9-2016

Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm

Matthew Warren Lowder
University of Richmond, mlowder@richmond.edu

Fernanda Ferreira

This is a pre-publication author manuscript of the final, published article.

Prediction in the Processing of Repair Disfluencies:

Evidence from the Visual-World Paradigm

Matthew W. Lowder and Fernanda Ferreira

Department of Psychology

University of California, Davis

Final copy published in the J*ournal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 42, No. 9, pp. 1400-1416

Address correspondence to:

Matthew W. Lowder
Department of Psychology
Young Hall, One Shields Avenue
University of California, Davis
Davis, CA 95616

mlowder@ucdavis.edu

Abstract

Two visual-world eye-tracking experiments investigated the role of prediction in the processing of repair disfluencies (e.g., *The chef reached for some salt uh I mean some ketchup…*). Experiment 1 showed that listeners were more likely to fixate a critical distractor item (e.g., *pepper*) during the processing of repair disfluencies compared to the processing of coordination structures (e.g., *…some salt and also some ketchup…*). Experiment 2 replicated the findings of Experiment 1 for disfluency versus coordination constructions and also showed that the pattern of fixations to the critical distractor for disfluency constructions was similar to the fixation patterns for sentences employing contrastive focus (e.g., *…not some salt but rather some ketchup…*). The results suggest that similar mechanisms underlie the processing of repair disfluencies and contrastive focus, with listeners generating sets of entities that stand in semantic contrast to the reparandum in the case of disfluencies or the negated entity in the case of contrastive focus.

Keywords: disfluencies, repairs, contrastive focus, coordination, prediction, eye movements

Everyday speech is far from perfect. It is estimated that for every 100 words of spontaneous speech, a speaker produces between six and ten disfluencies (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995), which include filled pauses (e.g. *uh*, *um*), repetitions (e.g., *The—the man*), repairs (e.g., *The man, I mean the woman*), and other deviations from the pristine input that has traditionally been used in psycholinguistic studies of utterance comprehension. Indeed, given that disfluencies are such a widespread phenomenon in spoken language, it is critical that theories of language comprehension be able to adequately explain how disfluencies are processed and what sorts of cognitive mechanisms are available to the listener to handle this imperfect input.

Research over the past decade or so has begun to provide evidence demonstrating that speech disfluencies do affect sentence processing, with most of this work tending to focus on filled pauses such as *uh* or *um*. For example, the presence of filled pauses in speech leads listeners to expect the speaker to refer to something new or unfamiliar (Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Hudson Kam, & Tanenhaus, 2007; Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Barr & Seyfeddinipur, 2010), even among listeners as young as two years of age (Kidd, White, & Aslin, 2011). In addition, listeners can more easily integrate an unpredictable word with its sentence context when that word is preceded by a filled pause than when it is not (Corley, MacGregor, & Donaldson, 2007; see also Collard, Corley, MacGregor, & Donaldson, 2008; Corley & Hartsuiker, 2011), and the presence of filled pauses can influence the parsing and ultimate acceptability of garden-path sentences (Bailey & Ferreira, 2003, 2007; Maxfield, Lyon, & Silliman, 2009).

While most psycholinguistic research on speech disfluencies has focused on filled pauses, there has been much less work on other types of disfluencies, such as repairs. Consider, for example, the utterance "*Please pass the salt uh I mean the pepper,*" in which a word or phrase referred to as the reparandum (e.g., *the salt*) is ultimately replaced by a repair (e.g., *the pepper*), often with some sort of editing phrase intervening between the two (e.g., *uh I mean*). Sentences like these are particularly interesting because they contain lexical material that is not intended to be part of the final interpretation of the sentence. That is, somehow the listener must replace the information contained in the reparandum with the information contained in the repair. This raises at least two fundamental research questions. First, once the listener encounters the repair, how complete is the process of overwriting the reparandum from the representation of the sentence? One possibility is that the old material is completely filtered out and replaced with the new material. Alternatively, it might be the case that information about the reparandum lingers and continues to influence the listener's ultimate interpretation of the sentence. There is previous work that addresses this first question (discussed below), but a second question which is completely unexplored involves understanding the extent to which information about the reparandum affects processing even before the listener encounters the repair. In other words, are there predictive effects associated with the reparandum? A major goal of the current work is to begin to address this second question.

Lingering Effects of the Reparandum

Research by Ferreira and colleagues (Ferreira & Bailey, 2004; Ferreira, Lau, & Bailey, 2004; Lau & Ferreira, 2005) demonstrates that indeed the reparandum is not completely erased from the listener's representation of the sentence, but rather traces of the reparandum linger and continue to influence interpretation (at least offline). Initial evidence for this came from Ferreira et al., who presented participants with spoken sentences like those in (1), which manipulated the argument structure of the critical verbs. That is, whereas a verb like *drop* requires two arguments (i.e., subject and object) and can optionally take a third argument (i.e., goal), a verb like *put* requires all three. Participants listened to fluent versions of these sentences (1a and 1c), as well as disfluent versions in which the reparandum verb had a different argument structure from the repair verb (1b and 1d). The participants' task was to judge whether each sentence was grammatical or not, and they were explicitly instructed not to base these judgments simply on the presence or absence of a disfluency. Results showed that in the case of a two-argument verb, the presence of a reparandum with a more complex argument structure (1b) made these sentences less acceptable than the fluent condition (1a). Importantly, however, in the case of the three-argument verb, the presence of a reparandum with a less complex argument structure (1d) made these sentences *more* acceptable than the fluent condition (1c). In other words, the pattern of grammaticality judgments tended to move in the direction that would have been correct if the reparandum had actually been the intended verb. These results were followed-up with work by Lau and Ferreira (2005) who used a similar design to demonstrate that the structural ambiguity of the verb in the reparandum affects the acceptability of garden-path sentences.

