NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Timofei A. Berezner, Elena S. Gorbunova

THE PRESENCE OF IMAGES DOES NOT MODERATE THE DISFLUENCY EFFECT

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PSYCHOLOGY

WP BRP 127/PSY/2021

Timofei A. Berezner[1], Elena S. Gorbunova[2]

THE PRESENCE OF IMAGES DOES NOT MODERATE THE DISFLUENCY EFFECT[3]

The disfluency effect is an improvement in the retention of information if its processing is difficult. Disfluency is usually attained by changing perceptual characteristics, and in the case of texts, fonts are made less legible. The disfluency effect has been demonstrated in several studies, but the number of unsuccessful replications is significantly higher. Among the possible reasons why the disfluency effect often cannot be detected, we consider the influence of other factors acting as moderators. This study examines the presence of images accompanying text information as a moderator. Based on dual-coding theory, we assume that the disfluency effect will manifest itself only when there are no images, since their presence contributes to better memorization.

Two experiments were conducted in which information written either in Arial (fluent condition) or Comic Sans (disfluent condition) was presented to participants. The information was also accompanied by images or not. In the first experiment, the participants were asked to memorize 40 short words, in the second 7 fictional facts about the Earth. In both cases, the hypotheses were not confirmed—the disfluency effect did not manifest under any conditions.

JEL Classification: Z

Keywords: disfluency effect, processing fluency, fonts, images, memory, learning, desirable difficulties

[1] National Research University Higher School of Economics, Laboratory for Cognitive Psychology of Digital Interfaces Users, Research Assistant, E-mail: tberezner@hse.ru

[2] National Research University Higher School of Economics, Laboratory for Cognitive Psychology of Digital Interfaces Users, Laboratory Head, E-mail: esgorbunova@hse.ru

[3] This study was implemented within the framework of the HSE University Basic Research Program in 2021

Introduction

Processing fluency is defined as a metacognitive, subjective experience of the ease of information processing (Alter & Oppenheimer, 2009). It is considered that a sense of fluency leads to making more intuitive judgments (Alter et al., 2007), to evaluating information as more plausible (Reber & Schwarz, 1999), pleasant (Reber et al., 2004), typical (Oppenheimer & Frank, 2008) and frequent (Alter & Oppenheimer, 2009). Disfluency, that is, difficulty in processing, contributes to the deeper analytical processing of information (Alter et al., 2007).

It has been suggested that disfluency, due to this deeper processing of information, may also lead to better retention, which was shown in a laboratory experiment and in school settings (Diemand-Yauman et al., 2011), on the basis of which the authors proposed using disfluency as a way to improve the educational outcomes in pedagogical practice. The explanation of the disfluency effect lies in Bjork's approach of desirable difficulties. He showed that productive learning is possible when students experience certain difficulties, overcoming which contributes to the better assimilation of the material (Bjork, 1994; Bjork & Bjork, 2011). This is because overcoming difficulties leads to deeper information processing, consequently, better long-term memorization, which is predicted within the Levels of Processing model (Craik & Lockhart, 1972; Craik & Tulving, 1975). Hence, disfluency, acting as a signal of the difficulty of learning, provokes deeper processing of the information, which leads to better memorization. That is why disfluency is considered an example of a desirable difficulty in learning.

In the initial study in which this effect was discovered (Diemand-Yauman et al., 2011), disfluency was produced using fonts of a certain type: Comic Sans, Bodoni, **Haettenschweiler**, *Monotype Corsiva*, as well as using italics and changing the font from black to gray. In other studies (see them below), disfluency was created by a variety of manipulations: 1) other fonts, such as *Brush*, *Brush Script*, Lucida Blackletter, *Edwardian Script*, *Mistral*, Sans Forgetica; 2) font size, italics, gray text; 3) the imitation of illegible handwriting; 4) a combination of font color and background color; 5) the imitation of a printer with not enough toner; 6) blurred words; and 7) 180 degrees inverted words.

After the discovery of the disfluency effect, many researchers tried to replicate the original results from Diemand-Yauman et al. (2011) in various contexts. For instance, in a study by French et al. (2013), the effect was tested on children with dyslexia. The effect was found for the control group and for children with dyslexia, furthermore, for the latter the improvement in retention due to the deterioration of the font was even greater. Retention improvement was stable for 6 weeks, as demonstrated in a follow-up study by French (2013). Weissgerber & Reinhard (2017) also showed this long-term effect—perceptual disfluency contributed to better information retention after 2 weeks. The disfluency effect was also tested in several other studies: for inverted words (Sungkhasettee et al., 2011); for students learning Japanese as a foreign language, in special fonts for Japanese (Lee, 2013); a 5 pt font led to better memorization than 18 pt (Halamish, 2018); and a U-shaped pattern was found, that is, average disfluency led to better memorization than weak or strong disfluency (Seufert et al., 2016).

However, in a large number of studies the influence of disfluency on memorization or learning performance was not statistically significant (Ball et al., 2014; Berezner & Gorbunova, 2021; Eitel et al., 2014; Eitel & Kuhl, 2016; Faber et al., 2017; Geller et al., 2018; Geller et al., 2020; Halamish et al., 2018; Katzir et al., 2013; Magreehan et al., 2016; Meyer et al., 2015 Miele et al., 2013; Pieger et al., 2016; Pieger et al, 2017; Rhodes & Castel, 2008; Rummer et al., 2016; Sanchez & Jaeger, 2015; Sanchez & Naylor, 2018; Sirota et al., 2021; Strukelj et al., 2016; Taylor et al., 2020; Yue et al., 2013). A meta-analysis of many of these studies (39 experiments from 25 articles involving 3,135 participants) showed a null effect of perceptual disfluency on learning outcomes (Xie et al., 2018). These problems in replicating the disfluency effect require some theoretical explanation. Skeptical assessments were made of the claim that disfluency is a desirable difficulty (Bjork & Yue, 2016). It is possible that disfluency does not actually affect high-level processes, does not lead to deeper processing of information, but only slows down the speed of reading. Other researchers have suggested that the disfluency effect may have moderators (Kuhl et al., 2014; Oppenheimer & Alter, 2014).

