

This is the accepted version of the following article: Diachek, E. & Brown-Schmidt, S. (2023). The effect of disfluency on memory for what was said. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(8), 1306–1324, which has been published in final form at <https://doi.org/10.1037/xlm0001156>

The Effect of Disfluency on Memory for What was Said

Evgeniia Diachek

Sarah Brown-Schmidt

Vanderbilt University

Corresponding Author:

Evgeniia Diachek

Vanderbilt University Department of Psychology and Human Development

230 Appleton Place · Nashville, TN 37203, USA

857-265-4471

evgeniia.diachek@vanderbilt.edu

Abstract

Disfluencies such as pauses, *ums* and *uhs* are common interruptions in the speech stream. Previous work probing memory for disfluent speech shows memory benefits for disfluent compared to fluent materials. Complementary evidence from studies of language production and comprehension have been argued to show that different disfluency types appear in distinct contexts and as a result serve as a meaningful cue. If the disfluency-memory boost is a result of sensitivity to these form-meaning mappings, forms of disfluency that cue new upcoming information (fillers and pauses) may produce a stronger memory boost, compared to forms which reflect speaker difficulty (repetitions). If the disfluency-memory boost is simply due to the attentional-orienting properties of a disruption to fluent speech, different disfluency forms may produce similar memory benefit. Experiments 1 and 2 compare the relative mnemonic benefit of three types of disfluent interruptions. Experiments 3 and 4 examine the scope of the disfluency-memory boost to probe its cognitive underpinnings. Across the four experiments, we observed a disfluency-memory boost for three types of disfluency that were tested. This boost was local and position-dependent, only manifesting when the disfluency immediately preceded a critical memory probe word at the end of the sentence. Our findings reveal a short-lived disfluency-memory boost that manifests at the end of the sentence but is evoked by multiple types of disfluent forms, consistent with the idea that disfluencies bring attentional focus to immediately upcoming material. The downstream consequence of this localized memory benefit is better understanding and encoding of the speaker's message.

Keywords: disfluency, memory, pause, repetition, filler

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). The present research examines the impact of this common feature of every-day language use on enduring representations of what was said. Patterns of disfluencies in speech show that speakers tend to produce different types of disfluencies in different contexts (Bortfeld 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 spoken material has already been planned and articulated. 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. Other research examining patterns of different disfluency types in naturalistic speech revealed converging evidence that fillers and silent pauses tend to occur between units of language planning. For example, Clark and Fox Tree (2002) found that fillers were more likely to occur between sentence boundaries (i.e., prior to large new chunks of information) as opposed to within sentence boundaries. In the case of silent pauses, Butterworth (1980) found that hesitations in the form of silent pauses in spontaneous speech coincided with the clause boundaries in the transcripts of speech. Taken together, if different forms of disfluency are used in different contexts, this raises the possibility that different forms of disfluency cue 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.

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 electrophysiological evidence indicates that disfluencies can help listeners predict or recognize upcoming information (Arnold et al., 2004), reduce surprisal for unpredictable words (Corley et al., 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 offer a meaningful cue about upcoming information (such as whether it is unpredictable or new) and by doing so, can guide predictive language processing.

Indeed, some prior work shows that listeners track the distributional properties of disfluencies in speech and later, utilize this knowledge during language processing. For example, Bosker et al. (2019) presented participants with spoken passages with either a typical distribution of

disfluencies (*uhs* before low-frequency words) or an atypical distribution (*uhs* before high-frequency words). They found that listeners increased their anticipatory fixations either on the low-frequency or high-frequency words depending on which distribution they were exposed to. These findings suggest that listeners track and update the distribution of disfluencies in natural speech, and that this distributional information subsequently modulates predictive processes during language comprehension. In contrast, across multiple experiments, Karimi, Brothers, and Ferreira (2019) collected judgements of sentence naturalness and sentence continuations for the stimuli that contained phonological and semantic repairs following either disfluency or focus constructions. They found that participants preferred phonological repairs following disfluency (e.g., *a hammer, uh I mean a hammock*), but semantic repairs following focus constructions (e.g., *not the hammer but the nail...*). Interestingly, in two eye-tracking studies of online language processing reported in the same paper, the authors found that participants did not use this information to guide their real-time predictions. These findings indicate that even when listeners possess the knowledge about the distributional properties of disfluencies, they do not necessarily use it to adjust their predictions in language comprehension. While it is not clear under which circumstances listeners do use the knowledge of distributional regularities to modulate their predictions during language processing, together, these studies indicate that listeners sometimes track and represent knowledge about the distribution of disfluencies in natural speech.

In addition to playing a role in online language processing, preliminary evidence also suggests that disfluencies impact memory for what was said (Corley et al., 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 (*ums* and *uhs*), 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 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 et al., 2007), and words preceded by a pause (MacGregor, Donaldson, & Corley, 2010). However, in a separate study, no mnemonic advantage 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, Schoepfer, and Lickley (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.

Attention, Memory, and Disfluency

The underlying cognitive mechanism that links disfluencies and enhanced memory remains poorly characterized in the literature. One mechanism we can consider and rule out relates to the fact that disfluent words are inherently distinct in the speech stream. It is true that primary distinctiveness (von Restorff, 1933) enhances attention. However, distinctiveness focuses attention on the deviant stimulus, improving memory (von Restorff, 1933), speeding recognition (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). Following this logic, if the disfluency itself captures attention, it should also create an attentional blink for the subsequent stimulus (Collard et al., 2008; Shapiro, Arnell, & Raymond, 1997). In fact, in studies of auditory attentional capture, Dalton and 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. If disfluencies capture auditory attention in the same way as auditory oddballs, this would predict that disfluency would *impair* recognition of and memory for the following material. The fact that previous studies examining the effect of disfluency on memory report the opposite pattern points to a different mechanism. Indeed, neurophysiological evidence shows that the ERP components typically elicited by oddball stimuli – the mismatch negativity (MNN) and P300 – are *reduced* (rather than enhanced) following disfluency, offering clear evidence against the idea that disfluency creates an attentional blink (Collard et al., 2008).

Here we consider two alternative explanations of the disfluency-related memory boost. The first possibility is that disfluency simply *orients* auditory attention (Addleman & Jiang, 2019) to the

speech stream. Unlike attention capture – a phenomenon when attention diverts from the primary cue by an irrelevant or unexpected event -- attention orienting, on the contrary, improves accuracy and speeds up reaction times in response to perceptual cues (Posner, 1980; Posner & Petersen, 1990; Posner, Snyder, & Davidson, 1980). Indeed, attentional capture and attentional orienting have been classified as two distinct types of attention shifts with the former belonging to high-level abilities, and the latter – to low level abilities (James, 1890). More evidence suggesting that disfluency might orient but not capture attention was offered by the study from Fraundorf and Watson (2011). The authors found that disfluency improved memory for spoken passages even when it appeared in unpredictable or atypical locations. They ruled out the hypothesis that listeners predict upcoming information using their knowledge about the distributional properties of disfluencies and concluded that disfluencies simply orient attention to the upcoming linguistic material. If so, the disfluency may act as a type of auditory orienting cue (Collard et al., 2008; also see Quinlan & Bailey, 1995; Spence & Driver, 1994) that for a period of time directs attention to the unfolding auditory stimulus. Under this attentional-orienting view, disfluency enhances attention to the upcoming word, resulting in improved perception and encoding of that word into memory (see Gazzaley & Nobre, 2012). Consistent with the 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.

An alternative possibility is that listeners track the distributional properties of disfluencies, learning where they are likely to occur in speech, and then use this distributional knowledge of form-meaning mappings as a cue to guide expectations about the unfolding speech stream.

Recall that Fraundorf and Watson (2014) found that different disfluency types occur in different contexts, suggesting that they might be used by listeners as a cue to upcoming meanings. Indeed, previous research shows that listeners are sensitive to statistical regularities in the linguistic input, shaping expectations about upcoming material (e.g., Altmann & Kamide, 1999; Ryskin et al., 2017). Further, listeners are sensitive to the distributional properties of disfluencies (Bosker et al., 2019; Karimi et al., 2019). Potentially consistent with this meaning-based account of the disfluency-memory boost is work reporting memory benefits for disfluent fillers and pauses (Corley et al., 2007; MacGregor et al., 2010), but not disfluent repetitions (MacGregor et al., 2009). Taken together, it is plausible that listeners recognize disfluent pauses and fillers to be a cue that new information is upcoming. Under this meaning-based view, when the listener hears a filler or pause, these form-meaning mappings cue the listener that new information is upcoming, supporting processing and encoding of this information. The downstream consequence of such improved encoding is better memory.

Present Research

The present research probes the basis and scope of the memory boost associated with disfluent speech. While prior work reports a memory benefit for disfluent fillers and pauses (Corley et al., 2007; MacGregor et al., 2010; Fraundorf & Watson, 2011), but no such benefit for disfluent repetitions (MacGregor et al., 2009), to our knowledge, the mnemonic advantages of the three disfluency types have yet to be directly compared in a well-powered study. Additionally, the cognitive mechanism that drives this disfluency-related memory boost remains debated in the literature. Here, we consider two competing hypotheses to explain the disfluency-memory boost. According to the attentional-orienting account, disfluency orients attention to the upcoming speech stream, improving recognition of subsequent words, and consequently, boosting memory for them. Alternatively, according to the meaning-based account, different disfluency types are associated with different meanings based on their distributional properties in language

production and as a result, have differential processing and mnemonic benefits. To test these hypotheses, we conducted two experiments. In Experiment 1 we directly compare three types of disfluency: fillers (*her um leg*), disfluent pauses (*her ... leg*), and disfluent repetitions (*her ... her leg*). If the mnemonic benefits are meaning-based, we expect to find better recognition memory for fillers and pauses compared to disfluent repetitions since these two types of disfluency are associated with new information, whereas repetition-type disfluencies are not associated with new information. If disfluency simply orients attention to the upcoming speech, all three types of disfluency would be expected to boost memory for the words that follow. Experiment 2 was a pre-registered direct replication of Experiment 1.

