). We choose this effect size as our SESOI due in part to the small effect sizes seen in
95 actual classroom studies (Bulter et al., 2014). Therefore, assuming an alpha of .05 and a
96 desired power of 90%, a sample size of 270 is required to detect whether an effect size of
97 $.35$ differs from zero. After excluding participants After excluding participants who 1) did
98 not complete every phase of the experiment, 2) started the experiment multiple times, 3)
99 reported experiencing technical problems did not indicate that they were fluent in English
100 [^2]: This question was not asked during the experiment., or 5) reported seeing our stimuli
101 before we were left with 115 participants per group.

102 Materials and Procedure

103 We preregistered two studies (<https://osf.io/d2vy8/>). All materials, data, and anal-
104 ysis scripts can be found at that website. Analyses are computationally reproducible on
105 our github by going here: (https://github.com/jgeller112/SF_Expt1; [https://github.com/](https://github.com/jgeller112/SF_Expt2)
106 [jgeller112/SF_Expt2](https://github.com/jgeller112/SF_Expt2)).

Participants were presented with 22 weakly related cue-target pairs taken from Carpenter et al., 2012^[1]: Two cue-target pairs () had to be thrown out as they were not preseted due to a coding error. The cue-target pairs were all nouns, 5–7 letters and 1–3 syllables in length, and high in concreteness (400–700) and frequency (at least 30 per million).

For half the participants, targets were presented in SF while the other half were presented in Arial font; for the other half of participants, targets were presented with missing letters (vowels were replaced by underscores) and the other half were intact (Arial font). After a short 2 minute distractor task (anagram generation), they completed a cued recall test. During cued recall, participants were presented 24 cues one at a time and asked to provide the target word. After they were thanked and debriefed.

In Experiment 1 participants were presented with 40 weakly related cue-target pairs. The pairs were nouns, 5–7 letters, 1–3 syllables in length, and high in concreteness (400–700) and frequency (at least 30 per million). For half of the participants, half of the targets were presented in SF while the other half were presented in Arial font; for the other half of participants, the targets were presented with missing letters (vowels were replaced by underscores) and the other half were intact (Arial font). After a short 2 minute distractor task (anagram generation), they completed a cued recall test.

Spell checking was automated with the hunspell package in R (Ooms, 2018) using spellCheck.R. At the next step we manually examined the output to catch incorrect suggestions and to add their own corrections. Becasuse participants were recruited in the United States, we used the American English dictionary. A nice walkthrough on how to use the package can be found in Buchcamam, De Deyne, and Montefinese (2019). Using the package, each response was corrected for misspelings. Corrected spellings are provided in the most probable order, therefore, the first suggestion is selected as the correct answer. Answers were marked correct if they provided the exact response. In order for a response to be judged correctly, the response had to match the correct answer.

