Would disfluency by any other name still be disfluent? Examining the disfluency effect with cursive handwriting

Jason Geller¹ · Mary L. Still² · Veronica J. Dark³ · Shana K. Carpenter³

Published online: 18 June 2018
© Psychonomic Society, Inc. 2018

Abstract

When exposed to words presented under perceptually disfluent conditions (e.g., words written in Haettenschweiler font), participants have difficulty initially recognizing the words. Those same words, though, may be better remembered later than words presented in standard type font. This counterintuitive finding is referred to as the disfluency effect. Evidence for this disfluency effect, however, has been mixed, suggesting possible moderating factors. Using a recognition memory task, level of disfluency was examined as a moderating factor across three experiments using a novel cursive manipulation that varied on degree of legibility (easy-to-read cursive vs. hard-to-read cursive). In addition, list type and retention interval between study and test were manipulated. Across all three experiments, cursive words engendered better memory than type-print words. This memory effect persisted across varied list designs (blocked vs. mixed) and a longer (24-hour) retention interval. A small-scale meta-analysis across the three experiments suggested that the cursive disfluency effect is moderated by level of disfluency: easy-to-read cursive words tended to be better remembered than hard-to-read cursive words. Taken together, these results challenge extant accounts of the disfluency effect. The theoretical and practical implications of these findings are discussed.

Keywords Word recognition · Recognition memory · Disfluency · Cursive

An important insight in the domain of learning and memory is the concept of *desirable difficulties* (Bjork, 1994; McDaniel, & Butler, 2011). Desirable difficulties are select manipulations or learning conditions that make the initial learning process more difficult, but have the added effect of improving long-term retention. Notable examples of desirable difficulties include having participants generate information from word fragments instead of passively reading intact words (e.g., Slamecka & Graf, 1978), spacing out study sessions instead of massing them (e.g., Carpenter, 2017), and having participants engage in retrieval practice after studying instead of simply restudying the information (Kornell & Vaughn, 2016).

Desirable difficulties can also arise from the perceptual characteristics of the stimulus. It has been shown that providing

individuals with words or text in a way that makes the text harder to read can lead to better long-term memory for that material (e.g., Diemand-Yauman, Oppenheimer, & Vaughan, 2011). The benefit associated with processing difficulty brought forth by changing perceptual characteristics of stimuli has been coined the *disfluency effect*. The disfluency effect is quite surprising as many traditional models of memory posit that encoding operates within a limited capacity channel (e.g., Atkinson & Shiffrin, 1968; Baddeley, 1981). In these models, increased demands on working memory would increase cognitive load, and therefore should be more likely to hurt rather than to help memory, especially if working memory capacity is low.

Nevertheless, the disfluency effect has been shown across a variety of perceptual disfluency manipulations. In a highly publicized paper with more than 200 citations to date, Diemand-Yauman et al. (2011) examined the effects of perceptual disfluency in both the laboratory and classroom using atypical fonts. In Experiment 1, participants studied a list of characteristics associated with space aliens in either a disfluent font (Comic Sans MS and Bodoni MT) or a more common, fluent typescript (Arial). At test, participants showed better memory on a cued-recall task for characteristics presented in a disfluent typescript compared with a fluent typescript (see French et al.,

---

✉ Jason Geller
jgeller1@uab.edu

¹ University of Alabama–Birmingham, 415 Campbell Hall 1530 3rd Avenue South, Birmingham, AL 35294-1170, USA

² Old Dominion University, Norfolk, VA, USA

³ Iowa State University, Ames, IA, USA

2013; Weltman & Eakin, 2014, for replications of this effect). In Experiment 2, Diemand-Yauman et al. (2011) manipulated the typescript (Monotype Corsiva, *Comic Sans Italicized*, and Haettenschweiler) of PowerPoint slides and handouts for a single unit in advanced placement (AP) English, history, chemistry, or physics classes. Similar to Experiment 1, the disfluent typescripts resulted in better memory than the fluent typescripts on a unit exam. Several other perceptual disfluency manipulations have shown a similar benefit: masking (i.e., presenting stimuli briefly, approximately 100 ms, and backward masking them, typically with a row of hash marks; e.g., Hirshman & Mulligan, 1991; Hirshman, Trembath, & Mulligan, 1994; Mulligan, 1996; Nairne, 1988), word inversion (Sungkhasettee, Friedman, & Castel, 2011), and high-level blurring (Rosner, Milliken, & Davis, 2015).

Theoretical mechanisms of the disfluency effect

Two explanations have been put forth to account for the memory benefit that arises from perceptual disfluency. According to a postlexical account, after the word has been recognized, the experience of disfluency functions as a subjective metacognitive cue to engage in deeper or more effortful processing of the word, and it is this processing that leads to enhanced memory (e.g., Alter, 2013; Alter, Oppenheimer, Epley, & Eyre, 2007; Diemand-Yauman et al., 2011).¹ This postlexical interpretation of the disfluency effect is related to the traditional levels-of-processing framework (Craik & Lockhart, 1972) or the more contemporary System 1 and System 2 dual-route model (Evans, 2016; Kahneman, 2011). Within these frameworks, words that are hard-to-read elicit deeper processing, or involve more System 2 processing, than easy-to-read words. It is this deeper processing that occurs postlexically, after a word has been identified, that drives the disfluency effect (Alter et al., 2007).

Alternatively, the compensatory processing account (Hirshman et al., 1994; Mulligan, 1996), which is based off the highly influential interactive-activation framework of word recognition (McClelland & Rumelhart, 1981), ties the disfluency effect directly to processes involved in word recognition itself (see Yap & Balota, 2015, for a review on visual word recognition). Within this account, attempting to identify a perceptually disfluent word evokes increased higher-level (i.e., lexical/semantic) processing to aid in word recognition. It is the increased interaction between lower and higher levels during word processing that enhances memory for disfluent stimuli. Although the accounts differ in the point at which it occurs, in both cases, some type of additional processing is

thought to be responsible for engendering better memory for perceptually disfluent items.

Failed replications of the disfluency effect

Although the aforementioned suggest that disfluency effect is a veridical phenomenon, it is important to note there have been several failed demonstrations of the disfluency effect. For instance, in simple memory paradigms, perceptual disfluency manipulations such as small font size (Rhodes & Castel, 2008), low color contrast (Hirshman et al., 1994), low-level blurring (Yue, Castel, & Bjork, 2013), atypical font style (Magreehan, Serra, Schwartz, & Narciss, 2016), and hard-to-hear auditory information (Rhodes & Castel, 2009) have all failed to enhance memory retention. These failures have also been observed with more complex study materials, such as learning about the mechanisms of a toilet flush (e.g., Eitel, Kuhl, Scheiter, & Gerjets, 2014) or how a plane achieves lift off (Strukelj, Scheiter, Nyström, & Holmqvist, 2015).

When a finding fails to replicate, it is important to examine why. The fact that perceptual disfluency does not always promote positive learning outcomes raises the question as to whether disfluency is a desirable difficulty. Given the mixed evidence, it is important to systematically examine the conditions producing and not producing the disfluency effect, as this may have important theoretical implications (Oppenheimer & Alter, 2014).

The discrepant findings may reflect the precise way in which disfluency manipulations affect word processing. For instance, during word recognition, manipulations can affect lower (e.g., featural and orthographic) and higher (e.g., lexical and semantic) levels of processing. Two perceptual manipulations that have shown the disfluency effect, masking (Hirshman et al., 1994; Mulligan, 1996) and word inversion (Sungkhasettee et al., 2011), seem to evoke increased top-down, higher-level (lexical) processing during word recognition. By comparison, two lower-level perceptual manipulations failed to demonstrate the disfluency effect. Specifically, Hirshman et al. (1994, Experiment 5) showed that a low luminance manipulation (i.e., words presented in dark-gray vs. white font on a black background) resulted in processing costs, denoted by longer naming latencies, but did not confer a memory benefit. Similarly, Yue et al. (2013) used a low-level blurring manipulation across several studies and did not find a disfluency effect. Critically, luminance (Hirshman et al., 1994) and low-level blurring (Yue et al., 2013) have been shown to exert an additive influence on word recognition at lower levels of processing, but not necessarily higher levels (e.g., Gomez & Perea, 2014; Reingold & Rayner, 2006; Sanchez & Jaeger, 2015; Yap & Balota, 2007). It is possible, then, that the perceptual disfluency effect occurs only when the perceptual manipulation leads to increased higher-level processing.

