

Fortune is fickle: null-effects of disfluency on learning outcomes

Ralf Rummer¹ · Judith Schweppe¹ · Annett Schwede¹

Received: 10 February 2015 / Accepted: 1 November 2015 / Published online: 12 November 2015
© Springer Science+Business Media New York 2015

Abstract Recently, Diemand-Yauman et al. *Cognition*, 118, 114–118 (2011) demonstrated that learning with disfluent (hard-to-read) materials is more effective than learning with easy-to-read materials – a study that has since stipulated a number of follow-up studies (with mixed results). However, there is a potential confound in the original experiments: The disfluent materials were not only disfluent but also rather unusual. Therefore, they might have been particularly distinctive and have attracted more attention, which then resulted in better learning. We conducted three experiments to address this confound, all of them slightly modified replications of Diemand-Yauman et al.’s Experiment 1. Participants received five lists, either at their own pace on a computer screen (Experiment 1) or experimenter-paced on paper (Experiments 2 & 3). In Experiments 1 and 2, participants either received one fluent and four disfluent lists or they received four fluent lists and one disfluent list. The position of the distinct list varied across participants. In Experiment 3, the distinct list was always the penultimate one. In none of the experiments, learning performance was affected by any of the experimental manipulations. Our results question the generality of the disfluency effect with respect to learning.

Keywords Disfluency effect · Learning · Memory · Replicability

Dual route theories (e.g., Kahneman 2011) suggest that information processing usually is based on quick and shallow processes, which do not cause much cognitive load and are associated with positive feelings (system 1 processing). This type of processing is mainly automatic and its outcomes are usually good enough but not excellent (for a different perspective see Gigerenzer et al. 1999). Only if necessary, a second slower and less convenient type of processing is applied. This optional system 2 is characterized by the contribution of a capacity-limited attentional (executive) control system (e.g., Norman and Shallice 1986) and supports deeper information processing. System 2 processing is thus more demanding than

✉ Ralf Rummer
ralf.rummer@uni-erfurt.de

¹ Department of Psychology, University of Erfurt, P. O. Box 900 221, 99105 Erfurt, Germany

system 1 processing but (usually) leads to more precise and more adaptive outcomes (e.g., Kahneman 2011; but see again Gigerenzer et al. 1999).

It is suggested that a condition that triggers system 2 processing is that the individual categorizes the current task as important for his or her purposes (e.g., Norman and Shallice 1986). System 2 processing is obligatory when the individual perceives a situation as highly important. For learning situations, this could be the case in a consequential exam. However, learning situations in psychological experiments, usually do not meet this criterion. Thus, learners try to restrict their efforts, which means they use system 1. This is particularly the case when the learning task is simple. System 2 comes into play when the *central* requirements of a task are perceived as difficult. According to theories on metacognition and learning, the evaluation of the task's difficulty is accomplished by a *monitoring system* which judges the *ease of task processing* (e.g., Bjork et al. 2013; Dunlosky et al. 2007; Nelson and Narens 1990). When the monitoring process signals that the task is demanding, a *control process* switches to system 2. However, when the monitoring process judges the task as easy, processing takes place in system 1.

The architecture of the metacognitive system also suggests another trigger for system 2 processing, even when the central requirements of the task are rather simple. Individuals can be tricked into system 2 processing *peripherally*. According to research on the *disfluency effect*, even rather simple tasks call for the contribution of system 2 when the perception of the task causes a *feeling of difficulty* (e.g., Alter et al. 2007). Such a feeling can, for instance, be induced by presenting instructions or materials in a hard-to-read font rather than an easy-to-read font. Alter et al. (2007) argue that the feeling of disfluency while reading indicates processing difficulties, as a consequence of which the task is processed by the attentionally demanding system 2. This (usually) leads to better performance as compared to a presentation in an easy-to-read font, which does not activate the attentional system and thus leads to shallow system 1 processing. In sum, disfluency is not only a consequence of the actual task difficulty. It can just as much be caused by surface characteristics of the materials that suggest difficulty such as hard-to-decipher fonts (Alter and Oppenheimer 2009).

In a recent study by Diemand-Yauman et al. (2011), the processing of hard-to-read learning materials resulted in better learning outcomes. In their first experiment, they presented three lists of characteristics describing unknown (invented) species of aliens. Each of these lists included the name of the alien, and six characteristics such as its color, nutrition, and habitat. The lists were presented on a piece of paper. Learning time was 90 s. Half of the participants received fluent materials, and half of them disfluent materials. In the fluent condition, the lists were presented in 16-point Arial pure black font; in the disfluent condition, they were presented in Comic Sans MS 12-point or Bodoni MT 60 % grayscale font. After a delay of 15 min, participants were asked seven (out of 21 possible) questions. Learning of disfluent lists was better than learning with fluent lists: In the fluent condition, participants answered 72.8 % of the questions correctly; in the disfluent condition, they were successful in 86.5 % of the cases.

