The Effect of Disfluency on Memory for What was Said

By

Evgeniia Diachek

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Psychology

August 31, 2020

Nashville, Tennessee

Approved:

Dr. Sarah Brown-Schmidt, Ph.D

Dr. Lisa Fazio, Ph.D.

ACKNOWLEDGEMENTS

This material is based on work supported by National Science Foundation Grant BCS 19-21492 to Sarah Brown-Schmidt. Thank you to Sarah Brown-Schmidt for the invaluable guidance and support throughout the first two years of grad school. Thank you to Jasmine Aggarwal and Jordan Zimmerman for their assistance in recording the experimental stimuli. Thank you to the faculty of the Department of Psychology and Human Development for the useful discussions in relation to this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Disfluencies, or interruptions in the fluent speech stream, are prevalent in spontaneous speech, occurring ~4-10 times per 100 spoken words (Branigan, Lickley, & McKelvie, 1999; Bortfeld et al., 2001). Analyses of how disfluencies pattern in speech show that speakers tend to produce different types of disfluencies in different contexts (Bortfield et al., 2001; Shriberg, 1996). For example, in a naturalistic language production study, Fraundorf and Watson (2014) found that repetitions (e.g., *Alice doesn't think* *that cats that cats* *can grin*) tend to be used when the planned material is available and articulated, and as a result, can be repeated. On the other hand, pauses (e.g., *She notices … a small … box that says "EAT ME"*) and fillers (e.g., *She grabs the fan and* *uh* *one pair of gloves*) are used in the planning of a new message before the articulation of that new message has begun. If different forms of disfluency are used in different contexts, this raises the possibility that different forms of disfluency signal different meanings (Arnold et al., 2004; Arnold, et al. 2003; Arnold et al., 2007; Bosker et al., 2019; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997; Corley, MacGregor, & Donaldson, 2007; Fox Tree, 2001; Grosman, 2015; Walker, Risko, & Kingstone, 2014; Watanabe et al., 2008), and therefore act as a potential source of information to guide language understanding.

Indeed, complementary work in language comprehension shows that while listeners may not remember having heard a disfluency (Bard & Lickey, 1997), disfluencies inform language processing both on and offline (Fox Tree, 1991, 1995; Bailey & Ferreira, 2003; Corley & Stewart, 2008). For example, Bailey and Ferreira (2003) presented participants with temporarily ambiguous garden-path sentences which contained disfluencies either before or after an ambiguous noun phrase. The results indicated that when the disfluency occurred immediately before an ambiguous noun phrase, listeners were more likely to interpret the noun phrase as the start of a new clause, preventing the garden path interpretation and improving grammaticality ratings. In addition to shaping grammatical processing, disfluencies also affect word recognition: Fox Tree (1991) examined the effect of fillers (*ums* and *uhs*) on recognition of the subsequent word. Participants were presented with a word probe followed by an auditory sentence that contained a disfluency. Participants were instructed to press a button when they heard the target word. The results indicated that recognition of a word was faster if it was preceded by *uh* (no such benefit was observed for words preceded by *um*, possibly because *um* is associated with longer delays in speech). Converging behavioral and neuroimaging evidence indicates that disfluencies can help listeners predict or recognize upcoming information (Arnold et al., 2004), reduce surprisal for unpredictable words (Corley, MacGregor, & Donaldson, 2007), facilitate lexical access (Fox Tree, 1995), direct attention to unfamiliar objects (Arnold et al., 2007), shape inferences about the speaker's metacognitive states (Brennan & Williams, 1995) and influence hiring recommendations (Brosy, Bangerter, & Mayor, 2016). Together, these studies suggest that disfluencies signal whether the upcoming information is unpredictable or new and by doing so, guide predictive language processing.

Preliminary evidence suggests that disfluencies also impact memory for what was said (Corley, MacGregor, & Donaldson, 2007; Fraundorf & Watson, 2011). For example, Fraundorf and Watson (2011) presented participants with spoken passages that were fluent, or that contained disfluencies prior to some (but not all) critical plot points. They found that participants recalled the gist of significantly more plot points when the passage contained disfluent fillers (*um*s and *uh*s), compared to fluent passages. A control condition in which the passages contained coughs resulted in *worse* memory, indicating that the disfluency-memory boost was not simply due to the presence of an interruption. Even more interesting is the following finding: the disfluency boost increased memory for *all* plot points -- not just the ones preceded by disfluency, suggesting that the disfluency boost spreads to improve memory for the gist of the entire narrative.

In addition to boosting memory for the gist of a story, disfluency also appears to boost memory for individual words in sentences. In one line of research, participants were presented with a series of unrelated sentences, some of which contained disfluencies (e.g. *Everyone's got bad habits and mine is biting my, er, nails*). A recognition memory test followed and probed memory for the words that were immediately preceded by disfluency (e.g. "nails"). The results revealed a clear memory benefit for words preceded by fillers (Corley, MacGregor, & Donaldson, 2007), and words preceded by a pause (MacGregor, Donaldson, & Corley, 2010). However, in a separate study, no memorial benefit was observed for words preceded by a disfluent repetition (MacGregor, Corley, & Donaldson, 2009). Together, these studies point to the possibility that different disfluency types have different effects on enduring memories. However, these studies did not provide a direct comparison of the different memory benefits for different disfluency types, leaving multiple open questions regarding the cognitive underpinnings that guide the disfluency-memory boost.

We note that in at least one prior study, which examined the effect of speech rate and disfluent repetitions on memory, Donahue and Lickey (2017) found that recall for passages that contained disfluencies was significantly *worse* compared to fluent passages. The contrasting findings may owe to the fact that they manipulated the materials in a way that resembled stuttering, which might be interpreted differently by listeners than other disfluent repetitions or restarts.

1.1 Attention, Memory, and Disfluency

The link between disfluent fillers and pauses, and increased memory for what was said (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010; Fraundorf & Watson, 2011) may be due to attentional processes cued by the disfluency itself. One mechanism may be that disfluent words are inherently distinct in the speech stream, and it is this primary distinctiveness (von Restorff, 1933) that enhances attention, in much the same way that a distinct auditory targets are recognized faster (Dalton & Lavie, 2004), and auditory "oddballs" remembered better (Fabiani, Karis, & Donchin, 1986).

A related possibility, based on evidence from event-related brain potentials (ERPs) (Collard et al., 2008) is that disfluent fillers act to orient attention to upcoming speech (rather than to the disfluency itself). If so, the disfluency may act as a type of auditory orienting cue (Quinlan & Bailey, 1995; Spence & Driver, 1994) that for a period of time directs attention to the auditory stimulus. Consistent with this idea that disfluency enhances attention to immediately following words are findings of enhanced *recognition* for words immediately following disfluency (Fox Tree, 2001; Corley & Hartsuiker, 2011). Given that spoken words are better remembered when attended (Christensen, et al 2012; Wallace, et al 2001; Bentin, Kutas & Hillyard, 1995), the downstream consequence of directed attention and enhanced word recognition, then, may be the subsequent memory boost for words following disfluency.

Alternatively, the mechanism driving the disfluency-memory boost may simply be that the presence of disfluency creates a delay in the unfolding speech stream and it is this delay that confers the memory benefit. Corley and Hartsuiker (2011) found that word recognition was improved when the critical word was immediately preceded by either a disfluent filler or a non-linguistic tone, which they argue is consistent with a delay-based account. We note, however, that it is not immediately clear why a delay-based memory boost would be observed when the delay contains a tone, but not when it contains a cough (Fraundorf & Watson, 2011).