¹ This is commonly called the metacognitive account of disfluency, but herein we refer to this as a postlexical account because the disfluency effect is thought to arise after the word has been recognized.

The suggestion is that, given the putative theoretical mechanism(s) of the disfluency effect, one potential moderating factor of the disfluency effect is the nature of the disruption produced by the disfluency manipulation. It may be that a manipulation affecting higher levels of processing (i.e., lexical or semantic) is required to elicit the disfluency effect. For word stimuli, the perceptual disfluency manipulation must elicit increased feedback from lexical or semantic levels either during (e.g., Mulligan, 1996) or after (e.g., Alter et al., 2007) word recognition. If the disfluency effect reflects additional higher-level processing, it might explain why some designs (particularly those that rely on lower-level manipulations of disfluency) consistently fail to find the effect. This possibility is examined in the present studies by manipulating the format in which words are presented—clear type print, easy-to-read cursive, and hard-to-read cursive.

Educational realism of cursive text

It has been argued (see Sierra, 2016) that fluency manipulations used in the laboratory rarely occur in a naturalistic setting such as the classroom, making them educationally unrealistic (but see Toftness et al., 2017). One solution to this problem is to use perceptual manipulations that are more likely to occur in an educational setting. One novel perceptual disfluency manipulation that is more educationally realistic is handwritten cursive. In a classroom setting it is quite common for instructors to present students with cursive written information, either on the chalkboard/whiteboard or on the projector. Furthermore, it has been shown that students benefit more from writing and studying handwritten notes (Mueller & Oppenheimer, 2014).

So what makes cursive disfluent? Handwritten cursive differs in two key ways from type-print: it is more ambiguous and it is nonsegmented. In cursive writing, the same nominal characters change their physical forms across contexts; very similar forms may even signal different intended characters in different contexts. When individuals write in cursive, their letters connect, creating potential problems in letter segmentation. These perceptual features create considerable difficulty in recognizing handwritten words. This is evinced by the fact that there is additional recruitment of frontal lobe areas responsible for attention when attending to handwritten stimuli (Qiao et al., 2010), as well as increased response latencies to handwritten stimuli compared with type-printed stimuli (Barnhart & Goldinger, 2010, 2015; De Zuniga, Humphreys, & Evett, 1991; Perea, Gil-López, Beléndez, & Carreiras, 2016). Given that cursive is both educationally realistic and naturally disfluent, it seems like an ideal ecologically valid manipulation to examine the disfluency effect. The question from a theoretical perspective then becomes, at what level does the disfluency produced by cursive writing occur?

The comparative difficulty associated with recognizing a word in cursive relative to type-print is thought to reflect

increased higher-level processing (Barnhart & Goldinger, 2010). Indeed, De Zuniga et al. (1991, Experiments 3 and 4) and Barnhart and Goldinger (2010, 2015) provided evidence for increased higher-level contributions by showing that certain lexical and semantic effects (i.e., word frequency, imageability, regularity, bidirectional consistency, and orthographic neighborhood size) are magnified during the recognition of handwritten cursive words compared with recognition of type-printed words. This suggests a strong top-down component to recognizing cursive words.

Perea, Gil-López, et al. (2016) along with Perea, Marcet, Uixera, and Vergara-Martínez (2016) further suggested that the recognition of handwritten words is moderated by legibility, with hard-to-read cursive recruiting more top-down processing during word recognition than easy-to-read cursive. Examining RTs using quantile analyses, Perea, Gil-López, et al. observed not only shifts in the RT distribution, but also skewing (i.e., higher RTs and more errors) in the cursive conditions. However, hard-to-read cursive affected higher-level processing more—the word frequency effect was larger for hard-to-read cursive words compared with easy-to-read cursive words. These legibility differences have also been observed during normal reading. Using eye tracking, Perea, Marcet et al. (2016b) found a substantial reading cost (i.e., more fixations, longer fixations, and longer reading times) for sentences written in cursive compared with type-print. However, at the local (word) level, hard-to-read cursive words (compared with easy-to-read cursive) more strongly affected late measures of eye movements—such as total reading time and go past times—that are presumed to reflect lexical and postlexical processes.

The purpose of the current study was to examine how the processing of perceptually disfluent stimuli affects later memory using an educationally realistic fluency manipulation: cursive handwriting. In the disfluency literature, it is quite common for only one level of disfluency to be used. It is possible, then, that examining multiple levels of disfluency might present a more nuanced view of the disfluency effect. Indeed, Diemand-Yauman et al. (2011) posited that differing levels of disfluency might follow a U-shaped curve insofar that illegible or hard-to-read stimuli might lead to a memory disadvantage, while more moderate disfluency would be the best for learning. Toward this aim, Experiment 1 examined whether a disfluency effect can be obtained for cursive writing, and whether memory outcomes differ between different levels of cursive legibility (i.e., easy-to-read vs. hard-to-read). Experiments 2 and 3 replicated and extended the findings from Experiment 1.

Experiment 1

In Experiment 1, three script types (easy-to-read cursive, hard-to-read cursive, and type-print) were used to examine whether cursive produces a disfluency effect. At study, we included

two objective measures of disfluency (accuracy and naming latency) to serve as a manipulation check to ensure that the cursive manipulation was disfluent. We also included a metacognitive measure of disfluency (aggregate judgments of learning [JOLs]). Aggregate JOLs required participants to give a subjective estimate of the strength of their learning associated with each script type immediately after the study phase. The inclusion of aggregate JOLs also allowed us to examine the interplay between subjective disfluency and control and regulatory processes leading to better memory (Pieger, Mengelkamp, & Bannert, 2016). Moreover, aggregate JOLs do not interfere with processing that occurs during encoding (Besken & Mulligan, 2013, 2014).

At test, an old-new recognition task assessed memory differences across the three script types.

An old-new recognition test was used as the explicit memory task because it appears to be more sensitive to the effects of disfluency (e.g., Hirshman & Mulligan, 1991; Mulligan, 1996, 1998; Rosner et al., 2015), thus increasing the odds of detecting a disfluency effect, if it exists.

While extant accounts of the disfluency effect would predict an overall disfluency effect for cursive words, they would predict a different pattern of results for easy-to-read and hard-to-read cursive words. Within the compensatory processing account (Mulligan, 1996), the disfluency effect reflects increased lexical and/or semantic processing. The memory benefit for processing hard-to-read cursive words would be larger than for easy-to-read cursive words, because hard-to-read cursive not only slows down reading, it also elicits increased interactivity between lower and higher levels during word recognition (Perea, Gil-López, et al., 2016; Perea, Marcet, et al., 2016). In contrast, it is less clear what postlexical accounts would predict. On the surface, it seems that subjectively, hard-to-read cursive should be more disfluent and therefore result in better memory. But as hypothesized by Diemand-Yauman et al. (2011) “it seems possible that the influence of disfluency on retention follows a U-shaped curve, and the exact parameters of this function remain to be determined. With this in mind, the most effective disfluency manipulations would likely be those that are within the bounds of the normal variation of fonts and materials that could reasonably appear in a classroom” (p. 4). From this perspective, it could be predicted that easy-to-read cursive would produce better memory than hard-to-read cursive.

Method

Participants

Thirty undergraduate students from Iowa State University participated for course credit. All were native speakers of English, with self-reported normal or corrected-to-normal vision.

Stimuli and apparatus

Stimuli were 198 nouns drawn from the English Lexicon Project database (Balota et al., 2007). Both frequency (all words were high frequency; mean log HAL frequency = 9.2) and length (all words were four letters in length) were controlled (see <https://osf.io/td5h7/> for the stimulus list).