In addition to this laboratory study, Diemand-Yauman et al. (2011) conducted a field experiment at a high school. The teachers whose students participated in this study always gave two classes of the same class level. One of these classes received their study materials as usual (i.e., in a fluent version), the other course was presented with disfluent materials. Disfluency was realized either by a hard-to-read font (i.e., Haettenschweiler, Monotype Corsiva, Comic Sans Italized) or – when electronic documents were unavailable – the study materials were copied disfluently (by moving the paper up and down during copying). Even under these field conditions, the students who received the disfluent materials performed better in their final exam than those in the fluent condition. Diemand-Yauman et al. (2011) interpreted their results such that the perceived surface characteristics of the text tricked the

learners into deeper processing (Craik and Tulving 1975) and thus led to a richer and more durable representation of the text content. Based on these two studies, Diemand-Yauman et al. (2011) suggested that the disfluency effect extends to real world scenarios and can be easily applied to all kinds of learning materials relevant in educational settings.

The results of these two experiments are of particular interest because instructional psychologists as well as practitioners usually suggest a negative relation between difficulties induced by the design of learning materials and the learning outcome: The higher the *extraneous cognitive load* (i.e., the cognitive load caused by the way the to-be-learned information is presented to the learner), the lower is the learning outcome (e.g., Sweller and Chandler 1994; Sweller et al. 2011). Thus, Diemand-Yauman et al.'s study stipulated a number of follow-up studies (e.g., Eitel et al. 2014; French et al. 2013). For instance, Eitel et al. (2014) published a series of four experiments, in which they applied the disfluency manipulation to multimedia materials including written text and pictorial information. Surprisingly, only in one of these four experiments a disfluency effect occurred.

The unsystematic appearance and nonappearance of a disfluency effect in learning in this study is mirrored in the overall pattern of studies conducted so far: Whereas a number of researchers demonstrated disfluency effects under different conditions (e.g., Diemand-Yauman et al. 2011; French et al. 2013; Sungkhasettee et al. 2011), others failed to do so (e.g., Carpenter et al. 2013; Eitel et al. 2014; Yue et al. 2013). Thus, the generality and/or robustness of the disfluency effect is questionable. Consequently, Eitel et al. (2014), Kühl et al. (2014) as well as Oppenheimer and Alter (2014) call for the search for so far unknown boundary conditions.

We investigate one possible boundary condition: Diemand-Yauman et al. (2011, p. 117) themselves noted that there is an alternative explanation for their findings, namely that “hard-to-read fonts were more distinctive and that the effects were driven by distinctiveness” rather than disfluency. In other words, the disfluent materials were not only disfluent but also rather unusual. Novel and unexpected stimuli are known to capture attention (e.g., Hunt and Worthen 2006; Jenkins and Postman 1948). Therefore, the materials in an unusual font might have been particularly *distinctive* and attracted more attention. This could induce system 2 processing just as perceived difficulty is assumed to do.

To sum up, surface characteristics of a memory task can be decisive for the metacognitive processes that result in either system 1 or system 2 processing. We assume that tasks' surface characteristics do not only affect these metacognitive processes via *disfluency* but also via *distinctiveness*.

Whether learning benefits from disfluency or just from distinctiveness makes a crucial difference for the practical implications of these findings. A disfluency effect which is based on the distinctiveness of hard-to-read fonts would suggest using disfluent fonts only rarely to emphasize particularly important information; in contrast, a general disfluency effect would call for a more comprehensive use of hard-to-read learning materials. It is thus essential to contrast these two explanations.

To investigate whether disfluency or distinctiveness caused the better learning outcomes for the disfluent lists in Diemand-Yauman et al.'s (2011) first experiment, we conducted a modified replication of their study. To this end, we manipulated both disfluency and distinctiveness experimentally such that two conditions resulted: Half of the participants received most of the lists in an easy-to-read font and only one list was hard-to-read; the other half received most of the lists in a hard-to-read font and only one list was easy-to-read. Distinctiveness is thus not operationalized quasi-experimentally in terms of frequency of occurrence outside the experiment but rather experimentally as being a singleton versus being “the usual” within the experiment.

If the disfluency effect is mainly driven by distinctiveness, the single hard-to-read list should benefit more from disfluency than the hard-to-read lists in the condition in which most lists are disfluent. More precisely, it should be more probable for a single hard-to-read list to attract attention than for several hard-to-read lists. This attentional advantage should also result in a memory advantage. If disfluency itself causes the learning benefit, there should be a main effect for disfluency that is independent of the distinctiveness manipulation. In other words, disfluent lists should be processed more deeply (and therefore remembered better) than fluent lists irrespective of the amount of disfluent lists presented.

Experiment 1

The experiment was based on a 2×2 mixed design. The within subject factor ‘fluency’ was manipulated on a local level (i.e., a list was presented either in fluent or disfluent font) and the between subjects factor ‘distinctiveness of disfluency’ was manipulated on a global level (i.e., participants received the majority of lists either in a fluent font or in a disfluent font). The correctness of answers to questions concerning the characteristics of the aliens served as the dependent variable. In order to test whether the disfluent font actually increases reading times, we recorded reading times per list in a self-paced reading paradigm.

Method

Participants

63 undergraduates of the University of Erfurt (52 of them female), aged between 18 and 38 ($M_{age}=22.81$; $SD=3.58$), participated in exchange for partial fulfillment of course requirements or payment.

