HOW WE, UH, PERCIEVE DISFLUENCIES: THE EFFECT OF LINGUSITIC DISFLUENCIES ON JUDGEMENTS OF LEARNING, ATTENTION-ORIENTATION AND ACADEMIC MATERIALS

by

Brenna Mauro

Submitted to the Faculty of

University Honor's College in fulfillment

of the requirements for the degree of

Bachelor of Philosophy in Psychology

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This thesis was presented

by

Brenna Mauro

It was defended on

November 6, 2017

and approved by

Dr. Melissa Libertus, Associate Professor, University of Pittsburgh Department of Psychology

Dr. Tessa Warren, Associate Professor, University of Pittsburgh Department of Psychology, Linguistics, and Communication Sciences and Disorders

Dr. Jennifer Arnold, Professor, University of North Carolina, Chapel Hill Department of Psychology

Thesis Advisor: Dr. Scott Fraundorf, Assistant Professor, University of Pittsburgh Department of Psychology

HOW WE, UH, PERCIEVE DISFLUENCIES: THE EFFECT OF LINGUSITIC DISFLUENCIES ON JUDGEMENTS OF LEARNING, ATTENTION-ORIENTATION AND ACADEMIC MATERIALS

Brenna Mauro

University of Pittsburgh, 2017

It has been proposed that linguistic disfluencies actually function to aid the listener's memory for what they hear. When speech contains disfluencies, the listener has higher rates of recall for the speech's content (Fraundorf & Watson, 2011). This phenomenon is known as the *disfluency effect* (Seufert et al., 2016). The current study examines how linguistically filled pauses (e.g. *uh, um*) affect judgments of learning (JOLs), and if they lead to better memory for sentences and longer academic discourses—if so, if this is due to an attention orientating effect of disfluencies. In two experiments conducted online, we hypothesized that disfluencies would act as a cue to the listener that the speaker is having difficulty with the topic, and that disfluent speech would be perceived as more difficult for the participant to understand and lead to increased memory. In both experiments, participants rated and perceived disfluent speech as more difficult for the speaker remember, and as more difficult for themselves to remember later. These findings reflect that disfluencies *do* alert the listener that the speaker could be having difficulty remembering or understanding the speech. However, we did not find the hypothesized disfluency effect, in that disfluencies did not aid later memory. This is an another example of a case when metamemory, or people's judgments and beliefs about memory, fails and people incorrectly judge what variables will affect their memory (Kornell et al., 2011; Kornell & Bjork, 2007; Yan, et al., 2017). Often, it is the sense of fluency that affects people's perceptions of their confidence in and ease of understanding the content. This is what we saw—disfluent speech led listeners to predict difficulty with later memory, even though there was no actual difference. Further, the degrees of disfluency matters for people's judgments of knowledge. People can perceive a difference in disfluency levels.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

In daily discourse, we hear thousands and thousands of words a day, and our brains must help us choose what is important and deserves our selective attention; cues such as the duration and tone of language help orient our attention (Lim, Wöstmann & Obleser, 2015; Givón, 1992). For example, beat gestures, which often accompany pitch accents in language, have been found to aid memory memory for spoken language (Morett & Fraundorf, 2017). This perspective is also mirrored in academia—academic materials need to be created in a format that works to eliminate inessential demands on memory (Carlson, Chandler & Sweller, 2003; Lehmann, Goussios & Seufert, 2016; Zhang, Zhang & Yang, 2016). The current study examines how linguistically filled pauses (e.g. *uh, um*) affect judgments of learning (JOLs), and if they lead to better memory for sentences and longer academic discourses—if so, if this is due to an attention orientating effect of disfluencies.

In general, a disfluency is the subjective, metacognitive experience of difficulty associated with cognitive tasks (Diemand-Yauman, Oppenheimer & Vaughan, 2010). Linguistic disfluencies include components of speech such as repetitions, and filled pauses such as *uh* and *um*. As well, language can also contain non-linguistic disfluencies, such as coughs. As both speakers and listeners, we encounter linguistic disfluencies constantly in natural conversations, the natural rate of disfluency in speech has been suggested to be as low as two per 100 or as high as six per 100 words (Fraundorf & Watson, 2008). Fraundorf and Watson (2014) propose that filled pauses, in particular, signal that the speaker is struggling to explain a concept.

It has been proposed that linguistic disfluencies actually function to aid the listener's memory for what they hear; in fact, when speech contains disfluencies, the listener has higher rates of recall for the speech's content (Fraundorf & Watson, 2011). This phenomenon is known as the *disfluency effect* (Seufert et al., 2016). A primary inference behind the process that leads disfluencies to improve memory is that disfluencies act as a cue to the listener that not only the material is difficult, but the speaker is engaging in a lot of effort to explain it (Collard, Corley, MacGregor & Donaldson, 2008; Fraundorf & Watson, 2011; Arnold, Kam & Tanenhaus, 2007).

The positive effect disfluencies have on later memory cannot simply be attributed to any additional processing time they allow for. Fraundorf and Watson (2011) manipulated audio files of a speaker reading pieces from Alice and Wonderland; before key plot points, the audio was manipulated to either contain linguistic disfluencies, non-linguistic disfluencies (i.e. coughs), or had no manipulation. Linguistic fillers facilitated recall: Speech with linguistic disfluencies resulted in a higher rate of correctly recalling a plot point. However, non-linguistic fillers led to lower recall. This key difference in recall rates between linguistic disfluencies and non-linguistic fillers illustrates that the positive effect disfluencies have on later memory cannot be attributed to more processing time—if this was true, there should have been no difference found between the two. Rather, these findings suggest linguistic disfluencies may serve an attentional-orienting purpose; disfluencies may act as a cue to the listener that not only the material is difficult, but also that the speaker is engaging in a lot of effort to explain it. Thus, this signals our brains to pay more attention since we are anticipating difficult material.

Further, linguistically disfluent speech leads to differences in language processing relative to fluent speech (Corley, MacGregor & Donaldson, 2007). Corley et al. (2007), conducted research on the effects of linguistic disfluencies on changes in brain functioning. The effects between fluent

and disfluent utterances on listeners' language comprehension was examined through an event-related potential (ERP) and memory test. Specifically, this study looked at the N400 effect, which represents the relationship between neural language processing and semantics; the N400 effect is reduced when the meaning of the language is predictable, and higher when the meaning is surprising (Kutas and Hillyard, 1980; Kutas & Federmeier, 2011). Speech preceded by a hesitation disfluency was found to have a reduced N400 effect and be better remembered. These findings not only suggest that disfluencies lead to differences in immediate language processing that has both long and short term effects for listeners, but also give credibility to the claim that disfluencies do act as a signal to our brains and give us a heads-up that something odd or unpredictable might be coming up in the conversation.