In a study by Lehmann et al. (2016), the capacity of working memory was considered a possible moderator. The authors hypothesized that the disfluency effect will manifest in those subjects who have a larger working memory, because the processing of disfluent information leads to a greater cognitive load, to the involvement of more cognitive resources, and then, only if there is a sufficient capacity of working memory, will the subjects be able to benefit from the disfluency of information. This assumption was confirmed, the disfluency effect was observed when the working memory capacity was higher. Strukelj et al. (2016) did not replicate these results. They found that the disfluency effect did not manifest under any conditions, concluding that the working memory capacity was not a moderator. They also used an eye-tracker, which made it was possible to identify that the first two parts of the text presented to the subjects (text was divided into 4 parts) were read faster, and the remaining two were read slower when the text was presented in a disfluent condition compared to the fluent condition. This pattern correlates with what Yue et al. (2013) found, where the disfluency effect did not appear immediately. Based on this, Strukelj et al. suggested that disfluency becomes a desirable difficulty only after a considerable period spent on lengthy educational materials, while when learning shorter materials, it simply serves as a metacognitive cue. Thus, the disfluency effect should be studied using longer texts and longer periods of time, as in the studies of Diemand-Yauman et al. (2011), French (2013), and Weissgerber & Reinhard (2017).

Other studies failed to find a disfluency effect using other possible moderators, for example, test expectancy (Eitel & Kuhl, 2016). The predicted result was found that for high test expectancy, educational outcomes increase, whereas there was no effect of test expectancy on the disfluency effect. The meta-analysis by Xie et al. (2018), mentioned above, not only considered the presence of the disfluency effect, but also its possible moderators, among which were prior knowledge, learning material domain, the pacing of presentation, study design, and the use of distraction tasks. The influence of none of the possible moderators were significant. A recent study showed that the level of English as a foreign language is also not a moderator of the disfluency effect (Berezner & Gorbunova, 2021).

In this study we consider the influence of another possible moderator, which has received little attention in previous studies. We check whether the presence of images is a moderator for the disfluency effect. A significant number of texts in the modern world are accompanied by illustrations, while it is widely known that their presence has a positive effect on the memorization of information (Glenberg & Langston, 1992; Paivio, 1975; Paivio, 2006), which is explained both by Paivio's dual-coding theory and the idea of images contributing to the formation of mental models of the content of texts. In some studies of the disfluency effect, texts were accompanied by images for reasons of environmental validity (e.g., Eitel & Kuhl, 2016), but the authors did not take this fact into account in any way when interpreting the data. Thus, the purpose of our research is to study the interaction of the disfluency effect and the presence of images in the memorizing of verbal information.

Experiment 1

In the first experiment, the subjects were asked to memorize verbal information presented fluently or disfluently, which was accompanied by images or not. In accordance with methodology widely used in studies of the disfluency effect, which consists in using three stages in the experiment—the presentation of stimuli, distraction tasks, and tests—we conducted an online experiment described below. The manipulation of perceptual disfluency is also common: the Arial font as a fluent, Comic Sans as a disfluent (considered as disfluent in Diemand-Yauman et al., 2011; Faber et al., 2017; Rummer et al., 2016). These fonts were chosen as they belong to the same group of sans serif fonts, consequently, the serif element, which can affect reading speed and other parameters, was absent. This is an important methodological property, since typographers have criticized many works in the field of disfluency studies for comparing fonts from different families (Thiessen et al., 2020). We put forward the following experimental hypotheses:

1) Information accompanied by images will be remembered better than information without them (according to the dual-coding theory by Paivio).

2) If images are present, there will be no differences in the memorization of information between fonts of different types, that is, the images are the moderator of the disfluency effect (according to the moderated disfluency hypothesis).

3) If there are no images, then the disfluent information will be remembered better (according to the consideration of disfluency as a desirable difficulty).

Procedure

Experiment 1 took place online on the platform 1ka.si. The subjects were asked to memorize words, then solve a distraction task, and then reproduce as many words as they could remember. The experiment began with detailed instructions and filling out an informed consent form. The subjects agreed with the confidentiality conditions and confirmed that they understood the instructions. Then the subjects reported their gender and age, which was followed by the presentation of words. Each word was presented for 5 seconds, the subjects had no opportunity to return to the previously read word. They were also forbidden to record the content of words in any way (for example, by taking a screenshot). In total, 40 Russian words were presented, denoting common objects (for example, household items or animals). After presenting the words, the subjects had to wait for 2 minutes (a time interval standard for research in this area, e.g., Ball et al., 2014; Eitel & Kuhl, 2016; Geller et al., 2018; Sungkhasettee et al., 2011), perform a distraction task—the addition or subtraction of 5 pairs of two-digit numbers. Finally, the testing phase followed. As part of the free recall, the subjects had to list as many words as possible that they remembered. For this they had to enter the words in a field on the screen. It should be noted that the subjects were warned in the instructions that the memorization of words would be checked. Thus, all the subjects had a high degree of test expectancy.

After that, the subjects were thanked for their participation and offered, if desired, to write to the experimenter with any questions about the goals or results of the study.

The study design was between-subject, in accordance with the levels of two independent variables (fluency and the presence of images), 4 experimental groups were formed (Table 1). Each subject was randomly assigned to one of the experimental groups. The dependent variable was the number of words memorized by the subjects.