Materials

The materials used in this experiment were similar to those presented in Diemand-Yauman et al.’s study 1. Participants were instructed to learn about five species of aliens, each of which had seven features, for a total of 35 features that needed to be learned. Materials were presented on a computer screen with the program PsychoPy (Peirce 2007). Depending on condition, each alien description was presented in either fluent or disfluent font. Disfluent lists were presented in 12-point 60 % grayscale Comic Sans MS font; fluent lists were presented in 16-point black Arial font. Across participants, each list occurred equally often in a fluent and in a disfluent font. The distinctiveness manipulation affected the distribution of fluent and disfluent lists per participant. Half of the participants received a majority of fluent lists with four of five lists in fluent font (Fig. 1a); the other half read a majority of disfluent lists with four of five lists in disfluent font (Fig. 1b). The first and the last lists were always presented in the majority font. Across participants, the position of the minority list was distributed equally across positions 2, 3, and 4. Assignment of the lists to position was randomized.

(a)	<p>Der Lerondo</p> <ul style="list-style-type: none"> • 5 cm groß • runde Kopfform • hat grüne Augen • frisst Insekten und Schnecken • schwarz-rötliche Hautfarbe • wird bis zu 14 Jahre alt • lebt im Wald 	<p>Der Sarino</p> <ul style="list-style-type: none"> • 25 cm groß • pyramidenförmige Kopfform • hat graue Augen • frisst Käse und Samen • orange-blaue Hautfarbe • wird bis zu 17 Jahre alt • lebt in der Wüste 	<p>Der Klalec</p> <ul style="list-style-type: none"> • 15 cm groß • zylindrische Kopfform • hat gelbe Augen • frisst Baumrinde und Moos • beige-braune Hautfarbe • wird bis zu 24 Jahre alt • lebt in sumpfigen Gebieten 	<p>Der Purome</p> <ul style="list-style-type: none"> • 10 cm groß • dreieckige Kopfform • hat blaue Augen • frisst Fische und Krebse • braun-graue Hautfarbe • wird bis zu 22 Jahre alt • lebt im Süßwasser 	<p>Der Bulgat</p> <ul style="list-style-type: none"> • 20 cm groß • Kopfform oval • hat braune Augen • frisst Blätter und Blüten • gelblich-grüne Hautfarbe • wird bis zu 11 Jahre alt • lebt in bergigen Gebieten
	(b)	<p>Der Sarino</p> <ul style="list-style-type: none"> • 25 cm groß • pyramidenförmige Kopfform • hat graue Augen • frisst Käse und Samen • orange-blaue Hautfarbe • wird bis zu 17 Jahre alt • lebt in der Wüste 	<p>Der Klalec</p> <ul style="list-style-type: none"> • 15 cm groß • zylindrische Kopfform • hat gelbe Augen • frisst Baumrinde und Moos • beige-braune Hautfarbe • wird bis zu 24 Jahre alt • lebt in sumpfigen Gebieten 	<p>Der Bulgat</p> <ul style="list-style-type: none"> • 20 cm groß • Kopfform oval • hat braune Augen • frisst Blätter und Blüten • gelblich-grüne Hautfarbe • wird bis zu 11 Jahre alt • lebt in bergigen Gebieten 	<p>Der Lerondo</p> <ul style="list-style-type: none"> • 5 cm groß • runde Kopfform • hat grüne Augen • frisst Insekten und Schnecken • schwarz-rötliche Hautfarbe • wird bis zu 14 Jahre alt • lebt im Wald

Fig. 1 Two examples for the materials presented in Experiments 1 and 2. **(a)** A trial example in which the majority of the lists is fluent; **(b)** an example in which the majority of the lists is disfluent

Procedure

Participants were tested in individual sessions. The learning time per list was self-paced, but limited to 120 s. Participants were then distracted for approximately 15 min with unrelated tasks. Finally, participants received a paper sheet with 21 questions addressing the 21 features of those three aliens presented in position 2, 3, and 4 (e.g., “What is the eye color of the Klalec?”). The first and the last list were never tested. This was to avoid primacy and recency effects (which would favor the majority condition). Depending on condition, either two disfluent lists and one fluent list or two fluent lists and one disfluent list were tested per participant. The seven features of each alien were tested in randomized order, without mixing features of different aliens. The experiment lasted approximately 35 min.

Subjects’ responses to the questions were scored by a rater who was blind to the experimental condition. One point was awarded for each correctly recalled feature (out of 21). For each participant, we calculated the percentage of correctly recalled features per condition (i.e., out of 14 for the majority font and out of 7 for the minority font).