To create the cursive stimuli, we asked seven individuals to write five sentences in cursive to be judged for penmanship. Forty students rated the level of penmanship of the seven individuals on a 1–7 Likert scale (1 = legible; 7 = illegible). The individuals with the best and worst penmanship (average score of 2.2 vs. 4.0, respectively) were chosen to write the stimuli for the experiment (see Perea, Gil-López, et al., 2016a, for a similar procedure).

The hard-to-read and easy-to-read cursive stimuli were created using a Livescribe digital pen. The writing instrument resembles a normal pen, with the addition of a small camera protruding from the tip. The camera reads a fine dot pattern on special paper, generating a digital trace of each pen stroke. The stimulus images were matched in size as closely as possible to the computer-generated words (44-point Courier New font). Figure 1 depicts an example of each of the three script types. All stimuli used for this experiment can be found on the Open Science Framework (OSF; <https://osf.io/td5h7/>).

Stimuli were presented via E-Prime experimental software (Version 1.1; Schneider, Eschman, & Zuccolotto, 2002) on a Dell computer with an LCD monitor. A microphone attached to the E-Prime SR response box recorded participants' responses and naming latencies. Participants used a keyboard to record their JOLs and the SR response box to record their old/new recognition responses.

Design and procedure

Type of script (type-print, easy-to-read cursive, and hard-to-read cursive) was manipulated within-subjects and aggregate JOLs were collected under incidental learning conditions. Participants were tested individually in a small, well-lit room, seated approximately 65 cm from the computer screen. Before the experiment, all participants were informed that they would be naming words and that some of the stimuli would appear as hard-to-read (perceptually disfluent) words and some as easy-to-read (perceptually fluent) words. Instructions made no mention of a final recognition memory test. Participants engaged in four practice trials to familiarize themselves with the procedure before the experiment proper.

A total of 198 words were presented, 99 at study (33 in each script condition) and 198 at test (99 old and 99 new). This resulted in six counterbalanced lists. Lists were assigned to participants so that across participants each word occurred equally often in the six possible conditions: type-print old,

Fig. 1 Examples of type-print (left), easy-to-read cursive (center), and hard-to-read cursive (right) used across experiments

easy-to-read cursive old, hard-to-read cursive old, type-print new, easy-to-read cursive new, and hard-to-read cursive new.

Each trial began with a fixation cross appearing at the center of the screen for 1,000 ms. The fixation cross was then replaced by a word in the same location for 2,000 ms. To ensure that all items were identified and to assess whether perceptually disfluent items were indeed disfluent, participants were instructed to name each word as quickly and accurately as possible. Naming latency and actual response were recorded on each trial. After the naming response, there was a 1,000 ms blank interstimulus interval before the next trial.

After all 99 words were named, participants completed aggregate JOLs wherein they were told that 33 easy-to-read cursive, 33 hard-to-read cursive, and 33 type-print words had appeared in the list, and were asked to estimate how many words from each condition they expected to recognize on a later memory test. The order of the memory judgments (easy-to-read cursive vs. hard-to-read cursive vs type-print) was counterbalanced across participants.

After the study phase, a short 3-minute distractor task was administered in which participants wrote down as many United States capitals as they could. Afterward, participants took an old-new recognition test. At test, a fixation cross appeared in the center of the screen for 1,000 ms and was followed by a word that either had been presented during study (“old”) or had not been presented during study (“new”). Old words occurred in their original script, and following the counterbalancing procedure, each new word was presented in one of the three scripts. For each word presented, participants used the SR button box to input their responses, using a button labeled “old” to indicate that they had named the word during study, and a button labeled “new” to indicate they did not remember naming the word. Words stayed on the screen until participants gave an “old” or “new” response. All words were individually randomized for each participant during both the study and test phases. After the experiment, participants were debriefed. The entire experiment took about 30 minutes to complete.

Analysis strategy

The data for all experiments reported here are located on OSF (<https://osf.io/bstfd/>). For all experiments, an alpha level of .05 was used for significance testing. Cohen’s *d*, eta-squared (η^2), and partial eta-squared (η_p^2) are reported as effect-size measures. In addition, alongside traditional analyses that utilize null hypothesis significance testing, Bayes factors, calculated

with the freeware software program JASP (Version 0.8.6; <https://jasp-stats.org>) for null findings (noted as BF_{01}) are reported (see Jarosz & Wiley, 2014, for a review on Bayes factors). A Bayes factors of 3 or greater is indicative of strong or positive evidence in favor of the null (Jefferys, 1961).

Results and discussion

In each experiment, items with more than 40% error rates across participants were not included in any analyses. Any participant who identified fewer than 75% of the items during study was replaced. Trials associated with microphone malfunctions (trials where no audio was detected) were excluded from the latency analysis. Trials associated with naming latencies faster than 150 ms or slower than 2.5 times the standard deviation for each participant were excluded from the latency analysis.

In Experiment 1, no participants were replaced. Two items (*vise* and *liar*) were removed from all analyses (4% of the data) due to high error rates. Microphone malfunctions occurred on 8% of trials. For all analyses, the main effect of script type was further divided into two orthogonal contrasts: One assessed whether cursive produces a disfluency effect and the other whether the memory benefit was larger for easy-to-read or hard-to-read cursive.

Study phase

Naming accuracy and latency Mean naming accuracy and latency are shown in Table 1. Examining accuracy, a one-way repeated-measures ANOVA indicated an effect of script, $F(2, 52) = 56.35$, $p < .001$, $\eta^2 = .66$. Participants were less accurate when naming cursive words compared with type-print words, $F(1, 29) = 74.76$, $p < .001$, $\eta^2 = .72$. Further, participants were less accurate naming hard-to-read cursive compared with easy-to-read cursive words, $F(1, 29) = 42.52$, $p < .001$, $\eta^2 = .60$.

Table 1. Mean naming accuracy (in proportions), naming latency (in milliseconds), and JOLs (in proportions) as a function of script type in Experiment 1

Script	Naming accuracy	Naming latency	JOLs
Type-print	.99 (0.00)	588 (17)	.76 (0.05)
Easy-to-read	.94 (0.01)	739 (26)	.73 (0.04)
Hard-to-read	.85 (0.02)	825 (30)	.54 (0.04)

Note. Standard errors are shown in parentheses

A similar analysis of naming latency indicated an effect of script, $F(2, 52) = 72.26, p < .001, \eta^2 = .71$. Participants named cursive words more slowly than type-print words, $F(1, 29) = 107.92, p < .001, \eta^2 = .79$. Further, hard-to-read cursive words were named more slowly than easy-to-read cursive words were, $F(1, 29) = 22.41, p < .001, \eta^2 = .44$. Together, the naming accuracy and latency results confirm that the cursive stimuli were perceptually disfluent, with hard-to-read cursive being most disfluent.

JOLs Two participants were excluded from the JOL analyses because they did not provide JOLs for each script type. Results indicated a significant effect of script, $F(2, 54) = 16.07, p < .001, \eta^2 = .37$. Participants predicted that they would recognize fewer cursive words than type-print words, $F(1, 27) = 7.45, p = .011, \eta^2 = .22$. Further, participants predicted that they would recognize fewer hard-to-read cursive words than easy-to-read cursive words, $F(1, 27) = 39.42, p < .001, \eta^2 = .59$.

Test phase

Conditionalized analysis Recognition memory was calculated only for words that were correctly named during encoding. Full ANOVA results for the unconditionalized d' analysis across each experiment is presented in Appendix A. It is important to note that the unconditionalized analysis closely matched our conditionalized analysis. The mean proportion of “old” responses across item type and script type are presented in Fig. 2. Memory sensitivity (d') values are displayed in Fig. 3 as a function of script type. A one-way repeated-measures ANOVA on d' values indicated a

significant effect of script type, $F(2, 58) = 5.36, p = .007, \eta^2 = .16$. Performance was better for cursive compared to type-print words, $F(1, 29) = 7.86, p = .009, \eta^2 = .21$. There was no significant difference between easy-to-read cursive and hard-to-read cursive words, $F(1, 29) = 2.15, p = .153, \eta^2 = .07$, $BF_{01} = 1.96$, however, the Bayes factor indicates only weak support for the null hypothesis.