Tab. 1. Combination of conditions in experimental groups

	Arial	Comic Sans
Images	1	3
No images (gray rectangles)	2	4

Stimuli

All the words presented to the subjects were of 2 syllables and 4 letters/phonemes. Balancing words in all length parameters was necessary so that the subjects spent the same amount of time reading each word. The words were not balanced for frequency or representability as the list of 40 words presented to each group of subjects was the same, that is, differences in these parameters between the lists of words could not affect the results. Each word was a noun. All the words were taken from a special database of stimuli for psycholinguistic research (Akinina et al., 2014). The words in the fluent condition were written in 20 pt Arial font. The words in the disfluent condition were written in 20 pt Comic Sans font. All other characteristics of fonts, including their color (black), brightness, tilt, etc. were identical. In groups 1 and 3, the words were accompanied by images. These images were taken from the same stimuli database as the words (Akinina et al., 2014), and were black-and-white drawings of objects designated by the words. The size of the images was 152 by 200 pixels in each trial. All other image characteristics (brightness, saturation, etc.) were identical between trials and between experimental groups. All the words were presented in the center of the screen, the images were located below them (see Figure 1). In groups 2 and 4, instead of images, gray rectangles of 152 by 200 pixels were presented, in order to level the conditions among themselves by the presence of a certain nonverbal pattern of gray under the word.

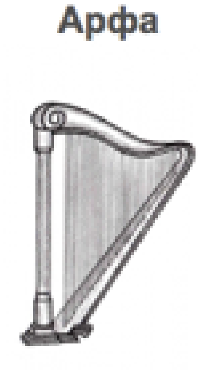

Fig. 1. Example of a trial in Experiment 1 (word in Arial font + image)

Participants

The study involved 144 participants (114 women) aged 17 to 30 years (M = 20.146, SD = 2.075). The subjects participated voluntarily and were recruited via social networks. Some of the subjects participated without inducement, while others had the opportunity to receive additional points for the Experimental Psychology or Usability course, at the Department of Psychology, HSE University.

Results

Data processing and analysis were carried out in the JASP program (version 0.13.1). First, we considered whether the subjects solved the distracting tasks. On average, 4.757 tasks (SD = 0.532) out of 5 were solved correctly. Thus, all subjects were sufficiently distracted by the tasks after the word presentation. On average, the subjects memorized 14.653 words (SD = 5.830) out of 40. Before the main data analysis, the variances of the experimental groups were checked by the Leven test for homogeneity—the Leven test showed homogeneity of variances ($F = 0.559$, $p = .643$). Then a two-factor 2 (fluency: fluency/disfluency) x 2 (images: yes/no) ANOVA was performed. A significant influence of the presence of images was found $F(1, 140) = 5.830$, $p = .017$, $\eta_p^2 = 0.040$—words accompanied by images were remembered significantly better (M = 15.747, SD = 5.775) than words accompanied by a gray rectangle (M = 13.464, SD = 5.695). No significant influence of the font fluency factor was found $F(1, 140) = 0.508$, $p = .477$, $\eta_p^2 = 0.004$. The interaction of factors was also insignificant $F(1, 140) = 0.154$, $p = .695$, $\eta_p^2 = 0.001$. In Figure 2, a comparison of the average values between the groups can be seen, the error bars here and further denote 95% confidence intervals.

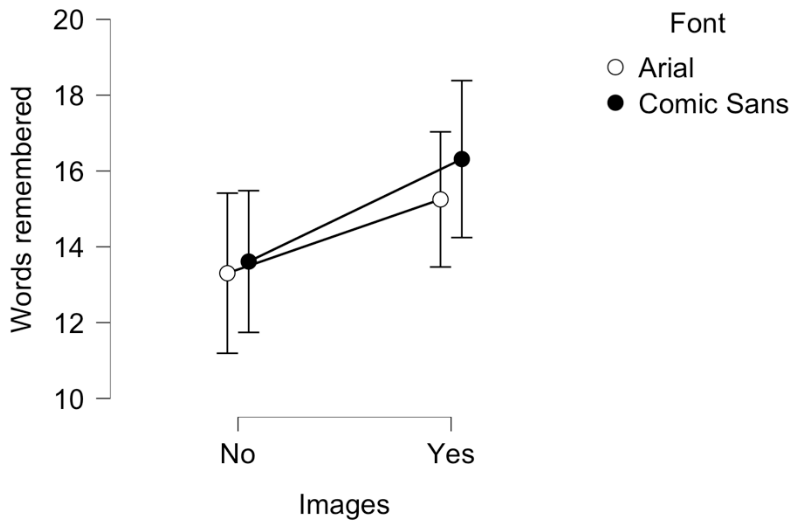

Fig. 2. The average values for the groups in Experiment 1

In this experiment, the subjects also made mistakes when reproducing words—these were both small errors (plural instead of singular: harps for harp), and recording words that were not in the list presented. For exploratory purposes, we considered whether the font type or the presence of images affected the number of errors—here only extra words. On average, the subjects made 0.417 mistakes each, 104 people made no mistakes, 26 people made 1 mistake, and 14 made 2 or more (one subject made 4 mistakes). Two Mann-Whitney U-tests were performed, since the Shapiro-Wilk test showed that the distribution of this dependent variable differs from normal ($p < .05$). Neither the images ($U = 2467, p = .541$) nor the font type ($U = 2860, p = .173$) affected the number of errors.

Discussion

In the first experiment, the subjects had to memorize the words presented to them, which were written either in a fluent or disfluent font and accompanied by an image of the objects or a gray rectangle. A significant effect of the presence of images on memorization was found, namely, words were remembered better if their images were presented with them. This result confirms our first hypothesis and is consistent with the predictions of dual-coding theory. The subjects from the groups with images had the opportunity to memorize objects with both systems at once: verbal and imagery. Indirectly, the idea that the subjects in the process of free reporting relied on drawings is confirmed by the presence of errors by some subjects. In fact, some errors could arise as an association with the drawing, for example, "pomegranate" could be recognized in the drawing "kiwi", and "tree" could be seen in the drawing for the word "burrow". Almost all errors were words consisting of a larger number of letters, that is, it is unlikely that they arose due to consonance.