1.2 Present Research

The present research probes the scope and basis of the memory boost associated with disfluent speech (Corley, MacGregor, & Donaldson, 2007; Fraundorf & Watson, 2011). In Experiment 1 we directly compare three types of disfluency: fillers (*her um leg*), disfluent pauses (*her … leg*), and disfluent repetitions (*her … her leg*). While prior work reports memory benefits for disfluent fillers and pauses (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010), but not disfluent repetitions (MacGregor, Corley, & Donaldson, 2009), to our knowledge, the three types of disfluency have yet to be directly compared in a well-powered study. If fillers and pauses do, in fact, confer a more pronounced memory benefit, the reason may be the fact that in language production, fillers and pauses signal the planning of a new message (Fraundorf & Watson, 2014; Clark & Fox Tree, 2002). Listeners are sensitive to statistical regularities in the linguistic input that link meaning and form, shaping expectations about upcoming material (e.g., Altmann & Kamide, 1999, Ryskin et al., 2017), thus this form-meaning mapping may shape how a disfluent utterance is interpreted. Alternatively, delays of any type may enhance attention to immediately following words (Corley & Hartsuiker, 2011); if so, all three types of disfluency would be expected to boost memory for the speech that follows. Experiment 2 was a pre-registered direct replication of Experiment 1.

A separate question concerns how long or at what grain size the disfluency-related memory boost operates. In Experiment 3, we therefore probe the longevity of the disfluency-related memory boost by manipulating where in the sentence the disfluency is placed. While studies probing word-level memory have demonstrated highly local effects, whereby disfluency boosts memory for the immediately following word (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010), studies probing gist memory for plot points (Fraudorf & Watson, 2011) finds a general memory boost throughout the story. Thus whether disfluencies boost memory for words in sentences regardless of where that disfluency occurs in the sentence, remains an open question.

CHAPTER 2

EXPERIMENT 1

2.1 Methods

Participants

To determine our target sample size, we conducted a power analysis using G*Power (Faul et al., 2007) to replicate the disfluency memory boost reported in Corley, MacGregor, and Donaldson (2007) with at least 95% power. According to the analysis, our target sample size was 100 participants. A total of 102 participants (50 female; mean age 36.1; range 21-62) were recruited on Amazon Mechanical Turk through the research platform FindingFive (FindingFive Team, 2019). Criteria for participation were: >95% acceptance rate, and participating in the US. Participants received $4 for the experiment which took ~30 minutes to complete. We oversampled in order to get an equal number of participants across the experimental lists (see below for details about the lists). At the beginning of the experiment, we collected basic demographic information (native language, gender, age). All participants reported themselves as native speakers of English. Informed consent was obtained in accordance with the Vanderbilt University IRB guidelines.

Materials

The corresponding author of the paper by Corley, MacGregor, and Donaldson (2007) graciously provided the written materials from their Experiment, which we used to record our auditory stimuli. In their study, Corley and colleagues were interested in the effects of disfluency and predictability on the ERP components, so their stimulus set included sentences with predictable (mean close probability 0.84 as reported in Corley, MacGregor, & Donaldson, 2007) and unpredictable (mean close probability 0) final words. Our materials included 160 unique sentences that used only the predictable final words (as memory for unpredictable words was not one of our research questions). Because the original study was conducted in the United Kingdom, we slightly changed the stimuli by substituting some words to follow American Standard English (e.g., holiday -> vacation, fashions -> trends). In addition, in the original stimuli, three words were repeated in the final position. We slightly altered these three sentences so that the final word in each sentence was unique and only appeared in the final position once. We created 4 versions of each sentence (fluent, disfluent with a filler (*um*), disfluent with a pause, disfluent with a repetition) resulting in 640 sentence versions (**Table 1**). We created eight experimental lists that counterbalanced sentences across conditions using a modified Latin square design. Each list contained 80 sentences, one version of each of the 160 unique sentences (20 in each condition). Each participant was randomly assigned to a single experimental list (~13 participants per list).

The auditory stimuli were recorded by a female research assistant with a North American accent of English. Fluent and disfluent sentences were recorded at a natural speaking rate. The research assistant was instructed to record the sentences to sound as natural as possible. Following Corley, MacGregor, and Donaldson (2007), Collard et al. (2008), MacGregor, Corley, and Donaldson (2009), and MacGregor, Corley, and Donaldson (2010), the disfluencies always preceded the final word of the sentence, which was used as the memory probe word.

Table 1. Example Stimulus Set for Experiment 1. The probe word is the final word of each sentence (e.g. "leg").

Condition	Sentence

Fluent	*My sister had a skiing accident and she broke her leg*
Filler	*My sister had a skiing accident and she broke her um leg*
Pause	*My sister had a skiing accident and she broke her … leg*
Repetition	*My sister had a skiing accident and she broke her her leg*

Procedure

At the beginning of the experiment, participants also completed five microphone check questions to ensure that their headphones worked properly and were set to a comfortable volume. For each of the five microphone check questions, participants heard one word and were asked to type it out. There were two phases of the experiment (**Figure 1**). In the first phase, participants listened to 80 fluent and disfluent audio sentences. Once the participant began listening to a given sentence, they could only advance to the next listening trial once the current sentence had finished playing. Participants also had to answer 8 comprehension questions about the sentences. These questions appeared in random locations throughout the first phase of the experiment, and were used to ensure that participants were paying attention to the stimuli. Participants were instructed that they needed to answer at least 85% of the questions correctly. Phase 2 of the experiment was a surprise recognition memory test. Participants viewed 160 visual word probes, one word at a time. Half of the words had appeared in the first phase of the study, and half were new (80 old + 80 new). New probes were the last words in the other half of the experimental sentences (i.e. ones that were not presented to participants on the current list). All memory probes were unique, meaning that, across all stimuli, each probe occurred in the final position only once. Participants were asked to click to indicate whether the probe word was old (present in the sentences they just heard), or new.

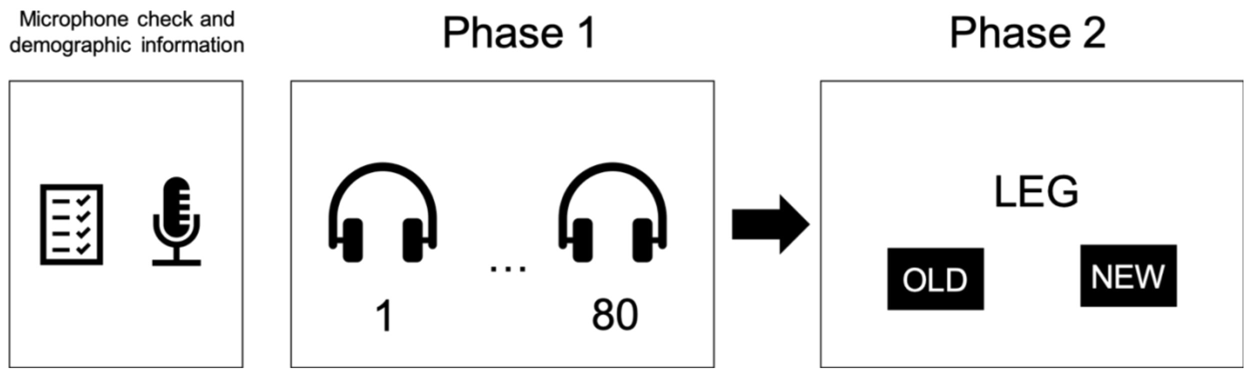

*Figure 1***.** Procedural details for Experiments 1 and 2. At the beginning of the experiment, participants completed a basic demographic questionnaire and a microphone check test. In Phase 1 of the experiment, participants listened to 80 audio sentences. In Phase 2, participants were asked to click on the button to indicate if the probe was old or new (present in the sentences they had just heard or not).

2.2 Predictions

If disfluencies are in fact a type of signal to meaning (Clark & Fox Tree, 2002), the probabilistic link between new information and disfluent pauses and fillers may result in stronger attentional cueing, and therefore a stronger memory enhancement for words preceded by pauses and fillers, compared to repetitions. Alternatively, if the disfluency-related memory boost is simply due to the fact that disfluency is distinctive (e.g. von Restorff, 1933; Dalton & Lavie, 2004; Fabiani, Karis, & Donchin, 1986), or because it creates a delay (Corley and Hartsuiker, 2011), the disfluency-related memory boost may be observed for any type of disfluent interruption, regardless of its probabilistic link to upcoming meanings.