Results

Reading times

Figure 2 shows mean reading times per list as a function of ‘disfluency’ (fluent vs. disfluent font) and ‘distinctiveness of disfluency’ (1 list disfluent vs. 4 lists disfluent). The mean reading times per condition were submitted to a 2×2 mixed analysis of variance (ANOVA). This analysis revealed a main effect of distinctiveness, $F(1,61)=5,14, p=.03, \eta^2=.078$, demonstrating that participants spent more time reading all the three mid lists (disfluent and fluent) when two of them were presented in a disfluent font

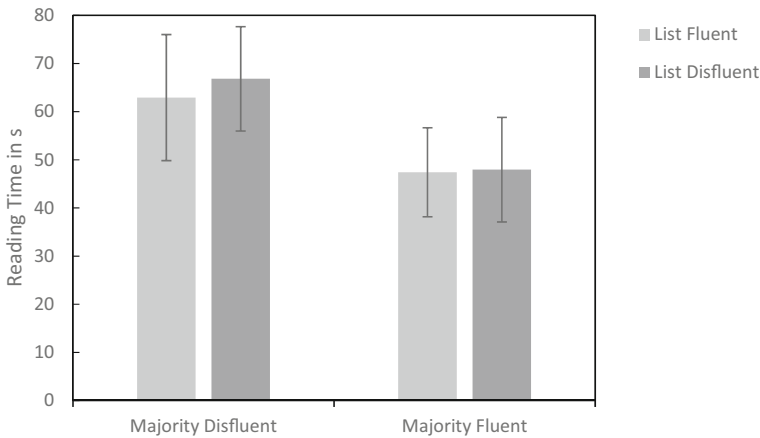


Fig. 2 Mean reading times per list as a function of ‘disfluency of list’ (disfluent vs. fluent) and ‘distinctiveness’ (majority disfluent vs. majority fluent) in Experiment 1. The error bars depict confidence intervals

(64.87 s) as compared to the condition in which one of them was presented in a fluent font (47.68 s). However, there was neither a main effect of disfluency, nor an interaction between disfluency and distinctiveness, both $F_s < 1$.

Test performance

The mean proportion of features recalled on the final test is shown in Fig. 3. The results were submitted to a 2×2 mixed ANOVA, with disfluency (fluent or disfluent font) and distinctiveness of disfluency (1 list disfluent vs. 4 lists disfluent) as independent variables. This analysis neither revealed a main effect of fluency nor a main effect of distinctiveness of disfluency; the interaction between the two factors was also not significant, all $F_s < 1$.

Discussion

This study was a slightly modified replication of Diemand-Yauman et al.’s (2011) first demonstration of a disfluency effect with learning. Unlike Diemand-Yauman et al. (2011, Exp. 1) we did not observe a learning advantage for lists written in a disfluent font. Moreover, none of the manipulations affected learning outcomes. That is, disfluent lists were never recalled better than fluent ones, neither when most of the lists were disfluent nor when there was one particularly distinct disfluent list.

As a treatment check, we measured reading times per list. Reading times were significantly higher for those subjects who were presented with four disfluent lists (and one fluent list) as compared for those who were presented with four fluent lists (and one disfluent lists). However, there was no main effect of disfluency and no interaction. This implies that not only the four disfluent lists but also the single fluent list were read more slowly. These findings are thus hard to interpret. It may be that being confronted with many hard-to-read lists resulted in learners perceiving the entire task as difficult, because of which they also devoted more time to the single easy-to-

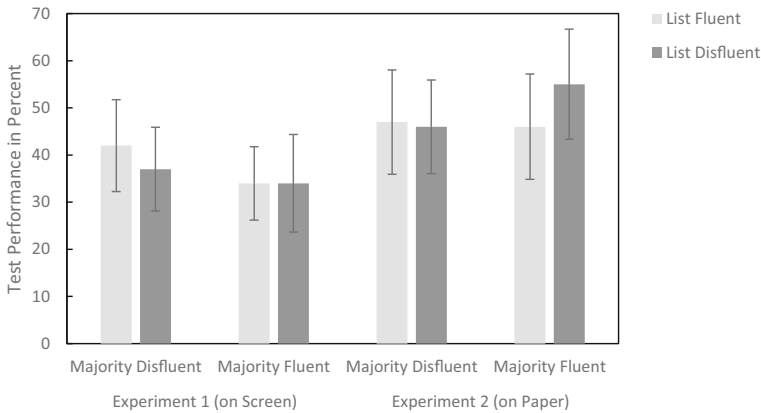


Fig. 3 Proportion of correct answers (in percent) as a function of ‘disfluency of list’ (disfluent vs. fluent) and ‘distinctiveness’ (majority disfluent vs. majority fluent) for Experiments 1 and 2. The error bars depict confidence intervals

read list in this condition.¹ However, the fact that participants spent more time reading the lists in this condition, did not translate to better learning outcomes.

The present study differs from Diemand-Yauman et al.’s original experiment in at least two respects: First, Diemand-Yauman et al. (2011, Exp. 1) presented their study materials on paper rather than on a computer screen. As it is plausible that readability of font sizes differs between reading on paper or screen, this might be a reason why our disfluency manipulation did not result in better learning. In addition, pacing in Diemand-Yauman et al.’s study was fixed, but self-determined in our study. In Experiment 2, we therefore replicated the first experiment in a paper-pencil version with fixed learning times.

Experiment 2

The experiment was based on the same design as Experiment 1. We manipulated the factor ‘fluency’ within participants and the factor ‘distinctiveness of disfluency’ between participants. The dependent variable was the correctness of answers to questions concerning the characteristics of the aliens.