A related question concerns how long or at what grain size the disfluency-memory boost operates. Previous studies examining attentional orienting found that this effect is transient, dissolving rapidly after the stimulus onset (Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Spence & Driver, 1994; Mulckhuyse & Theeuwes 2010). If attentional orienting is the mechanism underlying the disfluency-related memory boost, it should be short lived. In Experiment 3, we therefore probe the scope of the disfluency-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 et al., 2007; MacGregor et al., 2010), studies probing gist memory for plot points (Fraundorf & Watson, 2011) find a general memory boost throughout the story. Thus, whether disfluencies boost memory for words in sentences regardless of where disfluency occurs in the sentence, remains an open question. Finally, Experiment 4 evaluates if the magnitude of the disfluency-related memory boost is equivalent across multiple positions in the sentence. According to findings from Fraundorf and Watson (2011), disfluency improves memory for linguistic content regardless of its position, suggesting that the disfluency memory boost can occur at any place in the sentence. Simultaneously, the serial position effect

documents increased memorability of the list-initial and list-final words (Deese & Kaufman, 1957; Murdock, 1962) suggesting that the position of disfluency in the sentence might modulate the memory boost effect.

Experiment 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 et al. (2007) with at least 95% power. According to the analysis, our target sample size was 100 participants¹. A total of 110 participants 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. 8 participants failed to answer sound and / or comprehension questions correctly and were excluded from final analysis, leaving 102 participants (50 female; mean age 36.1; range 21-62). Informed consent was obtained in accordance with the Vanderbilt University IRB guidelines.

¹ To determine the target sample size, the reported effect size for the disfluency-memory boost ($\eta^2 = 0.031$) was converted to Cohen's $d=0.3577$; to achieve 95% power with a dependent-samples t-test at $\alpha=.05$ would require $N=86$ (one-tailed), and $N=104$ (two-tailed). Note that because a different analytic technique is used in the present research (mixed-effects models on trial-level data) as compared to ANOVA, this is only a rough estimate.

Materials. The corresponding author of the paper by Corley et al. (2007) kindly 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 et al., 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, 20 in each of the four experimental conditions. 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 rate². The research assistant was instructed to record the sentences to sound as natural as possible. Following Corley et al. (2007), Collard et al. (2008), MacGregor et al. (2009), and MacGregor et al. (2010), the disfluencies always preceded the final word of the sentence, which was used as the memory probe word.

² A sample of 20 sentences indicated that the speech rate was, on average, 198.45 words per minute (standard deviation = 35.88).

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

Condition	Sentence
Fluent	<i>My sister had a skiing accident and she broke her leg</i>
Filler	<i>My sister had a skiing accident and she broke her um leg</i>
Pause	<i>My sister had a skiing accident and she broke her ... leg</i>
Repetition	<i>My sister had a skiing accident and she broke her her leg</i>

Procedure. At the beginning of the experiment, participants also completed five audio check questions to ensure that their headphones worked properly and were set to a comfortable volume. For each of the five audio check questions, participants heard one word and were asked to type it out. There were two phases of the experiment. 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 but not about the probe word. 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 1 lasted approximately 15-20 minutes; immediately after Phase 1, participants began Phase 2 which was a surprise recognition memory test. Participants viewed 160 visual word probes in a random order, 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). Old and new probes were counterbalanced across the experimental lists. 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.

Predictions

If disfluencies are in fact a type of cue to meaning (Clark & Fox Tree, 2002), the probabilistic link between new information and disfluent pauses and fillers may result in improved encoding and better memory for words preceded by pauses and fillers compared to repetitions. Alternatively, if the disfluency-related memory boost is simply due to the fact that it orients attentions to the upcoming speech stream (Collard et al., 2008), the disfluency-related memory boost may be observed for any type of disfluent interruption, regardless of its probabilistic link to upcoming meanings.

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 started the model selection procedure with a maximal random effects structure following the recommendations by Barr et al. (2013). However, the maximal random effects model resulted in convergence failure, so we removed random slopes starting with the ones that explained the least amount of variance until the model converged successfully. The analyses were performed in R (R Core Team, 2020) through the RStudio interface (RStudio Team, 2018) using the “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.

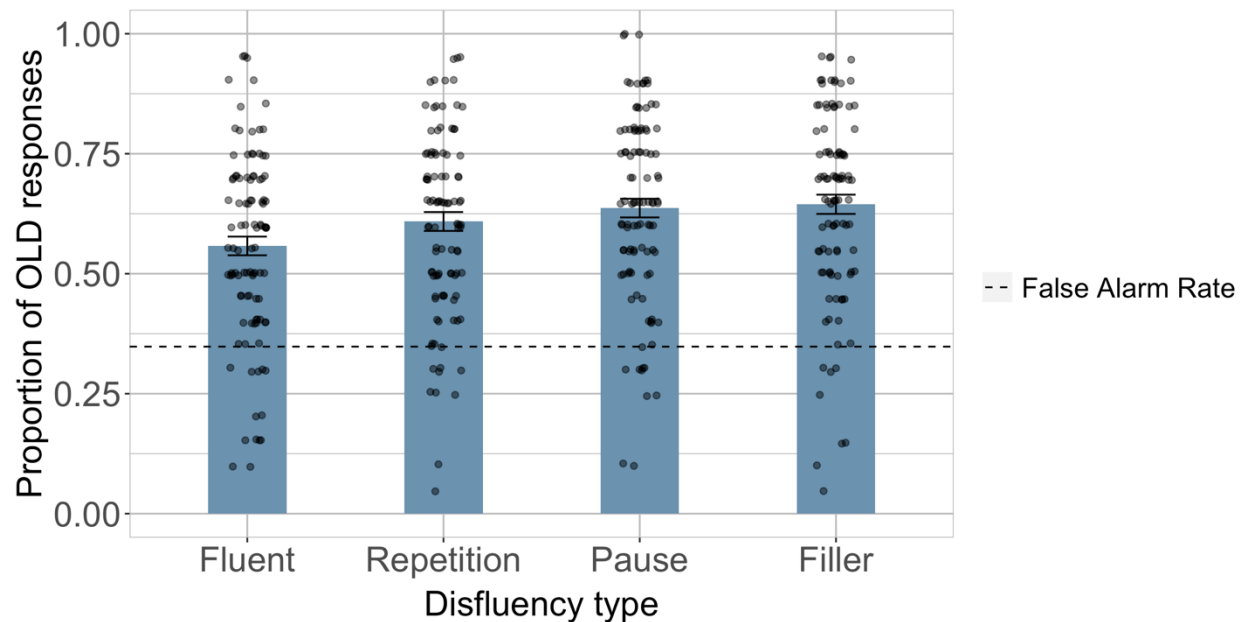
response ~ 1 + memory + disfluency + disfluency.type + fillers.vs.pauses + (1+ memory + disfluency | participant) + (1+ memory + disfluency | item)

Fixed Effects		Estimate	SE	z-value	p-value
	Intercept	-0.14	0.11	-1.19	0.23
	Memory effect (old (fluent, filler, pause, repetition) = each 0.5, new = -0.5)	1.34	0.12	11.51	< 10 ⁻¹⁵
	Disfluency effect (new = -0.125, fluent = -0.625, filler, pause, repetition = each 0.375)	0.35	0.07	4.82	< 10 ⁻⁵
	Disfluency type effect (new = -0.0625, fluent = -0.0625, filler, pause = 0.4375, repetition = -0.5625)	0.18	0.06	2.84	< 0.001
	Fillers vs. pauses (filler = 0.5, pause = -0.5)	0.06	0.07	0.77	0.44

Random Effects		Variance	St. Dev.	Correlations	
Item (Intercept)		0.62	0.79		
	Memory	1.01	1.00	-0.29	
	Disfluency	0.16	0.40	-0.13	-0.06
Participant (Intercept)		0.91	0.95		
	Memory	0.55	0.74	-0.30	
	Disfluency	0.06	0.24	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 successful recognition of the old probe words ($b=1.34$, $z=11.51$, $p<10^{-15}$). In addition, we observed significantly better recognition for probes from disfluent vs. fluent sentences ($b=0.35$, $z=54.82$, $p<10^{-5}$), and better recognition for pauses and fillers vs. repetitions ($b=0.18$, $z=2.84$, $p=0.004$) (Figure 1).

Figure 1. Memory results for Experiment 1. Error bars represent by-subject SEM. Data points represent mean accuracies for each participant. Means per condition are 56%, 61%, 64%, 64% for fluent, repetitions, pauses, and fillers respectively.



Discussion

The results from Experiment 1 replicate previous findings that disfluencies improve memory for linguistic material (Corley et al., 2007; MacGregor et al., 2010; Fraundorf & Watson, 2011). The odds of correctly recognizing the memory probes were 1.42 times higher when preceded by disfluency compared to fluent utterances. In addition, pauses and fillers resulted in a larger memory boost compared to disfluent repetitions. The odds of correctly recognizing a probe preceded by a filler or a pause were 1.20 times higher than if preceded by a repetition.

Compared to previous studies examining the effect of disfluency on memory (Corley et al., 2007; MacGregor et al., 2009; MacGregor et al., 2010), 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 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.

