Testing Expectancy, but not Judgements of Learning, Moderate the Disfluency Effect

Jason Geller (jgeller1@uab.edu)
Department of Psychology, 1530 3rd Avenue South.
Birmingham, AL 35205 USA
Mary L. Still (mstill@odu.edu)
Department of Psychology, 250 Mills Godwin Life Sciences Building.
Norfolk, VA 23529 USA

Abstract

Do students learn better with material that is perceptually harder-to-process? Previous research has been equivocal concerning this question. To clarify these discrepancies, the present study examined two potential boundary conditions to determine when disfluent text is, and is not, beneficial to learning. The two boundary conditions examined were: type of judgement of learning (JOLs) and testing expectancy. Boundary conditions were examined in separate Group (incidental aggregate JOLs vs. intentional aggregate JOLs vs. item-by-item JOLs) by Disfluency (Masked vs. Nonmasked) mixed ANOVAs. Results revealed that type of JOL did not moderate the disfluency effect, but testing expectancy did. These results bring forth questions pertaining to the utility of disfluency on learning.

Keywords: Disfluency; Testing Expectancy; JOLs; Desirable difficulties; Learning and Memory

Introduction

The desirable difficulty principle (Bjork, 1994) puts forth the paradoxical idea that making learning harder (not easier) should have the desirable effect of improving long-term retention. One simple strategy to make learning harder (and thereby improve memory), in both the classroom and laboratory, is to make material more perceptually disfluent by changing the material's perceptual characteristics (Diemand-Yaumen, Oppenheimer, & Vaughan, 2011; French et al., 2013). Visual material that is masked (Mulligan, 1996), inverted (Sungkhasette, Friedman, & Castel, 2011), uses atypical fonts (Diemand-Yaumen et al., 2011), or has high-level blurring (Rosner, Davis, & Milliken, 2015) have all been shown to produce memory benefits. The desirable effect of perceptual disfluency on memory has been called the disfluency effect (Diemand-Yaumen et al., 2011).

Although appealing as a pedagogical strategy, there have been several experiments that failed to find a memory benefit for perceptually disfluent materials (e.g., Magreehan, Serra, Schwartz & Narciss, 2016; Rhodes & Castel, 2008, 2009; Rummer, Scheweppe, & Schewede, 2016; Yue, Castel, & Bjork, 2013), casting doubt upon the veracity of the disfluency effect. Much of the literature examining the disfluency effect consists of conceptual replications using a wide variety of manipulations and tasks; this variety makes it difficult to know whether a replication fails due to procedural differences or due to the absence of a true disfluency effect. Given the equivocal findings in the literature, the current research aims to identify important moderating or boundary conditions of the effect (Oppenheimer & Alter, 2014). The present research examined two potential boundary conditions: type of judgment of learning (JOLs) and testing expectancy. Understanding the boundary conditions of the disfluency effect has important theoretical and practical implications.

One common feature in studies that did not obtain the disfluency effect is the use of item-by-item JOLs. JOLs are subjective predictions made by participants indicating the probability that they would later judge a word as being remembered. JOLs are used by researchers to examine how various cues, such as fluency, may be used by students to regulate the allocation of attentional resources during study and re-study (Metcalfe & Finn, 2008). When the JOLs are made item-by-item, participants immediately provide a judgment after each studied word on a scale of 0%-100%, with a JOL of 0% indicating that participants believe they will not be able to remember a word at a later time, and 100% indicating that they believe they will definitely recall the word at a later time.

Because the use of item-by-item JOLs is ubiquitous in the disfluency effect literature, it is quite possible, that the elicitation of item-by-item

JOLs themselves may have had an impact on actual memory. Therefore, the act of producing the item-by-item JOLs could have an impact on memory that distorts or even masks the disfluency effect. Besken and Mulligan (2013) provide two possible explanations as to why the elicitation of item-by item JOLs might attenuate the disfluency effect. One possible explanation is that item-by-item JOLs require the participant to retrieve the stimulus again, inducing deeper, more elaborative processing for *both* disfluent and fluent stimuli. Another possible explanation is that the elicitation of JOLs results in the allocation of more effort to the JOLs task, thereby leading to less post-lexical processing of the disfluent stimulus than would otherwise occur. Regardless of which account is correct, the elicitation of item-by-item JOLs appears to influence processing and the subsequent strength of the memory representation. Thus, it is important to examine the influence of type of JOL when examining disfluency.