Method

Participants

This second study took place in classroom settings in two introductory psychology courses at the University of Erfurt. 58 undergraduates (47 of them female) aged between 18 and 48 ($M_{age}=21.45$; $SD=4.54$) participated in exchange for partial fulfillment of course requirements. Half the participants were presented with mostly disfluent lists and half of them with mostly fluent lists.

¹ We thank an anonymous reviewer for this suggestion.

Materials and procedure

We used the same materials as in Experiment 1, but we presented the to-be-learned texts in a paper booklet, with a separate page for each list. Reading time for each list was restricted to 90 s. After learning, participants were distracted for approximately 15 min with unrelated tasks. The final test was the same paper-pencil-test as in Experiment 1. Scoring was analogous to Experiment 1.

Results

The mean proportion of lists recalled on the final test is shown in Fig. 3. The results were submitted to a 2×2 mixed ANOVA, with disfluency (fluent or disfluent font) and distinctiveness of disfluency (majority of lists disfluent or singleton disfluent) as independent variables. This analysis neither revealed a main effect of disfluency nor a main effect for distinctiveness, both $F_s < 1$. The two-way interaction between disfluency and distinctiveness was also not significant, $F(1,56) = 1.75$, $p = .19$, $\eta^2 = .030$.

Discussion

Even though we presented the materials on paper with fixed presentation times, as did Diemand-Yauman et al. (2011, Exp. 1), neither a disfluency nor a distinctiveness effect was found. It may be that the disfluency effect is smaller than expected based on Diemand-Yauman et al.'s findings. If this is the case, the power in our experiments may have been too low. To increase statistical power, we thus conducted a joint analysis of Experiments 1 and 2 with the additional between experiments factor 'presentation mode' (on paper vs. on screen). In this joint mixed ANOVA there was again neither a main effect of 'disfluency' nor of 'distinctiveness of disfluency', both $F_s < 1$, nor an interaction between the two factors, $F(1,117) = 2.08$, $p = .15$, $\eta^2 = .017$. The other two-way interactions as well as the three-way interaction were also not significant, all $F_s < 1.2$. That is, even with doubled number of participants, disfluency did not significantly affect learning outcomes. The only significant effect was a main effect for the between-experiments variable 'presentation mode', $F(1,117) = 6.92$, $p = .01$, $\eta^2 = .056$. Participants answered the questions more accurately when the lists were presented on paper rather than with self-paced reading on a computer screen (48.34 vs. 36.82 %, respectively). This may be due to the different times on task in the two studies. Whereas participants in Experiment 1 spent 56.14 s with the lists, learning time for the participants in Experiment 2 was 90 s per list.

These results suggests that – in contrast to our predictions – neither the list nor the context manipulations did cause any effect on learning outcomes. It is particularly surprising that no disfluency effect was observed, not even in the condition in which there was one distinctive disfluent list. However, we presented the minority lists equally often in position 2, 3, and 4. This may have affected the effectiveness of the distinctiveness manipulation. When the singleton was presented in position 2, it was in fact not particularly distinct because at this time, participants had only read one fluent list before encountering the single disfluent one. That is, its singularity could only be noticed in retrospect while at the time of presentation it was not distinctive in terms of Jenkins and Postman's (1948) account. Yet when the single disfluent list was preceded by three lists in fluent font, it should have been a lot more unexpected and thus more salient. Moreover, none of our conditions was directly comparable

to Diemand-Yauman et al.'s (2011, Exp. 1) since there was no condition in which lists were presented solely in a fluent font or solely in a disfluent font. It may be that we never found a disfluency effect because our distinctiveness manipulation overrode the influence of disfluency. We therefore changed our procedure such that the single disfluent list should be maximally unexpected and that there were both mixed font and single font conditions. That is, we manipulated orthogonally the font type of lists 1, 2, 3 and 5 (fluent or disfluent) and the font type of list 4 (fluent or disfluent).

Given the disfluency hypothesis is correct, memory for those lists presented in disfluent font should be better than memory for the fluent lists. This should particularly be the case in the homogeneous conditions, because they most closely resemble Diemand-Yauman et al.'s materials. If the distinctiveness hypothesis is correct, one should expect that list 4 should be remembered better when it is presented in a different font than the rest of the lists. In addition, it might be that only a distinct *and* disfluent list is processed more deeply. This would result in an interaction between disfluency and distinctiveness such that list 4 is remembered more accurately only when it is the single (and first) list in a disfluent font.

Experiment 3

We again manipulated disfluency by presenting lists in either hard-to-read or easy-to-read fonts. To manipulate the factor distinctiveness, we now presented either only the fourth list in a different font or we presented the fourth list in the same font as the other lists. This resulted in four conditions (1) all lists were disfluent, (2) all lists were fluent, (3) all but one (pos. 4) lists were disfluent, and (4) all but one (pos. 4) lists were fluent. The main dependent variable was correctness for list 4; in addition, we analyzed the influence of disfluency on correctness for lists 2 and 3.