(1a) *John says you should drop the ball.*

(1b) *John says you should put—uh drop the ball.*

(1c) *John says you should put the ball.*

(1d) *John says you should drop—uh put the ball.*

Findings such as these demonstrating that the reparandum has lingering effects on sentence interpretation led to the development of the Overlay model of disfluency processing (Ferreira & Bailey, 2004; Ferreira et al., 2004). The Overlay model takes as its starting point the assumption that sentence processing proceeds incrementally and operates within a lexicalized tree-adjoining grammar framework, according to which each word in the language is associated with an elementary syntactic tree that represents the possible structures it licenses. Under normal processing circumstances, the syntactic trees of individual words are combined through operations called Substitution and Adjunction that ultimately give rise to a global syntactic representation of the sentence. But what happens when the processor encounters a local ungrammaticality triggered by a repair disfluency (e.g., *put—uh drop*)? The Overlay model proposes that syntactic trees for both the reparandum and repair are retrieved, and the processor first tries to combine them using standard operations of Substitution or Adjunction. When these operations fail, the repair tree is superimposed over the reparandum tree through an operation called Overlay. Importantly, when the reparandum and the repair have different argument structures, the result is imperfect overlap. In these cases, features of the reparandum lurk in the background and continue to influence processing. For example, when the repair verb *drop* replaces the reparandum verb *put*, the node for an obligatory goal argument is still present in the representation of the sentence, which can lead the processor to judge the sentence as ungrammatical when this slot is never filled. In this way, the process of reanalyzing and reinterpreting repair disfluencies can be described as incomplete or "good enough," akin to the offline good enough representations that comprehenders derive in the processing of garden-path sentences (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Patson, 2007).

Predictive Effects of the Reparandum

As we have described previously (see Lowder & Ferreira, 2016), the Overlay model can be thought of as "backward-looking" in the sense that the Overlay operation is not initiated until a local ungrammaticality is detected; that is, the process of reinterpreting the utterance does not begin until two syntactic trees have been retrieved and the processor's attempts to combine them through standard operations have failed. However, it seems reasonable to expect that listeners may in some cases predict a repair before it is spoken. Consider again the utterance "*Please pass the salt uh I mean…*" Here, it seems likely that the listener will deduce from the editing phrase "*uh I mean*" that the speaker has made an error, and will then use relevant information to predict the upcoming repair before it is spoken. In this example, the close semantic relationship between *salt* and *pepper* may be particularly useful. However, the listener may also use information about the context to generate predictions. For example, if the speaker is already holding the salt shaker and says "*Please pass the salt…*", the listener may notice the error, mark *salt* as a reparandum, and predict the repair (e.g., *pepper*) even before the speaker becomes overtly disfluent.

Such a "forward-looking" theoretical approach to the processing of repair disfluencies can be derived from basic assumptions of Noisy Channel models of language comprehension (Gibson, Bergen, & Piantadosi, 2013; Gibson, Piantadosi, Brink, Bergen, Lim, & Saxe, 2013). A Noisy Channel framework argues that linguistic input is inherently imperfect, and thus a major goal of the comprehender is to reconcile the actual linguistic signal with semantic and pragmatic knowledge, as well as relevant knowledge about the speaker and the context. The comprehender may then mentally correct any perceived error into a more plausible alternative. Although Noisy Channel accounts of language comprehension were not developed with the explicit purpose of explaining the processing of repair disfluencies, the major assumptions of this theoretical framework easily lend themselves to the domain of disfluencies, offering a range of testable hypotheses.

The active, anticipatory nature of the Noisy Channel framework factors into a much broader trend in psycholinguistics that examines the role of prediction during online language comprehension. Indeed, a large body of research (see Kuperberg & Jaeger, 2016, for a recent review) suggests that comprehenders rapidly generate predictions about upcoming input from lower levels of sublexical and lexical representations (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Brothers, Swaab, & Traxler, 2015; DeLong, Urbach, & Kutas, 2005; Dikker, Rabagliati, Farmer, & Pylkkänen, 2010; Farmer, Christiansen, & Monaghan, 2006; Kamide, Altmann, & Haywood, 2003; Kim & Lai, 2012; Rayner, Slattery, Drieghe, & Liversedge, 2011), up to higher levels of representation associated with event structures and schematic knowledge (e.g., Knoeferle & Crocker, 2006, 2007; Matsuki, Chow, Hare, Elman, Scheepers, & McRae, 2011; McRae, Hare, Elman, & Ferretti, 2005; Metusalem, Kutas, Urbach, Hare, McRae, & Elman, 2012; Paczynski & Kuperberg, 2012). This work has tended to examine the extent to which comprehenders generate predictions under "normal" processing circumstances—that is, when the linguistic input is encountered at very fast speeds that are typical of natural reading or listening. Even at such a rapid pace, most work suggests that prediction is an inherent part of the language processing system that helps the comprehender to engage in efficient and accurate interpretation. It therefore seems possible that listeners may be even more likely to generate predictions when the speaker signals that he has made an error, especially when the stream of linguistic input is interrupted by filled pauses that provide the listener with extra time to generate these predictions.

We have recently argued (Ferreira & Lowder, in press; Lowder & Ferreira, 2016) that the anticipatory processing that occurs in language comprehension is supported—at least in certain cases—by a mechanism that generates a set of predictions about the upcoming input. We conceptualize this mechanism as similar to the creation of an alternate set during the processing of linguistic focus (e.g., Jackendoff, 1972; Rooth, 1992), such as when a speaker says "only x" or "not x," thus cueing the listener to generate a set of candidates that stand in semantic contrast to "x." Further, we have proposed that the items in the alternate set are weighted according to their probabilities, which are determined by the linguistic content as well as relevant contextual knowledge. An intriguing possibility that we investigate in the current paper involves understanding whether a prediction mechanism of this sort operates in the case of repair disfluencies, such that the editing phrase "uh I mean" might cue the listener to generate an alternate set, similar to what happens in the case of contrastive focus.

There is very little previous work investigating the possible role of prediction in the context of repair disfluencies. One study by Corley (2010) took as its starting point early demonstrations that listeners use information about the selectional restrictions of a verb to predict its upcoming object (Altmann & Kamide, 1999). Specifically, Corley showed that these selectional restrictions are overridden to some extent when a reparandum verb is replaced by a repair verb (e.g., *The boy will eat uh move the cake*). Although this work suggests that listeners can rapidly and incrementally use a repair to update their predictions about how a sentence is likely to unfold, it does not provide evidence about how listeners might use information about the reparandum to generate predictions about the repair before the repair is spoken. In two visual-world eye-tracking experiments (Tanenhaus, Spivey-Knowlton, Eberhardt, & Sedivy, 1995), the current study directly tests for predictive effects of the reparandum during the processing of repair disfluencies. Further, we examine the extent to which the mechanisms of prediction operating in disfluency contexts are similar to or different from the mechanisms of prediction in contexts of noun coordination and contrastive focus.