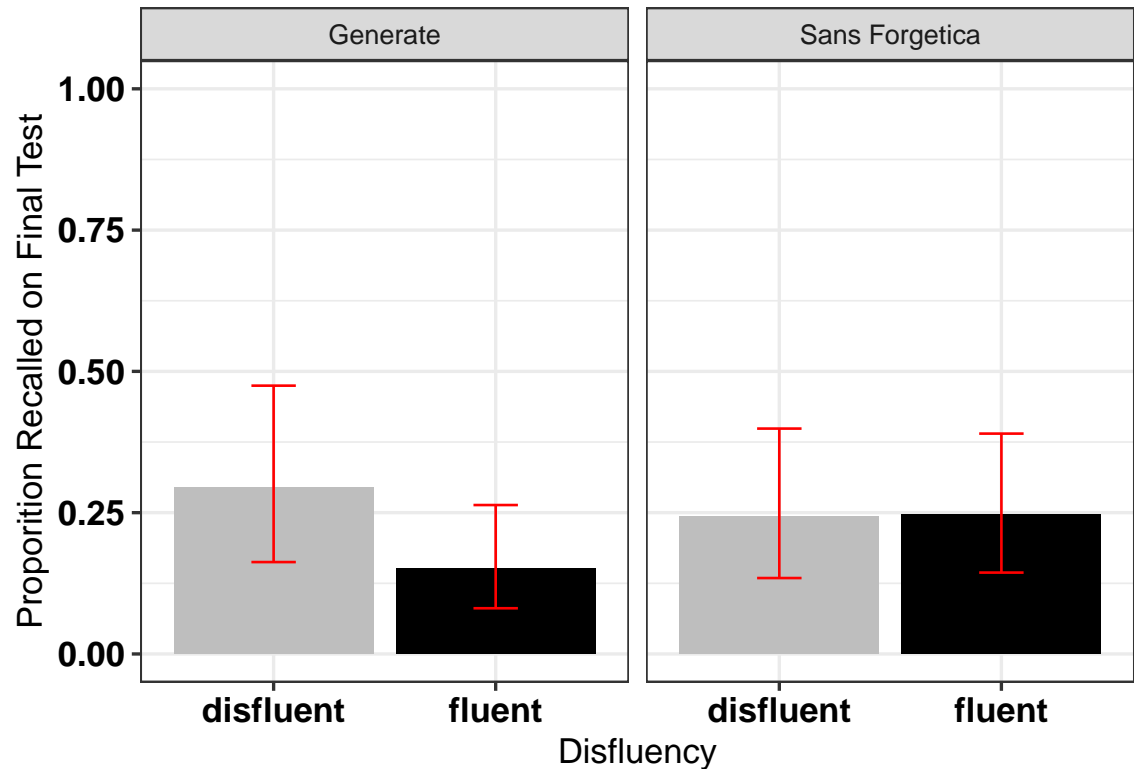
What did we find?

In Experiment 1 we found a sizeable generation effect, which replicated past work. However, we did not find a SF effect (See figure below)

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## condition*dis effect
##              dis
## condition      disfluent    fluent
## Generate      0.2952348 0.1507106
## Sans Forgetica 0.2429061 0.2469149
##
## Lower 95 Percent Confidence Limits
##              dis
```

```
147 ## condition      disfluent      fluent
148 ##   Generate      0.1625725 0.08088777
149 ##   Sans Forgetica 0.1343397 0.14399034
150 ##
151 ##   Upper 95 Percent Confidence Limits
152 ##           dis
153 ## condition      disfluent      fluent
154 ##   Generate      0.4747772 0.2635241
155 ##   Sans Forgetica 0.3987912 0.3898998
```



156

157

Experiment 2

158 The procedure in Experiment 1 could be said to lack educational realism. It is possible that Experiment 2 examined the SF in a more educationally realistic scenario. We presented participants a passage on ground water where some of the material was either: pre-highlighted, presented in SF, or presented with no changes. This was a between-subjects manipulation. Specifically participants read a passage about ground water (856 words) from the U.S. Geological Survey website (Yue, Storm, Kornell, Bjork, 2014). Eleven critical phrases¹ each containing a different keyword, were selected from the passage (e.g., the term

¹originally we had 12 critical phrases but a pilot test after the pretest showed that one of the questions was repeated twice so we removed one of them and also added a manipulation check question to

recharge was the keyword in the phrase: Water seeping down from the land surface adds to the ground water and is called recharge water.) and were either presented in SF, highlighted, or unchanged. Then, 11 fill-in-the blank questions were created from these phrases by deleting the keyword and asking participants to provide it on the final test (e.g., Water seeping down from the land surface adds to the ground water and is called _____ water).

Experiment 2 examined the SF in a more educationally realistic scenario. We presented participants a passage on ground water where some of the material was either: pre-highlighted, presented in SF, or presented with no changes. This was a between-subjects manipulation. Specifically participants read a passage about ground water (856 words) from the U.S. Geological Survey website (Yue, Storm, Kornell, Bjork, 2014). Eleven critical phrases² each containing a different keyword, were selected from the passage (e.g., the term *recharge* was the keyword in the phrase: Water seeping down from the land surface adds to the ground water and is called recharge water.) and were either presented in SF, highlighted, or unchanged. Then, 11 fill-in-the blank questions were created from these phrases by deleting the keyword and asking participants to provide it on the final test (e.g., Water seeping down from the land surface adds to the ground water and is called _____ water).

Participants

Participants were 528 undergraduates who participated for partial completion of course credit. Sample size was calculated based on the smallest effect of interest (Lakens & Evers, 2014). In this case, we were interested in powering our study to detect a medium-sized effect size ($d = .35$). Therefore, assuming an alpha of .05 and a desired power of 90%, a sample size of 170 is required to detect whether an effect size of .35 differs from zero. After excluding participants based on our preregistered exclusion criteria, we were left with unequal group sizes. Because of this, we decided to run six more participants per group, giving us 176 participants in each of the three conditions.