Experiment 2

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

Methods

Experiment 2 was a direct pre-registered replication of Experiment 1. The preregistration is available on the Open Science Framework (link: <https://osf.io/upkh2/>).

Participants. To determine our target sample size, we conducted an a-priori power analysis using data simulation package “simr” (Green & MacLeod, 2016) in R (R Core Team, 2020) through the RStudio interface (RStudio Team, 2018) to replicate the effect of disfluency type (pauses and fillers vs. repetitions, $b = 0.12$; odds ratio 1.13 to 1) 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 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.

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 finding would point to the meaning-based hypothesis as driving the disfluency-memory boost. Alternatively, if the disfluency-related memory boost is simply due to the fact that disfluency orients attention, all three types of disfluency would be expected to produce a similar memory benefit.

Results

Frequentist analysis. 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. Random effects included by-subject and by-item intercepts, and random slopes were included in the model if the model converged with them (for more details on model specification see Experiment 1). The analyses were performed in R (R Core Team, 2020) through the RStudio interface (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.

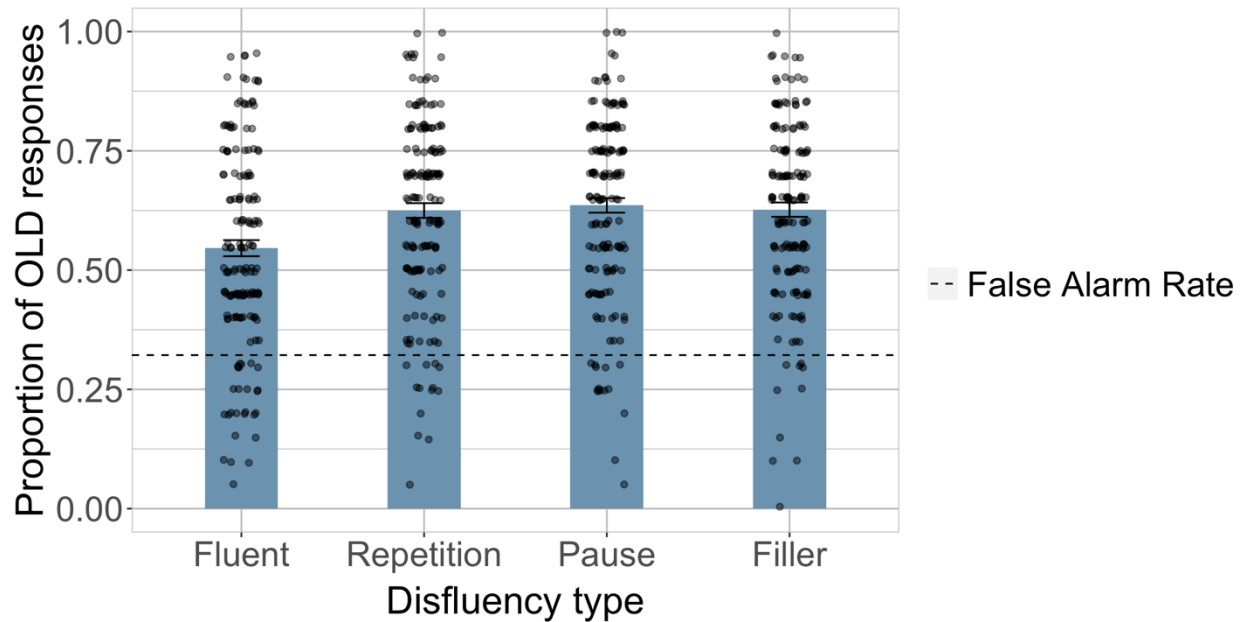
$$\text{response} \sim 1 + \text{memory} + \text{disfluency} + \text{disfluency.type} + \text{fillers.vs.pauses} + (1 + \text{memory} | \text{participant}) + (1 + \text{memory} + \text{disfluency} | \text{item})$$

Fixed Effects	Estimate	SE	z-value	p-value
Intercept	-0.19	0.08	-2.26	0.02
Memory effect (old = each 0.5, new = -0.5)	1.38	0.10	14.37	< 10 ⁻¹⁵
Disfluency effect (new = -0.125, fluent = -0.625, filler, pause, repetition = each 0.375)	0.42	0.06	6.73	< 10 ⁻¹⁰
Disfluency type effect (new = -0.0625, fluent = -0.0625, filler, pause = each 0.4375, repetition = -0.5625)	0.05	0.05	0.90	0.37
Fillers vs. pauses (new, fluent, repetition = each 0, filler = 0.5, pause = -0.5)	-0.04	0.06	-0.71	0.48

Random Effects	Variance	St. Dev.	Correlations
Item (Intercept)	0.29	0.54	
Memory	0.84	0.92	0.31
Disfluency	0.26	0.51	-0.21
Participant (Intercept)	0.75	0.87	
Memory	0.47	0.68	-0.11

The results of this analysis are shown 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 recognition of the critical words was significantly above chance ($b=1.38$, $z=14.37$, $p<10^{-15}$), 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 2**). Surprisingly, however, words preceded by pauses and fillers were not significantly better recognized than those preceded by repetitions ($b=0.05$, $z=0.90$, $p=0.37$), thus failing to replicate the disfluency *type* effect observed in Experiment 1.

Figure 2. Memory results for Experiment 2. Error bars represent by-subject SEM. Data points represent mean accuracies for each participant. Means by condition are 55%, 62%, 64%, 63% for fluent, repetitions, pauses, and fillers respectively.



Combined analysis. We conducted an additional frequentist analysis on the combined data from Experiments 1-2. 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. Random effects included by-subject and by-item intercepts, and random slopes were included in the model if the model converged with them (for more details on model specification see Experiment 1). A combined frequentist analysis of Experiments 1-2 yielded significant effects of memory ($b = 1.12$, $z = 19.67$, $p < 10^{-15}$) and disfluency ($b = 0.29$, $z = 7.61$, $p < 10^{-13}$). Importantly, the analysis additionally revealed a significant form-specific effect ($b = 0.06$, $z = 2.12$, $p = 0.03$), but the estimated effect size was small and impractical to pursue: it would require 300 participants to achieve only 65% power at $\alpha = .05$ in a replication attempt. Further, a power analysis based on the combined data across the two studies indicates that Experiment 1 ($N=102$) only had

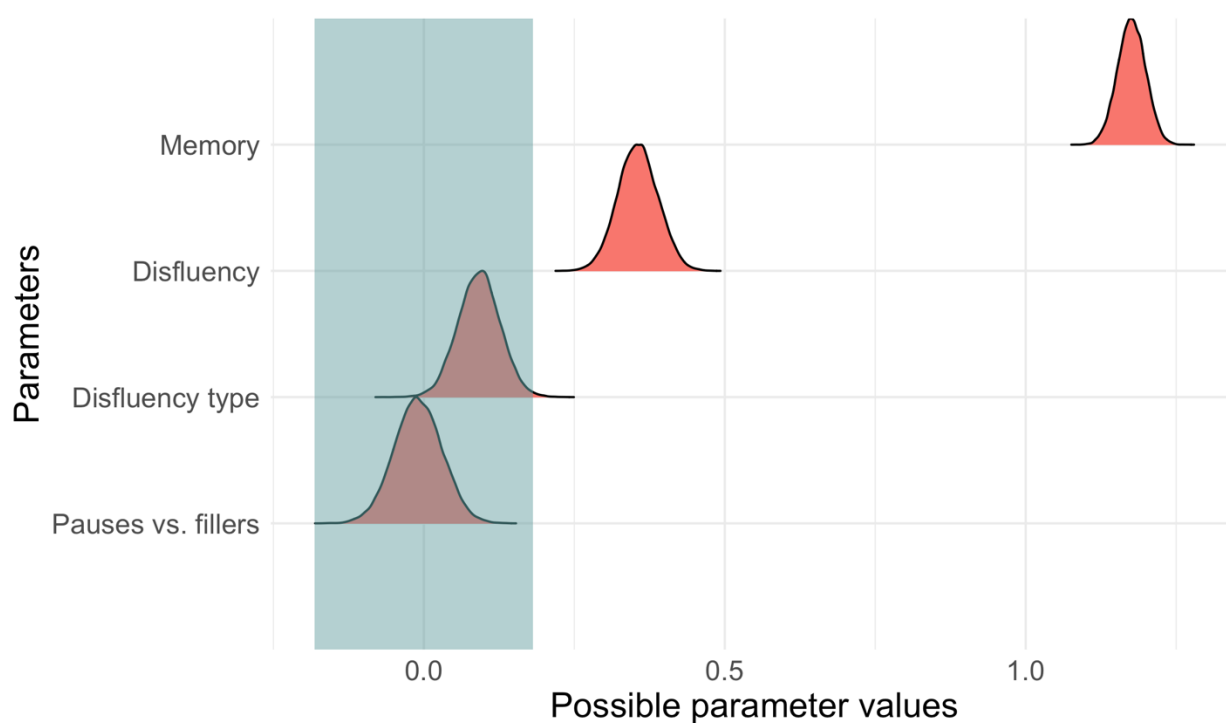
~26.2% power to detect the form-specific effect, indicating that Experiment 1 was underpowered to detect the effect, if it in fact exists (see relevant discussion in Simonsohn, 2015 and Simonsohn, Nelson, & Simmons, 2014). Finally, the sample size from the combined datasets from Experiments 1-2 was relatively large ($n = 263$). A large sample size in frequentist analyses can lead to the Meehl's paradox (Meehl, 1967, 1997) – a phenomenon when it is easier to *confirm* than to *disconfirm* a hypothesis as a result of the increased estimation precision.