Although the second hypothesis also found empirical confirmation, there is no reason to assert that the presence of images is a moderator of the disfluency effect. There were no differences in the memorization of words depending on the fluency of the font if the words were accompanied by images or not. The third hypothesis was not confirmed, the effect of disfluency was not detected under any conditions, which means that for groups with images, its absence cannot be explained by images being a moderator. It is more logical to say that the disfluency effect was not observed in principle in this experiment, regardless of the images. This result is consistent with many studies and the meta-analysis that showed a zero effect of disfluency (Xie et al., 2018), however this can be explained not only by there being in effect. First, the subjects were presented with 4-letter words. Perhaps, for the positive effect of disfluency on memorization, in order to cause deeper processing, verbal information of a much larger volume is needed—complete texts or, at least, sentences. Secondly, each word was presented only for 5 seconds. Although this time is sufficient for the subjects to read a short word, it may not be sufficient to produce deeper processing. Thirdly, the Comic Sans font itself might not be disfluent. Arial and Comic Sans fonts are quite similar, which eliminates as much as possible the influence of various typographic characteristics (for example, the presence of serifs), however the difference in the fluency of their processing may be insignificant.

In order to overcome these limitations and further study the interaction of the fluency of verbal information and the presence of images, a second experiment was conducted. It presented several sentences to check whether the disfluency effect would manifest itself in longer texts. The presentation of information was increased to 30 seconds so that the subjects had the opportunity to read the sentences several times. However, the Comic Sans font was not replaced as it could be disfluent and improve the memorization of sentences. Nevertheless, in order to check whether the subjects consider this font poorly legible, they were asked additional questions.

Experiment 2

Procedure

The procedure of experiment 2 was similar to experiment 1. Instead of 40 words, the subjects were presented with 7 facts about the Earth, also in Russian. Each fact was presented for 30 seconds.

The distraction task was the same but with different numbers. One of the key experimental differences was that instead of a free report, the subjects were given a test of 14 yes/no questions at the end of the experiment.

Stimuli

Facts about the Earth were chosen as stimuli for several reasons: first, this is consistent with previous studies in which subjects were also asked to memorize previously unknown, usually scientific or technical, information (Diemand-Yauman et al., 2011; Eitel & Kuhl, 2016); secondly, the task allows us to check the memorization of information of various types: names, numbers, dates, processes, etc. To minimize the possibility of prior knowledge of the participants, all the “facts” were false. However, in order for the facts to seem plausible to the participants, they were created based on real facts by changing some details. Initially, 16 facts were prepared, after which only facts of average plausibility were selected. The selection of facts was based on the degree of their readability and comprehensibility. For this purpose, a pilot study was conducted, described below. After selecting 7 facts, 14 statement questions were compiled for the yes/no test, 2 for each fact. 7 of these statements were false and 7 were true.

Pilot study

Ten people took part in the pilot study. They were presented with the 16 facts compiled as potential stimuli material. On a 7-point Likert scale, they evaluated the plausibility of each fact and its readability. The final list did not include those facts whose plausibility estimates were higher or lower than average for all the facts by one standard deviation. Since the subjectively assessed readability of all the remaining facts was high and did not identify those that should be excluded, the readability of the facts was calculated separately.

The Flesch index and the Gunning fog index, based on similar procedures for calculating the proportion of complex words among all words and words in sentences, were used to assess readability—the ease of reading and understanding. The Flesch index is traditionally interpreted as follows: a score of 0 corresponds to the lowest readability, the text can only be understood by university graduates, a score of 100 corresponds to easy readability, a 10-year-old child can easily understand the text. The Gunning fog index has a slightly different interpretation—a score of 6 corresponds to a text comprehensible by a 6th grader, and a score of 12 or more corresponds to a text comprehensible by a school graduate. It is considered optimal for most people to understand a text, its Flesch index should be 40 or more, and the Gunning fog index less than 8. The number of sentences and words was also counted in all the facts. The final list of 7 facts consisted of two sentences. On average the facts had 19.14 words (SD = 1.46), a Flesch index of 60.18 (SD = 14.82), which corresponds to the level of understanding by about 10th graders, and a Gunning index of 6.24 (SD = 1.09), which corresponds to 7th grade level. Thus, the selected facts had sufficient readability and were balanced by this indicator. Their average plausibility was 4.47 (SD = 0.39), that is, the facts were sufficiently plausible, they did not seem to be obvious falsities, but they did not give the impression of absolute truth.

Main experiment

In experiment 2, the subjects had to memorize the content of the facts presented. In the fluent condition, the facts were written in Arial font 20 pt, in the disfluent Comic Sans 20 pt. All other font characteristics were identical. The facts were balanced for plausibility, readability and length. They were not balanced in terms of the frequency of words used, syntactic structure or other linguistic indicators, since the same set of facts was presented to all experimental groups. In groups 1 and 3, the facts were accompanied by specially selected color images, 200 by 200 pixels (see an example in Fig. 3). By analogy with experiment 1, for groups 2 and 4, a blue square of the same size was presented. The images were presented under the facts in the center of the screen. As noted earlier, after presenting stimuli and a distraction task for 2 minutes, the subjects answered 14 questions. The subjects were warned in the instructions that their memorization of facts would be checked.

Fig. 3. Example of images used in Experiment 2

In addition, at the end of the study, the subjects were asked to evaluate the font with which the facts were written by two indicators: pleasantness and legibility. It has been suggested that if the Comic Sans font is indeed disfluent, then the subjects will rate it as less legible and, possibly, as less pleasant, since fluency is associated with greater attractiveness (Alter & Oppenheimer, 2009). The subjects were also asked to name the facts that they allegedly knew before the start of the study.

Another essential difference between Experiment 2 and Experiment 1 was that after presenting each fact, the subjects were asked to evaluate its plausibility on a 10-point scale. Such manipulation was done for two reasons: first, to additionally verify that each fact was considered to be plausible in a large sample, and second, to validate the data of previous studies, according to which fluent information is considered more plausible (Reber & Schwarz, 1999; Alter & Oppenheimer, 2009), as is verbal information without images (Blinova & Shcherbakova, 2020).

We put forward the following experimental hypotheses.