Experiment 1

Experiment 1 was designed to measure the strength of the online predictions listeners make during the processing of repair disfluencies. Specifically, we tested the four conditions illustrated in (2). Participants listened to these sentences while they viewed images on a computer screen. The array of images was identical across the four conditions and always consisted of an image representing the first noun phrase (NP1; e.g., salt), an image representing the second noun phrase (NP2; e.g., ketchup), an image representing a critical unnamed distractor (e.g., pepper), and an image representing a random distractor (e.g., milk). NP1 and the critical distractor (e.g., salt and pepper) always represented a pair of nouns high in semantic relatedness, designed such that listeners would be more likely to predict the critical distractor than NP2. We recorded participants' eye movements to these pictures as they listened to sentences representing a Disfluency condition (2a), a Coordination condition (2b), an NP1 Only condition (2c), and an NP2 Only condition (2d).

(2a) *The meat was pretty bland, so the chef reached for some salt uh I mean some ketchup, which made it much more flavorful.* (Disfluency)

(2b) *The meat was pretty bland, so the chef reached for some salt and also some ketchup, which made it much more flavorful.* (Coordination)

(2c) *The meat was pretty bland, so the chef reached for some salt, which made it much more flavorful.* (NP1 Only)

(2d) *The meat was pretty bland, so the chef reached for some ketchup, which made it much more flavorful.* (NP2 Only)

The design of this experiment thus allows us to examine the time course of listeners' processing of repair disfluencies, including predictions made during the editing phrase that intervenes between the reparandum and the repair. Any increases in looks to the critical distractor item for the Disfluency condition versus the Coordination condition would provide evidence for predictive effects of the reparandum above and beyond any baseline semantic priming effects. Further, examination of the NP1 Only and NP2 Only conditions provides an indication of listeners' tendencies to shift their gaze to the named entity—that is, to ensure that listeners are not biased toward looking at one of these images over the other.

Method

Participants. Twenty-eight students at the University of South Carolina participated in this experiment in exchange for course credit. They were all native English speakers and reported normal or corrected-to-normal vision.

Materials. Forty sets of experimental items were created (see example 2). The starting point for each of these sets was a pair of nouns that were judged by the experimenters to be high in semantic relatedness (e.g., *salt-pepper*; *bread-butter*; *dog-cat*). Sentence contexts for the Disfluency and Coordination conditions (2a & 2b) were constructed such that the first noun from the pair (e.g., *salt*) served as NP1 in the sentence (i.e., either the reparandum or the first conjunct), but the second noun from the pair never appeared in the sentence. Instead, the noun that actually served as NP2 (e.g., *ketchup*) was always a plausible continuation of the sentence, but never as plausible as the unnamed semantic associate (see more information about plausibility pretesting below). The only difference between the Disfluency and Coordination conditions was the words that intervened between NP1 and NP2 (i.e., *uh I mean* versus *and also*). The NP1 Only and NP2 Only conditions were then constructed by inserting only one of the NPs into the sentence while holding the rest of the sentence constant (see 2c & 2d). Each set of experimental items was associated with a visual display that consisted of four color images (see Figure 1) representing NP1 (e.g., salt), NP2 (e.g., ketchup), the critical distractor (e.g., pepper), and a random distractor (e.g., milk). See Appendix for the full set of experimental stimuli.

To validate our judgments of plausibility differences among possible continuations of our experimental items, we collected plausibility ratings from 36 native English speakers recruited through Amazon's Mechanical Turk. Items were counterbalanced among three lists that included the Disfluency version of each item in written form, but with different repairs that represented the critical distractor (e.g., *The meat was pretty bland, so the chef reached for some salt uh I mean some pepper*), NP2 (e.g., *The meat was pretty bland, so the chef reached for some salt uh I mean some ketchup*), and the random distractor (e.g., *The meat was pretty bland, so the chef reached for some salt uh I mean some milk*). Raters were instructed to indicate how likely they felt someone might actually say each sentence on a scale from 1 (highly unlikely) to 7 (highly likely). The mean ratings for each condition were 6.1 (critical distractor), 4.2 (NP2), and 1.5 (random distractor). Differences between each pairwise comparison were highly reliable, ts > 11.5, ps < .001, and in fact all 40 sets of items showed the same pattern of mean plausibility ratings with the critical distractor being the most plausible continuation, the random distractor being the least plausible, and NP2 falling between these two extremes. These ratings thus establish the unspoken critical distractor as a highly plausible offline continuation of the utterance, suggesting that listeners of these utterances may also actively predict the critical distractor during online processing.

The materials also included a set of 45 filler items representing a variety of different sentence types, each with a corresponding visual array of four images. Ten of these filler items contained a repair disfluency in which the actual repair was highly predictable based on the reparandum (e.g., *Tammy wanted fresh breath, so she bought a pack of gum uh I mean mints from the grocery store*). The purpose of these items was to discourage participants from developing strategies based on a realization that the most predictable repair in the critical items was never actually spoken. These ten filler items were based on a separate cloze norming study in Mechanical Turk, in which participants were presented with written sentence fragments up to, but not including the repair (e.g., *Tammy wanted fresh breath, so she bought a pack of gum uh I mean* ________), and were asked to fill in the blank with a plausible continuation. Ten items in which the repair had a cloze probability of greater than 50% were chosen as fillers in the current experiment. Notably, these ten predictable filler items were all repair disfluencies, and there were no predictable filler items that took on a coordination structure. We return to this point in Experiment 2.

The sentences were recorded by the first author—a male native speaker of American English. Sentences were read at a normal speaking rate and with prosodic features the speaker judged to be as natural as possible. The 40 sets of experimental sentences and their corresponding visual displays were counterbalanced across four lists so that each participant was presented with only one version of each item and so that each participant was presented with the same number of items from each of the four conditions.

Procedure. Eye movements were recorded with an EyeLink 1000 system (SR Research) at a sampling rate of 1,000 Hz, which was calibrated at the beginning of each session and was recalibrated throughout the session as needed. A chinrest was used to minimize head movement. Viewing was binocular, but only the right eye was tracked. Participants were told that they would view images on the computer screen while also listening to sentences. Participants were explicitly told that "some of the sentences might contain errors," but were instructed to just do their best to understand each sentence; there was no explicit task other than to try to understand the sentences. At the start of each trial, a fixation point was presented in the center of the screen. When the participant's gaze was steady on this point, the experimenter pressed a button that presented the visual display for the trial, and 3,000 ms later the sentence was presented via headphones. After the sentence finished playing, the images disappeared and the fixation point for the next trial appeared.

Participants were first presented with five of the filler sentences. After this warm-up block, the remaining sentences were presented in a fixed pseudorandom order under the constraint that no more than two trials from the same condition could be presented consecutively. Further, no more than two trials containing a repair disfluency (including the ten predictable filler trials) could appear consecutively. The locations of the four images within a visual display were randomized on each trial.