```
## [1] "/Users/gellr/SF_Expt2"
```

```
##
```

```
## FL_149_D0 effect
```

```
## FL_149_D0
```

```
## Highlight    Normal    Passage
```

```
## 0.4779431 0.3847635 0.4001508
```

```
##
```

```
## Lower 95 Percent Confidence Limits
```

```
## FL_149_D0
```

sure participants were paying attention

²originally we had 12 critical phrases but a pilot test after the preregistration showed that one of the questions was repeated twice so we removed one of them and also added a manipulation check question to sure participants were paying attention

```

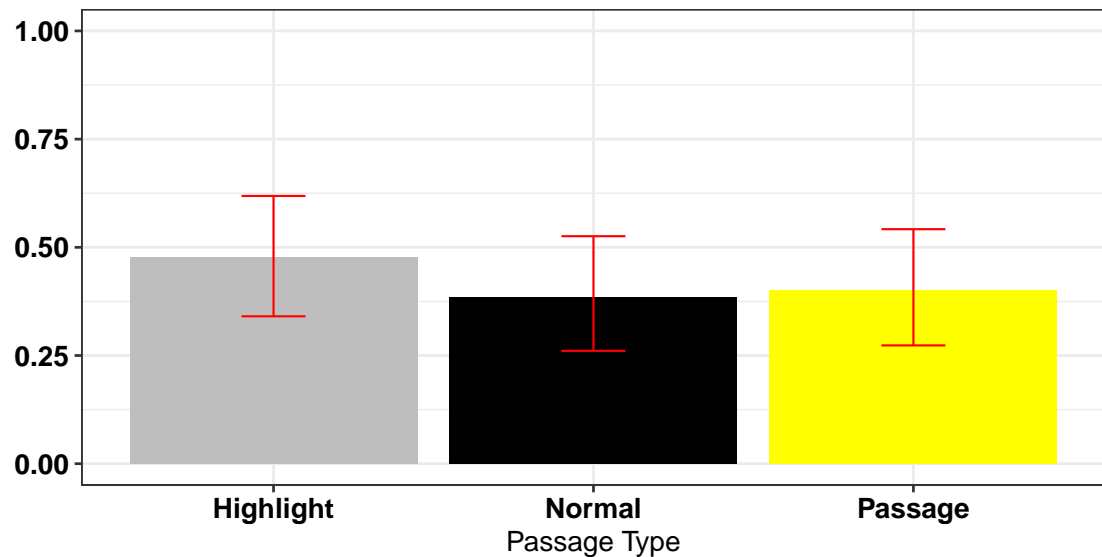
201 ## Highlight      Normal      Passage
202 ## 0.3405863 0.2608469 0.2733378
203 ##
204 ## Upper 95 Percent Confidence Limits
205 ## FL_149_D0
206 ## Highlight      Normal      Passage
207 ## 0.6187186 0.5256812 0.5419213

```