1) Facts accompanied by images will be remembered better than facts without them (according to dual-coding theory).
2) If images are present, then there will be no differences in memorizing facts between fonts of different types, that is, images are a moderator of the disfluency effect (according to the moderated disfluency hypothesis).
3) If there are no images, then disfluent facts will be remembered better (according to the consideration of disfluency as a desirable difficulty).
4) Fluent facts will be evaluated as more plausible.
5) Facts with images will be evaluated as less plausible.

Participants

The study involved 142 participants (104 women) aged 16 to 47 years (M = 21.338, SD = 3.852). The subjects participated voluntarily; they were recruited via social networks. Some of the subjects participated without inducement, while others could receive additional points for the Experimental Psychology or Usability course, at the Department of Psychology, HSE University.

Results

As in Experiment 1, we first considered whether the subjects solved the distraction tasks. On average, 4.811 examples were correctly solved (SD = 0.459). Thus the subjects were sufficiently distracted by the tasks after the fact presentation. Then we analyzed how many people answered each of the 14 questions correctly. Since the subject could score either 0 or 1 point for each question, the average score was in the range from 0 to 1. On average the subjects scored 0.786 points for each question (SD = 0.195). We checked whether any question could be considered an outlier by constructing a box plot (see figure 4).

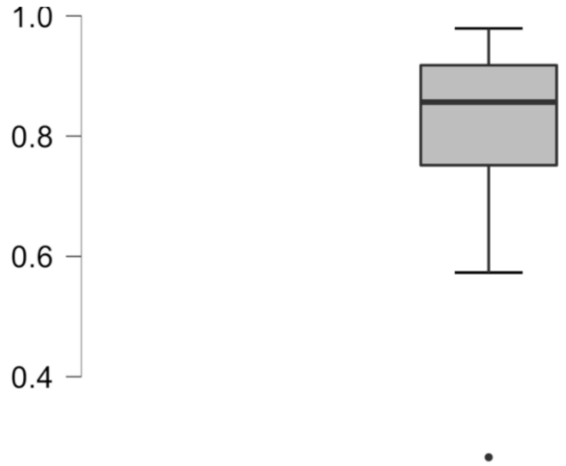

Fig. 4. A boxplot for scores on the questions of the test in the Experiment 2

The first question was an outlier (only 26.6% answered it correctly), and it was removed from further analysis. All other questions remained. On average, the subjects scored 10.8 points (SD = 1.68), which indicates that they coped well with the test, and the information (including numerical) was remembered normally. The possible influence of the subjects' prior knowledge of the facts was also considered. Only 13 people said that they had heard something about certain facts. No respondents indicated that they had known any of the numerical data before the experiment, and they also appealed to the content of the facts, and not to reality (for example, some indicated that they had heard about the extinction in the Jurassic period, which, in fact, never happened). Thus, none of these 13 people had genuine prior knowledge of the content of the facts.

The homogeneity of the variances was checked. Leven's test showed that the variances of the experimental groups were homogeneous $F = 1.396$, $p = .247$. Then a two-factor 2 (fluency: fluency/disfluency) x 2 (images: yes/no) ANOVA was carried out. As in Experiment 1, a significant influence of the image presence factor was found $F (1, 138) = 6.429, p = .012, \eta_p^2 = 0.045$. However, in this experiment, the subjects scored more points in the condition when there were no meaningful images (M = 11.159, SD = 1.511) than when there were such images (M = 10.452, SD = 1.772). The influence of the fluency factor was not significant $F(1, 138) = 0.821, p = .366, \eta_p^2 = 0.006$. The interaction of fluency factors and images was also not significant $F(1, 138) = 0.726, p = .396, \eta_p^2 = 0.005$. Figure 5 shows a comparison of average values for groups. Error bars indicate a 95% confidence interval.

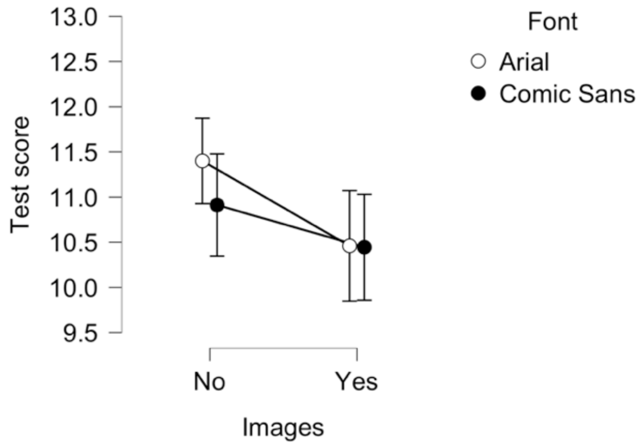

Fig. 5. The average values for the groups in Experiment 2

In order to test the fourth and fifth hypotheses, the influence of font fluency and the presence of images on the plausibility assessment was also considered. Before that, we checked whether the subjects evaluated any fact as significantly more or less plausible than others. The average plausibility score was 5.589 (SD = 0.966) out of 10, that is, the facts were assessed by the subjects as marginally plausible. This means that during the pilot study, facts corresponding to the goals of the main experiment were selected. There were no outliers in terms of its plausibility.

To assess the fluency, and the impact of the presence of images on the plausibility, the scores of all facts were used. To compare the plausibility estimates, t-tests were conducted, since the Shapiro-Wilk test showed that this dependent variable is normally distributed (p = .127). For font fluency, the t-test showed no significant differences *t(140) = -0.330, p = .742, d = -0.055*. Similarly, a t-test was conducted to compare plausibility estimates depending on the availability of images, which also showed no significant differences *t(140) = 0.750, p = .454, d = 0.126*. The fourth and fifth hypotheses were not confirmed; the facts were not evaluated as more or less plausible depending on the fluency of fonts or the presence of images (Figure 6).