Bayesian analysis. To overcome potentially inconclusive results associated with an increased sample size, we conducted a post-hoc Bayesian multilevel analysis to estimate the amount of evidence for the null model without disfluency type as a fixed effect against the alternative model with the effect of disfluency type. The analysis was conducted in R (R Core Team, 2020) through the RStudio interface (RStudio Team, 2018) using the “brms” package (Bürkner, 2017). For each parameter as well as the intercept, we used normal priors with the mean of 0 and standard deviation of 1, and for the random effects - default priors. We used the MCMC estimation algorithm with 4 Markov chains and 10,000 iterations per chain (2,000 iterations for the warmup). Both models converged as indicated by the $\hat{r}$ statistic values all equal to 1. The prior and posterior predictive checks indicated good model fit and are available on the project OSF page. The model comparison estimated the Bayes factor in favor of the null model compared to the alternative model to be equal to 2.31. In other words, the data are only 0.43 times more likely under the alternative model with the disfluency type as a fixed effect compared to the null model without the effect of disfluency type.

To further explore the evidence for and against the null hypothesis, we used an alternative approach to assess the null value of the disfluency type parameter with the Bayesian posterior distribution. More specifically, we tested the region of practical equivalence (ROPE), i.e., a region of values that are practically equal to the null value, against the 89% highest density

interval (HDI), i.e., the range of parameter values that includes 89% of most credible values in the posterior distribution (note that while 95% HDI is accepted in the field, 89% HDI is considered more stable, see Kruschke, 2010, 2011, 2014). If the ROPE completely excludes the 89% HDI, then the null hypothesis is rejected, if the ROPE completely includes the 89% HDI, then the alternative hypothesis is rejected, and if the ROPE and HDI partially overlap, then the decision remains inconclusive. Following Kruschke's (2018) recommendations for the ROPE range on the binary parameters, we used a ROPE from -0.18 to 0.18 (if a distribution of a parameter value overlaps largely with this range, the parameter is practically equal to zero). The resulting posterior distributions of the effect size for each parameter value of interest are illustrated in **Figure 3**. From the posterior distribution of the disfluency type parameter, we calculated that 89% of the most credible values lie between 0.03 and 0.14, and 99.5% of these values overlap with the ROPE. These findings suggest that the effect of disfluency type is practically equivalent to zero.

Figure 3. Posterior distributions on effect size for the fixed effects marked with 89% HDI and ROPE. The ROPE limits, the percentage of the posterior distribution that falls below, within, and above the ROPE are shaded in transparent green. The 89% HDI falls entirely inside the ROPE for the effect of disfluency type and the effect of pauses vs. fillers, and there is 99.5% and 100% probability respectively that these effect sizes are practically equivalent to zero.



Discussion

In Experiment 2, we found that disfluencies boost memory for immediately following words, replicating the same finding from Experiment 1 and similar findings in prior work (Corley et al., 2007; MacGregor et al., 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 demonstrate that the three different types of disfluency that we tested have similar memory benefits.

A post-hoc analysis using the combined data across the two experiments replicated the effect of disfluency on memory. Additionally, the combined analysis revealed that words preceded by pauses and fillers are recognized significantly better than those preceded by repetitions,

however, according to both the frequentist and Bayesian analyses, this observed effect of disfluency *type* on memory was practically equal to zero, corresponding to an odds ratio of only 1.09 in favor of pauses and fillers over repetitions.

Taken together, these findings provide evidence *against* the hypothesis that listeners make form-specific predictions with regard to different disfluency types in the course of language processing. The fact that all three types of disfluency boosted memory to similar degrees is, however, consistent with the hypothesis that disfluencies orient listener's attention to the upcoming context, and as a result, improve memory for the words that immediately follow (Collard et al., 2008).

Experiment 3

In Experiments 1-2, we found that all three types of disfluency that were tested improved memory for spoken words, consistent with an interpretation by which disfluency orients attention to the upcoming linguistic material. Note that in these studies, 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. Critically, if the locus of the disfluency-memory boost for words is an attentional orienting effect, it should be short-lived, as orienting effects tend to fade out 100-300 ms after stimulus onset (Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Spence & Driver, 1994; Mulckhuyse & Theeuwes 2010). Following this logic, if attentional orienting is the mechanism driving the memory benefit associated with disfluency, then it should not last long and primarily extend to the word that immediately follows disfluency. On the other hand, recall that Fraundorf and Watson (2011) reported 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. These findings, along with the results of the present Experiment 2 that all three types of disfluency boosted memory, suggest that the disfluency may yield more general processing benefits.

To further develop the attentional-orienting account of the disfluency-related memory boost, in Experiment 3, we examined the scope 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 late), 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.

Methods

This study was pre-registered on the Open Science Framework (link: <https://osf.io/upkh2/>).

Participants. In planning this experiment, we initially conducted a power analysis using data simulation package “simr” (Green & MacLeod, 2016) in R (R Core Team, 2020) through the RStudio interface (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 we 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 late). 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 $b = 0.18$; odds ratio 1.20 to 1) 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 exists.

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 and 85% accuracy on 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 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 (early, middle, and late). We created 7 versions of each of the 160 items (fluent, disfluent pause early, disfluent pause middle, disfluent pause late, disfluent filler early, disfluent filler middle, disfluent filler late), 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. Note that we did not test memory for words immediately following the disfluencies in the beginning and middle of the sentences because many of these words were repeated across the items. For example, the word “sister”, which occurred after the early disfluency in Item 1 also appeared in two other sentences in our stimulus set. As a result, it would be difficult to tease apart the effect of disfluency on memory for these words vs. the effect of repetition of these words throughout the experiment.

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. The last word in each sentence, e.g. “leg”, was always the critical memory probe across the conditions.

Condition	Sentence
Fluent	<i>My sister had a skiing accident and she broke her leg</i>
Filler Early	<i>My um sister had a skiing accident and she broke her leg</i>
Pause Early	<i>My ... sister had a skiing accident and she broke her leg</i>
Filler Middle	<i>My sister had a skiing um accident and she broke her leg</i>
Pause Middle	<i>My sister had a skiing ... accident and she broke her leg</i>
Filler Late	<i>My sister had a skiing accident and she broke her um leg</i>
Pause Late	<i>My sister had a skiing accident and she broke her ... leg</i>

Procedure. At the beginning of the experiment, participants completed five audio check questions to ensure that their headphones worked properly and to set the volume at a comfortable level. For each of the five audio 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 in a random order, 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. Old and new probes were counterbalanced across the experimental lists. Phase 1 lasted approximately 15-20 minutes; immediately following Phase 1, participants completed Phase 2. The entire study took approximately 30 minutes to complete.

Predictions

If the disfluency boost to word memory is due to short-lived attentional orienting, we predict better memory for the probe words when they are immediately preceded by a disfluency (late position), compared to when the disfluency occurs early or in the middle of a sentence.

Alternatively, the disfluency boost may confer broader memory benefits in spoken language (Fraundorf & Watson, 2011); if so, we predict a similar recognition advantage for the sentence-final probe words regardless of the position of the disfluency in a sentence.

Results

As in Experiments 1-2, we used a signal-detection theoretic mixed-effects analysis (Wright, Horry, & Skagerberg, 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. Because of the unequal number of sentences in each condition, we used a weighted contrast coding function (Fraundorf, 2021). Random effects included by-subject and by-item intercepts, and random slopes were included in the model if the model converged with them (for more details on model specification see Experiment 1). Because disfluency type and position are undefined for new and fluent items, the planned Model 2 analyzed old items only and included disfluency type as a centered fixed effect (pauses vs. fillers). Disfluency position was coded with the late position as the reference level, which allowed for a direct comparison of the two new disfluency positions (early and mid) with the late disfluency position tested in the prior studies. Both analyses were performed in R (R Core Team, 2020) through the RStudio interface (RStudio Team, 2018) using the “lme4” (Bates et al., 2015) package.

The results of Model 1 are shown in **Table 5a**. A significant intercept term reflected a response bias to say "new" ($b=-0.18$, $z=-2.32$, $p=0.02$); participants also successfully recognized the probe words above chance levels ($b=1.23$, $z=16.72$, $p<10^{-15}$). In contrast, the effect of disfluency on recognition 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 5b**. 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 6**). This analysis revealed that the disfluencies at the end of the sentence significantly improved memory for sentence-final words, 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 5a. 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.

$$\text{response} \sim 1 + \text{memory} + \text{disfluency} + (1 + \text{memory} + \text{disfluency} \mid \text{participant}) + (1 + \text{memory} + \text{disfluency} \mid \text{item})$$

Fixed Effects		Estimate	SE	z-value	p-value
(Intercept)		-0.18	0.08	-2.32	0.02
Memory effect		1.23	0.07	16.72	< 10 ⁻¹⁵
(fluent, disfluent = 0.5 each, new = - 0.5)					
Disfluency effect		0.11	0.07	1.66	0.10
(fluent = -0.625, disfluent = 0.375, new = -0.125)					
Random Effects		Variance	St. Dev.	Correlations	
Item (Intercept)		0.88	0.94		
	Memory	0.26	0.51	-0.24	
	Disfluency	0.03	0.18	-0.09	-0.39
Participant (Intercept)		0.1806	0.18	0.43	
	Memory	0.52	0.72	0.35	
	Disfluency	0.45	0.67	-0.08	-0.16

Table 5b. Experiment 3 memory results: Mixed effect model for old items only, 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.

response ~ 1 + disfluency.type + early.vs.late + middle.vs.late + (1 + early.vs.late + middle.vs.late | participant) + (1 + early.vs.late + middle.vs.late | item)