Experiment 1 examined whether cursive script, a type of educationally relevant perceptual disfluency, would produce a disfluency effect. The results were clear: cursive words were better remembered than type-print words, indicating that cursive script serves as a desirable difficulty. There was less clarity when considering theoretical explanations of the disfluency effect because there was no recognition memory difference between easy-to-read and hard-to-read cursive. This is problematic for both the postlexical and compensatory processing accounts of disfluency. While hard-to-read cursive was objectively and subjectively more disfluent, the additional disfluency did not produce better recognition memory. The compensatory processing account would predict that hard-to-read cursive words should produce better memory, while the postlexical account would predict better memory for easy-to-read cursive. Experiment 2 examined whether the pattern of results in Experiment 1 can be replicated, and if so, whether factors other than disfluency may be contributing.

Experiment 2

As noted earlier, the disfluency effect is thought to arise as a result of deeper processing given to perceptually disfluent

Fig. 2 Mean proportion of “old” responses for Experiment 1. Error bars reflect the standard error of the mean, corrected for between-subject variability (Morey, 2008)

Fig. 3 Memory sensitivity (d') as a function of script. Error bars reflect the within-subject standard error of the mean (Morey, 2008)

items. However, it has previously been noted that disfluency is confounded with distinctiveness (e.g., Diemand-Yauman et al., 2011; Rummer et al., 2016). That is, not only are cursive words disfluent compared with type-printed words, but they are also unusual. Thus, it could be that the memory benefit for cursive words is due to perceptual distinctiveness and not additional higher-level processing driven by disfluency.

Most studies examining the disfluency effect have used mixed designs in which a single list contains two levels of memoranda—a fluent level and disfluent level (e.g., Hirshman et al., 1994; Mulligan, 1996; Nairne, 1988; Sungkhasettee et al., 2011). In a mixed list, individuals may use a naïve (subjective) theory about what they believe requires more processing, and employ a strategy that involves preferential processing of the more disfluent stimuli over the more fluent stimuli, resulting in better memory for the former. It may be that only when disfluent words occur among fluent words, and thus stand out as distinctive, do they receive the additional processing that produces better memory.

One way to test this is by manipulating the design of the lists. If primary distinctiveness is removed as a cue by presenting fluent versus disfluent items in blocked (as opposed to mixed) fashion, distinctiveness cannot be used to support differential processing, and the disfluency effect should be attenuated. By contrast, if a disfluency effect is still observed when the possibility of using distinctiveness as a cue is minimized by mixing the presentation of items (as in Experiment 1), this would support an account based on deeper processing.²

In Experiment 2, two groups were compared: a mixed-list group and a blocked-list group. The mixed-list group is an exact procedural replication of Experiment 1. We expected

to replicate the disfluency effect for cursive words but were unsure as to whether a memory difference between hard-to-read and easy-to-read cursive would be found. Additionally, if list composition is indeed important, then an interaction should be observed between script (easy-to-read, hard-to-read, and type-print) and list-type (mixed vs. blocked). The presence of an interaction would indicate that visual distinctiveness may contribute to the disfluency effect.

Method

Participants

A one-tailed power analysis was conducted (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007) using the effect size obtained from the comparison between cursive and type-print in Experiment 1 ($d_z = 0.46$), and indicated that 31 participants were needed to detect an effect with 80% power. Given the six counterbalanced lists, 36 participants were assigned to each group and received course credit for participating.

Materials, procedure, and design

The materials were identical to Experiment 1, and the procedure for the mixed-list group was identical. In the blocked-list group the presentation order of items during study was randomized by participant, with the constraint that all type-print, easy-to-read cursive, and hard-to-read cursive were shown in separate blocks. Block order was counterbalanced across participants. Blocks were separated by a short break, in which the participants were told to take a break before proceeding to the next block. Thus, list type (blocked vs. mixed) was manipulated as a between-subjects variable and script (type-print vs. easy-to-read cursive vs. hard-to-read cursive) as a within-subjects variable. During the recognition test for both groups, all items were intermixed randomly for each individual.

² As pointed out by reviewers, blocking would serve to get rid of primarily distinctiveness, but not necessarily secondary distinctiveness (Schmidt, 1991), which is how items deviate from stored representations in long-term memory. It is possible that undergraduate participants may have less experience with cursive in general, so that cursive words retain secondary distinctiveness even in blocked lists.

Results and discussion

No participants were replaced due to low accuracy. Three items (4% of the data; 2% in the blocked group and 2% in the mixed group) had to be discarded due to low accuracy (*noon*, *liar*, and *flux*). Microphone errors occurred on 11% of the trials.

Study phase

Naming accuracy and naming latency Mean naming accuracy and latency are shown in Table 2. These were submitted to 2×3 mixed-factor ANOVAs, with list type (blocked vs. mixed) as a between-subjects variable and script (type-print vs. easy-to-read cursive vs hard-to-read cursive) as a within-subjects variable.

The accuracy analysis (Greenhouse–Geisser corrected) revealed a main effect for script, $F(2, 140) = 179.97, p < .001, \eta_p^2 = .72$. As in Experiment 1, participants were less accurate naming cursive words compared with type-print words, $F(1, 70) = 230.00, p < .001, \eta_p^2 = .77$, and also less accurate naming hard-to-read cursive words compared with easy-to-read cursive words, $F(1, 70) = 147.09, p < .001, \eta_p^2 = .68$. The main effect for list type and the interaction between script and list type were not significant, both $F_s < 1.50$, both $p_s > .10$, and $BF_{01} > 3$.

The latency analysis (Greenhouse–Geisser corrected) revealed a significant effect for script type, $F(2, 140) = 215.16, p < .001, \eta_p^2 = .76$. Cursive words were named more slowly than type-print words, $F(1, 70) = 312.18, p < .001, \eta_p^2 = .82$, and hard-to-read cursive words were named more slowly than easy-to-read cursive words, $F(1, 70) = 77.29, p < .001, \eta_p^2 = .53$. There was no main effect of list type, $F(1, 70) = .541, p = .464, \eta_p^2 = .01, BF_{01} = 4.18$. There was a marginal interaction between list type and script, $F(2, 140) = 2.58, p = 0.079, \eta_p^2 = .04$, with somewhat faster naming of type-print in the blocked list. Together, the naming accuracy and latency confirm that the cursive stimuli were perceptually disfluent, with hard-to-read cursive being most perceptually disfluent.

Table 2. Mean naming accuracy (in proportions), naming latency (in milliseconds), and JOLs (in proportions) as a function of script type and list type in Experiment 2

Script	Naming accuracy	Naming latency	JOLs
Blocked list			
Type-print	.99 (0.00)	595 (17)	.83 (0.04)
Easy-to-read	.94 (0.01)	783 (22)	.75 (0.04)
Hard-to-read	.76 (0.02)	905 (27)	.57 (0.04)
Mixed list			
Type-print	.99 (0.00)	648 (16)	.79 (0.04)
Easy-to-read	.93 (0.01)	801 (24)	.71 (0.04)
Hard-to-read	.79 (0.02)	896 (27)	.52 (0.04)

Note. Standard errors are shown in parentheses

JOLs Mean aggregate JOLs are shown in Table 2. One participant was removed in this analysis for not providing JOLs for each script type. The JOLs analysis revealed a main effect of script, $F(2, 138) = 66.53, p < .001, \eta_p^2 = .49$. As in Experiment 1, participants predicted that they would recognize fewer cursive words compared to type-print words, $F(1, 69) = 68.04, p < .001, \eta^2 = .50$, and that they would recognize fewer hard-to-read cursive compared to easy-to-read cursive words, $F(1, 69) = 65.00, p < .001, \eta^2 = .49$. The main effect of list type and the interaction between list type and script were not significant, both $F_s < 0.913, p_s > .34$, and $BF_{s01} > 2.03$. The model with script type was weakly preferred over a model with list type ($BF = 2.03$), but strongly preferred over a model with the interaction between script and list type ($BF = 21.34$).