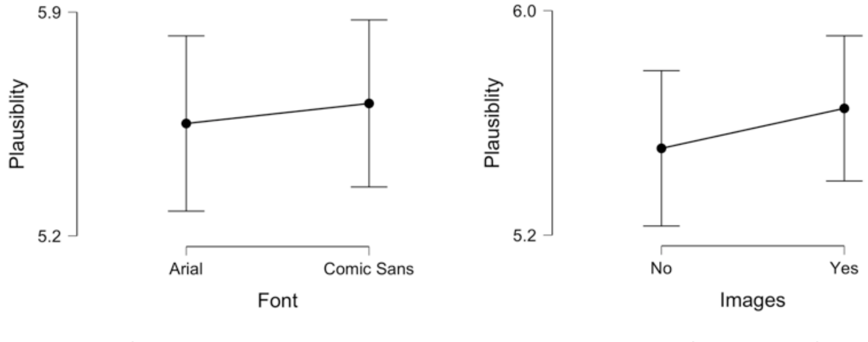

Fig. 6. Comparison of plausibility estimates depending on the fluency of the font (left) and the presence of images (right)

Finally, we checked whether the subjects rated the Comic Sans font as disfluent. This was checked by the Mann-Whitney criterion, since the Shapiro-Wilk test showed that legibility and pleasantness were not distributed normally. A comparison was made of the ratings of legibility and pleasantness of Arial and Comic Sans fonts. The Arial font was noted by the subjects as more legible (M = 8.514, SD = 1.839) than Comic Sans (M = 7.429, SD = 3.058), but this difference was not statistically significant ($U = 2910, p = .095$). The Arial font was also noted by the subjects as more pleasant (M = 7.292, SD = 2.079 vs. M = 6.686, SD = 2.932), but this difference was not statistically significant ($U = 2714.500, p = .424$).

Discussion

The second experiment was conducted in order to consider the interaction of the disfluency effect and the presence of images in a more complex task—memorizing information from small texts. Since we have proposed three possible explanations at once why the disfluency effect did not manifest itself in the first experiment, we focus on their consideration. This time, the subjects were presented not with short words, but facts in two sentences and about 20 words that were longer than 4 letters. The subjects were also given sufficient time—30 seconds—to read each fact. Finally, it was additionally checked whether the subjects considered the Comic Sans font to be disfluent. It was rated by subjects as less legible and pleasant than Arial, but these differences were not significant. Based on this, we can conclude that in the second experiment, a fluent and an almost disfluent font were compared. Due to the insignificant difference, it cannot be argued that Comic Sans was definitely a more disfluent font, and this will be discussed further. It is possible that in the first experiment, the subjects would have evaluated the legibility of Comic Sans in the same way as Arial.

The results obtained from the ANOVA, however, neither confirm the first hypothesis nor coincide with the results of the first experiment. It was shown that the subjects scored higher on the test when the facts were not accompanied by images. This can be explained as follows. In the first experiment, the image exactly corresponded to the word in meaning, it showed the object that was designated by the word, which gave an additional advantage when memorizing. In the second experiment, the image only thematically corresponded to the content of the fact, but did not reflect information—names, numbers, etc. Perhaps the subjects were distracted by looking at the images instead of carefully reading the facts, but the images themselves could not help them remember the information. In future studies, the hypothesis that the presence of images that do not carry a semantic load hinders memorization should be tested separately. This could be done using eye-tracking to understand the proportion of time spent by the subjects looking at the image instead of reading, and the overall dynamics of eye movements. If, instead of simple illustrations, facts presented in the form of infographics, this could improve memorization.

The second hypothesis was confirmed. There were no differences in memorizing facts if there were images. However, again there is no reason to say that this happened because the presence of images moderated the disfluency effect, as it has a stronger beneficial effect on memory. On the contrary, it has been shown that the presence of images affects memorization negatively. The disfluency effect was not detected even in the absence of images, there were no differences in memorization depending on the fluency of the font. The fluent Arial font in the absence of images even led to better memorization than Comic Sans, but this difference was not significant. The third hypothesis was not confirmed. In the two experiments we conducted, the disfluency effect did not manifest itself. The possible reasons for this are discussed below.

The fourth and fifth hypotheses concerned making judgments about plausibility, an aspect that could not be considered in the first experiment, since it presented only words. As noted in previous studies, processing fluency affects a wide range of diverse judgments, including attractiveness, frequency and—the subject of our consideration—truth (Alter & Oppenheimer, 2009). Since it has been shown that fluently processed judgments are evaluated as more plausible (Reber & Schwarz, 1999), we expected to find the same in our experiment. However, the hypothesis was not confirmed—the facts written in Arial or Comic Sans were evaluated as equally plausible. Although the subjects found the Comic Sans font less legible, perhaps it was not disfluent enough to lead to the assessment of facts as false. An indirect confirmation of this is the absence of differences between fonts in terms of their pleasantness. Reber and Schwarz (1999) manipulated fluency by contrasting the font color with the background, and secondly, within-subject design was used. If we had used a within-subject design, the differences might also have appeared—the subjects could have compared the fluency of different facts and evaluated disfluent ones as less plausible. In future studies, it is worth testing this assumption by replicating the experiment with a within-subject design. The fifth hypothesis concerned the effect on the plausibility assessment of the presence of images. The differences were insignificant and the hypothesis was not confirmed. Once again, it is worth further testing this hypothesis in an experiment with a within-subject design, when each subject could assess the plausibility of the facts by comparing those with and without illustrations. However, in previous studies, differences were obtained in the experiment with a between-subject design (Blinova & Shcherbakova, 2020). Accordingly, there must be other explanations for our results. The key difference between our experiment and previous studies was that the facts were created by us on the basis of real facts about the Earth, they were not completely fictional. Perhaps it was the similarity of the content to what is happening on our planet, that influenced the plausibility assessment much more than the presence of images (and the fluency of the font). In order to further test hypotheses about the beneficial effect of fluency on the assessment of plausibility and the negative impact of the presence of images, accurate replications of previous studies with completely made-up facts are necessary.