Although item-by-item JOLs may moderate the disfluency effect by inducing deeper processing of all words, the use of item-by-item JOLs also requires that participants receive intentional learning instructions; that is, individuals know that they will have to remember the words for a later test. It may be that testing expectancy, and not type of JOL, moderate the disfluency effect. The current experiment was designed to explicitly examine the role of type of encoding instruction (incidental versus intentional) without the confound of the presence of item-by-item JOLs.

Three groups were examined: an intentional item-by-item JOLs group, an intentional aggregate JOLs group, and an incidental aggregate JOLs group. The item-by-item group received instructions that alluded to a recognition memory test and were required to provide item-by-item JOLs during encoding. The aggregate JOLs intentional group received instructions that alluded to a recognition memory test and were required to provide aggregate JOLs after the whole list was studied. Finally, the aggregate JOLs incidental group were not told about a memory test but were required to provide aggregate JOLs. In the fluent condition, items were presented for 2 s; in the disfluency condition, items were presented for 80 ms and masked by a row of hashmarks (####).

Naming latencies, accuracy, and JOLs during study were used to ensure that masked words were in fact disfluent compared to unmasked words. The disfluency effect was indexed by the results of a recognition memory task. To examine whether type of JOL moderates the disfluency effect, a 2 (Group: item-by-item intentional JOLs vs. aggregate intentional JOLs) x 2 (Disfluency: masked vs. nonmasked) mixed ANOVA was performed. If an interaction arises between the two factors, then it would provide evidence that type of JOL moderates the disfluency effect. To examine whether testing expectancy moderates the disfluency effect, a 2 (Group: incidental aggregate JOLs vs. aggregate intentional JOLs) x 2 (Disfluency: masked vs. nonmasked) mixed ANOVA was performed. If an interaction arises between the two factors, then it would provide evidence that testing expectancy moderates the disfluency effect. Specifically, we predicted that participants who expected a later memory test would be less likely to benefit from the disfluency manipulation.

Method

Participants and Design

Eighty-four undergraduate students from Iowa State University participated for course credit; 28 students were assigned to each of the three groups. All participants were native speakers of English and self-reported normal or corrected-to-normal vision.

The within-subject variable was whether or not the words were masked. There were three between-subject groups: item-by-item JOLs with intentional instructions, aggregate JOLs with intentional instructions, and aggregate JOLs with incidental instructions.

Materials

Four counterbalancing lists were constructed from 200 four-letter, high-frequency (mean HAL frequency = 9.7) nouns taken from the English Lexicon Project (Balota et al., 2007). First, two separate 100-word lists were created, one to be used during study and test (old items) and one to be used only during test (new items). Next, two versions of each of these lists were created. Half the items were assigned to the perceptually disfluent (masked) condition and half to the perceptually fluent (nonmasked) condition. Lists were assigned to participants so that across participants each word occurred equally often in the four possible conditions: masked old, nonmasked old, masked new, and nonmasked new. It is important to note that each *new* item was categorized as masked or nonmasked for counterbalancing purposes. All items on the test were presented without a mask.

Procedure

Participants were tested individually in a small, well-lit room, seated approximately 65 cm from the computer screen. There were two different types of instructions – one for the intentional encoding groups and one for the incidental encoding group.

The instructions in in the intentional encoding groups were:

> "During this experiment, you will be presented with 100 words. Half of the words will be presented normally, while the other half will be presented very quickly, and followed by a row of hashmarks (####). Your task is to name each stimulus, as quickly and accurately as possible. Do your best to remember the words because your memory will be tested later."

The instructions in in the incidental encoding group were:

> "During this experiment, you will be presented with 100 words. Half of the words will be presented normally, while the other half will be presented very quickly, and followed by a row of hashmarks (####). Your task is to name each stimulus, as quickly and accurately as possible."