```

208 \begin{figure}

```



```

209 {

```

```

210 }

```

```

211 \caption{Accuracy on Cued Recall Test. Error bars are 95% CI dervied from the GLMER
212 model.} \end{figure}

```

213 Exploratory

214 In Experiment 2 we also asked students about their metacognitive awarness. Specifically
 215 we asked them: “How likely is it that you will be able to recall material from the passage
 216 you just read on a scale of 0 (not likely to recall) to 100 (likely to recall) in 5 minutes?”
 217 Initials analyses suggest that the normal passage was given higher JOLs ($M = 57.4$, $SE =$
 218 1.97) than the pre-highlighted passage ($M = 50.3$, $SE = 1.97$), $t(525) = -7.08$, $p = .023$.
 219 There were no reliable differences between the pre-highlighted passage and Sans Forgetica
 220 ($M = 53.8$, $SE = 1.97$), $t(525) = -3.52$, $p = .415$ or between the passage in Sans Forgetica
 221 and the passage presneted normally, $t(525) = 3.56$, $p = .406$.

222 One potential reason for pre-highlighted information recieving lower JOLs than the
 223 normal passage is that pre-highlighted information served to focus participants attention
 224 specific parts of the passage. Given the question, pariticipants might thought this would
 225 hinder them if tested over the passage as a whole. Future research should

| | contrast | estimate | SE | df | t.ratio | p.value |
|-----|---------------------|-----------|--------|-----|-----------|-----------|
| 226 | Highlight - Normal | -7.079546 | 2.7792 | 525 | -2.547332 | 0.0299152 |
| | Highlight - Passage | -3.517046 | 2.7792 | 525 | -1.265488 | 0.4153929 |
| | Normal - Passage | 3.562500 | 2.7792 | 525 | 1.281844 | 0.4060534 |

```

227 ## List of 14
228 ## $ text :List of 11
229 ## ..$ family : NULL
230 ## ..$ face : NULL
231 ## ..$ colour : NULL
232 ## ..$ size : num 10
233 ## ..$ hjust : NULL
234 ## ..$ vjust : NULL
235 ## ..$ angle : NULL
236 ## ..$ lineheight : NULL
237 ## ..$ margin : NULL
238 ## ..$ debug : NULL
239 ## ..$ inherit.blank: logi FALSE
240 ## ..- attr(*, "class")= chr [1:2] "element_text" "element"
241 ## $ axis.title.x :List of 11
242 ## ..$ family : NULL
243 ## ..$ face : NULL
244 ## ..$ colour : NULL
245 ## ..$ size : num 16
246 ## ..$ hjust : NULL
247 ## ..$ vjust : NULL
248 ## ..$ angle : NULL
249 ## ..$ lineheight : NULL
250 ## ..$ margin : NULL
251 ## ..$ debug : NULL
252 ## ..$ inherit.blank: logi FALSE
253 ## ..- attr(*, "class")= chr [1:2] "element_text" "element"
254 ## $ axis.title.y :List of 11
255 ## ..$ family : NULL
256 ## ..$ face : NULL
257 ## ..$ colour : NULL
258 ## ..$ size : num 16
259 ## ..$ hjust : NULL
260 ## ..$ vjust : NULL
261 ## ..$ angle : NULL
262 ## ..$ lineheight : NULL
263 ## ..$ margin : NULL
264 ## ..$ debug : NULL
265 ## ..$ inherit.blank: logi FALSE
266 ## ..- attr(*, "class")= chr [1:2] "element_text" "element"

```