Test phase

Conditionalized analysis The mean proportions of “old” responses across item type and script type are presented in Fig. 4. Memory sensitivity (d') values are displayed in Fig. 5. An analysis of d' was conducted using the same 2×3 mixed-factor ANOVA described earlier. The analysis revealed a main effect of script, $F(2, 140) = 4.92, p = .009, \eta_p^2 = .07$. Recognition memory for cursive words was better than for type-print, $F(1, 70) = 7.52, p = .008, \eta^2 = .10$. As in Experiment 1, memory for easy-to-read cursive was numerically higher than for hard-to-read cursive, but the difference was not significant, $F(1, 70) = 1.95, p = .167, \eta_p^2 = .03, BF_{01} = 3.01$. The main effect of list type and the interaction between list type and script were not significant, both $F_s < 0.13, p_s > .72, BF_{01} > 3$. The model with script type was preferred over a model with list type ($BF = 3.10$) and to a model with the interaction between script and list type ($BF = 35.34$).

The results of Experiment 2 showed the same pattern found in Experiment 1 for all measures, and there was no effect of mixed versus blocked lists. Cursive words were more disfluent than type-print words, with hard-to-read cursive being the most disfluent. Cursive stimuli were also better recognized than type-print were. There was no significant difference between memory for easy-to-read and hard-to-read cursive, although we note that, once again, the easy-to-read stimuli showed the numerically highest memory.

The statistically indistinguishable disfluency effect for blocked and mixed lists suggests that perceptual distinctiveness is not driving the cursive disfluency effect. Ruling out an alternative explanation provides indirect support for an explanation in which depth of processing induced by the difficulty of processing at encoding produces the memory benefit. The results are in accord with other studies demonstrating that distinctiveness plays only a small role in the disfluency effect (Hirshman et al., 1994; Sungkhasettee et al., 2011). However, it remains possible that the mnemonic benefit of cursive words is due to encoding surface-level features that aid later

Fig. 4 Mean proportion of “old” responses for Experiment 2. Error bars reflect the within-subject standard error of the mean (Morey, 2008)

recognition memory. In this study, context was reinstated at test (i.e., cursive items named at study were presented in cursive at test). This makes it difficult to completely rule out a memory benefit based on visuospatial characteristics.

To test this in future research, one could present words aurally at test and follow this with a source monitoring test that asks participants to identify the format of the words studied at test (see Yue et al., 2013, Experiment 2b). This

Fig. 5 Memory sensitivity (d') as a function of list type and script in Experiment 2. Error bars reflect the within-subject standard error of the mean (Morey, 2008)

would control for visual-spatial characteristics serving as superior memory cues.

Experiments 1 and 2 showed that cursive words can act as a desirable difficulty for an immediate recognition memory test. In order to be a true desirable difficulty, however, an effect must persist across longer retention intervals. Indeed, some desirable difficulties (e.g., spacing) become more robust after longer delays (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Experiment 3 assessed whether the disfluency effect persists across a longer retention interval of 24 hours.

Experiment 3

A major limitation in the research on perceptual disfluency is the use of short retention intervals (RIs). In simple list-learning paradigms, the disfluency effect is commonly examined without a significant time delay between study and test (~3 min; e.g., Hirshman et al., 1994; Sungkhasettee et al., 2011; Yue et al., 2013). Immediate performance on a test, however, is not an accurate indicator of long-term retention (Bjork, 1994). Thus, it is important to examine the disfluency effect across more educationally realistic RIs.

In an educational context, the time lapse between study and test could be hours, days, or weeks. Other desirable difficulties that promote learning (e.g., testing and spacing) persist and become more robust across longer RIs, spanning weeks (Carpenter, Pashler, Wixted, & Vul, 2008) or months (Carpenter, Pashler, & Cepeda, 2009; Larsen, Butler, & Roediger, 2009). Using more complex materials, a few studies have examined the disfluency effect across more ecologically valid RIs (e.g., Diemand-Yauman et al., 2011; French et al., 2013; Weltman & Eakin, 2014). The longest retention interval was used by Diemand-Yauman et al. (2011) when they tested the effects of disfluency by putting PowerPoint presentations in a hard-to-read font. In their Experiment 2, the final test occurred weeks or months after participants were exposed to the material. However, subsequent research has not been able to replicate Diemand-Yauman et al.'s initial findings (Rummer et al., 2016), leaving open the question of whether the disfluency effect persists over time, or is a short-lived phenomenon. In Experiment 3, we examined whether the cursive disfluency effect can extend to a more educationally realistic 24-hour RI.

Method

Participants

Thirty-six undergraduate students from Iowa State University participated for course credit. All participants were native speakers of English and self-reported normal or corrected-to-normal vision.

Table 3. Mean naming accuracy (in proportions), naming latency (in milliseconds), and JOLs (in proportions) as a function of script type in Experiment 3

Script	Naming accuracy	Naming latency	JOLs
Type-print	.98 (0.01)	625 (15)	.76 (0.04)
Easy-to-read	.92 (0.01)	779 (20)	.73 (0.04)
Hard-to-read	.79 (0.02)	865 (22)	.54 (0.04)

Note. Standard errors are shown in parentheses

Materials, procedure, and design

The materials and procedure were identical to Experiment 1, except that the length of the RI was extended to 24 hours. After the study phase, individuals were told to come back 24 hours later for the second part of the experiment. Participants were not debriefed until the end of the second session.

Results and discussion

Two participants were replaced due to accuracy less than 75%. One word (*noon*) was discarded due to error rates greater than 40%, resulting in the exclusion of 4% of the data. Microphone errors occurred on 7% of the trials.

Study phase

Naming accuracy and naming latency Naming accuracy and latency are shown in Table 3. Examining accuracy, the ANOVA (Greenhouse–Geisser corrected) indicated a main effect of script, $F(2, 70) = 79.19, p < .001, \eta^2 = .69$. Participants were less accurate naming cursive words compared with type-print words, $F(1, 35) = 78.46, p < .001, \eta^2 = .69$, and less accurate naming hard-to-read cursive words than easy-to-read cursive words, $F(1, 35) = 80.13, p < .001, \eta^2 = .70$.

Examining naming latency, the analysis (Greenhouse–Geisser corrected) revealed a significant effect of script, $F(2, 68) = 128.09, p < .001, \eta^2 = .79$. Participants took longer to name cursive words compared with type-print words, $F(1, 34) = 171.18, p < .001, \eta^2 = .83$, and took longer to name hard-to-read cursive words than easy-to-read cursive words, $F(1, 34) = 46.50, p < .001, \eta^2 = .58$.³

JOLs Mean JOLs are shown in Table 3. The ANOVA (Greenhouse–Geisser corrected) indicated a significant effect of script, $F(2, 70) = 27.59, p < .001, \eta^2 = .44$. Participants predicted that they would recognize fewer cursive words than type-printed words, $F(1, 35) = 18.79, p < .001, \eta^2 = .35$, and predicted that they would recognize fewer hard-to-read cursive words compared with easy-to-read cursive words, $F(1, 35) = 37.20, p < .001, \eta^2 = .52$.

³ One participant was removed from the latency analysis due to the microphone not recording latency for a majority of the trials.

Fig. 6 Mean proportion of “old” responses for Experiment 3. Error bars reflect the standard error of the mean, corrected for between-subject variability (Morey, 2008)

Overall, consistent with the results of the previous two experiments, the cursive stimuli used in Experiment 3 were objectively and subjectively disfluent, with hard-to-read cursive being most disfluent.

Test phase

Conditionalized analysis The mean proportion of “old” responses across item type and script type are presented in Fig. 6. Memory sensitivity (d') for each script type is displayed in Fig. 7. As would be expected with a 24-hour delay, accuracy was lower than in previous experiments. The ANOVA on d' values indicated a significant effect of script type, $F(2, 70) = 9.21$, $p < .001$, $\eta^2 = .21$. Participants had better recognition memory for cursive than type-print words, $F(1, 35) = 14.60$, $p < .001$, $\eta^2 = .29$. There was no reliable difference between easy-to-read cursive words and hard-to-read cursive words, $F(1, 35) = 2.51$, $p = .122$, $\eta^2 = .07$, $BF_{01} = 1.79$.