2.3 Results

We used a signal-detection theoretic mixed-effects analysis (Wright, Horry, & Skagerberg, 2009) for the response data. We fit a logistic mixed effect regression model to the participants' old-new responses, with item status (whether the item was actually old vs. new) as a factor, and then for old items, the presence and type of disfluency as predictors. These fixed effects were coded using orthogonal Helmert contrasts (**Table 2**). The model included random intercepts and slopes for the memory and disfluency effects by subject and item. We used a maximal random effects structure following the recommendations by Barr, et al. (2013); in each model we present in the paper, the maximal model converged successfully. The analyses were performed in RStudio (RStudio Team, 2018) using lme4 (Bates et al., 2015) package.

Table 2. Experiment 1 memory results: Mixed effect model with item status (old vs. new) and type of disfluency as fixed effects. The dependent measure is binary - whether the participant responded "old" or "new" on the memory test. Values in bold indicate significant results at an alpha level of .05.

		Estimate	SE	z-value	p-value
Fixed effects	(Intercept)	-0.13652	0.11477	-1.190	0.23421
	Memory effect (old (fluent, filler, pause, repetition) = each 0.5, new = -0.5)	**1.33822**	**0.11622**	**11.514**	**< 2e-16**
	Disfluency effect (new = -0.125, fluent = -0.625, filler, pause, repetition = each 0.375)	**0.34989**	**0.07243**	**4.818**	**1.45e-06**
	Fillers and pauses vs. repetitions (new = -0.0625, fluent = -0.0625, filler, pause = 0.4375, repetition = -0.5625)	**0.17893**	**0.06295**	**2.842**	**0.00448**
	Fillers vs. pauses (filler = 0.5, pause = -0.5)	0.05612	0.07338	0.765	0.44442

		Variance	St. Dev.	Correlations	
Random effects	Item (Intercept)	0.61860	0.7865		
	Memory effect	1.00538	1.0027	-0.29	
	Disfluency effect	0.16175	0.4022	-0.13	-0.06
	Subject (Intercept)	0.90618	0.9519		
	Memory effect	0.54604	0.7389	-0.30	
	Disfluency effect	0.05562	0.2358	0.16	0.35

The results of this analysis are reported in Table 2. A negative intercept term (b = -0.14, z = -1.19, p=0.23) was due to a non-significant response bias to say "new" (coded as 0) rather than "old" (coded as 1). A significant effect of item type (actually old vs. new), indicated good memory for the probe words (b=1.34, z=11.51, $p<10^{-15}$). In addition, we observed significantly better memory for probes from disfluent vs. fluent sentences (b=0.35, z=54.82, $p<10^{-5}$), and better memory for pauses and fillers vs. repetitions (b=0.18, z=2.84, p=0.004) (**Figure 2**).

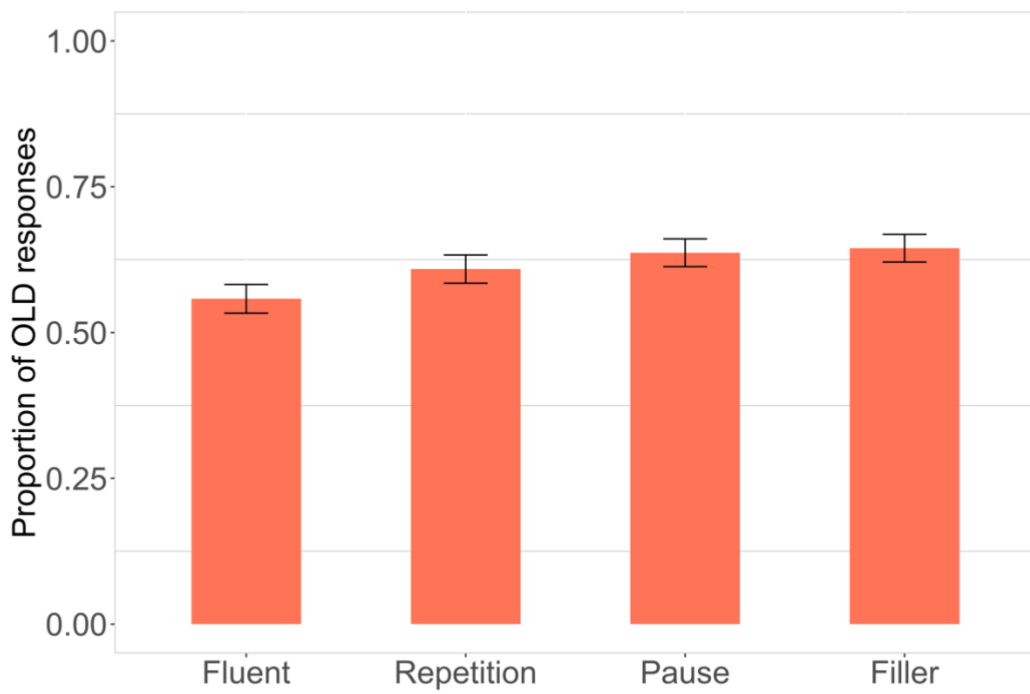

Figure 2. Memory results for Experiment 1. Error bars represent by-subject SEM.

2.4 Discussion

The results from Experiment 1 replicate previous findings that disfluencies improve memory for linguistic material (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson 2010; Fraundorf & Watson, 2011). In addition, pauses and fillers resulted in a larger memory boost compared to disfluent repetitions. Compared to previous studies looking at the effect of disfluency on memory (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2009; MacGregor, Corley, & Donaldson, 2010; Fraundorf & Watson, 2014; Donahue & Lickey, 2017), we used a larger sample size, which may have contributed to our ability to detect the relatively small effect of disfluency type (i.e. the larger memory benefit for pauses and fillers vs. repetitions). However, given that (to our knowledge), this is the first demonstration of this disfluency type effect, along with the fact that the effect is somewhat small (64% correct recognitions for pauses, 64% - for fillers, 61% - for repetitions, vs. 56% for fluent utterances), prompted us to test the reproducibility of this finding in a direct replication.

CHAPTER 3

EXPERIMENT 2

The primary goal of Experiment 2 was to replicate the findings of Experiment 1.

3.1 Methods

Experiment 2 was a direct pre-registered replication of Experiment 1. The preregistration is available on the Open Science Framework (https://osf.io/mzcrb).

Participants

To determine our target sample size, we conducted an a-priori power analysis using data simulation package simr (Green & MacLeod, 2016) in RStudio (RStudio Team, 2018) to replicate the differential effect of disfluency type (pauses and fillers vs. repetitions) on memory from Experiment 1 with at least 95% power (at an alpha level of 5%). According to the analysis, our target sample size was 160 participants. A total of 161 participants (70 female; mean age 38.6; range 21-69) were recruited on Amazon Mechanical Turk through the research platform FindingFive (FindingFive Team, 2019). Criteria for participation specified in the pre-registration were: >95% acceptance rate, participating in the US, and did not participate in Experiment 1. Participants received $4 for ~30 minutes of participation. At the beginning of the experiment, we collected basic demographic information (native language, gender, age). Nine participants reported themselves as non-native speakers of English and were excluded from further analyses, as specified in the pre-registration. An additional two participants failed to answer comprehension questions with at least 85% accuracy and thus, were excluded from further analyses, as specified in the pre-registration. To replace the excluded participants, we recruited an additional 12 participants, which resulted in oversampling by 1 participant. All participants included in the final dataset reported themselves as native speakers of English.

Materials

Materials were identical to Experiment 1.

Procedure

Procedure was identical to Experiment 1.

3.2 Predictions

The results of Experiment 1 indicated a small, but significant memory benefit for disfluent pauses and fillers compared to repetitions. Experiment 2 is well powered to detect this effect, if it is in fact as large as estimated in Experiment 1. Such a finding would point to form-meaning mappings in the input as driving the disfluency-memory boost. Alternatively, if the disfluency-related memory boost is simply due to the fact that disfluency is distinctive, or because it creates a delay, all three types of disfluency would be expected to produce a similar memorial benefit.

3.3 Results

The data were analyzed in the same way as in Experiment 1. We fit a logistic mixed effect regression model, with the item status (actually old vs. actually new) as a factor, and then for old items, the presence and type of disfluency as predictors, random intercepts and slopes for memory and disfluency by subject and item, and the response (old vs. new) as dependent. The analyses were performed in RStudio (RStudio Team, 2018) using lme4 (Bates et al., 2015) package.