Method

Participants

As Experiment 2, this study took place in two introductory psychology courses at the University of Erfurt. 80 undergraduates (47 of them female), aged between 18 and 31 ($M_{age}=20.72$; $SD=2.37$), participated. A quarter of the participants received the lists in each of the four conditions.

Materials and procedure

We used the same lists as in Experiments 1 and 2. However, the distribution of lists to conditions differed (as described above). The fourth list was always the "Punome". The procedure was analogous to Experiment 2 and scoring followed the same procedure as in Experiments 1 and 2.

Results

The results for the fourth list were submitted to a 2×2 between subjects ANOVA, with the factors fluency (fluent or disfluent font) and distinctiveness (same vs. different as preceding

lists). This analysis neither revealed a main effect of disfluency nor a main effect of distinctiveness; the two-way interaction was also not significant, all $F_s < 1$. The mean proportion correct for list 4 is shown in Fig. 4.

A t -test for the second and third lists also revealed no effect of disfluency ($M_{\text{disfluent}} = 37.05$ ($SD = 20.71$) vs. $M_{\text{fluent}} = 39.91$ ($SD = 24.25$); $t(78) = .57$, $p = .573$).

Discussion

Again, this experiment did not reveal any effect of disfluency or distinctiveness. Learning outcomes were not significantly affected by the experimental manipulations, neither by the readability of the fonts nor by their salience. This is particularly surprising, since this experiment included two conditions which very closely resembled Diemand-Yauman et al.'s (2011) first experiment. Even a comparison of these homogeneous disfluent and fluent lists did not result in a disfluency effect, $M_{\text{disfluent}} = 33.45$ ($SD = 19.41$) vs. $M_{\text{fluent}} = 41.84$ ($SD = 23.67$); $t(39) = 1.24$, $p = .224$. Rather, there was a descriptive advantage for the fluent lists.

Another surprising aspect of the findings is that distinctiveness did not improve recall accuracy, even though each distinctive list was preceded by three other lists in different font. We do not have a convincing explanation for this, but a difference to typical studies on the benefit of distinctiveness for memory is that an item in our study consisted of a list of seven features whereas the typical item is a single word. The grain size of the stimuli might thus affect their salience.

General discussion

The three studies reported in the present paper were originally motivated by the question whether the better memory performance for lists presented in a disfluent font is due to disfluency (as suggested by Diemand-Yauman et al. 2011) or due to the distinctiveness of

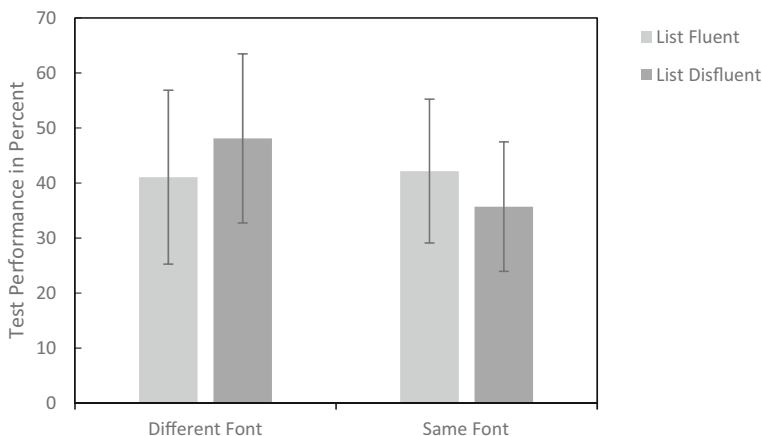


Fig. 4 Proportion of correct answers for list 4 (in percent) as a function of ‘distinctiveness’ (i.e., list 4 was either presented in the same font or in a different font than the other lists) and ‘disfluency’ (i.e., list 4 was either presented in a disfluent font or in a fluent font) in Experiment 3. The error bars depict confidence intervals

the hard-to-read font. We did not find any effect of distinctiveness. This is surprising since distinctiveness effects are known as very robust phenomena (Hunt and Worthen 2006). The most plausible explanation is that our operationalization of distinctiveness was not effective. We presented each list on a separate screen/page. Therefore, the differences between the fonts were less obvious than had all lists been presented on one screen/page. Potentially, the distinctiveness effect would show up when all lists were presented at once. More importantly, we did not observe an advantage of hard-to-read fonts over easy-to-read fonts in any of our three studies. This is particularly surprising since we used lists that were very similar to the original study materials by Diemand-Yauman et al. (2011, Exp. 1). Thus, our findings question the generality and/or the robustness of the disfluency effect in learning. In other words, the fact that some studies find a disfluency effect and others do not may be due to the influence of (so far unknown) moderators. Alternatively – or additionally, the disfluency effect in learning may be smaller than expected, because of which it is sometimes detected and sometimes not.

The current debate in the literature focuses on potential moderators (Alter 2013; Eitel et al. 2014; Köhl et al. 2014; Oppenheimer and Alter 2014). Among the moderators that have been suggested are learner characteristics (e.g., academic abilities, spatial abilities, prior knowledge, motivation), characteristics of the task and the materials (e.g., the degree to which the font is disfluent, the distinctiveness of the disfluent materials, the degree to which reasoning is necessary and possible, whether study or test time is restricted or not), and other situational influences (e.g., whether the experiment was run early or late in the semester). Here, two questions are relevant: (1) Which of these moderators can possibly account for the discrepancy between Diemand-Yauman et al.'s (2011, Exp. 1) data and ours? (2) Which moderators can account for the overall pattern of data in the literature on the disfluency effect?