The results from Experiment 3 are straightforward. Even after 24 hours, studying cursive words had a desirable effect on memory. There was not a reliable memory difference between easy-to-read cursive and hard-to-read cursive words. Similar to Experiments 1 and 2, there was a trend for easy-to-read cursive words to be better remembered than hard-to-read cursive words. Overall, it appears the cursive disfluency effect persists over longer delays.

Meta-analysis of Experiments 1–3

In all three experiments, presenting words in cursive for 2 s served as a desirable difficulty. That is, cursive words were better remembered than type-print words were. Furthermore, in all reported experiments, there was a tendency for easy-to-read cursive words to be better remembered than hard-to-read cursive words. In Experiments 1, 2, and 3, while the difference was not significant, the Bayes analyses did not support a strong conclusion that there was no difference. What is interesting is that this pattern is predicted by a postlexical account, but not by the compensatory processing account.

Given the theoretical importance of this comparison, a meta-analysis was done to examine the true nature of this effect.⁴ Because meta-analyses pool evidence from multiple studies, they are thought to be a more powerful indicator of true effects (McShane & Böckenholt, 2017). In light of this and recent recommendations (e.g., Braver, Thoenes, & Rosenthal, 2014; Cumming, 2014; Lakens & Etz, 2017; McShane & Böckenholt), a small-scale meta-analysis in R was conducted using the metafor package (Viechtbauer, 2010).⁵ In addition to examining the effect size between easy-to-read and hard-to-read cursive, an overall meta-analytic effect size is computed for the difference between

⁴ We also conducted a cross-experiment analysis to rule out an item selection effect that would disadvantage the hard-to-read cursive condition. This analysis is reported in Appendix B.

⁵ Cohen's d was calculated using the standardized mean change rate method for dependent samples (Morris, 2008).

Fig. 7 Memory sensitivity (d') as a function of script after 24 hours. Error bars reflect the within-subject standard error of the mean (Morey, 2008)

easy-to-read cursive and type print and hard-to-read cursive and type-print. For each comparison, a random effects model integrated effect sizes across Experiments 1, 2, and 3 ($k = 3$, $N = 144$).

The average effect sizes and the confidence intervals (CIs) for each comparison of interest are reported in Fig. 8. The meta-analysis revealed that the disfluency effect was present with easy-to-read cursive, $Estimate = 0.47$, $SE = 0.10$, $z = 4.48$, $p < .001$, and hard-to-read cursive, $Estimate = 0.28$, $SE = 0.08$, $z = 3.29$, $p < .001$. Further, easy-to-read cursive had a larger effect on memory than did

hard-to-read cursive, $Estimate = 0.20$, $SE = 0.08$, $z = 2.43$, $p < .05$. The magnitude of the effects fluctuated across studies. However, given that they were within the CIs, the fluctuations are likely due to sampling error (Cumming, 2014). Indeed, the heterogeneity of each effect of interest was not statistically significant (see Fig. 8). Overall, the meta-analysis showed a disfluency effect for both easy-to-read and hard-to-read cursive words and further that the difference between easy-to-read and hard-to-read cursive words was small, but nonzero, with easy-to-read cursive words producing better memory.

Fig. 8 Meta-analysis of Experiments 1, 2, and 3. **a** Standardized mean change between easy-to read cursive and type-print. **b** Standardized mean change between hard-to-read cursive and type, print. **c** Standardized mean change between easy-to-read cursive and hard-to-read cursive

General discussion

In the memory literature, an unexpected way to improve retention of material is to make the material more disfluent by changing the perceptual characteristics to make it hard to read. The afforded mnemonic benefit is called the disfluency effect (Diemand-Yauman et al., 2011). Research on perceptual disfluency has yielded equivocal results, so it is unclear whether perceptual disfluency could function as an educationally relevant desirable difficulty. A search for potential boundary conditions of the disfluency effect could clarify the situation (Oppenheimer & Alter, 2014).

The current experiments examined the influence of perceptual disfluency on recognition memory using cursive handwriting, an educationally relevant disfluency manipulation. We examined under what conditions cursive does and does not constitute a desirable difficulty. To this end, level of disfluency (i.e., legibility) was manipulated along with list composition and retention interval. All three experiments examined (1) whether there was an overall mnemonic benefit of cursive and (2) whether the memory benefit varied as a function of legibility of the cursive. Two types of cursive were chosen (easy-to-read and hard-to-read) because while both slow down processing and affect lexical processing, they do so to varying degrees. Hard-to-read cursive words take longer to recognize and are more error prone than easy-to-read cursive words, suggesting that hard-to-read cursive words exert a stronger influence on lexical and semantic processing (Perea, Gil-López, et al., 2016a; Perea, Marcet, et al., 2016b). In all three experiments, presenting words in cursive produced a disfluency effect. This did not differ as a function of list type (Experiment 2) or retention interval (Experiment 3). Corroborating this pattern of results, a small-scale meta-analysis showed a significant disfluency effect for easy-to-read ($d = .46$) and hard-to-read cursive ($d = .24$) words. This finding supports the hypothesis that perceptual disfluency can produce a disfluency effect. In the context of research showing that low-level perceptual disfluency like blurring (Yue et al., 2013) and luminance (Hirshman et al., 1994) does not produce a disfluency effect, we concluded that the perceptual disfluency manipulation must elicit an increase in higher-level processing to engender a disfluency effect. Further, the meta-analysis showed better memory for easy-to-read compared to hard-to-read cursive words. It appears that all perceptually disfluent stimuli are *not* created equal. The discovery that both easy-to-read and hard-to-read cursive can produce better memory, but that the effect for easy-to-read cursive is larger in magnitude, suggests that *how* disfluency is manipulated matters. These results provide important insights into the mechanisms behind the disfluency effect.

Mechanisms of the disfluency effect

Extant accounts of the disfluency effect place the phenomenon at one of two processing loci. According to postlexical accounts (e.g., Alter et al., 2007), the disfluency effect occurs after word recognition has taken place as a result of metacognitive control processes that enhance memory by way of deeper or more effortful processing. Alternatively, the compensatory processing account (e.g., Mulligan, 1996) posits that the disfluency effect arises during word recognition as a result of increased interaction between the letter level and word/semantic level. Both accounts predict that, as was found, naming words in cursive should produce better memory than naming words in type-print. However, they predict different pattern of results in regard to easy-to-read and hard-to-read cursive words. The compensatory processing account would predict a larger memory benefit for hard-to-read cursive compared to easy-to-read cursive due to the fact that hard-to-read cursive engenders greater top-down processing than easy-to-read cursive (Perea, Gil-López, et al., 2016a; Perea, Marcet, et al., 2016b). Contrary to this, in all three experiments and the meta-analysis, there was a memory advantage for easy-to-read cursive over hard-to-read cursive. Postlexical accounts, on the other hand, would posit that the results should follow a U-shaped curve (Diemand-Yauman et al., 2011), which was observed.

While the current results do provide some evidence for the post-lexical account of disfluency, there are some issues within the current instantiation of the theory that do not align well with the present findings. First, within a postlexical framework, misattribution as a result of experiential processing is thought to engender the mnemonic benefit of disfluency (Alter & Oppenheimer, 2009). That is, disfluency experienced during processing is used as a cue to allocate extra resources to the task. In the current experiments, participants were warned ahead of time about the varying levels of disfluency. Given this, they should have discounted disfluency as a valid cue (e.g., Oppenheimer, 2008; Oppenheimer & Frank, 2008). Despite this, a disfluency effect still held. Second, the postlexical account of disfluency is thought to be dependent on subjective evaluations of disfluency, which guide monitoring and control processes (Pieger et al., 2016). In the current experiments, hard-to-read cursive words were judged as subjectively more disfluent, but were not better remembered, while easy-to-read cursive words were judged as less disfluent, but were better remembered. Thus, rather than metamemory judgements guiding performance, it is possible that processing disfluency and not perceived disfluency mediates better recognition memory performance. Supporting this possibility, Sanchez and Jaeger (2015) showed that reading times of a passage (on-line measure

of difficulty) presented in an atypical font mediated performance whereas ratings of difficulty (i.e., perceived difficulty) did not. A simple metacognitive theory of disfluency is not tenable to explain the current pattern of results.