For each group, every trial began with a fixation cross appearing at the center of the screen for 1,000 ms. The fixation cross was replaced by a word in the same location. Words were presented in a 44-point Courier New font in black on a white background. Half of the words were presented under disfluent (masked) conditions and half under fluent (unmasked) conditions. Masked words appeared for 80 ms and were backward masked for 1,920 ms; unmasked words appeared for 2,000 ms. After each naming response, in the aggregate JOL groups, a blank 1,000 ms inter-stimulus interval (ISI) appeared. After all the words were named, participants were told that 50 clear and 50 masked words had appeared in the list, and were asked to estimate how many in each condition they expect to remember on a later test. The order of the two memory judgements was counterbalanced across participants. In the intentional item-by-item JOLs group, immediately after naming each item, participants used the keyboard to rate their confidence on a scale of 0 (not confident at all) to 100 (very confident) that would be able to remember the item they studied 5 minutes from now.

After a short distractor task, participants took an old-new recognition test. At test, a fixation cross appeared in the center of the screen for 1000 ms and was followed by a word that either had been presented during study ("old") or had not been presented during study ("new"). Words stayed on the screen until participants gave an "old" or "new" response on the button box. For masked words, study context was not reinstated at test.

Results

No participants had to be replaced and no items were discarded. An alpha level of .05 was used. η^2_p is reported as the effect size measure. In addition, alongside traditional analyses that utilize null hypothesis significance testing, Bayes' factors, calculated with the freeware software program JASP (Version 0.8.5; https://jasp-stats.org) for null findings (noted as BF_{01}) are reported (see Jarosz & Wiley, 2011, for a review). A Bayes factor of 3 or greater is indicative of strong or positive evidence in favor of the null.

Study Phase

For naming latencies, RTs faster than 150 ms or slower than 2.5 times the standard deviation for each participant were excluded from analysis. This outlier procedure resulted in the exclusion of 4% of the data. Trials in which there were microphone malfunctions (i.e., the microphone did not record a response) and errors were excluded (10%).

Naming accuracy, naming latency, and JOLs

As a manipulation check, naming latencies, accuracy, and JOLs were examined for each group (incidental aggregate JOLs, intentional aggregate JOLs, intentional item-by-item JOLs), respectively. Mean naming latencies, accuracy, and JOLs for each group are displayed in Table 1. Although accuracy was high in each condition, individuals in each group performed worse in the masked condition than the nonmasked condition, all *ts* > 2.29, *ps* < .03. Examining naming latencies, there were no differences for the intentional item-by-item and intentional aggregate group, all *ts* < .46, *ps* > .65, BF_{01} > 4. There was, however, a speed-accuracy tradeoff in the incidental aggregate group, *t*(27) = -2.27, *p* =.03, 95% *CI* [-27.24, -1.40], *d* = .43. Lastly, JOLs in each group were lower for the masked condition than the nonmasked condition, all *ts* > 3.05, *ps* < .01. Overall, although participants were not slower in naming masked words, they were more error prone (had lower accuracy) and gave lower JOLs than nonmasked words. This provides evidence that the masking manipulation was in fact disfluent.

Test Phase

Given the very high naming accuracy rates for both masked and nonmasked conditions in all three of the groups, we analyzed unconditionalized data. Memory sensitivity (*d'*) for each group is displayed in Figure 1. Separate 2 x 2 ANOVAs examined the moderating influence of type of JOL and testing expectancy on the disfluency effect. Examining the influence of type of JOL, there were no main effects or interaction, *Fs* < 2.91, *ps* > .09. The null model was preferred over the full model (*BF* = 11.39).

Examining testing expectancy, no reliable memory difference arose between masked and nonmasked items, *F*(1, 54) = .579, *p* = .450, η^2_p = .01. There was a marginal effect of group, *F*(1, 54) = 3.68, *p* = .061, η^2_p = .06. The intentional aggregate JOL group tended to have better memory than the incidental aggregate JOL group. Finally, there was a significant group by disfluency interaction, *F*(1, 54) = 4.06, *p* = .049, η^2_p = .07. A disfluency effect arose for the incidental aggregate group, but not the intentional aggregate JOL group.

Table1: Mean Naming Accuracy (in proportions), Naming Latencies (in milliseconds), and JOLs (in proportions) for Words in Experiment 1 as a Function of Masking and Group

Condition	Naming Accuracy	Naming Latency	JOLs
Incidental aggregate JOLs			
Nonmasked	.99 (.00)	589 (16)	.57 (.04)
Masked	.98 (.01)	592 (22)	.50 (.05)
Difference	.01	3	.07
Intentional aggregate			
Nonmasked	.99 (.00)	592 (16)	.49 (.00)
Masked	.97 (.01)	577 (15)	.46 (.01)
Difference	.02	-15	.03
Intentional item-by-item JOLs			
Nonmasked	.99 (.00)	811(43)	.54 (.04)
Masked	.98 (.01)	805(47)	.46 (.04)
Difference	.01	-6	.08

Note. Standard errors are shown in parentheses. * p < .05; ** p < .01; *** p < .001. Negative numbers signify that masked words were responded to faster than nonmasked words.