With respect to the discrepancy between Diemand-Yauman et al.'s (2011, Exp. 1) findings and ours, a number of factors can easily be discarded: (1) prior knowledge – all of these experiments used materials that were completely unknown to the participants, (2) academic abilities – participants were university students in all studies, (3) the degree of disfluency – the fonts and the font sizes we used were also used in the original study, (4) the time points of the experiments – our studies were conducted at different time points in the semester (Experiment 1 in the beginning and Experiments 2 and 3 towards the end) and never yielded a disfluency effect, and (5) factors related to materials, task, and procedure – all these were very similar between the original study and ours. However, there are some minor differences between Diemand-Yauman et al.'s original first experiment and our studies. Whereas their participants had to memorize only 21 features (of three aliens), our participants had to memorize 35 features (of five aliens). In addition, the testing conditions differed: Diemand-Yauman et al. (2011, Exp. 1) only asked seven questions, whereas we asked 21. Test performance differed between the original study and our replications (80 % correct answers vs. 40 % correct answers, respectively). Therefore, a possible candidate for a moderator in the learning domain might be task difficulty.² This idea is also theoretically plausible: Nelson and Narens' (1990) framework suggests that judgements of task difficulty affect the type of processing. Learning tasks that are perceived as difficult usually afford system 2 processing, whereas for simple tasks system 1 processing is sufficient. Thus, disfluency should particularly affect processing in tasks that are perceived as rather simple.

If this were the case, learning studies which observed a disfluency effect should differ systematically from studies which did not find an effect with respect to the overall level of

² We would like to thank an anonymous reviewer for this suggestion.

performance: Studies observing a disfluency effect should show high test performance, studies which did not observe a disfluency effect should show low rates. In fact, two (of three) studies which report a disfluency effect with respect to learning (i.e., Diemand-Yauman et al. 2011, Exp. 1; French et al. 2013) observed test rates of 80 % and more, whereas all studies which did not find an advantage of disfluent over fluent learning materials showed test rates of 50 % or even lower. However, there are also data which do not match the idea of difficulty as a boundary condition: Sungkhasettee et al. (2011), who observed a disfluency effect, report test rates lower than 40 %. Altogether, a systematic investigation of the difficulty assumption may be worth the effort.³

A second difference between our studies and Diemand-Yauman et al.'s experiment is that we never tested the first and the last lists. In contrast, the first and the last lists were included in the analysis by Diemand-Yauman et al. (2011) – in fact, they constituted two thirds of the lists. The disfluency effect in Diemand-Yauman et al.'s study might thus go back to better learning for the first list that was encountered in an unfamiliar font, which attracts attention because of its unfamiliarity. However, this explanation can be excluded based on our third experiment. Here, the distinctive list was always the first one in that particular font. If any of these ad-hoc explanations for a lack of a disfluency effect in our experiments constitutes an actual boundary condition, this suggests that the disfluency effect is not very robust and occurs only under very specific conditions.

A third difference between Diemand-Yauman et al.'s and our study might concern the motivational state of the participants. Neither study included a measure of motivation, so that we can only speculate about this issue. A factor that may affect participants' motivation is whether the experiment is conducted in isolation or whether multiple experiments are conducted within one session. We speculate that the latter procedure is not seldom applied. Since, unfortunately, this is usually not mentioned in methods sections, its influence cannot be assessed. If Diemand-Yauman et al.'s Experiment 1 was part of a multi-experiment session, previous experiments might have reduced participants' motivation to memorize the alien descriptions. Our experiments, however, took place in single sessions (which we did not report in the methods sections either). Therefore, the motivation of our participants might have been generally higher than that of Diemand-Yauman et al.'s (2011, Exp. 1). Based on the metacognitive framework outlined in the introduction, it is plausible to assume that higher motivation favors system 2 processing per se, so that no additional effect of disfluency is expectable.

Unfortunately, the picture gets even less clear and more inconsistent when looking at the literature on disfluency effects more generally. A survey of studies investigating the disfluency effect in different domains by Alter (2013, p. 438) includes a figure depicting the increase of performance when participants experienced disfluent rather than fluent materials. Interestingly, Diemand-Yauman et al.'s (2011) first experiment showed the smallest disfluency benefit (16 %) of all studies included. Even though this is only an observation, it perfectly matches the title of a recent article by Alter et al. (2013), which states that “disfluency prompts analytic thinking – but not always greater accuracy”. Following this interpretation, disfluency effects could be higher when more reasoning is necessary (and helpful) to comprehend the materials and to fulfill the task. This is also in line with the fact that, originally, the disfluency effect was a phenomenon in the reasoning literature (e.g., Alter and Oppenheimer 2008, 2009; Alter et al. 2007). Based on these considerations, one might assume that disfluency causes only small

³ A first attempt to test this hypothesis was undertaken by Eckart et al. (2015), who varied the number of the to-be-remembered aliens in the learning phase and the number of questions asked in the final test. In none of the conditions, a disfluency effect was found.

effects when analytic thinking does not play a major part. For the memory studies conducted here this means that disfluency benefits should be rather small and fragile.