Table 3. Experiment 2 memory results: Mixed effect model with item status (old vs. new) and type of disfluency as fixed effects. The dependent measure is binary - whether the participant responded "old" or "new" on the memory test. Values in bold indicate significant results.

		Estimate	SE	z-value	p-value
Fixed effects	**(Intercept)**	**-0.18574**	**0.08230**	**-2.257**	**0.024**
	Memory effect (old = each 0.5, new = -0.5)	**1.38343**	**0.09629**	**14.368**	**< 2e-16**
	Disfluency effect (new = -0.125, fluent = -0.625, filler, pause, repetition = each 0.375)	**0.42382**	**0.06299**	**6.729**	**1.71e-11**
	Fillers and pauses vs. repetitions (new = -0.0625, fluent = -0.0625, filler, pause = each 0.4375, repetition = -0.5625)	0.04513	0.05003	0.902	0.367
	Fillers vs. pauses (new, fluent, repetition = each 0, filler = 0.5, pause = -0.5)	-0.04108	0.05781	- 0.710	0.477

		Variance	St. Dev.	Correlations	
Random effects	Item (Intercept)	0.2948	0.5430		
	Memory effect	0.8381	0.9155	0.31	
	Disfluency effect	0.2619	0.5117	-0.21	-0.30
	Subject (Intercept)	0.7539	0.8683		
	Memory effect	0.4670	0.6834	-0.11	

The results of this analysis are show in Table 3. A significant intercept term reflected a response bias to say "new" (b=-0.19, z=-2.26, p=0.02) rather than “old”. In addition, we replicated the finding that memory for words was overall good (b=1.38, z=14.37, $p<10_{-15}$), and the finding that words preceded by disfluency were more likely to be correctly recognized ($b=0.42$, $z = 6.73$, $p<10^{-10}$) (**Figure 3**). Surprisingly, however, pauses and fillers were not significantly better remembered than repetitions ($b=0.05$, $z=0.90$, $p=0.37$), thus failing to replicate the disfluency *type* effect observed in Experiment 1.

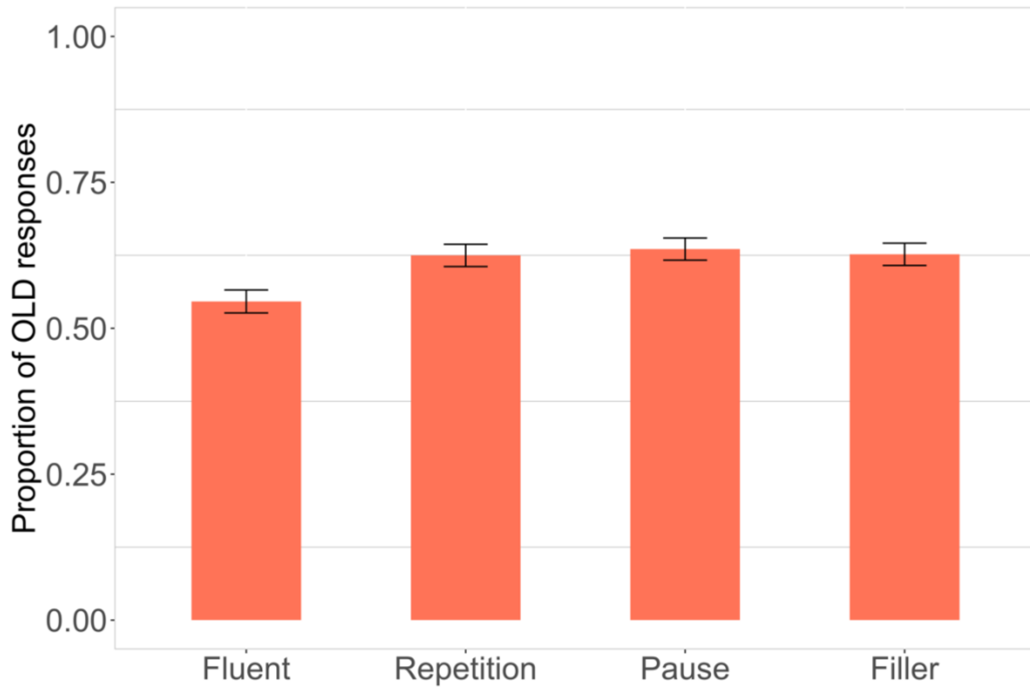

Figure 3. Memory results for Experiment 2. Error bars represent by-subject SEM.

3.4 Discussion

In Experiment 2, we found that disfluencies boost memory for immediately following words, replicating the same finding from Experiment 1 and prior work (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010). However, we failed to replicate the effect of disfluency type -- unlike Experiment 1, in Experiment 2 words that had been preceded by pauses and fillers were recognized at similar rates to words that had been preceded by disfluent repetitions (64% – for pauses, 63% – for fillers, 62% – for repetitions, vs. 55% for fluent utterances). These findings suggest that the three different types of disfluency that we tested have similar memorial benefits.

A post-hoc analysis using the combined data across the two experiments replicated the effect of disfluency ($b=0.37$, $z=7.23$, $p<10^{-12}$). Additionally, the combined analysis revealed that pauses and fillers are recognized significantly better than repetitions ($b=0.09$, $z=2.12$, $p=0.03$) however, the observed effect was small, corresponding to an odds ratio of only 1.09 in favor of pauses and fillers over repetitions.

Taken together, these findings provide only limited evidence in support for the hypothesis that different disfluencies signal different upcoming meanings. The fact that all three types of disfluency boosted memory to similar degrees is however, consistent with the attentional orienting hypothesis – the idea that disfluencies direct listener's attention to the upcoming context, and as a result, improve memory for the words that immediately follow.

CHAPTER 4

EXPERIMENT 3

Recall that prior work by Fraundorf and Watson (2011) found that plot points in spoken passages were more likely to be recalled when the passage contained disfluencies, and critically, that the disfluency boost was observed for *all* plot points -- not just the ones preceded by disfluency. Further, in a second experiment they found that this wide-scope disfluency boost to memory occurred even if the disfluencies were placed in atypical locations (in the middle of a plot point), as opposed to a more typical location for disfluency (right before a new plot point). Fraundorf and Watson (2011) concluded that disfluency acts to direct attention to the speech stream in general, but isn't necessarily used as a signal for listeners to predict specific upcoming material. These findings, along with the results of the present Experiment 2 that all three types of disfluency boosted memory, suggest that the disfluency boost in memory may not be strongly tied to form-meaning mappings in the input. While listeners may learn and use distributional cues related to disfluency when processing language in the moment (Arnold et al., 2007; Barr & Seyfeddinipur, 2010; Bosker et al., 2019; Yoon & Brown-Schmidt, 2020), the presence of disfluency may confer more general processing benefits.

In Experiments 1-2, the critical memory probe words were always immediately preceded by the disfluency, leaving open the question of whether the boost in word memory is localized to the word immediately following the disfluency, or whether the disfluency confers longer lasting word memory benefits. A critical feature of the present studies (also Corley et al., 2007; MacGregor, Corley, & Donaldson, 2010) is we probe recognition memory for individual words.

By contrast, Fraundorf and Watson (2011) measured gist recall of plot points in a passage. Potentially relevant here are findings from Corley and Hartsuiker (2011) who tested recognition of words directly preceded by disfluency, vs. cases where the disfluency occurred earlier in the sentence; they reported that word recognition was better when it was immediately preceded by the disfluency, compared to when the disfluency was earlier in the sentence. However, that study did not include a comparison case where sentences were fluent, making it difficult to know whether the early-in-the-sentence disfluencies also improved word recognition for sentence-final word compared to a fluent control. In sum, whether disfluency confers wide-scope memory benefits down at the level of individual words, is an open question. It is well known that sentence memory tends to be more gist-like, with participants remembering the gist and inferences more so than the individual words (Brewer, 1977; Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972; Sachs, 1974). Thus it remains plausible that the scope of the disfluency-memory boost may be limited to immediately following words for word memory, but broader in scope for gist-memory.