Fixed Effects		Estimate	SE	z-value	p-value
Intercept		0.64	0.09	7.28	< 10 ⁻¹²
Disfluency type					
(pauses = -0.5, fillers = 0.5)		-0.09	0.05	-1.66	0.10
Early vs. late					
(early pause and filler = 1, middle and late pause and filler = 0)		-0.23	0.07	-3.20	0.001
Middle vs. late					
(middle pause and filler = 1, early and late pause and filler = 0)		-0.14	0.09	-1.48	0.14
Random Effects		Variance	St. Dev.	Correlations	
Item (Intercept)		0.39	0.62		
	Early vs. late	0.27	0.52	-0.16	
	Middle vs. late	0.41	0.64	0.06	0.75
Participant (Intercept)		0.79	0.89		
	Early vs. late	0.18	0.42	-0.21	
	Middle vs. late	0.28	0.53	-0.08	0.91

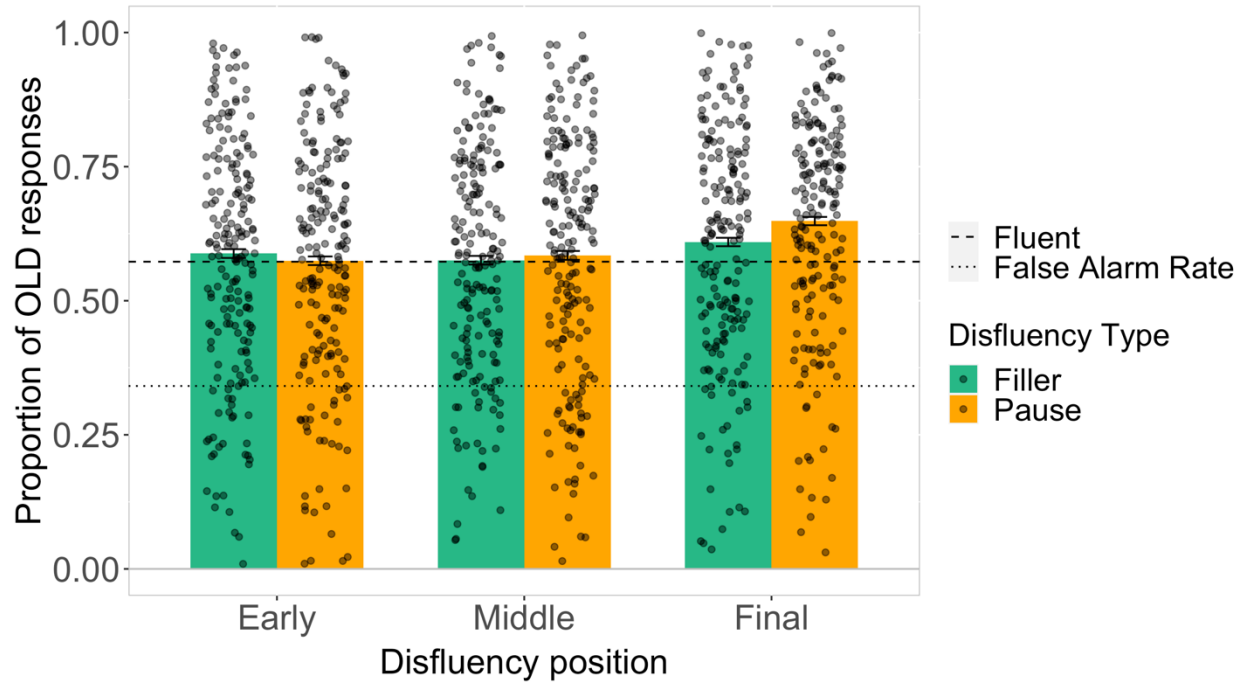


Figure 4. Memory results for Experiment 3. Error bars represent by-subject SEM. Data points represent mean accuracies for each participant.

Table 6. Experiment 3 memory results: Supplemental mixed effect model for old items only, 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.

$$response \sim 1 + early + middle + late + (1 + late | participant) + (1 + early + middle + late | item)$$

Fixed effects		<i>Estimate</i>	<i>SE</i>	<i>z-value</i>	<i>p-value</i>
	Intercept	0.36	0.10	3.68	0.0002
	Early	0.04	0.07	0.59	0.55
	(early = 1, middle, late, fluent = 0)				
	Middle	0.04	0.08	0.52	0.60
	(middle = 1, early, late, fluent = 0)				
	Late	0.28	0.09	3.21	0.001
	(late = 1, early, middle, fluent = 0)				

Random effects		<i>Variance</i>	<i>St. Dev.</i>	<i>Correlations</i>		
Item						
(Intercept)		0.67	0.82			
	Early	0.41	0.64	-0.49		
	Middle	0.49	0.70	-0.49	0.92	
	Late	0.63	0.79	-0.7	0.74	0.87
Participant (Intercept)		0.85	0.92			
	Late	0.14	0.38	-0.28		

Discussion

In Experiment 3, we tested the scope of the beneficial effect of disfluency on recognition memory for final words in sentences. Consistent with the findings from Experiments 1 and 2, we found that both pauses and fillers improved memory for the sentence-final word that followed the disfluency. Critically, however, we found that this effect was short-lived, consistent with previous literature showing that attentional orienting is transient (Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Spence & Driver, 1994; Mulckhuyse & Theeuwes 2010). Only disfluencies that immediately preceded the probe word significantly improved memory over fluent sentences. The odds of correct recognition of probes with late disfluency were 1.26 times higher than probes with early disfluency (and 1.32 times higher than fluent sentences). In contrast, when disfluencies appeared early or in the middle of the sentence, the recognition memory for the probe word was similar to fluent sentences.

Experiment 4

The results of Experiment 3 demonstrate a short-lived disfluency-memory boost for sentence-final words that is consistent with the idea that disfluency orients attention, and that orienting is short-lived (Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Spence & Driver, 1994; Mulckhuyse & Theeuwes 2010). An open question, then, is whether this short-lived memory boost is present throughout a sentence, or if it is localized to sentence-final words. One reason to think that the short-lived memory boost would occur throughout a sentence is that Fraundorf

and Watson (2011) observed a disfluency-memory boost for all plot points in stories, regardless of where in the story the disfluencies occurred. On the other hand, the serial position of words in word lists affects memory (i.e., primacy and recency effects, Deese & Kaufman, 1957; Murdock, 1962), and serial position can affect memory for words in sentences to a certain extent as well (Baddeley, Hitch, & Allen, 2009). Thus, some word positions may be more susceptible to a disfluency-related orienting of attention and subsequent memory boost. In Experiment 4, we address these questions by manipulating the position of the probe word in addition to position of disfluency in the sentence.

Methods

This study was pre-registered on the Open Science Framework (link: <https://osf.io/upkh2/>).

Participants. In planning this experiment, we conducted a power analysis using data simulation package “mixedpower” (Kumle et al., 2021) in R (R Core Team, 2020) through the RStudio interface (Rstudio Team, 2018) to determine the sample size needed to replicate the disfluency position effect from Experiment 3 (i.e., early vs. late probes; $b = -0.23$). According to the power analysis, in order to replicate this effect, we needed to recruit 200 participants to achieve 93% power at an alpha level of 5%.

A total of 293 participants were recruited on Amazon Mechanical Turk. Criteria for participation specified in the pre-registration were: > 95% acceptance rate, participating in the US, and did not participate in Experiment 1, Experiment 2, or Experiment 3. Participants received \$4 for ~30 minutes of participation. We excluded 145 participants because they failed to or didn’t attempt to answer sound check and / or comprehension questions, leaving 147 participants. We suspect the large number of the excluded participants reflects an increase in the number of bots (automated responses). In order to approach the planned sample size, an additional 44

participants from the Vanderbilt University participant pool were recruited in exchange for partial course credit. One participant completed the study twice, and their second study response was excluded from analysis. Recruitment of the for-credit participants was a change from the preregistration, and was undertaken due to data quality issues that arose on MTurk. All participants completed the study through the research platform FindingFive (FindingFive Team, 2019). At the beginning of the experiment, we collected basic demographic information (native language, gender, age). 8 participants reported themselves as non-native speakers of English, and were excluded from further analyses, leaving 183 participants in the final analysis (99 female; mean age 34.04; range 18-66; 140 from MTurk, 43 from the for-credit participant pool). While this sample size is less than the preregistration goal of 200, it was as close as we were able to achieve within a reasonable amount of time. Informed consent was obtained in accordance with the Vanderbilt University IRB guidelines.

Materials. The sentences were a subset of the materials used in Experiment 3. Recall that the materials were initially designed to test memory for the final word in each sentence, which were unique across the stimulus set. Because Experiment 4 additionally manipulated probe position (early and late), we first identified the 66 sentences with unique early probes. Each of the 66 sentences had a unique word that served as an early probe, and a unique word that served as a late probe (note that the early and late probes are necessarily different words as it is not possible to swap them without disrupting the sentence meaning) (**Table 7**). We recorded 5 versions of each of the 66 sentences (fluent, disfluent pause early, disfluent pause final, disfluent filler early, disfluent filler final), resulting in 330 total sentence versions.

Table 7. Example *Stimulus Set for Experiment 4. Early (“sister”) and late (“leg”) memory probe for each condition is underscored.*

Condition	Sentence
-----------	----------

Fluent	<i>My <u>sister</u> had a skiing accident and she broke her <u>leg</u></i>
Early disfluency	<i>My (um / ...) <u>sister</u> had a skiing accident and she broke her <u>leg</u></i>
Late disfluency	<i>My <u>sister</u> had a skiing accident and she broke her (um / ...) <u>leg</u></i>

All sentences were distributed across 9 experimental lists (~22 participants per list) following a modified Latin square design such that all sentences were counterbalanced across the conditions and each participant only heard a given sentence frame once. The novel manipulation in Experiment 4 was the position of the probe word in the sentence (early vs. late).