However, in a recent study, Meyer et al. (2015) pooled the data of 16 attempts (three of which had been published earlier by Thompson et al. 2013) to replicate the findings by Alter et al. (2007) on reasoning. Overall, there was no effect of disfluency (pooled $d=-0.01$). “Indeed, of the 17 experiments, only the study reported by Alter et al. (2007) finds significantly higher scores in the disfluent font condition” (Meyer et al. 2015, p. e17). The authors conclude that there is little or no evidence that disfluency affects analytic reasoning. However, they make a brief reference to disfluency effects in the learning domain that is in stark contrast to our speculation: “several research groups have found that disfluent fonts improve performance on memory tasks (Cotton et al. 2014; Diemand-Yauman et al. 2011; French et al. 2013; Lee 2013; Sungkhasettee et al. 2011; Weltman and Eakin 2014). Though some have also failed to replicate these effects (Eitel et al. 2014; Yue et al. 2013), the balance of evidence suggests that disfluent fonts may aid memory but not reasoning—presumably because reading words more slowly benefits memory, but not reasoning” (Meyer et al. 2015, e20). Meyer et al.’s (2015) and our study are two instances of failed attempts to replicate influential studies reporting disfluency effects, one in the reasoning domain and one in the learning domain. The devil’s advocate might conclude that whenever you look superficially, the disfluency effect seems rather robust, and whenever you scrutinize, it disappears.

However, from an applied perspective, there is one important shortcoming in our study. The fact that we could not replicate the findings of Diemand-Yauman et al.’s first experiment does not touch Diemand-Yauman et al.’s second study, which investigated the use of disfluent materials in classroom settings. Based on our findings, we question the basic mechanism which is assumed to be responsible for the disfluency effect (and which we described in the introduction section of the article). However, it is a different question (which needs further investigation) whether the findings of Diemand-Yauman et al.’s field experiment are replicable or not.

Based on the inconsistency of disfluency effects in learning contexts, we do not share the clearness with which at least some researchers recommend the application of disfluency manipulations in instructional contexts. A general recommendation for presenting learning materials in hard-to-read fonts as put forward by Diemand-Yauman et al. (2011) or French et al. (2013) is thus not tenable in our view.

From a theoretical point of view, research on disfluency and learning should emphasize the boundary conditions of the disfluency effect. It might be promising to identify the specific subdomains in which there are robust benefits of disfluency and to investigate what this implies for learning. From a more applied perspective, it is necessary to run studies which reinvestigate the disfluency effect in real classroom settings.

References

- Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science*, 22(6), 473–442.
- Alter, A. L., & Oppenheimer, D. M. (2008). Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle). *Psychological Science*, 19, 161–167.
- 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.
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136(4), 569–576.
- Alter, A. L., Oppenheimer, D. M., & Epley, N. (2013). Disfluency prompts analytic thinking - but not always greater accuracy: response to Thompson et al. (2013). *Cognition*, 128, 252–255.

- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.
- Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*, 1350–1356.
- Cotton, D., Joseph, E., Lede, M., & Ronan, D. (2014). The effect of font structure on memory and reading. Time Poster presented at the biannual evening of psychological science, University of Connecticut.
- Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology*, *104*(3), 268–294.
- 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*, 114–118.
- Dunlosky, J., Serra, M., & Baker, J. M. C. (2007). Metamemory applied. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky, & T. J. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 137–59). New York: Wiley.
- Eckart, H., Predatsch, Ch., Rauscher, V., Gregory, S., Nikolaev, B. & Rummer, R. (2015). Hard to read fonts do not improve memory (not even with very simple learning materials). Unpublished Manuscript, University of Erfurt.
- Eitel, A., Kühn, 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.
- French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, 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*, 301–304.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Hunt, R. R., & Worthen, J. B. (Eds.). (2006). *Distinctiveness and memory*. New York: Oxford University Press.
- Jenkins, W. O., & Postman, L. (1948). Isolation and spread of effect in serial learning. *American Journal of Psychology*, *61*, 214–221.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane and Farrar, Straus and Giroux.
- 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*, 805–806.
- Lee, M. H. (2013). Effects of disfluent kanji fonts on reading retention with e-book. Paper presented at the 2013 I.E. 13th International Conference on Advanced Learning Technologies (ICALT; pp. 481– 482). Beijing, China: IEEE.
- 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.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). New York: Academic.
- Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation*. New York: Plenum Press.
- Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, *28*(4), 502–504.
- Peirce, J. W. (2007). PsychoPy - psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.
- Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review*, *18*, 973–978.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185–233.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*, 237–251.
- Weltman, D., & Eakin, M. (2014). Incorporating unusual fonts and planned mistakes in study materials to increase business student focus and retention. *INFORMS Transactions on Education*, *15*, 156–165.
- 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.