In sum, the primary goal of Experiment 3 was to examine the durability of the disfluency effect on memory for words in sentences. To this end, we manipulated the position of a disfluency in the sentence (early, middle, and final), and probed memory for the final word of the sentence as before. If the effect of disfluency on word memory has a wide scope, we predict a memory benefit for the memory probe word regardless of the disfluency position in a sentence. On the other hand, the effect of disfluency on word memory is local, we predict an effect of the disfluency only when it immediately precedes the probe word.

4.1 Methods

This study was pre-registered on the Open Science Framework (https://osf.io/d6248).

Participants

In planning this experiment, we initially conducted a power analysis using data simulation package simr (Green & MacLeod, 2016) in RStudio (RStudio Team, 2018) to determine the sample size needed to replicate the effect of disfluency type (i.e., pauses & fillers vs. repetitions) from the combined data from Experiment 1 and 2 (b=0.09). According to the power analysis, in order to replicate the effect of disfluency *type*, we needed to recruit over 300 participants to achieve at least 65% power, which the authors judged as not cost effective. As illustrated in more detail below, we therefore designed Experiment 3 to focus on the two disfluency types with the numerically largest memory effect (pauses and fillers), and selected a sample size to detect an effect of disfluency *position* (early, middle, or final). Recall that the power analysis used in planning Experiment 2 indicated that in order to replicate the larger effect of disfluency type reported in Experiment 1 (effect size b = 0.18) at 95% power, we needed to recruit 160 participants. Using this as a benchmark, we selected a sample size of 200 participants to increase our chances of detecting an effect of position, if it is present.

A total of 200 participants were recruited on Amazon Mechanical Turk through the research platform FindingFive (FindingFive Team, 2019). Criteria for participation specified in the pre-registration were: >95% acceptance rate, and participating in the US, and did not participate in Experiment 1 and or Experiment 2. Participants received $4 for ~30 minutes of participation. In the beginning of the experiment, we collected basic demographic information (native language, gender, age). Two participants reported themselves as non-native speakers of English and were excluded from the further analyses, consistent with the exclusion criteria specified in the pre-registration. In our pre-registration, we specified that participants would be excluded if they did not reach a criterion of 85% accuracy on the comprehension questions or the sound check questions. However, we relaxed the sound check criterion post-hoc to 60% because after data collection we discovered that one of the five sound-check questions was unclear. Thus, participants were included in the analyses if they answered at least 3/5 sound check questions correctly (60% accuracy). Fifty participants failed to answer the comprehension questions with at least 85% accuracy or the sound check questions with at least 60% accuracy, and were excluded from further analyses. We therefore recruited an additional 52 participants in order to achieve the planned sample size of 200 participants (103 female; mean age 37.83; range 21-76). All participants included in the final dataset reported themselves as native speakers of English. Informed consent was obtained in accordance with the Vanderbilt University IRB guidelines.

Materials

The sentences were the same as in Experiment 1 and 2. In Experiment 3, we only included sentences with fillers and pauses since these two types of disfluencies had resulted in numerically better performance across Experiment 1 and 2, and it was not possible to fully cross the three types of disfluency with three sentence positions with the number of stimulus materials we had available to us. The novel manipulation in Experiment 3 was the position of the disfluency in the sentence (initial, middle, and final). We created 7 versions of each of the 160 items (fluent, disfluent pause initial, disfluent pause middle, disfluent pause final, disfluent filler initial, disfluent filler middle, disfluent filler final), resulting in 1120 total sentences (**Table 4**). All sentences were distributed across 16 experimental lists (~13 participants per list) following a Latin square design such that each participant only heard a given sentence frame once. Following the design of Experiments 1 and 2, the final word in the sentence was always used as a memory probe.

The auditory stimuli were recorded by a female research assistant with a North American accent of English (due to scheduling constraints, this was a different speaker from Experiments 1-2). Fluent and disfluent sentences were recorded at a natural speaking rate. The research assistant was instructed to record the sentences to sound as naturally as possible.

Table 4. Example Stimulus Set for Experiment 3.

Condition	Sentence
Fluent	*My sister had a skiing accident and she broke her leg*
Filler Initial	*My um sister had a skiing accident and she broke her leg*
Pause Initial	*My … sister had a skiing accident and she broke her leg*
Filler Middle	*My sister had a skiing um accident and she broke her leg*
Pause Middle	*My sister had a skiing … accident and she broke her leg*
Filler Final	*My sister had a skiing accident and she broke her um leg*
Pause Final	*My sister had a skiing accident and she broke her … leg*

Procedure

At the beginning of the experiment, participants completed five microphone check questions to ensure that their headphones worked properly and to set the volume at a comfortable level. For each of the five microphone check questions, they heard one word and were asked to type it. There were two phases of the experiment. In Phase 1, participants listened to 80 fluent and disfluent audio sentences (20 fluent + 10 disfluent per each disfluent condition, i.e. 60 disfluent sentences in total). Participants could only advance to the next listening trial once the current audio sentence finished playing. Participants also had to answer 8 comprehension questions about the sentences to ensure that they were paying attention to the stimuli. The comprehension questions were randomly presented during the first phase of the experiment. To ensure that participants were paying attention to the stimuli, we instructed them that they needed to answer at least 85% of the questions correctly. In Phase 2, participants viewed 160 single-word probes, one at a time (80 old + 80 new), and were asked to indicate whether the probe was old (present in the sentences they just heard), or new. The entire study took approximately 30 minutes to complete.

4.2 Predictions

If the disfluency memory boost has a wide scope (Fraudorf & Watson, 2011), then we predict a similar memorial benefit for the probe words regardless of the position of a disfluency in a sentence. Alternatively, if the disfluency memory boost is local, we predict better memory for the probe words that are immediately preceded by a disfluency (final position) and worse memory for the probe words from the sentences where disfluency occurs early or in the middle of a sentence.

4.3 Results

As in Experiments 1-2, we used a signal-detection theoretic mixed-effects analysis (Wright, et al., 2009) for the response data. Based on our pre-registration, we fit 2 separate logistic mixed effect regression models to probe the effect of disfluency and the effect of disfluency position on participant responses (old vs. new). Model 1 included item status (old vs. new) as a factor, then for old items, the presence of disfluency as a predictor. Random effects included by-subject and by-item intercepts, and random slopes were included in the model if the model converged with them. Model 2 analyzed old items only, and included disfluency type as a centered fixed effect (pauses vs. fillers). Position of the disfluency in the sentence was dummy coded, with late-position disfluencies as the reference level. Because of the unequal number of sentences in each condition, we used weighted contrast coding. Both analyses were performed in RStudio (RStudio Team, 2018) using lme4 (Bates et al., 2015) package.

The results of Model 1 are shown in **Table 5.** A significant intercept term reflected a response bias to say "new" ($b=-0.18$, $z=-2.32$, $p=0.02$); participants also had overall good memory for the probe words ($b=1.23$, $z=16.72$, $p<10^{-15}$). In contrast, the effect of disfluency on memory ($b=0.11$, $z=1.66$, $p=0.10$) (**Figure 4**) was not significant, though as we shall see, this is likely due to the fact that it includes all three disfluency positions. The results of Model 2 are shown in **Table 6**. Focusing on the old items only, Model 2 revealed a significant effect of disfluency position such that disfluencies at the end of the sentence produced better memory for the probe words than disfluencies at the beginning of the sentence ($b=-0.23$, $z=-3.20$, $p=0.001$). When comparing disfluencies at the end of the sentence with the disfluencies at the middle of the sentence, we do not find a significant difference ($b=-0.14$, $z=-1.48$, $p=0.14$). A supplemental analysis directly compared each disfluency position to the fluent sentences (**Table 7**). This analysis revealed that the disfluencies at the end of the sentence significantly improved memory relative to fluent control ($b=0.28$, $z=3.21$, $p=0.001$), thus replicating the effect of disfluency on memory found in Experiments 1 and 2. On the contrary, while sentences with disfluencies early ($b=0.04$, $z=0.59$, $p=0.55$) and in the middle ($b=0.04$, $z=0.52$, $p=0.60$) produced numerically better memory than fluent sentences, neither effect was significant.