A follow-up to this study was conducted to examine the attention orienting effects of hesitations in speech (Collard, et al., 2008). Using ERPs, researchers found evidence that was consistent with the findings in Corley et al. (2007): In both studies, hesitations preceding speech were found to affect the listener's attention. Specifically, the researchers found ERP deflections consistent with attention orientating when the listeners encountered hesitations. Words were better recalled when they were preceded by a disfluency than when they were not. Though this study concentrated on hesitations marked by the word *er*, and less on the filled linguistic disfluencies of *uh* and *um* that the current study is focused on, these findings give sufficient support that other filled linguistic disfluencies would likely serve an attentional-orienting purpose as well.

Disfluencies can affect how the speaker is perceived. Brennan and Williams (1995) examined the effect that filled linguistic disfluencies have on both the perception of how well another person understands a topic, the "feeling of another's knowledge" (FOAK), as well as one's own "feeling of knowledge" (FOK); that is, how well one feels they understand the topic.

Brennan and Williams found that when a long, filled disfluency occurred before an answer to a question, participants rated both FOAK and FOK lower. Although previous research has already established disfluency use can actually have a positive effect on memory, this finding suggests that disfluencies negatively affect the perception of knowledge. However, the evidence about disfluency use in language leading to more attention from the listener is still somewhat indirect. In Fraundorf and Watson's study (2011), only the participant's eventual memory was studied, and it was merely hypothesized that effect disfluencies had on memory were due to an attention-orientating cue of language. While Corley's (2007) study does include a measure of both initial processing & later memory, this was only studied for individual words not in a context of natural language production. What is missing is an assessment of how the effects of disfluency on long-term memory for entire sentences or discourses are mediated by initial processing and interpretation.

1.1 DISFLUENCIES IN ACADEMIA

An additional question Brennan and Williams's (1995) results pose is what effect a lower FOAK and FOK due to disfluency usage might have in an academic setting. Given that professors are expected to be experts in their field, would disfluency use in academia impede memory and reverse the disfluency effect? Thus, if disfluency use causes listeners to feel less confident that the speaker knows what they're talking about, this may have a bigger and backwards effect on listeners in academic settings; if professors seem unsure of their confidence in the material, this may cause students to lose confidence in their professor's knowledge, and thus could lower their memory for the subject matter.

Just as linguistic disfluencies are normal occurrences in natural language production, they are also apparent in academic settings—though their saliency and effect on educational outcomes has not been fully examined (Schachter, Christenfeld, Ravina & Bilous, 1991). Learners often hold misjudged metacognitive preconceptions surrounding which learning methods may actually increase learning and later memory; often, these misconceptions are pervasive and very difficult to eliminate (Yan, Bjork & Bjork, 2017). For example, learners often fail to take advantage of the substantial benefits of self-testing rather than re-reading study materials (Kornell & Bjork, 2007), and they believe that mass studying is more efficient than short study sessions spread out over a longer period of time (Kornell & Bjork, 2007). When it comes to language, we can infer people may hold similar misconceptions regarding what will aid us in our comprehension and memory; though not directly examined yet for disfluencies, the current study will examine if disfluency production in speech is perceived to help or hinder memory.

Current research on disfluencies in academia has been concentrated on the effects of perceptual disfluencies, such as *disfluent, difficult to read fonts*. Diemand-Yauman et al., (2011) studied the effects of perceptual disfluencies on academic outcomes; they hypothesized that disfluencies would lead to better recall because disfluent materials would lead to deeper and more effortful processing. In their first experiment, participants learned made-up taxonomic information regarding an alien race in either a fluent or disfluent font; the disfluent condition led to higher overall recall of correct answers. Their second experiment expanded on their previous findings and implemented perceptual disfluencies in a real classroom environment in a high school; researchers manipulated PowerPoint presentations and handouts for history and physics classes to

be written in either disfluent or fluent texts. Researchers found the disfluent condition led to better classroom assessment scores.

Research specifically focusing on linguistic disfluencies in academia consists only of examining the rates of disfluency production in academic settings (Schachter et al., 1991). Critically, research has not examined the impact disfluencies have on student's learning and memory. Disfluency production in different types of classes, such as natural sciences, social science and humanities, is not equal. Specifically, more disfluencies are used in lectures on subjects that have more "options". That is, in the natural sciences, these topics of study do not have many, if any, different answers to choose from; for example, e=mc^2 has no options, e cannot be mc^3 or mc^4. Correspondingly, lectures in the natural sciences had the lowest rate of disfluency production. By contrast, in classes such as those in the humanities, professors have multiple options to explain a topic; for example, an instructor has many different options to pick from in describing what Shakespeare could have meant in a particular passage. Thus, lectures in the humanities had the highest rate of disfluency production.

The differing rate of disfluency production across academic subjects poses the question of whether there will be a relationship between disfluencies used in certain subjects and the rates of recall for that content. Specifically, one could realistically predict disfluency use in "no-option" subjects, such as the natural sciences could either impair or increase recall. Since disfluencies are most uncommon in those topic areas, this could be positive or negative. For example, disfluency usage is so uncommon, it might especially orient one's attention. In contrast, since the speaker is not expected to be hesitant about the subject mater, this might significantly lower one's FOAK of the speaker, and perhaps decrease memory. Thus, the current study includes three different topics in two main subject areas for academic lectures—the natural sciences and the humanities—to

examine if disfluencies will lead to difference between recall for lectures on the humanities versus the natural sciences.

1.2 ABOVE-NATURAL-RATE DISFLUENCY

At least among nonlinguistic disfluency, there does appear to be a threshold for when material becomes too disfluent and the positive memory effect is reversed (Seufert, Wagner and Westphal, 2016). Seufert et al. (2016) examined different levels of perceptual disfluencies and the threshold at which material becomes too disfluent for an individual to read and comprehend. As the level of perceptual disfluency increased, learning and memory performance followed suit—however, once the text became too illegible, the disfluency effect was reversed. There has not been research conducted on if this threshold translates to linguistic disfluencies, but we might expect that there would also be a threshold at which speech becomes too disfluent for the listener to adequately comprehend, and thus the disfluency effect would be reversed. An aim of the current study is to determine whether above-natural-rate linguistic disfluency similarly impairs memory and if above-natural-rate disfluency is able to be perceived by participants and thus lower their judgments of learning.