Both theoretical accounts of disfluency effects face challenges in explaining the current pattern of findings. Cognitive load theory (Sweller, van Merriënber, & Paas, 1998), a different theoretical framework not developed to explain disfluency, could be useful. Within this framework, there are three types of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load is influenced by the inherent difficulty associated with the material that is to be learned. For example, low-frequency (infrequent) words are harder to read than high-frequency (frequent) words. Intrinsic cognitive load cannot be altered by the researcher. By contrast, extraneous cognitive load is influenced by how the information is presented and can be manipulated by the researcher. Cursive would influence extraneous cognitive load, as the only difference between words is what script it is in. Germane cognitive load is associated with the processes undertaken to process material at a deeper level. This type of cognitive load is thought to be beneficial for learning. Learning is thought to occur when processing capacity outweighs cognitive load. In our experiments, it is possible that easy-to-read cursive imposed a moderate level of extraneous load compared to hard-to-read cursive words, thereby allowing more beneficial types of cognitive load (germane) to facilitate learning.

Another way to consider the enhancement of memory due to perceptual disfluency is to place it within the conflict monitoring framework of cognitive control (Rosner et al., 2015). In the conflict monitoring framework (Botvinick, Braver, Barch, Carter, & Cohen, 2001), the up regulation and down regulation of attentional resources is modulated by the amount of conflict experienced during processing. Attentional control is carried out in the brain via activation of the anterior cingulate cortex (ACC), which detects conflict during processing. The activation of the ACC engenders increased activation of the dorsolateral prefrontal cortex—an area important for sustained or selective attention. The conflict monitoring model has been primarily used within the context of response inhibition (e.g., Stroop task) and the detection of errors.

Applied to the current results, it is possible that during the course of word processing, perceptual disfluency increases response ambiguity (i.e., not knowing what the perceptual input is), which in turn activates the ACC and increases cognitive control mechanisms leading to a memory benefit. Thus, the disfluency effect arises from an interaction between the quality of the sensory input that affects word recognition and the utilization of postlexical resources. With handwritten cursive, there appears to be an encoding cost that arises during word recognition itself. This disfluency experienced during word recognition serves a metacognitive cue to engage

in increased postlexical processing, leading to the overall memory benefit found. However, because cognitive control mechanisms are finite (e.g., working memory and attentional control; Evans & Stanovich, 2013), if the word is too hard to recognize, more resources will be utilized (see Geller, Still, & Morris, 2015), leading to a smaller memory benefit. Thus, according to the conflict-monitoring framework, if something is too ambiguous or hard to recognize, especially under time constraints, memory for that particular stimulus might be hindered, leading to poorer memory. Indeed, two recent studies (Seufert, Wagner, & Westphal, 2017; Weissgerber & Reinhard, 2017) showed that stimuli that are too disfluent (e.g., words presented in Haettenschweiler, 15% grayscale, 14-pt. font, or that are scrambled) produce poorer memory than minimally disfluent stimuli. In other words, perceptual disfluency seems most beneficial to memory when the perceptual disfluency manipulation does not require too many cognitive resources. Future research should more explicitly test the role of cognitive control in disfluency.

Conclusion

We believe that the current research makes important contributions to research on disfluency and desirable difficulties. The findings, which show the importance of type and level of disfluency, help to illuminate inconsistencies in the literature. Whether or when disfluency is desirable depends upon how disfluent the manipulation is. Thus, the disfluency effect is not as straightforward as placing something in a hard-to-read font. The conflict monitoring framework provides a useful explanation for the obtained disfluency effect. However, if extant theories were modified to take into account difficulty that arises during word recognition and the interaction with cognitive control mechanisms involved in memory formation, they could also account for the findings.

For the first time, cursive writing has been shown to act as a desirable difficulty, with better immediate and delayed recognition memory performance for cursive words than for type-printed words. However, there is an important caveat: cursive cannot be too disfluent. This aligns well other findings in the learning and memory domain suggesting a “sweet spot” for efficacious learning. To achieve the most desirable results, learning cannot be too easy, nor too hard. For example, when spacing out study, individuals benefit more from lags that are neither too short nor too long (see Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008, for an optimal lag-to-final-test ratio). Further, during study, it is most beneficial for individuals to study moderately difficult items as opposed to well-learned or difficult items (region of proximal learning; Kornell & Metcalfe, 2006).

As a recommendation to researchers, we feel it is important to ensure that the perceptual manipulation used be sufficiently strong to make encoding difficult, but not so strong that it consumes too many attentional resources needed for postlexical processing. To control for this, researchers should employ objective behavioral measures of difficulty (e.g., RTs and accuracy). Alternatively, more fine-grained methods such as eye tracking or mathematical modeling (e.g., distributional analyses) could be employed to examine levels affected by a perceptual manipulation (Balota & Yap, 2011; Reingold & Rayner, 2006). Future research should examine other moderating factors of the disfluency effect, and whether the cursive disfluency effect generalizes under more educationally realistic conditions, such as various testing formats and different encoding instructions.

Appendix A.

Table 4. Analyses of variance for unconditionalized d' values in Experiments 1, 2, and 3

Experiment	ANOVA	
Source	F	η^2
Experiment 1		
Script	$F(2, 58) = 5.18^{***}$	.15
Type-print < Cursive	$F(1, 29) = 2.52$	.08
Easy-to-read > Hard	$F(1, 29) = 9.64^{***}$	.24
Experiment 2		
Script	$F(2, 140) = 4.06^*$	.06
Type-print < Cursive	$F(1, 70) = 4.16^*$	.06
Easy-to-read > Hard	$F(1, 70) = 3.95^*$	.05
List	$F(1, 70) = .077$	.00
Script $\times$ List	$F(2, 140) = .406$	.01
Experiment 3		
Script	$F(2, 70) = 6.48^{**}$	.16
Type-print < Cursive	$F(1, 35) = 7.86^{**}$	.18
Easy-to-read > Hard	$F(1, 35) = 5.56^*$	.14

Note. $*p < .05$. $**p < .01$. $***p < .001$. Although there are slight differences in terms of statistical significance between comparisons, the direction of effects are consistent with the conditionalized analysis

Appendix B.

By-item analysis

In the previous three experiments and the meta-analysis, we observed that memory for hard-to-read cursive words was worse than for easy-to-read cursive. Because we ran analyses on conditionalized data, it could be that our interpretation is

biased insofar as more memorable items were in the hard-to-read cursive condition, thus leading to a memory disadvantage. To partially rule out the possibility that item selection effects were driving the results, we conducted analyses on the conditionalized and unconditionalized data in Experiments 1, 2, and 3. It is also important, though, to rule out an item selection effect that would disadvantage the hard-to-read cursive condition. If there were a disadvantage, we would expect to see recognition performance increase as error rates increase (i.e., positive slope estimate). For this purpose, a linear mixed-effect model was performed in R using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). For each experiment, a model with random by-item intercepts was fit with script, item error rates at study, and the interaction between the two predicting probability of recognizing the word on the test (hit rate). The pattern of findings observed across all three experiments does not suggest that hard-to-read words were disadvantaged (see Table 5). Error rates predicting recognition performance were similar across easy and hard scripts.