Analysis

The dependent variable was whether a particular image was fixated within a particular time window (coded as 1 if a fixation was made to the image and 0 if it was not). Separate analyses were conducted on each of the three images of theoretical importance: NP1, NP2, and the critical distractor. Because the data were binomial, analyses relied on logit mixed effects models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008) using the glmer function from the lme4 package in R (Bates, Maechler, & Dai, 2012). The models included experimental condition as the fixed effect (Disfluency versus Coordination), as well as subjects and items as crossed random effects. The random effects structure included the maximally appropriate random intercepts as well as by-subject and by-item random slopes for experimental condition. In cases where the model failed to converge, the random effects structure was sequentially simplified until convergence was achieved (Barr, Levy, Scheepers, & Tily, 2013). The final random effects structure for each model is presented in Table 1.

Within each sound file, we marked the onsets of target words that were then used to create time windows for analysis. For all four conditions, we marked the onset of the first critical noun as time point zero. For the Disfluency and Coordination conditions, we also marked the onsets of the immediately following word (i.e., *uh* versus *and*), as well as the onset of the second critical noun (i.e., the repair or the second conjunct). Analysis time windows for the Disfluency and Coordination conditions were constructed as follows: The first window measured from the onset of the first critical noun to the onset of *uh* or *and*; the second window measured from the onset of *uh* or *and* to the onset of the second critical noun; and the third window measured from the onset of the second critical noun until 800 ms had elapsed. Each time window was shifted forward by 200 ms to account for the time required to initiate a signal-based saccade.

As the utterances were recorded to sound as natural as possible, this led to systematic differences in the duration of the first and second time windows across conditions. Specifically, the first time window was longer for the Coordination condition (946 ms) than for the Disfluency condition (817 ms), $t(39) = 6.87$, $p < .001$, whereas the second time window was longer for the Disfluency condition (1058 ms) than for the Coordination condition (652 ms), $t(39) = 29.11$, $p < .001$. We return to this issue below. We also analyzed the maximum fundamental frequency (F0 max) of N1 (e.g., "*salt*") and found that F0 max did not differ significantly between the Disfluency (300 Hz) and the Coordination condition (272 Hz), $t(39) = 1.05$, $p > .30$.

Results and discussion

Figure 2 displays the probability of fixating each of the four picture types within each of the four experimental conditions across the time course of the utterance. The top half of Figure 2 shows fixation probabilities for the NP1 Only and NP2 Only conditions. These plots show that in both conditions participants rapidly shifted their gaze to the named entity (e.g., *salt* or *ketchup*), suggesting that there was no inherent bias to look at one of these pictures over the other.

The bottom half of Figure 2 shows fixation probabilities for the Disfluency and Coordination conditions, and Table 1 shows results of the mixed effects analyses. Visual inspection of the first time window (from the onset of *salt* to the onset of *uh*/*and*) shows that in both conditions participants tended to look at the named entity rather than the other three images. Consistent with this observation, statistical analyses of fixation probabilities within the first time window showed no differences between the Disfluency and Coordination conditions.

Visual inspection of the second time window (from the onset of *uh*/*and* to the onset of *ketchup*) shows a decrease in proportion of fixations to NP1, but an increase in proportion of fixations to the critical distractor—a pattern that appears stronger in the Disfluency condition compared to the Coordination condition. Statistical analyses revealed that indeed listeners were significantly more likely to fixate the critical distractor in the Disfluency condition versus the Coordination condition in this time window. In addition, there was a greater likelihood of fixating NP2 in the Disfluency condition versus the Coordination condition in this time window. This pattern suggests that listeners were more likely to generate predictions about the upcoming word during the "*uh I mean*" portion of a repair disfluency compared to the "*and also*" portion of a coordination construction[1].

[1] As noted above, there were systematic differences in the lengths of the analysis time windows between the Disfluency and Coordination conditions. In an effort to address this issue, we conducted supplemental analyses in which the duration of the first time window of the Coordination condition was truncated to equal the duration of this window for the Disfluency condition, whereas the duration of the second time window of the Disfluency condition was truncated to equal the duration of this window for the Coordination condition. The patterns of results for these analyses were largely similar to the patterns observed in the primary analyses. Of greatest theoretical importance, listeners were more likely to fixate the critical distractor in the Disfluency condition compared to the Coordination condition during the second time window (estimate = -0.660, $SE = 0.270$, $z = -2.447$, $p < .02$), as in the original analysis. This issue of unequal time windows is addressed more directly in Experiment 2.

A similar pattern emerged in the third time window (from the onset of *ketchup* until 800 ms had elapsed). Specifically, listeners were more likely to fixate the critical distractor, as well as NP2 (marginally significant), in the Disfluency condition compared to the Coordination condition. In contrast, listeners were more likely to fixate NP1 in the Coordination condition compared to the Disfluency condition.

As an additional post-hoc comparison, Figure 3 plots fixation probabilities for the ten disfluency filler items in which the actual repair was highly predictable based on the reparandum (e.g., *Tammy wanted fresh breath, so she bought a pack of gum uh I mean mints from the grocery store*). Although not directly comparable with the experimental Disfluency condition, visual inspection of this plot reveals a clear pattern in which listeners rapidly shift their gaze from the reparandum to the predicted repair, beginning at the onset of "*uh*," and continuing to gain strength through the realization of this prediction. Also note that there is no tendency for listeners to shift their gaze back to NP1 once it has been established as the reparandum. Thus, the pattern of prediction for these filler items resembles the pattern of prediction observed in the experimental Disfluency condition; however, in this latter case, the accumulating strength of the prediction is cut short when a less predictable repair is uttered instead.

The results of Experiment 1 demonstrate that listeners use semantic information about a reparandum to actively predict an upcoming repair during the editing phrase of a speech disfluency. Further, the prediction effect observed in the context of repair disfluencies was stronger than the prediction effect observed in the context of NP coordination. Specifically, listeners were more likely to fixate the critical distractor (e.g., *pepper*) in the Disfluency condition (e.g., *…the chef reached for some salt uh I mean…*) compared to the Coordination condition (e.g., *…the chef reached for some salt and also…*), which rules out a simple semantic-priming-based explanation of these findings. Listeners were also more likely to fixate the actual, less plausible NP2 (e.g., *ketchup*) in the Disfluency condition compared to the Coordination condition, suggesting that at least in some cases listeners may have been activating several possible candidates for the repair.