1.3 PRESENT WORK

The current study not only aims to examine why the disfluency effect is observed (e.g., if disfluencies are attention-orienting), but also to assess what effect linguistic disfluencies have in

academic settings. Thus, we have three main aims of this study: (1) explore if disfluencies will lead to increased later memory recall for academic materials. (2) Test if disfluent speech will be rated differently on JOL measures and thus reflect an attention-orienting purpose of disfluencies; importantly, we are going to measure attention—orientation through JOL measures. If attention mediates the disfluency effect, we will see more difficult/less confident JOL ratings for disfluent facts, showing that participants are attuned to a disfluent difficulty. (3) Examine if degrees of disfluency matter, and if an above the average rate of disfluency, or above-natural-rate disfluency, will affect perceptions of learning and memory. This will expand the current research on the effect of disfluencies on memory and education, and it will explore the implications regarding the most effective manner for professors to teach and lecture. This will deepen our understanding of the effects of linguistic disfluencies on memory and education. With online learning expanding, disfluency is important to consider. For example, professors may believe that it would be best to record lectures for online classes from a script and thus be extremely fluent if the disfluency effect is found to extend to academic content, then it will be important that this is known for the creation of online courses.

1.3.1 Experiment 1.

In Experiment 1, we tested what effect disfluencies have on perception of difficulty. Participants heard trivia facts, either spoken fluently or containing a disfluency at the beginning of the sentence, where they most naturally occur (Stolcke & Shriberg, 1996; Siu & Ostendorf, 1996) Critically, we also obtained participants' ratings of the perceived difficulty of the facts and of the speaker's level of knowledge through judgment of learning (JOL) questions. We hypothesized that content with disfluencies will be perceived to be more difficult and it is this perception of difficulty—and the

resulting deeper, more effortful processing—that mediates the effect of disfluencies on long-term memory.

1.3.2 Experiment 2.

In Experiment 2, we tested the effects of linguistic disfluencies in academic settings. Participants heard three minute recorded lectures, on topics in the humanities and natural sciences; these lectures were either fluent, contain the natural rate of disfluencies, or contain a greater-than natural rate of disfluencies (above-natural-rate disfluency). We hypothesized that disfluencies will lower confidence in JOL ratings. As well, that academic content containing a natural rate of linguistic disfluencies will lead to better recall than linguistically fluent language. However, we posit that when the rate of disfluencies production rises above the natural production rate, it will impair recall and decrease participants' confidence in their predicted memory.

2.0 EXPERIMENT 1

2.1 METHOD

2.1.1 Participants.

Participants were recruited from the Amazon Mechanical Turk (MTurk) website. Participants had to be over the age of 18 and native English speakers to be eligible for participation in this study. We first ran 9 participants through MTurk and used estimates of effect size derived from this preliminary data to conduct a power analysis. The power analysis was conducted on the JOL ratings to determine how many participants we would need for at least 80% power to detect a significant difference of the size observed in the pilot data. This analysis indicated we needed at least 25 participants; we had funds for 51 additional participants, this larger number of participants gave us power to look at other points of interest, such as the relation of the JOLs to memory recall.

2.1.2 Materials.

2.1.2.1 Recorded trivia facts The materials consisted of recorded speech of 70 relatively obscure trivia facts; of these, 50 are critical items, and 25 are filler items. A list of the critical trivia facts can be found in Appendix A. The filler items included a variety of different responses (e.g., “I definitely know that!” or “I can’t remember that one”) to make the speech sound more natural and

so that the disfluencies are not the only difference the participants may notice (potentially giving away the purpose of the experiment).

The critical items were presented in two different conditions—one disfluent and one fluent—which varied within-subjects (i.e. each participant heard a mix of both speech conditions). For the disfluent speech, the disfluency occurred at the beginning of the sentence, which is where disfluencies most often occur in natural speech (Stolcke & Shriberg, 1996; Siu & Ostendorf, 1996). The disfluent and fluent conditions were created using auditory splicing; speech was originally recorded with the disfluency, which was spliced out in the fluent condition. This procedure controlled for differences in the rest of the utterance; that is, the speaker enunciated and stressed on the same syllables, and talked at the same speed for both conditions.

2.1.2.2 Survey questions After hearing each trivia fact, participants were immediately given two JOL questions, one regarding their own feeling of understanding/difficulty of the material (self JOL), and the second about their feeling of the speaker's understanding/difficulty of the material (other JOL). The JOL questions in Experiment 1 are more akin to judgments of difficulty, rather than more typical judgments of confidence (see Table 2.1 below).

Table 1. JOL questions from experiment 1

JOL question asked:
Please rate how difficult you think the material was for the respondent to understand, from 1 being not at all difficult, to 6 being extremely difficult.
Please rate how difficult you think the material will be for you to remember, from 1 being not at all difficult, to 6 being extremely difficult.

The JOLs were rated on a scale from 1-6; a higher rating meant that the participant felt the fact was more difficult and thus was less confident. The JOLs were rated on a scale from 1-6; a higher rating meant that the participant felt the fact was more difficult and thus was less confident.

2.1.2.3 Filler task and recall test The results from the pilot suggested that the memory task was too easy; we tried to make it harder by increasing the retention interval. Before the participants took the recall test on the trivia facts they heard, they had to complete a filler task, taking about 10-15 minutes. The filler task consisted of 60 mental rotation questions. Participants saw an arrangement of blocks, and then had to choose which of two pictures matched the blocks rotated at 90 degrees. After the filler task, participants took a fill-in-the-blank recall test on all 50 critical facts. A list of all trivia facts and their respective recall question can be found in Appendix A.

2.1.3 Procedure.

Once participants were accepted to take the survey, MTurk directed participants to a Qualtrics survey. Participants had to complete all parts of the study to receive compensation. Once

completed, they received a confirmation code to enter on the MTurk website to confirm their completion of the study. For a cover story to explain why we showed participants disfluent speech, we informed participants that they were hearing speech from a previous participant trying to remember a fact from another study, so any disfluencies would be natural and not seem odd to include. Thus, before each trivia fact, a separate speaker asked a question to prompt the speaker to recite the full trivia fact.