Table 5. Experiment 3 memory results: Mixed effect model 1 with item status (old vs. new) and presence of disfluency as fixed effects. The dependent measure is binary - whether the

participant responded "old" or "new" on the memory test. Values in bold indicate significant results.

		Estimate	SE	z-value	p-value
Fixed effects	**(Intercept)**	**-0.17553**	**0.07574**	**-2.318**	**0.0205**
	Memory effect (fluent, disfluent = 0.5, new = -0.5)	**1.22738**	**0.07339**	**16.724**	**<2e-16**
	Disfluency effect (fluent = -0.625, disfluent = 0.375, new = -0.125)	0.11395	0.06885	1.655	0.0979

		Variance	St. Dev.	Correlations	
Random effects	Item (Intercept)	0.8835	0.9400		
	Memory effect	0.2576	0.5075	-0.24	
	Disfluency effect	0.0311	0.1764	-0.09	-0.39
	Subject (Intercept)	0.1806	0.4250		
	Memory effect	0.5208	0.7217	0.35	
	Disfluency effect	0.4462	0.6680	-0.08	-0.16

Table 6. Experiment 3 memory results: Mixed effect model 2 with disfluency type (filler vs. pause) and position of disfluency as fixed effects. The dependent measure is binary - whether the participant responded "old" or "new" on the memory test. Values in bold indicate significant results.

		Estimate	SE	z-value	p-value
Fixed effects	**(Intercept)**	**0.64324**	**0.08842**	**7.275**	**3.47e-13**
	Disfluency type (pauses = -0.5, fillers = 0.5)	-0.08517	0.05113	-1.66	0.09575
	Early vs. late (early pause and filler = 1, middle and final pause and filler = 0)	**-0.23381**	**0.07305**	**-3.201**	**0.00137**
	Middle vs. late (middle pause and filler = 1, early and final pause and filler = 0)	-0.13947	0.09434	-1.478	0.13930

		Variance	St. Dev.	Correlations	
Random effects	Item (Intercept)	0.3885	0.6233		
	Early vs. late	0.2712	0.5208	-0.16	
	Middle vs. late	0.4069	0.6379	0.06	0.75
	Subject (Intercept)	0.7943	0.8912		
	Early vs. late	0.1803	0.4246	-0.21	
	Middle vs. late	0.2796	0.5287	-0.08	0.91

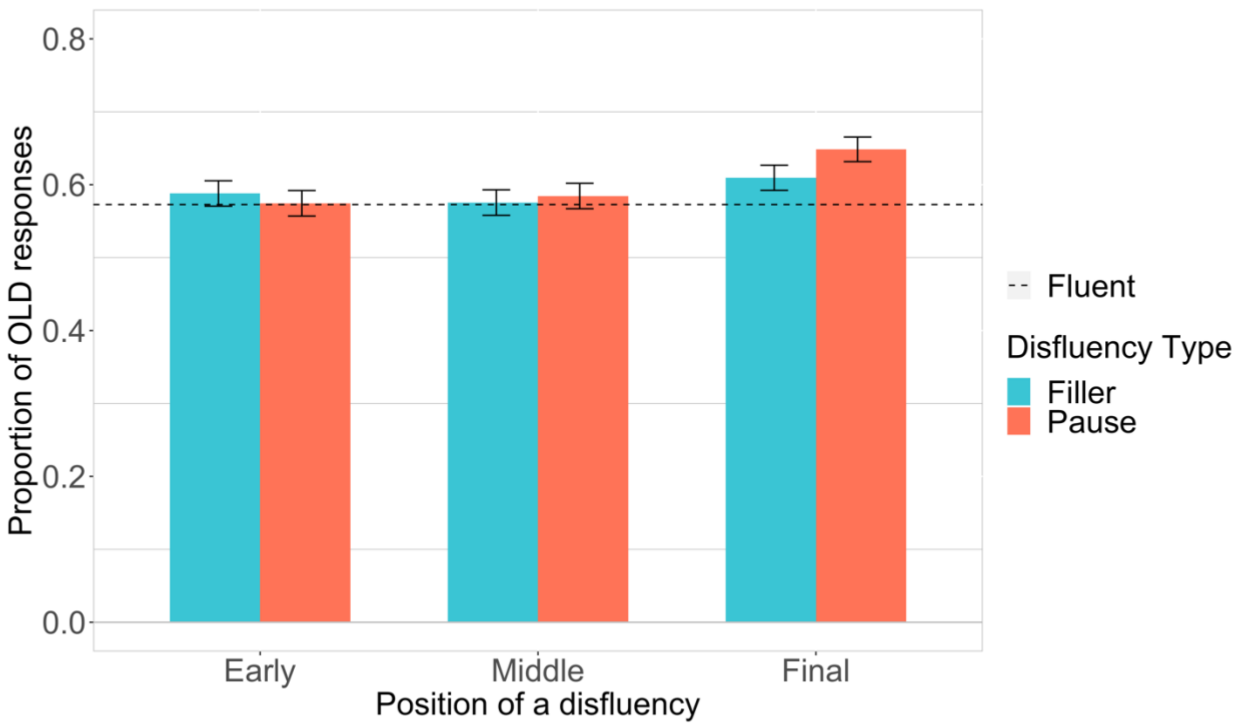

Figure 4. Memory results for Experiment 3. Error bars represent by-subject SEM.

Table 7. Experiment 3 memory results: Supplemental mixed effect model with position of disfluency as a fixed effect. The dependent measure is binary - whether the participant responded "old" or "new" on the memory test. Values in bold indicate significant results.

		Estimate	SE	z-value	p-value
Fixed	**(Intercept)**	**0.36480**	**0.09904**	**3.683**	**0.00023**

	Early (early = 1, middle, final, fluent = 0)	0.04268	0.07202	0.593	0.55341
	Middle (middle = 1, early, final, fluent = 0)	0.03928	0.07543	0.521	0.60252
	Final (final = 1, early, middle, fluent = 0)	**0.27647**	**0.08609**	**3.212**	**0.00132**

		Variance	St. Dev.	Correlations		
Random effects	Item (Intercept)	0.6706	0.8189			
	Early	0.4054	0.6367	-0.49		
	Middle	0.4861	0.6972	-0.49	0.92	
	Final	0.6303	0.7939	-0.70	0.74	0.87
	Subject (Intercept)	0.8520	0.9230			
	Final	0.1444	0.3800	-0.28		

4.4 Discussion

In Experiment 3, we were tested the scope of the beneficial effect of disfluency on memory for words in sentences. Consistent with our findings from Experiments 1 and 2, we found that both pauses and fillers improved memory for the subsequent word. Critically, however, we found that this effect was short lived. Only disfluencies that immediately preceded the probe word significantly improved memory over fluent sentences.

CHAPTER 5

GENERAL DISCUSSION

Disfluencies are ubiquitous in spontaneous speech. Their use in speech is patterned, with pauses and fillers occurring before the articulation of a new message and repetitions occurring when the articulation has already begun and the planned material is available to be repeated (Fraundorf & Watson, 2011). Disfluencies also inform language processing both online and offline (Fox Tree, 1995; Bailey & Ferreira, 2003; Corley & Stewart, 2008; Heller, et al 2015), and allow listeners to make predictions about upcoming material (Arnold et al., 2004; Arnold et al., 2007; Lowder & Ferreira, 2016a,b). Studies investigating the impact of disfluency on memory for language offered initial evidence that some but not all disfluencies have a beneficial effect on memory, with some studies showing a memory boost for fillers and pauses (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010), and others showing no such benefit for repetitions (MacGregor, Corley, & Donaldson, 2009). Here we show that three types of disfluency (pauses, fillers, and repetitions) confer a clear but short-lived memorial boost for subsequent words.