This pattern of results suggests that listeners are particularly sensitive to linguistic cues indicating that the speaker has made an error. One possible explanation for the results of Experiment 1 is that when listeners heard "*salt uh I mean*…", they immediately interpreted this to mean that the chef did *not* reach for the salt, but rather reached for something else instead, and they might then use semantic information about the reparandum to generate possible repairs that stand in contrast to the reparandum, in the sense of the type of contrast set implied by the presence of semantic focus (e.g., Rooth, 1992). Experiment 2 was designed to directly test this possibility, as well as to address some potential methodological shortcomings of Experiment 1.

Experiment 2

Experiment 1 demonstrated that listeners were more likely to generate predictions about the repair in the context of a repair disfluency than the second conjunct in the context of a noun coordination structure. However, this conclusion is complicated by two methodological issues. First, as described above, there were systematic differences in the duration of the time windows between the onsets of the critical words across the Disfluency and Coordination conditions. Most notably, there tended to be more time between the onset of "*uh*" and the onset of the repair in the Disfluency condition versus the onset of "*and*" and the onset of the second conjunct in the Coordination condition, and it could therefore be argued that the difference in fixations to the critical distractor stemmed entirely from listeners' having extra time to generate predictions in the Disfluency condition (though see Footnote 1). Second, recall that ten of the filler trials in

Experiment 1 were repair disfluencies in which the actual repair was the most predictable word. In contrast, there were no "special" filler trials that employed a coordination construction in which the second conjunct was the most predictable word. Thus, one could argue that listeners might on some level learn that their predictions sometimes turn out to be correct in disfluency contexts, whereas this is never the case in coordination contexts, and that this difference might bias listeners toward making stronger predictions in the case of disfluencies. Accordingly, one goal of Experiment 2 was to determine whether we could replicate the key findings from Experiment 1 while also addressing these two methodological points.

The second goal of Experiment 2 was to further explore the mechanism underlying prediction in the processing of repair disfluencies. Under a Noisy-Channel approach to language comprehension, listeners are particularly sensitive to speaker errors and use relevant linguistic and contextual information to mentally correct an error into a word or phrase that seems to be consistent with the speaker's intended meaning. One way this might occur is that the listener uses the error signal (e.g., *uh I mean*) to negate the information contained in the reparandum (e.g., *not the salt*) and to then generate a set of alternates that stand in semantic contrast to the reparandum and seem like plausible repairs (e.g., *pepper*, *ketchup*, *mustard*) (e.g., Rooth, 1992). To the extent that this is the case, we should find that listeners show similar tendencies to generate predictions during the processing of repair disfluencies (*…some salt uh I mean…*) and contrastive focus (*not some salt but rather…*).

A great deal of previous work has demonstrated that language comprehenders are sensitive to linguistic cues that mark the focus of the sentence, including manipulations of syntactic structure (e.g., Birch & Rayner, 1997, 2010; Lowder & Gordon, 2015; Morris & Folk, 1998), prosodic features (e.g., Arnold, 2008; Birch & Clifton, 1995, 2002; Cutler & Foss, 1977;

Dahan, Tanenhaus, & Chambers, 2002), and discourse context (e.g., Benatar & Clifton, 2014; Cutler & Fodor, 1979; Sturt, Sanford, Stewart, & Dawydiak, 2004). Similarly, recent work has shown that comprehenders are quite sensitive to the negation operator "*not*," which places linguistic focus on one entity and signals that a contrasting entity is forthcoming. Specifically, Orenes, Beltrán, and Santamaría (2014) presented participants with target utterances that contained negation (e.g., "*The figure is not red.*") while they viewed an array of four different-colored shapes. Critically, the verbal context that was spoken immediately before the target utterance instructed participants either that the target figure would be one of two colors (binary context; e.g., "*The figure is red or green.*"), or that the target figure could be any color (multary context; e.g., "*The figure is red or green or yellow or blue*."). In the case of the binary context, listeners shifted their gaze away from the negated entity and toward the alternate, whereas in the case of the multary context, listeners continued to fixate the negated entity. This pattern suggests that listeners use information about the negated entity to predict the contrasting entity when a strongly activated alternate is available.

We hypothesized that a similar mechanism underlies the generation of predictions during the processing of repair disfluencies, and Experiment 2 was designed to test this hypothesis. Participants listened to utterances representing the four conditions illustrated in (3) and viewed arrays of images depicting NP1, NP2, a critical distractor, and a random distractor, just as in Experiment 1. We again compared the Disfluency condition (3a) to the Coordination condition (3b) in an attempt to replicate the main findings of Experiment 1, but with crucial modifications to the materials (see Method section). We also compared the Disfluency condition (3a) to a Focus condition (3c) in which NP1 is negated and NP2 stands in semantic contrast to NP1. To the extent that similar mechanisms underlie the generation of predictions during the processing of repair disfluencies and contrastive focus, we should observe similar tendencies to predict the upcoming NP2 across these two constructions, as measured by fixations to the critical distractor. Finally, Experiment 2 also included an NP2 Only condition (3d), just as in Experiment 1, to examine baseline looking patterns to this less predictable NP in isolation.

(3a) *The meat was pretty bland, so the chef reached for some salt uh I mean some ketchup, which made it much more flavorful.* (Disfluency)

(3b) *The meat was pretty bland, so the chef reached for some salt and also some ketchup, which made it much more flavorful.* (Coordination)

(3c) *The meat was pretty bland, so the chef reached for not some salt but rather some ketchup, which made it much more flavorful.* (Focus)

(3d) *The meat was pretty bland, so the chef reached for some ketchup, which made it much more flavorful.* (NP2 Only)

Method

Participants. Twenty-eight students at the University of South Carolina participated in this experiment in exchange for course credit. They were all native English speakers and reported normal or corrected-to-normal vision. None had participated in Experiment 1.

Materials. The experimental materials were the same 40 sets of items used in Experiment 1, but the NP1 Only condition was eliminated and a Focus condition was created instead (see example 3). The Focus condition was constructed by inserting the word “not” before NP1 and then inserting “but rather” between NP1 and NP2. Otherwise, the Focus condition was identical to the Disfluency and Coordination conditions. The same visual displays used in Experiment 1 were used in Experiment 2. See Appendix for the full set of experimental stimuli.

The materials also included a set of 45 filler items representing a variety of different sentence types, each with a corresponding visual array of four images. Recall that in Experiment 1 ten of these filler items contained a repair disfluency in which the actual repair was highly predictable based on the reparandum (e.g., *Tammy wanted fresh breath, so she bought a pack of gum uh I mean mints from the grocery store*). In contrast, Experiment 2 included 12 filler items in which NP2 was highly predictable, but these represented disfluency, coordination, and contrastive focus structures equally. In other words, four filler items were repair disfluencies in which the repair was predictable, four were coordination structures in which the second conjunct was predictable, and four were focus structures in which the contrasted alternate was predictable. As in Experiment 1, the purpose of these items was to discourage participants from developing strategies based on a realization that the most predictable NP2 was never actually spoken in the experimental items; however, unlike Experiment 1, these special fillers resembled the three critical conditions in equal numbers. The sentences chosen as the 12 special filler items were based on the Mechanical Turk cloze norming study described in Experiment 1.