2.2 RESULTS AND DISCUSSION

We ran paired sample t-tests on each of the measures (e.g. recall test and JOL ratings) to determine whether the disfluent critical items were remembered better and whether the disfluent condition were rated as more difficult for the speaker and harder for the participant to remember.

We first looked specifically at the self and other JOL ratings. In Experiment 1, a higher JOL rating signifies *more* perceived difficulty with the material. For the other JOLs, the disfluent facts (M=2.02) were perceived as significantly more difficult than the fluent facts (M=1.78), $t(51) = 4.70$, $p < .0001$. While the self JOLs did not show such a large difference, the disfluent facts (M=3.05) were still rated as significantly more difficult than the fluent facts (M=2.93), $t(51) = 2.08$, $p = .02$.

Overall, participants did extremely well on the recall test, far better than anticipated. The results for memory accuracy of disfluent and fluent facts revealed a non-significant trend, $t(51) = -.08$, $p = .53$, with disfluent facts (M = .828) being remembered fractionally less than fluent facts (M = .829). The means of memory accuracy for disfluent and fluent facts were almost identical and overall seemed to display a ceiling effect.

To determine if facts rated as harder to remember were *actually* more difficult to remember, we examined the relationship of JOL ratings and memory accuracy by running two one-way ANOVAs. The first ANOVA examined whether memory accuracy differed across facts given different self JOL ratings, [$F(5, 264) = 15.58$, $p < .0001$]. The second ANOVA examined the same, but compared memory accuracy of facts given different other JOL ratings, [$F(5, 179) = 3.83$, $p = < .01$]. There was a clear, statistically significant trend with JOL rating and memory accuracy: When participants rated a JOL higher (i.e., they felt this fact was more difficult), there was lower memory accuracy for that fact—participants were correctly identifying which facts are more difficult. The trend for the self JOLs (see Figure 1) is clear and consistent, while the trend for the other JOLs (see Figure 2) is a little less so.

Since one of our aims of this research project was not only to examine how attention mediates the disfluency effect, but also to study the role of linguistic disfluencies in an educational context, we ran a second experiment. Experiment 2 looks at what effect linguistic disfluencies have on memory for educational content, and what effects above-natural rates of disfluencies has on memory.

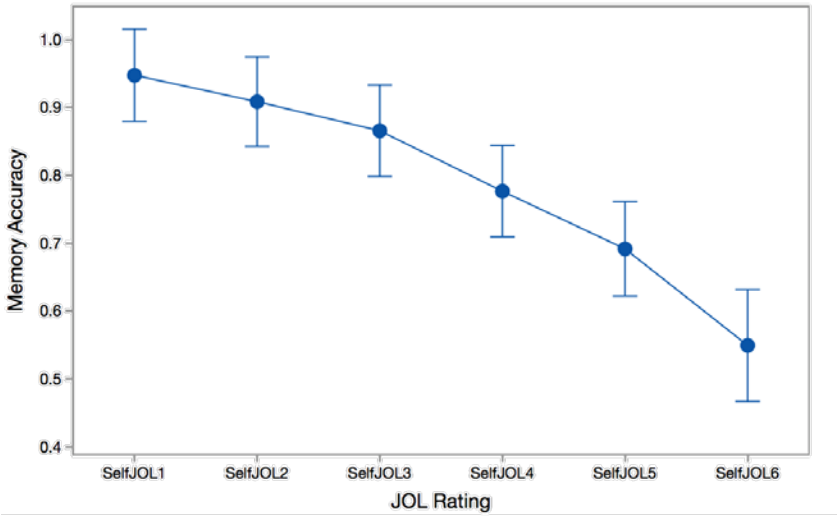

Figure 1. An interval plot of memory accuracy compared with participants' self JOL ratings.

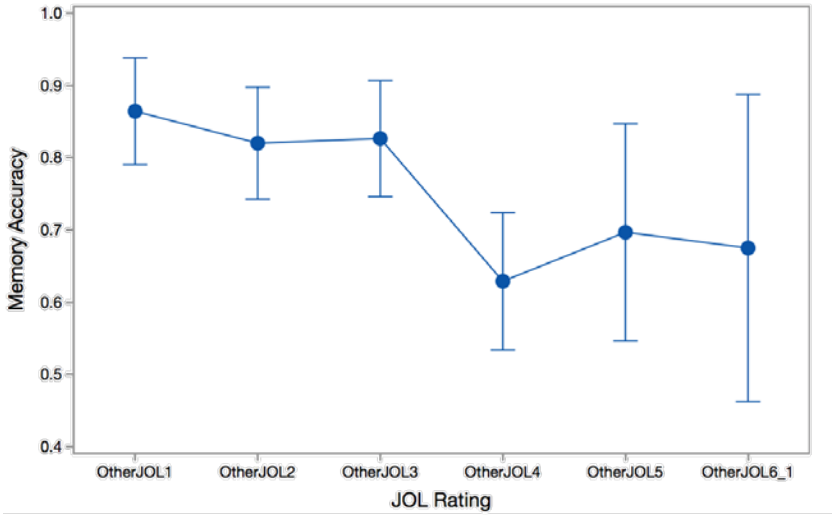

Figure 2. An interval plot of memory accuracy compared with participants other JOL ratings.

3.0 EXPERIMENT 2

3.1 METHOD

3.1.1 Participants.

Participants were recruited from the Amazon Mechanical Turk (MTurk) website. Participants had to be over the age of 18 and native English speakers to be eligible for participation in this study. Because this is a similar task with the same subject population, we felt confident that the power analysis conducted in Experiment 1 for the required sample size (at least 25 participants) would be similar for this experiment. We had funding for 60 participants and ran all 60 for data collection. 1 participant had to be excluded from the final analysis because they were not a native English speaker.

3.1.2 Materials.

3.1.2.1 Recorded lectures Participants heard short, three to four minute recorded lectures on academic subjects. There were six total lectures to hear. Three of these were humanities topics: (a) a lecture on the analysis of Robert Frost's poem, *The Road Not Taken,* (b) a lecture on the philosophy of happiness, and (c) a lecture on the analysis of Pablo Picasso's painting, *Les Demoiselles d'Avignon*. The other three lectures were natural science topics: (a) a lecture about the element Zirconium, (b) a lecture on gravitational waves, and (c) a lecture explaining infectious diseases. The lectures range from 467 words to 521 words. The lectures were recorded and played

as a sound file, participants also saw one relevant picture along with the recording; for example, for the Pablo Picasso lecture, participants saw the painting that is discussed. This procedure is intended to emulate the appearance of a real lecture, except that the use of only a single picture ensures that participant's recall of the information is based purely on the speech, not on the visual lecture materials.