Procedure. At the beginning of the experiment, participants completed five audio check questions to ensure that their headphones worked properly and to set the volume at a comfortable level. For each of the five audio check questions, they heard one word and were asked to type it. Next, there were the two phases of the experiment. In Phase 1, participants listened to 40 fluent and disfluent audio sentences (8 fluent + 8 disfluent per each disfluent condition, i.e., 32 disfluent sentences in total). Participants could only advance to the next listening trial once the current sentence finished playing. Participants also had to answer 4 comprehension questions during the sentence listening task to ensure that they were paying attention to the stimuli. The comprehension questions were randomly presented during the first phase of the experiment, and inquired about the sentence but not about the probe words. To ensure that participants were paying attention to the stimuli, they were instructed that they needed to answer at least 85% of the comprehension questions correctly. Immediately after completing Phase 1, participants began Phase 2, where they viewed 160 single-word probes in a random order, one at a time. Half of the probes were old, and half were new. For each probe, participants were asked to indicate whether the probe was old (present in the sentences they just heard), or new. The 80 old probes came from the 40 sentences participants listened to (2 probes from each sentence), 52 new probes came from the sentences that participants did not listen to (2 probes from each sentence). Note that which probes were among the 80 old probes,

and which were among the 52 new probes was counterbalanced across experimental lists. Finally, we included an additional 28 new probes from sentences that were not used in the design of Experiment 4 (but that had been used in Experiments 1-3), to ensure that participants saw an equal number of new and old probes at test. Because they were not counterbalanced across lists, these 28 additional new probes were excluded from the analyses. Phase 1 of the experiment lasted approximately 10-15 minutes, and Phase 2 lasted <10 minutes, such that the entire study took about 20 minutes to complete.

Due to changes in data quality on MTurk, after collecting responses from 177 respondents, approximately half of which had to be excluded for failing to answer either sound check or comprehension questions, we strengthened the participation criteria to > 99% acceptance rate, and > 300 completed HITs, and added four new catch-trials, which participants had to answer correctly to proceed with the study. The first three catch-trials appeared at the very beginning of the experiment before the sound check question, and asked participants to listen to two short (3s each) phrases and transcribe them. If participants failed to transcribe the phrases correctly, they could not advance to the actual experiment. The fourth catch trial appeared after completing Phase 1, but before starting Phase 2 of the experiment, and asked participants to answer a simple arithmetic question (*Add 2 to 4, and type the answer in lower-case word form*). If participants failed to answer the question correctly, they could not advance to Phase 2 of the experiment. For all four new catch trials, participants had an unlimited number of attempts.

Predictions

If disfluency briefly orients attention to upcoming linguistic material, and this effect obtains throughout the sentence, for sentence-final probe words, we would expect to replicate the finding from Experiment 3 that memory for the probe word is boosted when the disfluency immediately preceded the probe word, compared to when the disfluency occurred earlier in the

sentence (i.e., the late > early effect from Experiment 3). Critically, we would predict the *opposite* pattern for sentence-early probes, with better memory when the disfluency came early in the sentence as opposed to later. Alternatively, disfluency may orient attention to upcoming linguistic material, but only at certain points in the sentence. If so, we would expect to replicate the short-lived disfluency-memory boost for sentence-final probe words seen in Experiment 3, but this short-lived effect may be attenuated or absent for earlier words in the sentence.

Results

Similar to Experiments 1-3, we used a signal-detection theoretic mixed-effects analysis (Wright, Horry, & Skagerberg, 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 interaction between probe position and disfluency position. Model 1 included item status (old vs. new) as a factor, then for old items, the presence and position of disfluency as a predictor. Because of the unequal number of sentences in each condition, we used weighted contrast coding (Fraundorf, 2021). Random effects included by-subject and by-item intercepts, and random slopes were included in the model if the model converged with them (for more details on model specification procedures see Experiment 1). Model 2 analyzed old disfluent items only and included the main effects and the interaction term for probe position and disfluency position as fixed effects; random slopes were included in the model if the model converged with them. Both probe position and disfluency position were coded with early probes and early disfluencies as reference levels. Both analyses were performed in R (R Core Team, 2020) through the Rstudio interface (Rstudio Team, 2018) using the “lme4” (Bates et al., 2015) package.

The results of Model 1 are shown in **Table 8a**. A significant negative intercept term reflected a response bias to say “new” ($b = -0.21$, $z = -2.86$, $p = 0.004$); participants also successfully recognized the probe words above chance levels ($b = 1.74$, $z = 26.65$, $p < 10^{-15}$), and the effect of

disfluency on recognition memory was significant ($b = 0.12$, $z = 2.34$, $p = 0.02$). Finally, Model 1 revealed a main effect of probe position on memory ($b = 0.18$, $z = 3.62$, $p < 10^{-15}$) such that late probes were more memorable than early probes regardless of the position of disfluency. The results of Model 2 are shown in **Table 8b**. Focusing on the old disfluent items only, Model 2 revealed a significant disfluency position by probe position interaction ($b = 0.58$, $z = 3.11$, $p = 0.002$). The effect of disfluency in this model was not significant, indicating that for early probe words (which were coded as the reference level), the effect of disfluency position was not significant ($b = 0.04$, $z = 0.63$, $p = 0.53$). In contrast, when we flip the reference level for probe position to be late probes, the effect of disfluency position was significant: late probes were significantly more likely to be recognized when the disfluency occurred immediately before the probe word ($b = 0.29$, $z = 4.60$, $p = 0.002$) replicating the late > early effect from Experiment 3.

Table 8a. Experiment 4 Model 1 memory results: Mixed effect model with item status (old vs. new), presence and position of disfluency as fixed effects. The dependent measure is binary – whether the participant responded “old” or “new” on the memory test.

response ~ 1 + memory + disfluency + early.vs.late + (1 + memory + early.vs.late | participant) + (1 | item)

Fixed effects	Estimate	Std. Error	z value	Pr(> z)
Intercept	-0.21	0.07	-2.84	0.004
Memory effect (new = -0.61; fluent, early, late = 0.39 each)	1.73	0.07	26.12	< 10 ⁻¹⁵
Disfluency effect (new = -0.18; fluent = -0.68; early, late = 0.32 each)	0.12	0.05	2.70	0.007
Late > early effect (new, fluent = 0; early = -0.5; late = 0.5)	0.18	0.05	3.60	0.0003
Random effects	Variance	Std. Dev.	Correlations	
Participant (Intercept)	0.52	0.72		

	Memory	0.56	0.75	-0.17	
	Late > early	0.14	0.38	-0.18	0.13
Item (Intercept)		0.15	0.39		

Table 8b. Experiment 4 Model 2 memory results: Mixed effect model with item status (old vs. new), presence and position of disfluency as fixed effects. The dependent measure is binary – whether the participant responded “old” or “new” on the memory test.

*response ~ 1 + probe.position*disfluency.position + probe.position + disfluency.position + (1 + disfluency.position | participant) + (1 | item)*

Fixed effects		Estimate	Std. Error	z value	Pr(> z)
	Intercept	0.45	0.09	4.83	<10 ⁻⁵
	Probe x disfluency position	0.25	0.08	3.10	0.0002
	Probe position (early probe = 0, late probe = 1)	-0.04	0.06	-0.71	0.48
	Late > early disfluency for early probes (early disfluency = 0, late disfluency = 1)	0.04	0.06	0.63	0.53
	Late > early disfluency for late probes (early disfluency = 0, late disfluency = 1)	0.29	0.06	4.60	<10 ⁻⁵

Random effects		Variance	Std.Dev.	Correlations
Participant (Intercept)		0.64	0.80	
	Disfluency position	0.10	0.32	-0.40
Item (Intercept)		0.25	0.50	