As noted above, the sentences in Experiment 1 were recorded to sound as natural as possible, which had the consequence of introducing systematic variability across conditions in the durations of the time windows between critical word onsets. Thus, the sentences were rerecorded with the goal of sounding as natural as possible while also being mindful of the timing of the critical words (see analysis of time window durations below). The experimental sentences were counterbalanced across four lists as described in Experiment 1.

Procedure. All aspects of the procedure were identical to the procedure described in Experiment 1.

Analysis

Data were analyzed using logit mixed effects models as in Experiment 1. The models included experimental condition (Disfluency, Coordination, and Focus) as the fixed effect, as well as subjects and items as crossed random effects. The experimental conditions were compared using two a priori contrasts defined using treatment coding: the first contrast compared the Disfluency condition to the Coordination condition, and the second contrast compared the Disfluency condition to the Focus condition. As in Experiment 1, the random effects structure included the maximally appropriate random intercepts as well as by-subject and by-item random slopes. In cases where the model failed to converge, the random effects structure was sequentially simplified until convergence was achieved. The final random effects structure for each model is presented in Table 2.

As in Experiment 1, we marked the onsets of target words within each sound file that were then used to create time windows for analysis. For all four conditions, we marked the onset of the first critical noun as time point zero. For the Disfluency, Coordination, and Focus conditions, we also marked the onsets of the immediately following word (i.e., *uh* versus *and* versus *but*), as well as the onset of the second critical noun. Analysis time windows for the Disfluency, Coordination, and Focus conditions were constructed in a similar manner as described in Experiment 1: The first window measured from the onset of the first critical noun to the onset of *uh*, *and*, or *but*; the second window measured from the onset of *uh*, *and*, or *but* to the onset of the second critical noun; and the third window measured from the onset of the second critical noun until 800 ms had elapsed. As in Experiment 1, each time window was shifted forward by 200 ms.

In contrast to Experiment 1, there were no differences in the duration of these time windows across conditions (Window 1: Disfluency = 796 ms, Coordination = 791 ms, Focus = 803 ms, $F < 1$; Window 2: Disfluency = 1,053 ms, Coordination = 1,077 ms, Focus = 1,063 ms, $F < 1$). As in Experiment 1, we analyzed F0 max for N1 across conditions. We found that F0 max for the Disfluency condition (276 Hz) was significantly higher than for the Coordination condition (210 Hz), $t(39) = 2.23$, $p < .05$, and was marginally higher than for the Focus condition (228 Hz), $t(39) = 1.90$, $p < .10$. The Coordination and Focus conditions did not differ from one another, $t < 1$. We return to this issue in the Discussion.

Results and discussion

Figure 4 displays the probability of fixating each of the four picture types within each of the four experimental conditions across the time course of the utterance. The top-right portion of Figure 4 shows fixation probabilities for the NP2 Only condition. As in Experiment 1, participants rapidly shifted their gaze to the named entity (e.g., *ketchup*).

Statistical analyses examined differences in fixation probabilities to the pictures representing NP1, NP2, and the critical distractor between the Disfluency, Coordination, and Focus conditions within each of the three time windows. Results of these analyses are summarized in Table 2. In the first time window (from the onset of *salt* to the onset of *uh/and/but*), there were fewer fixations to NP1 in the Focus condition compared to the Disfluency condition, but no difference between the Disfluency and Coordination conditions. This finding is consistent with the idea that negation of NP1 led to a decreased likelihood that listeners would fixate this object. In this same time window, there were also more fixations to the critical distractor and NP2 in the Focus condition compared to the Disfluency condition, but no difference between the Disfluency and Coordination conditions. Thus, negation of NP1 led listeners to shift their gaze to objects that might be named in contrast to NP1.

Similar patterns were observed in the second time window (from the onset of *uh/and/but* to the onset of *ketchup*). Specifically, there were fewer fixations to NP1 in the Focus condition compared to the Disfluency condition, but more fixations to the critical distractor and NP2 in the Focus condition compared to the Disfluency condition. Importantly, however, there were also more fixations to the critical distractor in the Disfluency condition compared to Coordination condition, replicating the critical finding from Experiment 1.

In the third time window (from the onset of *ketchup* until 800 ms had elapsed), there were again fewer fixations to NP1 in the Focus condition compared to the Disfluency condition, but there were also fewer fixations to NP1 in the Disfluency condition compared to the Coordination condition. In addition, there were more fixations to the critical distractor in the Disfluency condition compared to the Coordination condition; however, there was no difference between the Disfluency condition and the Focus condition. There were no significant differences between conditions in fixations to NP2.

To summarize the results, at the onset of NP1 (e.g., *salt*), where there is no difference in the speech input between the Disfluency and Coordination conditions, there were also no differences in patterns of fixations between these two conditions. However, the presence of the negation operator (e.g., *not the salt*) led to decreased fixations to NP1 in the Focus condition and instead immediate generation of predictions about the upcoming NP2, as reflected by increased looks to the critical distractor (e.g., *pepper*) as well as NP2 (e.g., *ketchup*). Similar patterns for the comparison between the Focus condition and the Disfluency condition persisted into the second time window. Further, there were more fixations to the critical distractor during "*uh I* *mean*” in the Disfluency condition compared to during “*and also*” in the Coordination condition, suggesting that listeners used this error signal to begin generating predictions about the repair. In the window following the onset of NP2 (e.g., *ketchup*), there was no difference in fixations to the critical distractor between the Focus and Disfluency conditions, but there were more fixations to the critical distractor in the Disfluency condition than the Coordination condition.

As in Experiment 1, we conducted an additional post-hoc comparison of the critical filler items in which the most predictable NP2 was the word that was actually spoken, representing disfluency, coordination, and contrastive focus constructions. Fixation probabilities are plotted in Figure 5. Again, although not directly comparable with our experimental items, visual inspection of the plots reveals interesting patterns: In the Focus condition, the tendency to fixate NP1 is rather weak, but instead, listeners quickly begin to shift their gaze to the predicted NP2, which continues to gain strength through the utterance. Similarly, at the onset of “*uh*” in the Disfluency condition, listeners shift their gaze from the reparandum to the predicted repair, and this prediction continues to gain strength. In contrast, there appears to be only a weak tendency to predict the second conjunct in the Coordination condition.