In Experiments 1-2, we tested the hypothesis that different forms of disfluency, due to their different patterns of use in language production (Fraundorf & Watson, 2011), would differentially signal meanings to listeners. If so, we hypothesized that forms of disfluency that signal new information (pauses and fillers) would result in more attentional orienting, and a larger memory boost compared to forms associated with speaker difficulty for already planned material (repetitions). While the results of Experiment 1 were potentially consistent with this hypothesis, in Experiment 2 we did not observe a significant form-specific benefit. While a combined analysis of Experiments 1-2 did reveal a significant form-specific effect, the estimated effect size was small and impractical to pursue (it would require 300 participants to achieve only 65% power in a replication attempt). Further, the power analysis based on the combined data across the two studies indicates that Experiment 1 (N=102) only had ~26.2% power to detect the form-specific effect, indicating that Experiment 1 was underpowered to detect the effect, if it in fact exists (see relevant discussion in Simonsohn, 2015 and Simonsohn, Nelson, & Simmons, 2014). Taken together, then, what we can clearly conclude is that these three types of disfluency improve memory for immediately following words, regardless of the fact they are used in different ways by speakers. Across the two studies, the odds of correct recognition was 1.45 times higher for disfluent compared to fluent utterances. By contrast, based on the combined data for Experiments 1-2, fillers and pauses increased the odds of correct recognition by only 1.09 over disfluent repetitions. In Experiment 3 we determined that the scope of this disfluency-based memorial benefit is short-lived, primarily extending to the immediately following word.

Taken together, our findings are clear that disfluencies boost memory for the subsequent word, however we find only weak evidence for the hypothesis that the disfluency memory boost is related to form-meaning mappings in the input. Disfluencies certainly shape online processing of language (Fox Tree, 1991, 1995; Bailey & Ferreira, 2003, Arnold et al., 2004; Corley, MacGregor, & Donaldson, 2007), consistent with the idea that they act as a signal to meaning (Clark & Fox Tree, 2002). However, pauses and fillers did not confer a consistently stronger memory boost compared to disfluent repetitions, suggesting that the probabilistic link between new information and disfluent pauses and fillers, and already planned material with repetitions (Fraundorf & Watson, 2014) did not result in attentional or processing differences that shaped memory. While language comprehension often involves predictive processing (Altmann & Kamide, 1999; Pickering & Garrod, 2013; Dell & Chang, 2013; Federmeier, 2007), predictions many not be necessary or universal (Pickering & Gambi, 2018; Huetting & Mani, 2015), and predictions may be attenuated if they do not confer a clear processing benefit (Ryskin, Levy, & Fedorenko, 2019). Likewise, evidence from the directed forgetting literature indicates that participants can choose to temporarily ignore or set aside information that they are nonetheless able to later recognize if needed (Elmes, Adams, & Roediger, 1970). Thus, the lack of a strong disfluency type effect in the present work may be due to participants choosing not to use the cues, rather than a lack of awareness about them. The disfluency-related memory boost that we observed, rather than driven by form-meaning mappings in the input, may instead be due to the fact that disfluency is distinctive (e.g. von Restorff, 1933; Dalton & Lavie, 2004; Fabiani, Karis, & Donchin, 1986), or because it creates a delay (Corley and Hartsuiker, 2011).

Several open questions, however, remain. First, if it is the distinctiveness of the disfluency itself that captures attention, it would remain unexplained then, why it is the following word that is more likely to be correctly recognized (Fox Tree, 2001; Corley & Hartsuiker, 2011), and remembered (Corley, et al 2007; MacGregor, Corley, & Donaldson, 2010). Primary distinctiveness of stimuli tends to orient attention to the deviant stimulus, improving memory for that stimulus (von Restorff, 1933), speeding recognition of it (Dalton & Lavie, 2004), and changing the real-time processing of that stimulus in a way that is linked to subsequent memory benefits (Fabiani, Karis, & Donchin, 1986). If the disfluency itself captures attention, why does it not create an attentional blink for the subsequent stimulus (Shapiro, Arnell, & Raymond, 1997)? For example, in studies of auditory attentional capture, Dalton & Lavie (2004) demonstrated that irrelevant feature singletons captured attention, and were associated with behavioral costs as participants searched for a target auditory stimulus. They concluded that auditory attentional capture occurs in the presence of unique perceptual objects because they are acoustically distinct from the background noises and as a result, might indicate an important change in the environment. Following this logic, disfluencies should capture auditory attention, and impair recognition of and memory for the following material. The fact that we observe the opposite pattern -- that disfluency orients attention to the upcoming word without necessarily capturing it -- points to a different mechanism. Alternatively, if the observed disfluency-memory boost is due to the delay (Corley and Hartsuiker, 2011) or interruption caused by the disfluency, it is not clear why some studies find that non-linguistic delays do not yield a memory boost (e.g. coughs, Fraundorf and Watson, 2011). We also note that in at least one study that directly compared disfluencies of different lengths (Collard, 2009, Experiment 4), longer disfluencies were not associated with stronger attentional orienting (measured in a change-detection paradigm).

A second question is how the localized memorial boost observed in the present studies and prior work (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010) relates to the more general boost to gist memory for story plot points reported by Fraundorf and Watson (2011). One possibility is that disfluency does confer local benefits in word recognition and processing, and that these local benefits to word level processes have cascading effects, thereby boosting sentence or propositional level gist memory. If so, this would suggest that one could observe a disfluency-conferred boost in gist memory for a sentence when, if tested on a word-by-word basis, it is only the immediately following word that itself sees that disfluency benefit at a word memory level.

Third, our findings raise the possibility that disfluencies of different types might not be as distinct as previously thought. The results of Experiments 1 and 2 provide insufficient evidence for the differential effect of the three different disfluency types on memory. Consequently, in the context of memory, different disfluency types should be treated similarly. An open question is whether such a distinction might be more meaningful in the context of online language comprehension. For example, Fox Tree (1991) found differential effects of different fillers (*ums* and *uhs*) on subsequent word recognition, leaving open the possibility that different disfluency types indeed vary in terms of their impact on language processing.

Finally, our study focused on language comprehension and memory. Whether *producing* disfluency in spontaneous speech results in a similar beneficial effect on memory remains an open question. Fraudorf and Watson (2011) suggested that different disfluency types reflect different difficulties that speakers experience in language production. It is feasible that different production difficulties that result in different disfluencies might have a differential long-term effects for memory.

5.1 Conclusion

Considerable evidence now shows that disfluencies, rather than a nuisance to be ignored, are used in meaningful ways by speakers, and are used as a cue to guide online processing of language. An emerging body of research shows that in addition to these real-time processing effects, disfluencies also have downstream consequences, shaping the enduring memory of the discourse. The present research explored two key questions regarding the nature of the disfluency-related memory boost. Given that there exist regularities in form-meaning mappings in the way different forms of disfluency are used in spoken language, we asked if different forms of disfluency would result in different levels of a memory boost. Tentative evidence for such a link existed in the literature, with pauses and repetitions (which are linked to the planning of new information) linked to a memory boost, but not repetitions (which signal speaker difficulty). In a pair of well-powered studies we found only weak evidence at best for this hypothesis. Instead, disfluent pauses, fillers, and repetitions all conferred a memory boost to the immediately following word. The results of Experiment 3 clarified that this memory boost was short-lived, manifesting most strongly on the immediately following word. Taken together, our findings reveal a disfluency boost in memory for words that is short lived but evoked by multiple types of disfluent forms, consistent with the idea that disfluencies bring attentional focus to immediately upcoming material. We speculate that the downstream consequence of this localized memorial benefit to individual words that immediately follow disfluent interruptions, is better understanding and encoding of the speaker's message as a whole.

REFERENCES

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If You Say Thee uh You Are Describing Something Hard: The On-Line Attribution of Disfluency During Reference Comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(5), 914–930. https://doi.org/10.1037/0278-7393.33.5.914

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*(9), 578–582. https://doi.org/10.1111/j.0956-7976.2004.00723.x

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*(2), 183–200. https://doi.org/10.1016/S0749-596X(03)00027-5

Bard, E. G., & Lickley, R. J. (1997). On not remembering disfluencies. *Eurospeech'97 Proceedings*, *January*, 2855–2858.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441–455. https://doi.org/10.1080/01690960903047122

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. doi: 10.18637/jss.v067.i01.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*(3), 201–211. https://doi.org/10.1177/002383097902200301

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 54.

Bortfeld, H., Leon, Ê. S. D., Brennan, S. E., Bloom, J. E., & Schober, Ë. M. F. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, *44*(2), 123–147.