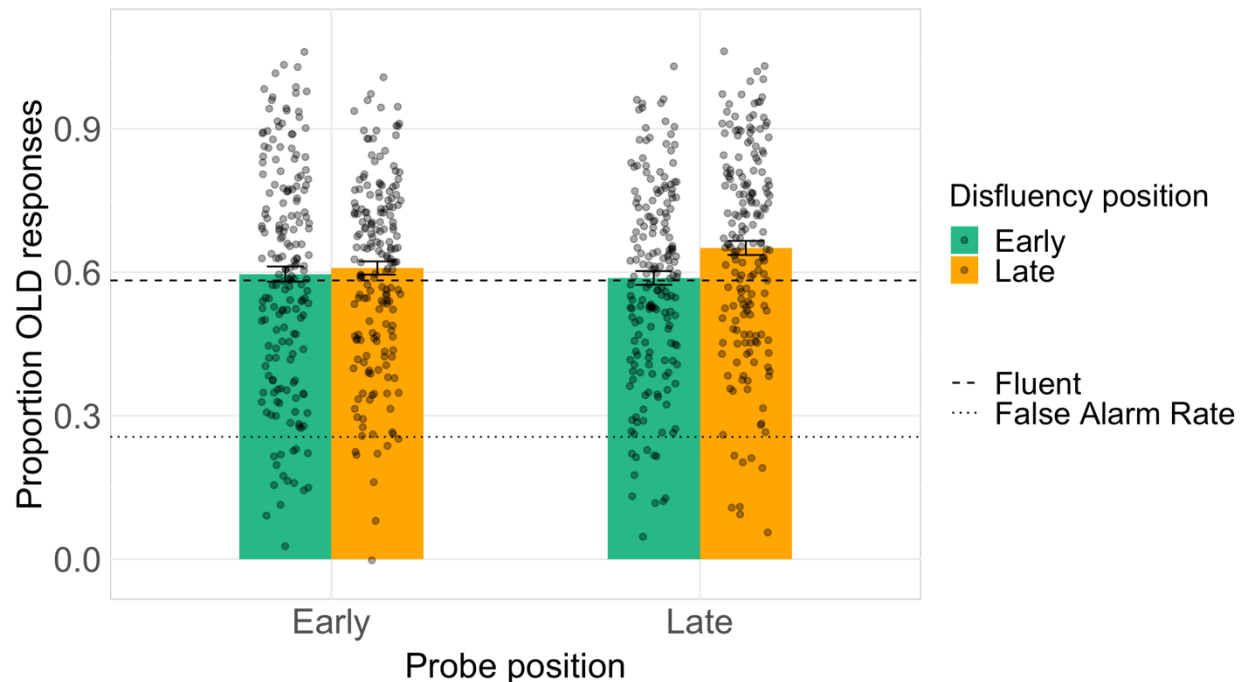


Figure 5. Memory results for Experiment 4. Error bars represent by-subject SEM. Data points represent mean accuracies for each participant. Means by condition are 58%, 59%, 61%, 59%, and 65% for fluent, Pr. Early – Dis. Early, Pr. Early – Dis. Late, Pr. Late – Dis. Early, Pr. Late – Dis. Late respectively.

Discussion

We hypothesized that if disfluency orients attention to upcoming linguistic material, and this effect obtains throughout the sentence, that we would observe a short-lived disfluency-memory boost both for probes early in the sentence, and probes later in the sentence. Replicating the key finding from Experiment 3, we observed that memory for sentence-final probe words was boosted when the disfluency immediately preceded the probe word, compared to when the disfluency occurred earlier in the sentence, consistent with a short-lived orienting of attention (Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Spence & Driver, 1994; Mulckhuyse & Theeuwes 2010). However, this short-lived effect was absent for probe words that appeared early in the sentence. For early probes, the odds of correct recognition were only 1.04 times

higher with early disfluency than with late disfluency. In contrast, for late probes, the odds of correct recognition were 1.34 times higher with late disfluency than with early disfluency. One explanation for the locus of the disfluency-memory boost being at the end of the sentence is that the penultimate position may be an atypical location for disfluency to occur (Shriberg, 1994; also see Beattie, 1979; Swerts, 1998). Thus, the atypicality of the disfluency position may have increased the overall salience of the disfluencies in our materials. Another possibility is that the disfluency-memory boost is enabled by more general sentence wrap-up effects (Warren, White, & Reichle, 2009). We also note that the materials we used were originally designed to test memory specifically for the final word. It was not possible to manipulate the position of individual probe words in this stimulus set without altering the semantic and syntactic structures in ways that could also affect sentence memorability (e.g., *My sister had a skiing accident and she broke her leg*; *The leg of my sister was broken in a skiing accident*). Thus, idiosyncratic differences between the early and late probe words may be in play; a future study with materials that afford a manipulation of probe position, holding other factors constant, may be worthwhile. In sum, the results of Experiment 4 replicate the findings of Experiment 3 and further demonstrate that this disfluency-related attentional orienting results in a detectable memory boost that may be localized to sentence-final words.

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, 2014). 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 et al., 2007; MacGregor et al., 2010), and others showing no such benefit for repetitions (MacGregor et al., 2009).

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, 2014), would differentially cue upcoming meanings to listeners. On this meaning-based view, we hypothesized that forms of disfluency that cue that new information is upcoming (pauses and fillers) would guide predictions regarding the type of the upcoming information (i.e., new and important) resulting in better encoding and subsequently, a larger memory boost. The meaning-based account also suggests that the forms associated with speaker difficulty for already planned material (repetitions) would serve as a cue that the articulated information is about to be repeated, and as a result, generating neither an encoding nor a mnemonic benefit. We contrasted this view with an attentional-orienting hypothesis that disfluency simply acts as a cue to orient attention to upcoming speech, resulting in improved processing and encoding of that speech, and as a result a memory boost that is not tied to the specific form-meaning mappings for different disfluent forms. In Experiments 3-4 we addressed the temporal and positional scope of the disfluency-based memory boost.

The results of Experiment 1 were potentially consistent with the meaning-based view, however the results of Experiment 2 were not: all three types of disfluency tested conferred a memory boost and the form-specific benefit was not significant based on the planned frequentist analyses. 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 at $\alpha = .05$ 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). Finally, the Bayesian analysis on the combined data indicated that the posterior distribution of the parameter values for the form-specific effect overlapped with the region of practical equivalence by 99.5%, demonstrating that the observed effect is practically equivalent to zero. In sum, we can clearly conclude 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 were 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 sum, here we show that all three types of disfluency (pauses, fillers, and repetitions) confer a clear and practically equivalent memory boost for subsequent words.

Given the findings of Experiments 1-2 support an attentional orienting account of the disfluency-memory boost, Experiments 3 and 4 probed the locus of this effect, focusing on how long the orienting effect lasts, and where it is likely to occur in a sentence. Using the same sentence materials as Experiments 1-2 but manipulating the disfluency location, Experiment 3 demonstrated that the scope of the disfluency-based memory benefit for sentence-final words is brief, primarily extending to the immediately following word. In Experiment 4 we further determined that the disfluency memory boost is position-dependent, and only manifested at the end of the sentence.

Attention and disfluency

Taken together, our findings are consistent with the hypothesis that disfluency orients auditory attention to upcoming speech. Across the four experiments, we found a significant memory boost for probe words immediately preceded by disfluency. Importantly, the recognition memory advantage was observed regardless of disfluency type, demonstrating that the probabilistic link between new information and disfluent pauses and fillers, and already planned material with repetitions (Fraundorf & Watson, 2014) does not differentially impact the cognitive processes that enhance memory for words in sentences. Evidence consistent with our findings include ERP evidence that disfluent fillers act to orient attention to upcoming speech (Collard et al., 2008), as well as findings that word *recognition* is enhanced 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. Notably, the acoustic distinctiveness (von Restorff, 1933) of disfluency does not appear to *capture* attention (Collard et al., 2008), but instead *orients* attention to upcoming words, improving encoding of those words into memory (see Addleman & Jiang, 2019; Gazzaley & Nobre, 2012). Lastly, the fact that the observed effect was short-lived, primarily manifesting on the word that immediately followed the disfluency, is consistent with previous findings establishing that attentional orienting is transient and dissolves rapidly 100-300 ms after the stimulus onset (Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Spence & Driver, 1994; Mulckhuyse & Theeuwes 2010), thus offering further support for the attentional-orienting interpretation of the observed disfluency-memory boost.

Other related findings show that rhythm in speech can create expectations, shifting attention to syllables that are expected to receive stress (Pitt & Samuel, 1990; Zheng & Pierrehumbert, 2010), and that the listener's attention can be attenuated by unpredictability in the speech

stream (Kakouros & Rasanen, 2016). Extended to disfluency, then, to the extent that disfluency is relatively surprising or unpredictable, it may function to orient attention to speech. While this account does not rule out the possibility that language users learn detailed, context-specific mappings between specific disfluent forms and speaker meanings (and possibly use them in processing speech), it does suggest that the mechanism underlying the disfluency-memory boost is considerably simpler and based on orienting of attention.

Our findings raise the possibility that disfluencies of different types might not be as distinct with respect to how they influence attentional processes as previously thought. The results of Experiments 1 and 2 provide sufficient evidence against the differential effect of the three different disfluency types on memory. Consequently, in the context of memory, different disfluency types should be treated similarly. Related to this point, recall that Fraundorf and Watson (2014) argued repetitions, unlike pauses and fillers, tend to occur when a message has already been planned and articulated. However, in a corpus analysis, Clark and Wasow (1998) report that repetition of an initial determiner was more likely when it was a part of complex rather than a simple noun phrase. These findings suggest that repetitions might indicate planning difficulty in some contexts. The discrepancy may owe to the fact that Clark and Wasow (1998) looked at the repetition of determiners (e.g., *the, the diesel*), while Fraundorf and Watson (2014) examined repetition of content words (e.g., *that cats that cats*). That said, additional evidence for the different patterns of usage of repetitions vs. fillers and pauses comes from findings showing that unlike repetitions, fillers and pauses do tend to occur at the linguistic (prosodic, syntactic, and semantic) boundaries (Butterworth, 1975; Shriberg, 1994; Swerts, 1998; Clark & Fox Tree, 2002) – the places where message planning is most likely to occur (Butterworth, 1975; Clark & Fox Tree, 2002). In sum, while pauses, fillers, and repetitions may be probabilistically linked to different contexts in language *production*, the present findings show

that the downstream consequences of these disfluent forms on memory following language *comprehension*, are similar.

Disfluency in online language processing

While our research questions focused on the impact of disfluency on enduring memory for what was said, the effect of disfluency on memory begins with how it is processed in the moment. It is well known that language comprehension involves predictive processing (Altmann & Kamide, 1999; Pickering & Garrod, 2013; Dell & Chang, 2013; Federmeier, 2007), yet, if and how listeners use distributional cues related to different disfluency types during language comprehension remains an open question. For example, Bosker and colleagues (2019) demonstrated that after being exposed to audio passages with typical and atypical disfluency distributions (typical: fillers before low-frequency words, no fillers before high-frequency words, atypical: fillers before high-frequency words, no fillers before low-frequency words), listeners modulated their predictions about the frequency of the upcoming words (i.e., learned to predict high-frequency words upon hearing disfluency). The authors proposed that the associations between disfluency and the frequency of following referents can be learned and can shape the disfluency bias. Additionally, Fox Tree (1991) observed differential effects of two filler types (*ums* and *uhs*) on subsequent word recognition with faster recognition of words preceded by *uhs* but not *ums*. Together these findings leave open the possibility that different disfluency types indeed vary in terms of their impact on language processing, even if they do not differentially affect memory.