The results of Experiment 2 replicate and extend the findings of Experiment 1. First, we again demonstrated that listeners are more likely to generate predictions during the processing of repair disfluencies than during the processing of coordination. Whereas this conclusion was complicated by potential methodological issues in Experiment 1, we observed an identical pattern in Experiment 2, which carefully controlled the duration of the analysis windows across conditions and also introduced predictable filler trials that represented all our structures of interest. Second, Experiment 2 examined listeners’ predictions during the processing of contrastive focus. These results demonstrated that just as listeners begin predicting an upcoming repair in response to "*uh*" in the processing of repair disfluencies, they also begin predicting an upcoming alternate NP in response to the negation in the processing of contrastive focus. This suggests that listeners show similar tendencies to generate predictions in the processing of repair disfluencies and contrastive focus; however, these two constructions display important differences regarding when relevant information becomes available to the listener, as well as how listeners might then use this information during processing. We discuss this in greater detail below.

In rerecording our materials to equate the conditions on duration of the analysis windows, we unintentionally introduced systematic variation in the pitch of N1. That is, whereas F0 max for this word did not differ between the Disfluency and Coordination conditions in Experiment 1, the pitch of this word in Experiment 2 tended to be higher for the Disfluency condition than the Coordination and Focus conditions. Although this difference is undesirable, there are several reasons to doubt that our eye movement results can be explained by these pitch differences. First, although the pitch of N1 for Disfluency versus Coordination differed between Experiments 1 and 2, the eye movement patterns to the picture representing N1 did not. That is, in both experiments listeners showed similar tendencies to direct their gaze to "*salt*" upon hearing this word, suggesting that this pitch difference did not affect listeners' predictions. Second, the critical difference between the Disfluency and Coordination conditions did not emerge until the onset of "*uh I mean*" versus "*and also*", and it would be difficult to explain how pitch differences in N1 that did not affect patterns of looks to the corresponding object could explain differences in looks to the critical distractor. Finally, the fixation differences that were observed in Experiment 2 do not pattern with the pitch differences. That is, whereas the fixation patterns revealed similar tendencies to generate predictions for the Disfluency and Focus conditions versus the

Coordination conditions, the prosody analysis showed that "*salt*" was spoken at a higher pitch in the Disfluency condition versus the Focus and Coordination conditions. We acknowledge that future work should consider more carefully the role prosody might play in shaping listeners' interpretation of disfluencies; however, we conclude that the unintentional prosodic variations observed in the current experiment do not explain the observed differences in fixation patterns.

General Discussion

Taken together, the two experiments reported in this article demonstrate that listeners engage in rapid prediction of an upcoming repair during the processing of repair disfluencies. In Experiment 1, listeners displayed a greater tendency to predict the upcoming repair when the speaker signaled that he had made an error (e.g., "…*some salt uh I mean…*"), compared to their tendency to predict the second conjunct when the speaker signaled the presence of a coordination structure (e.g., "…*some salt and also*…"). Experiment 2 replicated this finding using modified stimuli and also examined prediction during the processing of contrastive focus (e.g., …*not some salt but rather…*"). Results showed that listeners use information about the negated NP to immediately begin generating predictions about the alternate NP—a pattern that resembles listeners' tendency to generate predictions during the processing of repair disfluencies. In the remainder of this section, we first discuss more precisely how the mechanisms of prediction operating during the processing of repair disfluencies and contrastive focus are similar and different from one another. We then discuss implications of this work to models of disfluency processing and language prediction more generally.

In the case of both contrastive focus and repair disfluencies, the listener receives an overt cue that something is about to be said that will stand in contrast to another word or phrase in the sentence; however, the timing of these cues differs between the two constructions. Consider first the case of contrastive focus, in which the listener receives the negation cue early (e.g., "…*not the salt but rather…*"). Here, the negation operator "*not*" directly focuses one NP, signaling to the listener that an upcoming word will stand in contrast to it. We propose that listeners use semantic knowledge about the negated NP, as well as other sources of relevant information such as the visual context, to generate a set of alternate entities that are likely to be mentioned instead. In cases where a strongly activated alternate is available (see Orenes et al., 2014), listeners shift their attention to this item. Experiment 2 provided evidence for this idea, as listeners showed a decreased tendency to look at NP1 (e.g., *salt*) in the Focus condition during the first time window, and instead shifted their gaze to the critical distractor (e.g., *pepper*)—a tendency that grew stronger over the course of the second time window.

We propose that a similar mechanism underlies the process of generating predictions in the case of repair disfluencies. If a speaker says, "*Please pass the salt uh I mean…*", we propose that the listener interprets this to mean *not* to pass the salt, and then uses this information to predict a set of likely repairs that stand in semantic contrast to the reparandum. In both contrastive focus and repair disfluencies, the items in the alternate set will be weighted by their probabilities, determined by both the linguistic content as well as relevant knowledge (see Ferreira & Lowder, in press, for further discussion).

It is important to note, however, that in the case of repair disfluencies (unlike in the case of contrastive focus), the listener receives a signal that the speaker has made an error after having first processed the reparandum; thus, the listener has an extra challenge in having to decide which part of the utterance the speaker is about to correct. In the context of our experiments, listeners are of course presented with a limited array of items. However, the set of possible repairs is much larger in the real world. Consider again the example used in our experiments:

"*The meat was pretty bland, so the chef reached for some salt uh I mean*…" Although it seems most likely in this case that "*salt*" will be the reparandum, it could instead be the case that the speaker is about to go back and correct an earlier portion of the sentence (e.g., replace "*meat*" with "*fish*," or replace "*chef*" with "*sous chef*"). It is also possible that "*salt*" is indeed the intended reparandum but that the speaker has not made an error but rather intends to replace this word with a more specific term (e.g., "*kosher salt*," "*fleur de sel*") either to resolve a potential ambiguity or for another pragmatically appropriate reason (see Levelt's, 1983, distinction between error repairs and appropriateness repairs). Given the relatively sparse work on the processing of repair disfluencies, it is at this point unclear what information the listener uses to anticipate the site of the reparandum, as well as how mechanisms of prediction in the context of appropriateness repairs are similar or different to those operating in the context of error repairs (see Lowder & Ferreira, 2016, for further discussion).