As with Experiment 1, the speech was recorded with disfluencies, which were spliced out for the fluent condition. There were three fluency conditions: a fluent condition; a natural-disfluent condition, with seven disfluencies which is at the natural rate of 1.87 fillers per 100 words (Fraundorf & Watson, 2008); and an above-natural-rate-disfluent condition of 15 disfluencies, which is twice the natural rate. There were three versions of the survey counterbalanced so that each version had one fluent, one natural-disfluent, and one above-natural-rate-disfluent for each of the humanities and science lectures.

3.1.2.2 Survey questions After hearing each lecture, participants were given two JOL questions, one regarding their own feeling of understanding/confidence (self JOL), and the second about their feeling of the professor's understanding/confidence (other JOL) of the lecture material. In Experiment 2, these JOL questions are the more typical judgments of confidence in one's learning. The JOLs were rated on a scale from 1-6; a higher rating meant the participant felt more confidence in the learning of the lecture (see Table 2 below for JOL questions asked).

Table 2. JOL questions from experiment 2

JOL question asked:
Please rate how confident you think the professor felt about the material, from 1 being not at all difficult, to 6 being extremely difficult.
Please rate how confident you are that you will remember the material in this lecture, from 1 being not at all difficult, to 6 being extremely difficult.

3.1.2.3 Recall test After the same mental rotation filler-task of 10-15 minutes that we used in Experiment 1, participants were given a fill-in-the-blank recall test of 42 questions in total, 7 from each lecture. For the natural-disfluent condition, the parts of the lecture participants were tested on contained a disfluency preceding the sentence. When hearing the above-natural-rate-disfluent version of a lecture, participants were not additionally tested on the sentences where an above-natural-rate disfluency occurred. Appendix B shows a list of the recall questions the participants were prompted with.

3.1.3 Procedure.

Once participants were accepted to take the survey, MTurk directed the participants to a Qualtrics survey, where they completed the measures. Participants had to complete all parts of the study to receive compensation. Once finished, they received a confirmation code to enter on the MTurk website to confirm their completion of the study. For a cover story to explain why we are showing participants different lecture topics spoken by the same person, we informed participants that they

are hearing parts of lectures from a general studies course for freshman in college, where a professor touches on multiple topics from different fields.

3.2 RESUTS AND DISCUSSION

Since we had two independent variables, the lecture type (i.e. humanities or science) and the speech condition, we ran two-way ANOVAs to look at the interactions among these variables and each of the dependent measures—the other and self JOL ratings and memory recall. We first ran a two-way ANOVA looking at the interaction between lecture type, speech condition and the other JOL ratings (i.e. the perceived confidence in the professor's knowledge). There was a significant main effect for lecture-type on other JOL ratings, [$F(1, 348) = 25.2$, $p<.0001$]. In their other JOL ratings, participants rated humanities lectures (M=4.44) higher in confidence than science lectures (M=3.78). Additionally, there was a significant main effect for speech condition on the other JOL ratings, [$F(2, 348) = 40.41$, $p < .0001$]. The appearance of disfluencies in a lecture led to significantly lower confidence. We ran paired sample t-tests to further look at this relationship. The other JOL ratings for fluent lectures (M=4.79) were rated statistically significantly higher in confidence than natural-disfluent lectures (M=4.18), $t(118) = 4.48$, $p < .0001$. We ran another paired sample t-test on the other JOL ratings for natural-disfluent lectures (M=4.18) versus above-natural-rate-disfluent lectures (M=3.36). Participants rated above-natural-rate-disfluent lectures statistically significantly lower in confidence than the natural-disfluent lectures, $t(118) = 6.24$, $p < 0.0001$. However, the interaction between lecture type and speech condition was not significant.

We ran another two-way ANOVA to examine the relationship between lecture type, speech condition, and self JOL ratings. There was a significant main effect between lecture type and self

JOL ratings, [$F(1, 348) = 5.42$, $p = .02$]. Self JOL ratings for science lectures were statistically significantly lower in confidence ($M = 2.98$) than humanities lectures ($M = 3.32$). There was no significant main effect of speech condition on self JOL ratings, [$F(1, 348) = 0.77$, $p = .46$]. However, upon further analysis through planned paired sample t-tests, there was a marginally significant difference between self JOL ratings for natural-disfluent lectures (M = 3.22) and above-natural-rate-disfluent lectures (M = 3.03), $t(118) = 1.63$, $p = .05$.

A final two-way ANOVA examined if there was any interaction between the two IVs and participant's memory accuracy. There was no main effect for lecture type, [$F(1, 348)=1.60$, $p=.21$], nor for speech condition, [$F(1, 348)=.04$, $p=.96$] on participants' memory accuracy.

4.0 GENERAL DISCUSSION

The findings in Experiment 1 tell us two important points: disfluencies do alert the listener to a difficulty in speech and also mediate people's metacognition. We hypothesized that disfluencies would act as a cue to the listener that the speaker is having difficulty with the topic; moreso, that disfluent speech would be perceived as more difficult for the participant to understand. This is reflected in the participants' JOL ratings—in both experiments, participants rated and perceived disfluent speech as more difficult for the speaker remember, and as more difficult for themselves to remember later. These findings reflect that disfluencies *do* alert the listener that the speaker could be having difficulty remembering or understanding the speech.

While we found no overall effect of disfluency on memory, in Experiment 1, we did find some specific relationships between JOLs and memory that we were initially interested in. We found that when a participant anticipated personal difficulties understanding the material (as reflected in their JOL), they did poorer on later memory (see Figure 1). For example, when participants gave a self JOL rating of 1 (the lowest possible judgment of difficulty), they averaged 95% memory accuracy on the recall test (see Figure 1). Comparatively, when participants gave a self JOL rating of 6 (the highest possible judgment of difficulty), they averaged only about 55% memory accuracy on the recall test (see Figure 1). This was also the case with JOLs made for the speaker—when participants felt that the speaker had increasingly difficulty understanding the material, the participants also did increasingly worse on their later memory accuracy (see Figure 2). This shows that, as a whole, disfluencies mediated people's metacognitive understandings. In Experiment 2, participants also gave less confident JOL ratings of lectures with disfluencies,

providing further evidence that disfluency affects people's perceptions of their confidence in later memory.