General discussion

We conducted two experiments in which the disfluency effect and the presence of images as its possible moderator were studied. In the first experiment, the subjects were presented with words that were written in a fluent or disfluent font, and which were accompanied by images of the objects depicted by the words or simple gray rectangles. Only the presence of images was statistically significant, confirming that images improve memorization. This is consistent with the dual-coding theory (Paivio, 1975; Paivio, 2006) and the idea that images allow the formation of mental models of the texts content (Glenberg & Langston, 1992). The fluency of the font and the interaction of factors did not have a significant impact. Since short words were used in this experiment, an even simpler explanation can be found: those subjects who saw images of objects had two sources of information at once. They did not need to form mental models, they just saw an object and memorized it. It is possible that the subjects in those groups in which the images were presented looked at them longer than at the words. This is indirectly confirmed by the presence of errors in some subjects—naming those words that were not on the list. The fluency of the font did not affect these false memories, although in previous studies disfluency contributed to a greater number of them (Sanchez & Naylor, 2018), but still the task of memorizing words was simple and can hardly be interpreted in terms of false memories. Rather, they were just mistakes related to the fact that the subjects remembered the picture well, but not the word. To test the assumption that in such a cognitive task more attention is paid to images, it would be possible in future studies to use an eye-tracker. The results of the first experiment did not replicated the disfluency effect.

The second experiment was designed to evaluate memorization in the more complex context of sentences. Situations when we need to memorize individual words are rare—in most cases, people read complete texts, therefore, in Diemand-Yauman et al. (2011) and many subsequent studies, longer texts are used as stimuli. There are also studies which required the remembering of a word from a pair (for example, Taylor et al., 2020), but such a task was usually not the only one and was accompanied by an assessment of text memorization. Hence, in our second experiment, instead of words, the subjects were presented with facts written in the same fonts, and accompanied or not by pictures, as in experiment 1. The disfluency effect was not detected in this experiment either. The influence of images on memory contradicted our hypothesis and the results of the first experiment. We explained this phenomenon by the images we chose distracting the subjects, instead of helping to build mental models (Glenberg & Langston, 1992). In studies revealing the positive role of images, images contain information relevant to the text. In our experiment, the images corresponded to the facts thematically but did not contain any specific information. This could be what influenced the negative role of images in memorization in the second experiment. The second experiment also examined the effect of fluency on plausibility. Processing fluency theory predicts that fluent stimuli should be evaluated as plausible (Alter & Oppenheimer, 2009; Reber & Schwarz, 1999), however, this was not observed. Nor was there was any effect on the assessment of plausibility for the presence of images, although this was demonstrated in a previous study (Blinova & Shcherbakova, 2020). Why was the disfluency effect not replicated in both of our experiments?

It could be argued that our data confirm the results of numerous other studies, including their meta-analysis, in which there was also no disfluency effect. However, our results can be explained by our understanding of disfluency. We understood Comic Sans as a disfluent font, which may have been incorrect. There are reasons to believe that the Comic Sans font is not disfluent. We chose Comic Sans because it is similar to Arial (both are sans serif), and typographers have criticized fluency studies for using dissimilar fonts from different families, as this can introduce confounding variables (Thiessen et al., 2020). Nevertheless, Comic Sans is much more legible than many other fonts called disfluent in the literature, for example, *Monotype Corsiva* or **Haettenschweiler**. One of the limitations of our work is the fact that we did not check beforehand whether Comic Sans was actually rated as disfluent. We conducted a post hoc check in the second experiment, which showed that the subjects found it less legible and pleasant than Arial, but these differences were insignificant. Perhaps Comic Sans is still less fluent than Arial, but not enough to act as a desirable difficulty. In future studies, our experiments could be repeated using other fonts.

Another possible explanation for the results is an almost instantaneous memory testing. In accordance with established methodology, we used a distraction task between the presentation and testing phases, but this might not be sufficient for disfluency to become a desirable difficulty. According to Bjork (1994), the benefits of desirable difficulties are felt only after a while, perhaps several weeks. This may explain the unsuccessful replication of the disfluency effect, not only by us, but also by other authors. In the second experiment of Diemand-Yauman et al. (2011), the subjects interacted with disfluent fonts for several weeks, in the studies of French (2013) and Weissgerber &

Reinhard (2017) also showed the disfluency effect after a few weeks. This does not answer the question why in the first experiment of Diemand-Yauman et al. (2011), which took place in the laboratory, the disfluency effect was also detected, but suggests that sufficient time should pass for its occurrence. In future studies, it is worth accounting for this and testing the memorization of information by the subjects after a few weeks.

Another limitation of our work is that we warned our subjects in advance that their memorization of information would be tested. In the work of Eitel & Kuhl (2016), it was assumed that test expectancy is a moderator of the disfluency effect. Their hypothesis was not confirmed, but in general, very few studies have been conducted on the impact of the test expectancy on the disfluency effect. It would make sense to check the effect of this factor in the future. Finally, processing fluency is a metacognitive cue that affects a variety of judgments. For this reason, many studies have evaluated the impact of disfluency on making judgments of learning. In our work, we did not ask the subjects to predict their success during testing. It would be useful to study the effect of disfluency on learning judgments within the procedure we presented from the point of view that it could provide additional evidence about whether the subjects consider the Comic Sans font sufficiently disfluent or not. Perhaps it is more reasonable to use a within-subject design for such a study. Another development in fluency theory would be to study perceptual disfluency not only using texts, but also using disfluent images.

References:

Akinina, Yu., Iskra, E., Ivanova, M., Grabovskaya, M., Isaev, D., Korkina, I., Malyutina, S., Sergeeva, N. (2014). Stimulus database ‘Noun and object’: normative data collection for psycholinguistic variables. In A. Kibrik et al. (Eds.), *Sixth International Conference on Cognitive Science: Abstracts*. Kaliningrad, June 23-27 2014. Issue 6 (pp.112-114)

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(4), 569-576.

Alter, A.L., Oppenheimer, D.M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review,* 13(3), 219-235.