```

267 ## $ axis.text      :List of 11
268 ##   ..$ family      : NULL
269 ##   ..$ face         : NULL
270 ##   ..$ colour       : NULL
271 ##   ..$ size         : num 14
272 ##   ..$ hjust        : NULL
273 ##   ..$ vjust        : NULL
274 ##   ..$ angle        : NULL
275 ##   ..$ lineheight   : NULL
276 ##   ..$ margin       : NULL
277 ##   ..$ debug        : NULL
278 ##   ..$ inherit.blank: logi FALSE
279 ##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
280 ## $ axis.text.x      :List of 11
281 ##   ..$ family      : NULL
282 ##   ..$ face         : NULL
283 ##   ..$ colour       : NULL
284 ##   ..$ size         : NULL
285 ##   ..$ hjust        : NULL
286 ##   ..$ vjust        : num 0.5
287 ##   ..$ angle        : num 45
288 ##   ..$ lineheight   : NULL
289 ##   ..$ margin       : NULL
290 ##   ..$ debug        : NULL
291 ##   ..$ inherit.blank: logi FALSE
292 ##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
293 ## $ axis.line.x      :List of 6
294 ##   ..$ colour       : chr "black"
295 ##   ..$ size         : num 0.5
296 ##   ..$ linetype     : chr "solid"
297 ##   ..$ lineend      : NULL
298 ##   ..$ arrow        : logi FALSE
299 ##   ..$ inherit.blank: logi FALSE
300 ##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
301 ## $ axis.line.y      :List of 6
302 ##   ..$ colour       : chr "black"
303 ##   ..$ size         : num 0.5
304 ##   ..$ linetype     : chr "solid"
305 ##   ..$ lineend      : NULL
306 ##   ..$ arrow        : logi FALSE
307 ##   ..$ inherit.blank: logi FALSE
308 ##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
309 ## $ legend.text      :List of 11
310 ##   ..$ family      : NULL
311 ##   ..$ face         : NULL

```