However, in both these experiments disfluency did not have an overall effect on later memory accuracy. Thus, disfluency seemed to create an illusion of differences in memorability. This is an another example of a case when metamemory, or people's judgments and beliefs about memory, fails and people incorrectly judge what variables will affect their memory (Kornell et al., 2011; Kornell & Bjork, 2007; Yan, et al., 2017). This follows suit with a study by Kornell et al. (2011) on how font size and perceived fluency led to incorrect judgments of learning. Specifically, people often wrongly perceive variables of fluency, such as a larger font size, to predict memory. This relates to the *ease-of-processing heuristic:* When something appears fluent, people judge it have been learned well (Kornell et al., 2011). Thus, it can be the sense of fluency that affects people's perceptions of their confidence in and ease of understanding the content. This is what we saw—disfluent speech led listeners to predict difficulty with later memory, even though there was no actual difference. The choppiness, or lack of fluency that linguistic disfluencies bring, may make speech *seem* more difficult, when it really may not be.

Further, the degrees of disfluency matters for people's judgments of knowledge. The effects of above-natural-rate linguistic disfluency had not yet been studied, which is a major contribution this paper makes. We found that people can perceive a difference in disfluency levels—in Experiment 2, between natural-disfluent and above-natural-rate-disfluent lectures, there was a significant difference in both self and other JOL ratings. Participants could tell when the level of disfluency increased, and this was reflected in their less confident JOL ratings for above-natural-rate-disfluent lectures. While this was not found to affect memory, there does seem to be

a threshold of when disfluency is perceived to increase and negatively affect a person's metamemory.

We remain uncertain why we did not find the disfluency effect for later memory. In Experiment 1, there seemed to be a ceiling effect at play. Participants' memory for the trivia facts was far higher than anticipated. Participants averaged about 82% memory accuracy on the recall test, and multiple participants got 100% correct for both disfluent and fluent facts. This possible ceiling effect could be a reason why we did not find the hypothesized disfluency effect; in fact, the memory accuracy for fluent and disfluent facts were almost identical. In Experiment 2, we saw a low average recall averaging around the 30% mark. Perhaps six recorded lectures to listen to per participant was too high a demand on their memory.

Additionally, the methodology of this experiment could have been off target. In regards to Experiment 1, research on the disfluency effect in context of single sentences is extremely limited. As such, perhaps the disfluency effect is not salient when placed only in one sentence, and not within a larger passage or conversation. This could be possible, as using single sentences as our materials may have allowed the participants more time to focus their attention to the single fact. Since the speaker did not continue on immediately to discuss more facts, the participant did not have to continue listening and attenuating to different information; as such, the participants had more, individualized attention to each fact than if the facts were placed within a passage. However, this would not explain why we did not see the disfluency effect in Experiment 2. Further, in Experiment 1 we chose relatively novel trivia facts because we did not want participants to have prior knowledge of them. However, this may have caused the facts to be unusually interesting and striking, which may have contributed to the ceiling effect.

Future experiments should focus on different methodology. For example, future research could use less interesting facts, or try to implement this in a laboratory setting and not online, where the researchers would have more control over any confounding factors. Additionally, further research should focus on the salience of above-natural-rate disfluencies across multiple contexts and not just educational material. In this project, we saw the importance that fluency can play on metamemory and people's judgments of learning. Further, we highlight the significance of people's tendency to judge their memory on more superficial, processing-based variables, and how this can cause inaccuracies in people's metacognitive assumptions.

APPENDIX A

CRITICAL TRIVIA FACTS AND RECALL QUESTIONS USED IN EXPERIMENT 1

1. The archers at the ancient Olympic Games used doves as targets.
    a. The archers at the ancient Olympic Games used ____ as targets.
2. The raven was the first bird mentioned in the Bible.
    a. The ____ was the first bird mentioned in the Bible.
3. A group of foxes is called a skulk.
    a. A group of foxes is called a _____.
4. The 5-day 40-hour work week was introduced by the steel industry.
    a. The 5-day 40-hour work week was introduced by the _____ industry.
5. An emu's eggs are the color green.
    a. An emu's eggs are the color ____.
6. China is the world's largest tobacco producer.
    a. ____ is the world's largest tobacco producer.
7. Italy was the first country to produce lace.
    a. ____ was the first country to produce lace.
8. Libya is the only country with a flag that consists of a solid, single color.
    a. ____ is the only country with a flag that consists of a solid, single color.

9. The city of Zagazig is is located in Egypt.
    a. The city of Zagazig is located in _____.
10. The only product ever promoted by Elvis Presley in a TV commercial was donuts.
    a. The only product ever promoted by Elvis Presley in a TV commercial was _____.
11. In the human body the eyelids have the thinnest skin.
    a. In the human body the ____ has/have the thinnest skin.
12. In the United States, Nevada is the state most dependent upon tourism.
    a. In the United States, _____ is the state most dependent upon tourism.
13. In 1492, Columbus set sail for the New World on a Friday.
    a. In 1492, Columbus set sail for the New World on which day of the week? A _____.
14. The most remote weather station is located in Canada.
    a. The most remote weather station is located in _____.
15. Canada imports the most American cars.
    a. The country _____ imports the most American cars.
16. Norway consumes the most spicy Mexican food.
    a. The European country ____ consumes the most spicy Mexican food.
17. Grapes are the largest fruit crop in the world.
    a. _____ is/are the largest fruit crop in the world.
18. The world's largest herb is the banana.

a. The world's largest herb is the _____.

19. Leonardo da Vinci was born in the village of Vinci.

    a. Leonardo da Vinci was born in the village of ____.

20. The Datsun was the first Japanese car imported to the United States.

    a. The _____ was the first Japanese car imported to the United States.

21. Robert Fulton launched his first steamboat in Paris.

    a. Robert Fulton launched his first steamboat in the city ____.

22. The Carthaginians used snakes to defeat the Romans in 3 B.C.

    a. The Carthaginians used snakes to defeat the Romans in 3 B.C.

23. Poland has the last herd of bison in Europe.

    a. The European country ____ has the last herd of bison in Europe

24. When a lobster's blood is exposed to oxygen it turns blue.

    a. When a lobster's blood is exposed to oxygen it turns the color _____.