On the other hand, predictions may not be necessary or universal (Pickering & Gambi, 2018; Huettig & Mani, 2015), and may be attenuated if they do not confer a clear processing benefit (Ryskin, Levy, & Fedorenko, 2020). Recall that Karimi et al. (2019) found that participants possessed implicit knowledge that focus constructions tend to be followed by semantically

related words (e.g., *not the hammer but the nail*), but that repairs tend to be followed by phonologically related words (e.g., *a hammer, uh I mean a hammock*) based on norming studies. However, this knowledge did *not* modulate predictions during a study of online language comprehension, leading Karimi et al. (2019) to conclude that the language processing system is not “smart” or flexible enough to use this distributional knowledge to generate differential predictions. This disconnect between having knowledge but not using it resembles evidence from the directed forgetting literature that 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, another possibility is that 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.

What factors determine when and how listeners utilize distributional knowledge about form-meaning mappings during language processing? One possibility is that in the presence of other cues such as pragmatic knowledge, listeners discount distributional knowledge or re-weight attention to different cues. Consider that in Arnold et al.’s (2007) study, listeners made anticipatory fixations to low-frequency or discourse-new objects when hearing a disfluency. However, when listeners were informed that the speaker had a disorder of object naming, the effect of disfluency was attenuated. These findings suggest that in some cases, top-down cues (i.e., pragmatic inference) modulate consideration of bottom-up cues (i.e., distributional knowledge) in language processing. Differences in linguistic exposure may result in changes to the distributional knowledge that relates form to meaning (Levy, 2008; Ryskin, Kurumada, & Brown-Schmidt, 2019; Ryskin, Levy, & Fedorenko, 2020). In the case of disfluency, it may be perceived as a disruption to the speech stream that acts as a low-level orienting cue, but also a cue to speaker meaning in some cases.

Further questions

A few open questions remain. First, it is not clear how the localized memory boost observed in the present studies and prior work (Corley et al., 2007; MacGregor et al., 2010) relates to the more general boost to gist memory for story plot points reported by Fraundorf and Watson (2011). One possibility is that when disfluency appears in multiple places throughout a passage (as it did in Fraundorf & Watson, 2011), attention to the overall passage is heightened, which in turn, results in better memory for the meaning of the passage in general. Another possibility is that disfluency does confer local benefits in word recognition and processing, and that by improving memory for that one word, that word in turn serves as a retrieval cue, thereby boosting sentence or propositional level memory for the entire story. 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.

Second, our study focused on language comprehension and memory. Whether *producing* disfluencies in spontaneous speech results in a similar beneficial effect on memory remains an open question. Fraundorf and Watson (2014) suggested that different disfluency types reflect different difficulties that speakers experience in language production. It is possible that different production difficulties that result in different disfluencies might have differential long-term effects for memory.

Finally, our study used the materials that were meticulously designed for an earlier experimental study (Corley et al., 2007). The four experiments presented here, along with earlier work using sentence stimuli (Fox Tree, 1991; Corley et al., 2007; MacGregor, Corley, & Donaldson, 2009; MacGregor, Donaldson, & Corley, 2010) and recorded stories (Fraundorf & Watson, 2011) all find that disfluency is beneficial to memory for recorded linguistic material. Whether disfluency

improves memory in for unscripted speech and in interactive settings remains an open question. Related to this question, Toftness and colleagues (2017) examined the effect of a professor's fluency in recorded instructional videos on students' metacognitive judgments and learning. While students rated disfluent professors as less effective at teaching compared to the fluent professors, the actual learning outcomes did not differ between the two types of instruction. Future research, then, is needed to assess the generalizability of the disfluency-memory benefit across various forms of language use in natural and interactive contexts.

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 degrees 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 reflect speaker difficulty). In a pair of well-powered studies, we find that 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 local, manifesting most strongly on the immediately following word, consistent with an attentional-orienting account of the observed disfluency-memory boost. The results of Experiment 4 replicated this short-lived disfluency-related memory boost, and clarified that it primarily occurs at the end of the sentence. Taken together, our findings reveal a short-lived disfluency boost in memory for words that occur at the end of the sentences, and that is

evoked by multiple types of disfluency, consistent with the idea that disfluencies briefly bring attentional focus to immediately upcoming linguistic material. We speculate that the downstream consequence of this localized memory benefit to individual words that immediately follow disfluent interruptions, is better understanding and encoding of the speaker's message as a whole.

Open Practices Statement

The written materials for this study were obtained by contacting the corresponding author of Corley, MacGregor, and Donaldson (2007). The audio recordings for all experiments were based on Corley et al.'s written materials and are available from the first author (ED) upon request. Supplemental materials, including examples of the audio sentences for each condition and an additional analysis for Experiment 4 exploring the effect of probe proximity to disfluency can be found on the project's OSF page. The raw data for the review process are available at this anonymized review-only link: <https://osf.io/upkh2/>.

Acknowledgements

This material is based on work supported by National Science Foundation Grant 19-21492 to Sarah Brown-Schmidt. We thank Jasmine Aggarwal and Jordan Zimmerman for their assistance in recording the experimental stimuli.

References

- Addleman, D. A., & Jiang, Y. V. (2019). Experience-Driven Auditory Attention. *Trends in cognitive sciences*, 23(11), 927-937.
- 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>
- Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, 61(3), 438-456.
- 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](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., & 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>
- 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.

- 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](https://doi.org/10.18637/jss.v067.i01).
- Beattie, G. W. (1979). Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. *Linguistics*, *17*(1-2), 61-78.
- 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, Ê. 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., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive psychology*, *2*(4), 331-350.
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive psychology*, *3*(2), 193-209.
- 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>
- 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>

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

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

Bürkner, P. C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, 80(1), 1-28.

Butterworth B. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*. 1975; 4:75–87.

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](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>
- Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of experimental psychology*, 54(3), 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. *Proc. DISS*, 2017, 17-20.
- 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. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, 29(2), 320–326. <https://doi.org/10.3758/BF03194926>

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

Fraundorf, S. H. (2021). `contr.helmert.weighted.R` (Computer Software).

<https://github.com/sfraundorf/psycholing/blob/39d39a903524d488d76d47de488523f6d95de3c0/R/contr.helmert.weighted.R>

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>

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in cognitive sciences*, 16(2), 129-135.

- Green P, MacLeod CJ (2016). "simr: an R package for power analysis of generalised57eneralized linear mixed models by simulation." *Methods in Ecology and Evolution*, 7(4), 493–498. Doi: [10.1111/2041-210X.12504](https://doi.org/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.
- Huetig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, 31(1), 19-31.
- James, W. (1890). The perception of reality. *Principles of psychology*, 2, 283-324.
- Kakouros, S. & Rasanen, O. (2015). Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features. *Cognitive Science*, 40, 1739-1774.
- Karimi, H., Brothers, T., & Ferreira, F. (2019). Phonological versus semantic prediction in focus and repair constructions: No evidence for differential predictions. *Cognitive psychology*, 112, 25-47.
- Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.
- Kruschke, J. K. (2010). Bayesian data analysis. Wiley Interdisciplinary Reviews: Cognitive Science, 1, 658–676. Doi:10.1002/wcs.72

- Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, 6, 299–312.
- Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2), 270-280.
- Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
- 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 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>
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34, 103–115.
- Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by

- confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 395–425). Mahwah, NJ: Erlbaum.
- Miller, S. (2018). *Disfluency, prediction and attention in language*.
- Mulckhuyse, M., & Theeuwes, J. (2010). Unconscious attentional orienting to exogenous cues: A review of the literature. *Acta psychologica*, 134(3), 299-309.
- Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental psychology: Human perception and performance*, 15(2), 315.
- Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology*, 64(5), 482.
- Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision research*, 29(11), 1631-1647.
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, 36(04), 329-347.
- Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human perception and performance*, 16(3), 564.
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25. <https://doi.org/10.1080/00335558008248231>
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, 13(1), 25-42.
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, 109(2), 160.
- 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>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 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., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive science*, 43(8), e12769.
- 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.
- Shriberg, E. E. (1994). Preliminaries to a theory of speech disfluencies (Doctoral dissertation, University of California, Berkeley).
- Shulman, G. L., Remington, R. W., & McLean, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance*, 5(3), 522–526. <https://doi.org/10.1037/0096-1523.5.3.522>
- 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.
- Smith, V. L., & Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, 32, 25–38.
- Spence, C. J., & Driver, J. (1994). Covert spatial attention in audition: {E}xogenous and endogenous mechanisms. *J. Exp. Psychol.*, 20(3), 555–574.
- Swerts M. Filled pauses as markers of discourse structure. *Journal of Pragmatics*. 1998; 30:485–496.
- 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](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.

- Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior research methods*, 41(2), 257-267.
- Yoon, S. O. & Brown-Schmidt, S. (July, 2020). Partner-specific adaptation in disfluency processing. Proceedings of the Cognitive Science Society.
- Zheng, X., & Pierrehumbert, J. B. (2010). The effects of prosodic prominence and serial position on duration perception. *The Journal of the Acoustical Society of America*, 128(2), 851-859.