Given this pattern of results, type of JOL does not appear to be a moderator of the disfluency effect. However, testing expectancy does seem to moderate the disfluency effect. When not told about an upcoming test (i.e., incidental aggregate JOLs group), a disfluency effect was observed. Being told explicitly about an upcoming memory test (i.e., intentional aggregate JOLs group) appeared to attenuate the advantage in recognition for disfluent stimuli.

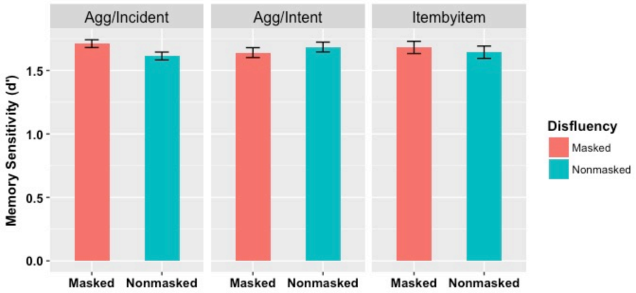

Figure 1: Memory sensitivity (*d'*) as a function of group (left - aggregate incidental group; center - aggregate intentional group; right - item-by-item intentional group (right)) and fluency (masked vs. masked). Error bars reflect the within-subject standard error of the mean (Morey, 2008).

Discussion

This study set out to examine two potential moderating factors of the disfluency effect: type of JOL and testing expectancy. We did not find any evidence that type of JOL moderates the disfluency effect. In the intentional item-by-item JOLs and intentional aggreagte JOLs groups, masked words were not better recognized than nonmasked words. This finding differs from other studies that did find that type of JOL moderates the disfluency effect (e.g., Besken & Mulligan, 2013, 2014). Although type of JOL did not moderate the disfluency effect in this study, it is important to note that in those studies testing expectancy was not examined as a potential moderating factor.

Eitel and Kuhl (2016) posited that testing expectancy may be an important moderator of the disfluency effect. They reasoned that if the disfluency effect arises because of deeper, more effortful, processing, telling participants about a memory test should eliminate the effect. This occurs because testing expectancy would countervail the effects of disfluency by eliciting additional processing for both fluent and disfluent stimuli. In contrast, incidental instructions are less likely to impact processing of individual items, leaving effects of processing difficulty on recognition memory intact. In their study, Eitel and Kuhl found that testing expectancy lead to better learning, overall, but they failed to find a disfluency effect, which makes it difficult to make inferences about potential moderators. In our study, we demonstrated that the disfluency effect is indeed modulated by testing expectancy. Consistent with this, a disfluency effect was observed only when aggregate JOLs were used in conjunction with incidental instructions.

Conclusion

Recent studies by Diemand-Yauman et al. (2010) and French et al. (2013) have recommended that teachers and students use perceptual disfluency to enhance learning. Although we have shown that perceptual disfluency can enhance learning in a very simplified context (i.e., list learning), its efficaciousness as a potential learning technique is tempered by the finding that testing expectancy can eradicate the effect. In an educational setting, students are always told about upcoming tests. Thus, disfluency might not be an effective manipulation to enhance memory in a more ecologically valid setting. What is clear from the current findings is that disfluency's usefulness as an educational intervention is not as straightforward as placing something in a hard-to-read font. Future research should continue to explore the boundary conditions of the disfluency effect.

References

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.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation betweenmetamemoryandmemory 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.

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

Eitel, A., & Kühl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, *11*(1), 107–121. https://doi.org/10.1007/s11409-015-9145-3

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.

Jarosz, 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.

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

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.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, *4*(2), 61–64.

Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, *28*(4), 502-504. http://doi.org/10.1002/acp.3023

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*(2), 57–70.

Sungkhasette, 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. http://doi.org/10.3758/s13423-011-0114-9

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.