25. The skin is the largest organ in the body.

    a. The _____ is the largest organ in the body.

26. The world's most expensive spice is saffron.

    a. The world's most expensive spice is _____.

27. The 'Q' in Q-tips stands for quality.

    a. The 'Q' in Q-tips stands for _____.

28. The giant squid has the largest eyes in the world.

    a. The ____ ____ has the largest eyes in the world.

29. Walt Disney was afraid of mice.

a. The animal Walt Disney was afraid of was ____.

30. 4,000 years ago, ice cream was invented in China.

    a. 4,000 years ago, ice cream was invented in ____.

31. The Pittsburgh Pirates baseball team almost built their stadium over a river.

    a. The _____ baseball team almost built their stadium over a river.

32. Goldilocks' original name was Silver Hair.

    a. Goldilocks' original name was ____ ____.

33. Olympic gold medals are actually made out of silver.

    a. Olympic gold medals are actually made out of the metal ____.

34. The state flower of Rhode Island is mistletoe.

    a. The state flower of Rhode Island is _____.

35. Ancient Egyptians used stones as pillows

    a. Ancient Egyptians used ____ as pillows.

36. The state vegetable of Oklahoma is the watermelon.

    a. The _____ is the state vegetable of Oklahoma.

37. One of the most popular pizza toppings in Brazil is green peas.

    a. One of the most popular pizza toppings in Brazil is _____.

38. Scissors were invented by Leonardo da Vinci.

    a. Scissors were invented by _____ _____.

39. There is a city named "Rome" on every continent.

    a. There is a city named "_____" on every continent.

40. The only mammal that can't jump is the elephant.

a. The _____ is the only mammal that can't jump.

41. Coca-Cola would be green if additional coloring was not added to it.

    a. Coca-Cola would be the color ____ if additional coloring was not added to it.

42. Softball was originally called Kitten Ball.

    a. Softball was originally called ____ ____.

43. The electric chair was invented by a dentist.

    a. A worker from the profession of a _____ invented the electric chair.

44. The neck of a turkey is known as a wattle.

    a. The neck of a turkey is known as a _____.

45. In English, "Thailand" means "land of the free".

    a. _____ is the English translation of "Thailand".

46. The mascot for the University of California Santa Cruz is the Banana Slug.

    a. The mascot for the University of California Santa Cruz is the _____ _____.

47. The official state beverage of Delaware is milk.

    a. The official state beverage of Delaware is ___.

48. The first electric traffic light was installed in Ohio.

    a. _____ is the state the first electric traffic light was installed in.

49. The planet with the strongest surface winds is Jupiter.

    a. The planet _____ Jupiter has the strongest surface winds.

50. The first sport to be filmed for an audience was boxing.

    a. The first sport to be filmed for an audience was _____.

APPENDIX B

CRITICAL LECTURE SENTENCES AND RECALL QUESTIONS USED IN EXPERIMENT 2

These are the critical sentences from the lectures that participants were asked on, and their respective recall question.

1. Frost’s work frequently used settings from rural life in New England.
    a. Robert Frost's work frequently used settings from rural life in this region _____ .
2. In fact, Frost was a farmer before he became a famous poet.
    a. Robert Frost worked as a ____ before he became a famous poet.
3. While writing about this landscape, Frost merges the traditional with the modern to become a writer who is simultaneously terrifying and comfortable.
    a. While writing about the rural landscape, Robert often merged the _____ with the modern.
4. By the end of the poem he makes his choice in a famous statement of individualism.
    a. By the end of "The Road Not Taken", Robert Frost makes a famous statement of _____.
5. One realizes that neither road is “less travelled by.”
    a. In "The Road Not Taken", one realizes that _____ road is "less traveled by."
6. We have to choose, and most terrifyingly, the choice may not actually matter.
    a. In "The Road Not Taken" , the speaker argues that the _____ may not matter.

7. Frost won four Pulitzer prizes during his life.
    a. Robert Frost won a total of ____ Pulitzer Prizes in his life.
8. Zirconium is a silver-gray transition metal, which is a type of element that is malleable and flexible.
    a. Zirconium is a silver-grey _____ metal.
9. Interestingly, zirconium has very low toxicity and it is estimated that humans ingest about 50 micrograms per day.
    a. Humans ingest about ____ micrograms of zirconium a day.
10. Lunar rocks appear to have a surprisingly high Zircon content compared to terrestrial rocks.
    a. More zircon is found in _____ rocks than terrestrial rocks.
11. Zircon comes closer to resembling a diamond than any other natural gem.
    a. Zircon comes closer to resembling a _____ than any other natural gem.
12. Rocks containing zircon that were found in Australia in 2000 were dated to be 4.4 billion years old.
    a. In 2000, rocks containing zircon were dated to be __.__ billion years old.
13. The use of zirconium compounds in medicine began in 1969 when it was used to manufacture hip prosthetics.
    a. The use of Zirconium compounds in medicine began in 1969 when it was used to manufacture ____ prosthetics.
14. One such study found that there might be a link between zirconium implants and some health problems, such as inflammation and skeletal and connective tissue disorders.
    a. There might be a link between Zirconium implants and some health problems, such as _____, skeletal and connective tissue disorders.

15. Plato argued that human flourishing comes not from material wealth and physical goods, but rather from something that might not occur to you: reflection and wisdom.
    a. Plato argued that human flourishing comes not from material wealth and physical goods, but from reflection and _____.
16. Hadit's first claim he calls the progress principle, which is the discovery that most of our pleasure comes not from the achievement of a goal, but from the process of achieving that goal.
    a. Hadit's claim of the _____ states that most of our pleasure comes nor from the achievement of a goal, but from the process of achieving that goal.
17. The second is that because we are more sensitive to changes in goods than to absolute levels of goods, in that more of something doesn't always make us happier.
    a. Hadit also claims that because we are more sensitive to changes in goods than to _____ of goods, that more of something doesn't always make us happier.
18. The hedonic treadmill is the idea that in order to maintain the same amount of happiness, if it's based on material goods, requires us to run to stay in the same place, to keep the same level of goods.
    a. The _____ is the idea that in order to maintain the same amount of happiness, if it's based on material goods, requires us to run to stay in the same place, to keep the same level of goods.
19. Aristotle begins with an argument called the argument in favor of the *summum bonum*, which is known as the highest good.
    a. Aristotle's argument of the *summum bonum* is known as the _____.