Bosker, H. R., van Os, M., Does, R., & van Bergen, G. (2019). Counting 'uhm's: How tracking the distribution of native and non-native disfluencies influences online language comprehension. *Journal of Memory and Language*, *106*(January), 189–202. https://doi.org/10.1016/j.jml.2019.02.006

Branigan, H., Lickley, R., & McKelvie, D. (1999, August). Non-linguistic influences on rates of disfluency in spontaneous speech. In *Proceedings of the 14th International Conference of Phonetic Sciences* (pp. 387-390).

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive psychology*, *3*(2), 193-209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive psychology*, *2*(4), 331-350.

Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition*, *5*(6), 673-678.

Brennan, S.E., & Williams, M. (1995). The Feeling of Another′s Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. In *Journal of Memory and Language* (Vol. 34, Issue 3, pp. 383–398). https://doi.org/10.1006/jmla.1995.1017

Brennan, S. E., & Schober, M. F. (2001). How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, *44*(2), 274–296. https://doi.org/10.1006/jmla.2000.2753

Brosy, J., Bangerter, A., & Mayor, E. (2016). Disfluent responses to job interview questions and what they entail. *Discourse Processes*, *53*(5-6), 371-391.

Christensen, T. A., Almryde, K. R., Fidler, L. J., Lockwood, J. L., Antonucci, S. M., & Plante, E. (2012). Modulating the focus of attention for spoken words at encoding affects frontoparietal activation for incidental verbal memory. *International journal of biomedical imaging*, *2012*.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H., & Wasow, T. (1998). Repeating Words in Spontaneous Speech. *Cognitive Psychology*, *37*(3), 201–242. https://doi.org/10.1006/cogp.1998.0693

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention Orienting Effects of Hesitations in Speech: Evidence From ERPs. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(3), 696–702. https://doi.org/10.1037/0278-7393.34.3.696

Corley, M., & Hartsuiker, R. J. (2011a). Hesitation in speech can. . . um. . . help a listener understand. *Human Communication Research*.

Corley, M., & Hartsuiker, R. J. (2011b). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019792

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Linguistics and Language Compass*, *2*(4), 589–602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668. https://doi.org/10.1016/j.cognition.2006.10.010

Dalton, P., & Lavie, N. (2004). Auditory Attentional Capture: Effects of Singleton Distractor Sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 180–193. https://doi.org/10.1037/0096-1523.30.1.180

Dell, G. S., & Chang, F. (2013). The P-chain: relating sentence production and its disorders to comprehension and acquisition. Philos. Trans. R. Soc. Biol. Sci. 369 (1634), 20120394. https://doi.org/10.1098/rstb.2012.0394.

Donahue, J., Schoepfer, C., & Lickley, R. (2017). *The effects of disfluent repetitions and speech rate on recall accuracy in a discourse listening task*. *August*.

Elmes, D. G., Adams, C., & Roediger, H. L. (1970). Cued forgetting in short-term memory: Response selection. *Journal of Experimental Psychology*, *86*(1), 103.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. Psychophysiology 44 (4), 491–505. https://doi.org/10.1111/ j.1469-8986.2007.00531.x.

Ferreira, F., & Patson, N. D. (2007). The "Good Enough" Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x

Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and Tree Adjoining Grammars. *Cognitive Science*, *28*(5), 721–749. https://doi.org/10.1016/j.cogsci.2003.10.006

FindingFive Team (2019). FindingFive: A web platform for creating, running, and managing your studies in one place. FindingFive Corporation (nonprofit), NJ, USA. https://www.findingfive.com

Fox Tree, J. E. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. In *Journal of Memory and Language* (Vol. 34, Issue 6, pp. 709–738). https://doi.org/10.1006/jmla.1995.1032

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*(2), 151-167.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, *29*(2), 320–326. https://doi.org/10.3758/BF03194926

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language*, *65*(2), 161–175. https://doi.org/10.1016/j.jml.2011.03.004

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, *29*(9), 1083–1096. https://doi.org/10.1080/01690965.2013.832785

Green P, MacLeod CJ (2016). "simr: an R package for power analysis of generalised linear mixed models by simulation." *Methods in Ecology and Evolution*, **7**(4), 493–498. doi: [10.1111/2041-210X.12504](10.1111/2041-210X.12504), https://CRAN.R-project.org/package=simr.

Green, T. J., & McKeown, J. D. (2001). Capture of attention in selective frequency listening. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1197–1210. https://doi.org/10.1037/0096-1523.27.5.1197

Grosman, I. (n.d.). *Complexity Cues or Attention Triggers? Repetitions and Editing Terms for Native Speakers of French*. http://www.disfluency.org/DiSS_2015/Programme_files/Grosman-DISS2015.pdf

Heller, D., & Chambers, C. G. (2014). Would a blue kite by any other name be just as blue? Effects of descriptive choices on subsequent referential behavior. *Journal of Memory and Language*, *70*(1), 53–67. https://doi.org/10.1016/j.jml.2013.09.008

Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K. (2015). Inferring difficulty: Flexibility in the real-time processing of disfluency. *Language and Speech*, *58*(2), 190-203.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience, 31*(1), 19-31.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(1), 29–53. https://doi.org/10.1177/00238309050480010201

Lowder, M. W., & Ferreira, F. (2016a). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1400.

Lowder, M. W., & Ferreira, F. (2016b). Prediction in the processing of repair disfluencies. *Language, cognition and neuroscience, 31*(1), 73-79.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2009). Not all disfluencies are are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language*, *111*(1), 36–45. https://doi.org/10.1016/j.bandl.2009.07.003

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, *48*(14), 3982–3992. https://doi.org/10.1016/j.neuropsychologia.2010.09.024

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, *3*(6), 427–438. https://doi.org/10.3758/BF03205750

Maxfield, N. D., Lyon, J. M., & Silliman, E. R. (2009). Disfluencies along the garden path: Brain electrophysiological evidence of disrupted sentence processing. *Brain and Language*, *111*(2), 86–100. https://doi.org/10.1016/j.bandl.2009.08.003

Miller, S. (2018). *Disfluency, prediction and attention in language*.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Quinlan, P. T., & Bailey, P. J. (1995). An examination of attentional control in the auditory modality: Further evidence for auditory orienting. *Perception & Psychophysics*, *57*(5), 614–628. https://doi.org/10.3758/BF03213267

Race, D. S., & Macdonald, M. C. (2003). *The use "that" in the Production and Comprehension of Object Relative Clauses*. *25*, 25. https://cloudfront.escholarship.org/dist/prd/content/qt5n3394p8/qt5n3394p8.pdf

RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, 107258.

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, *2*(1), 95-100.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, *60*(3), 362–367. https://doi.org/10.1037/0022-3514.60.3.362

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in cognitive sciences*, *1*(8), 291-296.

Shen, D., Ross, B., & Alain, C. (2016). Temporal cuing modulates alpha oscillations during auditory attentional blink. *European Journal of Neuroscience*, *44*(2), 1833–1845. https://doi.org/10.1111/ejn.13266

Shriberg, E. (1996, October). Disfluencies in switchboard. In *Proceedings of International Conference on Spoken Language Processing* (Vol. 96, pp. 11-14). Philadelphia, PA: IEEE.

Smith, V. L., & Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, *32*, 25–38.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, *143*(2), 534.

Spence, C. J., & Driver, J. (1994). Covert spatial attention in audition: {E}xogenous and endegenous mechnisms. *J. Exp. Psychol.*, *20*(3), 555–574.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94. https://doi.org/10.1016/j.jml.2005.02.003

van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology*, *19*(4), 403–440. https://doi.org/10.1016/0010-0285(87)90014-4

Walker, E. J., Risko, E. F., & Kingstone, A. (2014). Fillers as signals: Evidence from a question–answering paradigm. *Discourse Processes*, *51*(3), 264-286.

Wallace, W. P., Shaffer, T. R., Amberg, M. D., & Silvers, V. L. (2001). Divided attention and prerecognition processing of spoken words and nonwords. *Memory & cognition*, *29*(8), 1102-1110.

Yoon, S. O. & Brown-Schmidt, S. (July, 2020). Partner-specific adaptation in disfluency processing. Proceedings of the Cognitive Science Society.