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied*, *20*(4), 336-348

Berezner T.A., & Gorbunova E.S. (2021). Improving text memorization by changing fonts: the study of Sans Forgetica. *Psikhologicheskie Issledovaniya, 14*(78), 2.

Bjork, R.A. (1994). 'Memory and metamemory considerations in the training of human beings'. In *Metacognition: Knowing about knowing*, 185-205.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the real world: Essays illustrating fundamental contributions to society*, 56–64.

Bjork, R. A., & Yue, C. L. (2016). Commentary: is disfluency desirable? *Metacognition and Learning*, 11(1), 133–137.

Blinova, E., & Shcherbakova, O. (2020). When seeing is not believing: The role of illustrations in judging the reliability of information. In *Proceedings of the 4th International Conference "Neurobiology of Speech and Language"* (50-51).

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.

Craik, F.I.M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268-294.

Eitel, A. & Kuhl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, 11, 107-121.

Eitel, A., Kuehl, 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*(4), 488-501.

Diemand-Yauman C., Oppenheimer D.M., Vaughan E.B. (2011). Fortune favors the Bold (and the Italicized): effects of disfluency on educational outcomes. *Cognition*, 118 (1), 111-115.

Faber, M., Mills, C., Kopp, K., & D'mello, S. (2017). The effect of disfluency on mind wandering during text comprehension. *Psychonomic bulletin & review*, *24*(3), 914-919.

French M. M. J., Blood A., Bright N.D., Futak D., Grohmann M.J., Hsthorpe A., Heritage J., Poland R.L., Reece S., Tabor J. (2013). Changing fonts in education: how the benefits vary with ability and dyslexia. *The Journal of Educational Research*, 106(4), 301-304.

French, M. M. J. (2013). Changing fonts in education: The time dependence of the benefits. Retrieved from http://matthewfrench.net/pubs/Changing%20Fonts%20in%20Education%20Time%20Dependence.pdf

Geller, J., Davis, S. D., & Peterson, D. (2020). Sans forgetica is not desirable for learning. *Memory*, 28(8), 957-967.

Geller, J., Still, M. L., Dark, V. J., & Carpenter, S. K. (2018). Would disfluency by any other name still be disfluent? Examining the disfluency effect with cursive handwriting. *Memory & Cognition*, *46*(7), 1109-1126.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of memory and language*, *31*(2), 129-151.

Halamish, V. (2018). Can very small font size enhance memory?. *Memory & cognition*, *46*(6), 979-993.

Halamish V., Nachman H. and Katzir T. (2018). The effect of font size on children's memory and metamemory. *Frontiers in Psychology*, Vol. 9, Article 1577.

Katzir T., Hershko S., Halamish V. (2013). The effect of font size on reading comprehension on second and fifth grade children: bigger is not always better. *PLoS ONE* 8(9): e74061.

Kühl, T., Eitel, A., Scheiter, K., & Gerjets, P. (2014). A call for an unbiased search for moderators in disfluency research: reply to Oppenheimer and Alter (2014). *Applied Cognitive Psychology*, *28*(5), 805-806.

Lee, M. H. (2013). Effects of disfluent kanji fonts on reading retention with e-book. In *2013 IEEE 13th International Conference on Advanced Learning Technologies* (pp. 481-482). IEEE.

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: an aptitude-treatment-interaction study. *Metacognition and Learning*, *11*(1), 89-105.

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, *11*(1), 35-56.

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook G., Ackerman R., Thompson V.A. & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16-e30.

Miele D.B., Metcalfe J., Son L.K. (2013). Children's naive theories of intelligence influence their metacognitive judgments. *Children Development*, 84, 1879-1886.

Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, *28*(4), 502–504.

Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, *106*(3), 1178-1194.

Paivio, A. (1975). Coding distinctions and repetition effects in memory. In *Psychology of learning and motivation* (Vol. 9, pp. 179-214). Academic press.

Paivio, A. (2006). Dual coding theory and education. Draft chapter presented at the conference on Pathways to Literacy Achievement for High Poverty Children at The University of Michigan School of Education.

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency – Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, *44*, 31-40.

Pieger, E., Mengelkamp, C., & Bannert, M. (2017). Fostering analytic metacognitive processes and reducing overconfidence by disfluency: The role of contrast effects. *Applied Cognitive Psychology*, *31*(3), 291-301.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and cognition*, *8*(3), 338-342.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review*, *8*(4), 364-382.

Rhodes M.G. & Castel A.D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology*, 137, 615-625.

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: null-effects of disfluency on learning outcomes. *Metacognition and Learning*, *11*(1), 57-70.

Sanchez, C. A., & Jaeger, A. J. (2015). If it's hard to read, it changes how long you do it: Reading time as an explanation for perceptual fluency effects on judgment. *Psychonomic bulletin & review*, *22*(1), 206-211.

Sanchez, C. A., & Naylor, J. S. (2018). Disfluent presentations lead to the creation of more false memories. *PloS one*, *13*(1), e0191735.

Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science, 45*(2), 221-238.

Sirota, M., Theodoropoulou, A., & Juanchich, M. (2021). Disfluent fonts do not help people to solve math and non-math problems regardless of their numeracy. *Thinking & Reasoning*, *27*(1), 142-159.

Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning, 11*(1), 71-88.

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin and Review*, 18(5), 973–978.

Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are not desirable difficulties: the (lack of) effect of Sans Forgetica on memory. *Memory*, 28(7), 850-857.

Thiessen, M., Beier, S., & Keage, H. (2020). A Review of the Cognitive Effects of Disfluent Typography on Functional Reading. *The Design Journal*, *23*(5), 797-815.

Weissgerber, S. C., & Reinhard, M. A. (2017). Is disfluency desirable for learning? *Learning and instruction, 49*, 199-217.

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.

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 judjments and memory. *Memory & Cognition*, 41(2), 229-241.

Corresponding author:

Timofei A. Berezner
National Research University Higher School of Economics (Moscow, Russia), Laboratory for Cognitive Psychology of Digital Interfaces Users, Research Assistant
E-mail: tberezner@hse.ru