20. This leads to what Aristotle calls political science-- political science in the sense of the study of human beings as political, that is, social animals
    a. Aristotle calls _____ the study of humans as social animals.
21. So just as knives are great when they can cut well, humans are great because we can reason, and thus, according to Aristotle, reason is the key to happiness.
    a. Aristotle claims that _____ is the key to happiness.
22. Gravitational waves were Einstein's hypothesized ripples in space-time and were spotted for the first time in 2016.
    a. Einstein's hypothesized gravitational waves were spotted for the first time in ____.
23. They whirled about each other at half the speed of light and finally merged.
    a. The black holes had spiraled around each other at ____ the speed of light.
24. The collision sent a shudder through the universe and formed ripples in the fabric of space and time called gravitational waves.
    a. The collision sent ripples in the fabric of _____.
25. This recorded observation tests Einstein's theory of relativity, with unprecedented rigor and provides proof positive that black holes exist.
    a. The recorded observation of gravitational waves tests Einstein's theory of _____.
26. LIGO tested this by having two facilities with massive detectors, one in Washington and the other in Louisiana.
    a. LIGO could detect gravitational waves by having one facility in the state of Washington and another in the state of _____.
27. To detect a gravitational wave, researchers can compare the relative lengths of the two rulers to within 1/10,000 the width of a proton.

a. Researchers can compare the lengths of the two ultra-precise rulers to within 1/10,000 the width of a_____.

28. Only a black hole—which is made of pure gravitational energy and gets its mass through Einstein's famous $E=mc^2$ equation—can pack so much mass into so little space.
    a. A black hole is made up of pure _____ energy.
29. Viruses are smaller than bacteria, and viruses cause a multitude of diseases — ranging from the common cold to AIDS.
    a. _____ are smaller than bacteria.
30. Fungi are many skin diseases, such as ringworm and athlete's foot.
    a. Fungi are many ____ diseases.
31. A pathogen is a micro-organism that has the potential to cause disease.
    a. A _____ is a micro-organism that has the potential to cause disease.
32. The site at which they enter is known as the portal of entry.
    a. The site where microbes enter is known as the _____.
33. The first site is the respiratory tract, such as through the mouth and nose; this is how influenza, or the flu, gets into our bodies.
    a. The _____ is how influenza gets into our bodies.
34. The final site of entry is through cuts or breaks in the skin surface; this is where and how people contract the tetanus virus.
    a. Microbes entering though breaks in the skin surface is how people contract the _____ virus.
35. After attaching, microbes must multiply rapidly and obtain their nutrients from the host.

a. After attaching to a target site, microbes must multiply and obtain their ____ from the host.

36. Together with fellow artist Georges Braque, Picasso pioneered cubism.

a. Along with fellow artist _____, Picasso pioneered Cubism.

37. Primitive art typically referred to African masks and statuary.

a. Primitive art typically referred to _____ masks and statuary.

38. A characteristic of primitive art were striking shapes and contours.

a. Striking shapes and contours are characteristics of _____.

39. The above painting is *Les Demoiselles d'Avingon*, which was painted in 1907.

a. *Les Demoiselles d'Avingon* was painted in the year _____.

40. Its title, *Les Demoiselles d'Avingon,* translates to "The Young Women of Avignon."

a. *Les Demoiselles d'Avingon* translates to _____.

41. There is a noticeable lack of depth.

a. In *Les Demoiselles d'Avingon,* there is a noticeable lack of _____.

42. Over the course of his career, Picasso produced works that significantly shaped Surrealism and Expressionism.

a. Picasso's work greatly influenced _____ and Expressionism.

BIBLIOGRAPHY

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.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. Trends in Cognitive Sciences, 11(7), 280-289.

Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 68, 33-47.

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. Journal of Memory and Language, 34(3), 383-398.

Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. Journal of Educational Psychology, 95(3), 629-640. doi:http://dx.doi.org/10.1037/0022-0663.95.3.629

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.

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.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. Cognition, 118(1), 114-118.

Fraundorf, S. H., & Watson, D. G. (2008). Dimensions of variation in disfluency production in discourse. In J. Ginzburg, P. Healey, & Y. Sato (Eds.), Proceedings of LONDIAL 2008, the 12th Workshop on the Semantics and Pragmatics of Dialogue (pp. 131-138). London: King’s College London.

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. doi:http://dx.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. doi:http://dx.doi.org/10.1080/01690965.2013.832785

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. Linguistics, 30(1), 5-55. doi:http://dx.doi.org/10.1515/ling.1992.30.1.5

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219-224.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. Psychological Science, 22(6), 787-794. doi:10.1177/0956797611407929

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203-205. 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.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review of Psychology, 62, 621-647. doi:http://dx.doi.org/10.1146/annurev.psych.093008.131123

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. Metacognition and Learning, 11(1), 89-105.

Lim, S., Wöstmann, M., & Obleser, J. (2015). Selective attention to auditory memory neurally enhances perceptual precision. The Journal of Neuroscience, 35(49), 16094-16104. doi:http://dx.doi.org/10.1126/science.7350657

Morett, L.M., & Fraundorf, S.H. (2017, March). Top-down expectations affect integration of multiple cues to emphasis in discourse. Poster presented at CUNY 2017: Conference on Human Sentence Processing, Boston, MA.

Seufert, T., Wagner, F., & Westphal, J. (2016). The effects of different levels of disfluency on learning outcomes and cognitive load. Instructional Science, doi:http://dx.doi.org/10.1007/s11251-016-9387-8

Schachter, S., Christenfeld, N., Ravina,B.,& Bilous, F. (1991). Speech disfluency and the structure of knowledge. Journal of Personality and Social Psychology, 60, 362-367

Siu, M., & Ostendorf, M. (1996). Modeling disfluencies in conversational speech. 1 386-389 vol.1. doi:10.1109/ICSLP.1996.607135

Stolcke, A., & Shriberg, E. (1996). Statistical language modeling for speech disfluencies. 1 405-408 vol. 1. doi:10.1109/ICASSP.1996.541118

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. Journal of Experimental Psychology: General, 145(7), 918-933.

Zhang, X., Zhang, X., & Yang, X. (2016). A study of the effects of multimedia dynamic teaching on cognitive load and learning outcome. Eurasia Journal of Mathematics, Science & Technology Education, 12(11), 2851-2860. doi:http://dx.doi.org/10.12973/eurasia.2016.02308a