```

312 ## ..$ colour      : NULL
313 ## ..$ size        : num 16
314 ## ..$ hjust       : NULL
315 ## ..$ vjust       : NULL
316 ## ..$ angle       : NULL
317 ## ..$ lineheight  : NULL
318 ## ..$ margin      : NULL
319 ## ..$ debug       : NULL
320 ## ..$ inherit.blank: logi FALSE
321 ## ..- attr(*, "class")= chr [1:2] "element_text" "element"
322 ## $ legend.title   :List of 11
323 ## ..$ family      : NULL
324 ## ..$ face        : NULL
325 ## ..$ colour      : NULL
326 ## ..$ size        : num 16
327 ## ..$ hjust       : NULL
328 ## ..$ vjust       : NULL
329 ## ..$ angle       : NULL
330 ## ..$ lineheight  : NULL
331 ## ..$ margin      : NULL
332 ## ..$ debug       : NULL
333 ## ..$ inherit.blank: logi FALSE
334 ## ..- attr(*, "class")= chr [1:2] "element_text" "element"
335 ## $ legend.position : chr "right"
336 ## $ panel.border    : list()
337 ## ..- attr(*, "class")= chr [1:2] "element_blank" "element"
338 ## $ panel.grid.major: list()
339 ## ..- attr(*, "class")= chr [1:2] "element_blank" "element"
340 ## $ panel.grid.minor: list()
341 ## ..- attr(*, "class")= chr [1:2] "element_blank" "element"
342 ## $ plot.title      :List of 11
343 ## ..$ family      : NULL
344 ## ..$ face        : chr "bold"
345 ## ..$ colour      : NULL
346 ## ..$ size        : num 16
347 ## ..$ hjust       : NULL
348 ## ..$ vjust       : NULL
349 ## ..$ angle       : NULL
350 ## ..$ lineheight  : num 0.8
351 ## ..$ margin      : NULL
352 ## ..$ debug       : NULL
353 ## ..$ inherit.blank: logi FALSE
354 ## ..- attr(*, "class")= chr [1:2] "element_text" "element"
355 ## - attr(*, "class")= chr [1:2] "theme" "gg"
356 ## - attr(*, "complete")= logi FALSE

```

357 ## - attr(*, "validate")= logi TRUE

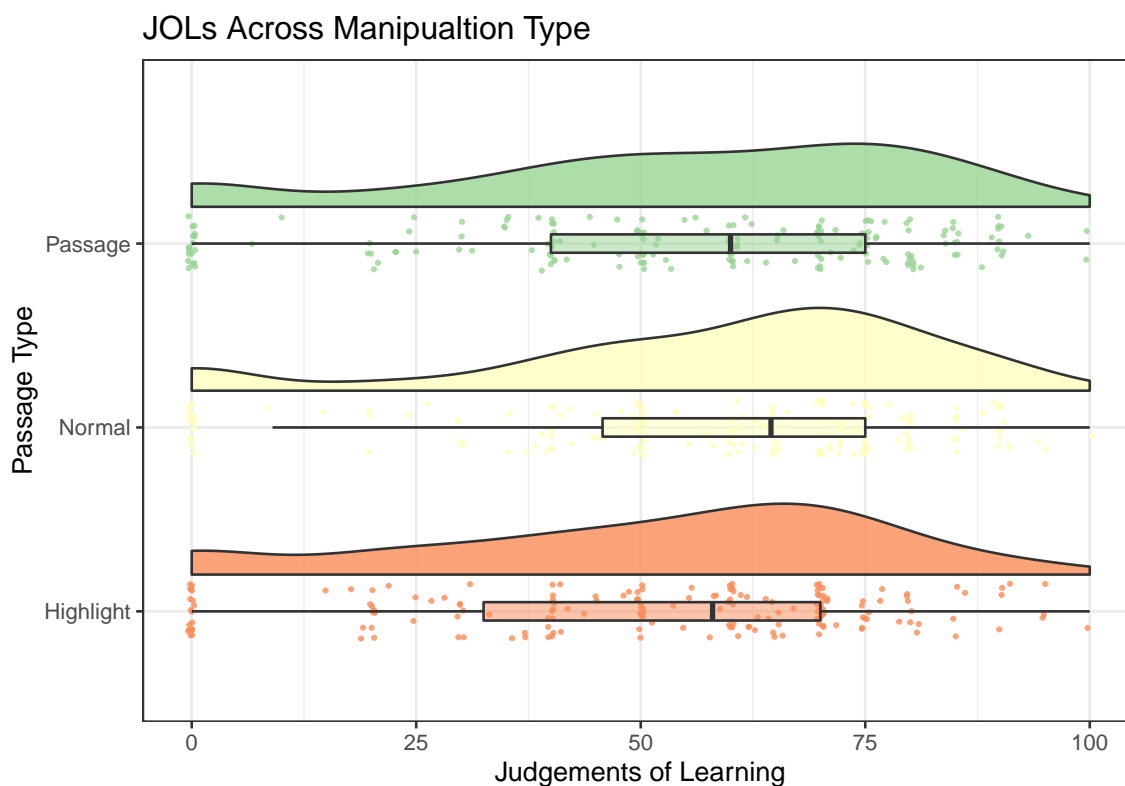


Figure 1

358 We hypothesized that sentences pre-highlighted or presented in sans forgetica would be
 359 better remembered than sentences presented normally. Further, we predicted that there
 360 would be no recall differences between the pre-highlighted and the sans forgetica
 361 conditions. Our hypotheses were only partially confirmed. We found that information that
 362 was pre-highlighted had better recall than passages presented normally, $Estimate =$
 363 $-.328$, $SE = .166$, $z = -1.97$, $p = .048$. Sentences that were pre-highlighted were also
 364 remembered marginally better than sentences presented in sans forgetica, $Estimate =$
 365 $-.307$, $SE = .167$, $z = -1.84$, $p = .066$. Looking at Bayes Factor for this comparison
 366 suggests that evidence for a difference between the two conditions is fairly weak. Critically,
 367 sentences presented in sans forgetica were not better remembered than sentences
 368 presented normally, $Estimate = -.328$, $SE = .166$, $z = -1.97$, $p = .048$, $BF =$.

369 Conclusions

370 The evidence contained herein suggests that SF does not have the mnemonic effects
 371 purported by its creators. Now it is possible that there is an effect of SF, but the effect
 372 size might be smaller than we could detect across our two studies. Our SESOI was $d =$
 373 $.35$. If so, it probably does not have any real educational benefit. It is our conclusion that
 374 SF is really forgettable and you should not be using it as a way to boost learning.

References