Table 5. By-item regression analyses for predicting hit rate at test using script type (easy-to-read vs. hard-to-read) cursive, item error rates, and the interaction between the two as predictors

Experiment	Estimate	Standard error
Experiment 1		
Easy-to-read cursive	.111***	.02
Hard-to-read cursive	.104***	.02
Item error rate	-.114	.44
Easy-to-read: mean item error	.015	.45
Hard-to-read: mean item error	-.003	.44
Experiment 2		
Blocked list		
Easy-to-read cursive	.164***	.02
Hard-to-read cursive	.166***	.02
Item error rate	.254	.29
Easy-to-read: mean item error	-.432	.30
Hard-to-read: mean item error	-.427	.30
Mixed list		
Easy-to-read cursive	.117***	.02
Hard-to-read cursive	.125***	.02
Item error rate	.658*	.29
Easy-to-read: mean item error	-.774*	.30
Hard-to-read: mean item error	-.810**	.30
Experiment 3		
Easy-to-read cursive	.132***	.02
Hard-to-read cursive	.141***	.02
Item error rate	.572*	.26
Easy-to-read: mean item error	-.710**	.27
Hard-to-read: mean item error	-.744**	.26

Note. $*p < .05$. $**p < .01$. $***p < .001$. Model: hit rate $\sim$ condition $\times$ error rate + (1|item), contrasts = list(condition = "contr.sum")

References

- Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science*, 22(6), 437–442. <https://doi.org/10.1177/0963721413498894>
- 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. <https://doi.org/10.1177/1088868309341564>
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136(4), 569.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation*, 2, 89–195.
- Baddeley, A. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition*, 10(1), 17–23.
- Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science*, 20(3), 160–166.
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39(3), 445–459.
- Barnhart, A. S., & Goldinger, S. D. (2010). Interpreting chicken-scratch: Lexical access for handwritten words. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 906–923.
- Barnhart, A. S., & Goldinger, S. D. (2015). Orthographic and phonological neighborhood effects in handwritten word perception. *Psychonomic Bulletin & Review*, 22(6), 1739–1745.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, 41(6), 897–903.
- Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 429–440.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: MIT Press.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.
- Braver, S. L., Thoenes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, 9(3), 333–342. <https://doi.org/10.1177/1745691614529796>
- Carpenter, S. K. (2017). Spacing effects in learning and memory. In J. T. Wixted & J. H. Byrne (Eds.), *Learning and memory: A comprehensive reference*. Cambridge: Academic Press.
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, 23(6), 760–771. <https://doi.org/10.1002/acp.1507>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438–448.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. <https://doi.org/10.1037/0033-2909.132.3.354>
- Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridge of optimal retention. *Psychological Science*, 19(11), 1095–1102. <https://doi.org/10.1111/j.1467-9280.2008.02209.x>
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29. <https://doi.org/10.1177/0956797613504966>
- De Zuniga, C. M., Humphreys, G. W., & Evett, L. J. (1991). Additive and interactive effects of repetition, degradation, and word frequency in the reading of handwriting. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 10–33). Hillsdale: Erlbaum.
- 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. <https://doi.org/10.1016/j.cognition.2010.09.012>
- Eitel, A., Köhl, 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.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Evans, J. S. B. T. (2016). Reasoning, Biases and Dual Processes: The Lasting Impact of Wason (1960). *Quarterly Journal of Experimental Psychology*, 69(10), 2076–2092
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., ... Tabor, J. (2013). Changing fonts in education: How the benefits vary with ability and dyslexia. *The Journal of Educational Research*, 106(4), 301–304.
- Geller, J., Still, M. L., & Morris, A. L. (2016). Eyes wide open: Pupil size as a proxy for inhibition in the masked-priming paradigm. *Memory & Cognition*, 44(4), 554–564. <https://doi.org/10.3758/s13421-015-0577-4>
- Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *The Quarterly Journal of Experimental Psychology*, 67(12), 2455–2466. <https://doi.org/10.1080/17470218.2014.937447>
- Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(3), 507–513. <https://doi.org/10.1037/0278-7393.17.3.507>
- Hirshman, E., Trembath, D., & Mulligan, N. (1994). Theoretical implications of the mnemonic benefits of perceptual interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 608–620.
- Jaros, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7(1), 2.
- Jefferys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Clarendon Press.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32, 609–622.

- Kornell, N., & Vaughn, K. E. (2016). How Retrieval Attempts Affect Learning: A Review and Synthesis. *Psychology of Learning and Motivation*, 65, 183–215. <https://doi.org/10.1016/BS.PLM.2016.03.003>
- Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science*. <https://doi.org/10.1177/1948550617693058>
- Larsen, D. P., Butler, A. C., & Roediger, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: A randomised controlled trial. *Medical Education*, 43(12), 1174–1181. <https://doi.org/10.1111/j.1365-2923.2009.03518.x>
- 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. <https://doi.org/10.1007/s11409-015-9147-1>
- McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). New York, NY, US: Psychology Press
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375.
- McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, 43(6), 1048–1063. <https://doi.org/10.1093/jcr/ucw085>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. *Reason*, 4(2), 61–64.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386.
- Mueller, P. A., & Oppenheimer, D. M. (2014). The Pen Is Mightier Than the Keyboard. *Psychological Science*, 25(6), 1159–1168. <https://doi.org/10.1177/0956797614524581>
- Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1067–1087. <https://doi.org/10.1037/0278-7393.22.5.1067>
- Mulligan, N. W. (1998). Perceptual interference at encoding enhances recall for high-but not low-imageability words. *Psychonomic Bulletin & Review*, 5(3), 464–469.
- Naime, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 248.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, 12(6), 237–241. <https://doi.org/10.1016/J.TICS.2008.02.014>
- Oppenheimer, D. M., & Frank, M. C. (2008). Effects of perceptual fluency on categorization. <https://doi.org/10.1016/j.cognition.2007.05.010>
- Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, 28(4), 502–504. <https://doi.org/10.1002/acp.3023>
- Perea, M., Gil-López, C., Beléndez, V., & Carreiras, M. (2016). Do handwritten words magnify lexical effects in visual word recognition? *Quarterly Journal of Experimental Psychology*, 69(8), 1631–1647. <https://doi.org/10.1080/17470218.2015.1091016>
- Perea, M., Marcet, A., Uixera, B., & Vergara-Martínez, M. (2016). Eye movements when reading sentences with handwritten words. *Quarterly Journal of Experimental Psychology* (2006), 1–9. <https://doi.org/10.1080/17470218.2016.1237531>
- 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. <https://doi.org/10.1016/j.learninstruc.2016.01.012>
- Qiao, E., Vinckier, F., Szwed, M., Naccache, L., Valabrègue, R., Dehaene, S., & Cohen, L. (2010). Unconsciously deciphering handwriting: Subliminal invariance for handwritten words in the visual word form area. *NeuroImage*, 49(2), 1786–1799.
- Reingold, E. M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the EZ Reader model. *Psychological Science*, 17(9), 742–746.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137(4), 615–625.
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, 16(3), 550–554.
- Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica*, 160, 11–22.
- 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.
- Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, 19(6), 523–542. <https://doi.org/10.3758/BF03197149>
- Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). E-Prime (Version 1. 1) [Computer software]. Pittsburgh, PA: Psychology Software Tools.
- Serra, M. (2016). *Does processing fluency really matter for metacognition in actual learning situations? (Part Two)*. Retrieved from <http://www.improvethelearning.com/processing-fluency-really-matter-metacognition-actual-learning-situations/>
- 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. <https://doi.org/10.1007/s11251-016-9387-8>
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592.
- Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2015). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning*, 1–18.
- 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(5), 973–978. <https://doi.org/10.3758/s13423-011-0114-9>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2017). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning*. <https://doi.org/10.1007/s11409-017-9175-0>
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Weissgerber, S. C., & Reinhard, M.-A. (2017). Is disfluency desirable for learning? *Learning and Instruction*, 49, 199–217. <https://doi.org/10.1016/j.learninstruc.2017.02.004>
- 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(1), 156–165.

- Yap, M. J., & Balota, D. A. (2007). Additive and Interactive Effects on Response Time Distributions in Visual Word Recognition, 56–58. <https://doi.org/10.1037/0278-7393.33.2.27>
- Yap, M. J., and Balota, D. A. (2015). Visual word recognition, in Oxford Handbook of Reading, eds A. Pollatsek and R. Treiman (New York, NY: Oxford University Press), 26–43.
- 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(2), 229–241.