As mentioned in the Introduction, a great deal of psycholinguistic work has argued that prediction is an inherent part of the language comprehension system. However, researchers have often struggled to demonstrate conclusively that such "prediction" effects actually arise from preactivation of linguistic information before the comprehender encounters the input, as opposed to an explanation that would attribute these effects instead to facilitated integration of a target word with its preceding context (see Van Petten & Luka, 2012, for a review of this debate). The methodological approach we employed here obviates this explanatory hurdle by measuring listeners' looks to a highly predicted entity that was never actually spoken. Thus, we believe these experiments provide an especially powerful demonstration of prediction during sentence processing, which is consistent with the notion that comprehenders preactivate an upcoming word before receiving the linguistic input. Nonetheless, one potential limitation of this design is that the sentence contexts were created to strongly bias the listener to consider NP1 and the critical distractor as the most likely items to be named (e.g., salt and pepper). Indeed, the fixation plots (see Figures 2 and 4) reveal that listeners tended to consider these two items rather than NP2 (e.g., ketchup) before the onset of the critical words. Thus, it could be argued that increased looks to the critical distractor reflect not the existence of an alternate set cued by the semantic contrast between the critical distractor and NP1, but rather a less sophisticated mechanism that just moves to the next most likely item once it becomes clear that the speaker is going to mention something else[2]. Although our data do not directly address this point, we propose that listeners make use of a variety of cues to generate their set of predictions, including the visual display and the left context of the sentence in addition to the negated NP or reparandum. Because the left contexts of these sentences were designed to lead the listener to salt and pepper, for example, it is not clear how the prediction mechanism would react if the critical words were globally unexpected but were nonetheless semantic associates of one another (e.g., *The meat was pretty bland, so the chef reached for some ketchup uh I mean mustard…*”). For now, we speculate that listeners would still use semantic information about the reparandum to predict the upcoming repair, although this process will likely be much more difficult when the earlier sentence context has led the listener to anticipate something else.

Importantly, the design of our experiments does allow us to rule out the possibility that the prediction effects we observed are simply effects of semantic priming. Although semantic priming likely played some role in driving listeners' fixations, our inclusion of the Coordination condition in both experiments acted as a control for any baseline semantic priming effects. Thus, the increased looks to the critical distractor in the Disfluency condition compared to the

[2] We thank an anonymous reviewer for raising this possibility.

Coordination condition demonstrates enhanced effects of prediction above and beyond any effects of semantic priming.

Still, it is interesting and perhaps a bit surprising that the proportion of looks to the critical distractor in the Coordination condition was quite low compared to the proportion of looks to the critical distractor in the Disfluency and Focus conditions. Although the current work does not directly address why this might be, we propose that the critical difference may stem from the fact that disfluency and focus constructions invite the listener to generate specific alternates to replace an entity in the utterance, either because the speaker signals that he has made an error ("…*uh I mean*…"), or because the speaker directly negates an entity and thus establishes a contrast set ("…*not…but rather…*"), whereas coordination constructions establish the two entities as a pair ("…*and also*…"). Another important property of the coordination structures used in our experiments is that the two conjuncts were combined using "*and also*" rather than the simple "*and*". It is possible that conjoined phrases using forms such as "*and also*", "*as well as*", etc. are less frozen than those using the simple conjunction "*and*", thus making the second conjunct less predictable. It therefore seems possible that other kinds of coordination constructions could promote some degree of prediction beyond what we observed in this study. Critically, however, listeners' tendencies to predict appear to be exceptionally robust in the case of disfluency and focus constructions.

As described in the Introduction, several experiments have shown that the reparandum is not "filtered out" or otherwise eliminated from the listener's representation of the utterance; instead, traces of the reparandum linger and continue to influence the listener's ultimate interpretation of the sentence. Whereas these lingering effects of the reparandum can be explained by the Overlay model of disfluency processing (Ferreira & Bailey, 2004; Ferreira et al., 2004; Lau & Ferreira, 2005), we propose that effects of prediction in the processing of both repair disfluencies as well as contrastive focus can be explained within a Noisy Channel framework. Recall that Noisy Channel models of language comprehension assume a “noisy” or imperfect linguistic environment in which the language comprehender is constantly reconciling the input with linguistic and contextual knowledge, actively correcting any perceived errors. The current work demonstrates that listeners are very sensitive to an explicit error signal—be it the negation cue “*not*” or the editing phrase “*uh I mean*”—and listeners then use semantic information about the negated entity or reparandum to predict the upcoming input. However, an important component of the Noisy Channel model is that the listener is constantly weighing the linguistic input against assumptions about the speaker’s meaning. Thus, this account predicts that listeners should be able to anticipate that a speaker has made an error before the speaker becomes disfluent, such as in cases where the speaker says something implausible or otherwise inconsistent with the listener’s knowledge about language, the speaker, or the context. Consider the example of sitting around the dinner table and being asked to “*Please pass the sugar*.” Given the context of eating dinner, which likely includes a table filled with savory foods, we propose that the listener would mark “*sugar*” as a probable reparandum, mentally correct this error, and possibly even begin to reach for the salt, all before the speaker becomes disfluent. Importantly, we propose that mental corrections of this sort, even in the absence of an explicit error signal, arise from the generation of sets of contrastive alternates in a way that is similar to the alternate sets that are generated during the processing of self-corrected repair disfluencies and contrastive focus structures of the sort examined here. Ongoing work in our lab is aimed at uncovering direct evidence for such “mental corrections” of speech errors, even in the absence of an overt correction from the speaker. Evidence of this sort will help determine the extent to which the predictions listeners generate during the processing of speech errors are adequately explained by current Noisy Channel models, or whether a new theoretical model is warranted. Given the data currently available (which we acknowledge is modest given the current state of our understanding of how disfluencies are processed), we have opted for a conservative approach in which we take advantage of an existing model that captures our findings rather than postulating a new model specifically to account for our results.

Finally, the work presented here contributes to the theoretical goal of understanding how language comprehenders process and understand speech errors in general and repair disfluencies in particular. Previous work examining comprehenders' offline metalinguistic judgments has established that the reparandum is not eliminated from the representation of the sentence, but rather continues to influence interpretation—an effect that may be attributable to the inherent limitations of the memory system. In contrast, the current work has demonstrated using online eye tracking methodology that comprehenders use available linguistic and visual information to actively anticipate an upcoming repair. Such a robust effect of prediction offers many exciting areas of future research, including documenting the range of linguistic and contextual information that listeners use to generate predictions during disfluency processing, establishing the time course of these effects, and gaining a more complete understanding of how the prediction mechanisms involved in the processing of speech disfluencies are similar or different to other linguistic constructions. We believe that uncovering this knowledge will be instrumental toward developing a more complete understanding of speech disfluencies, linguistic prediction, and language comprehension in general.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

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

Arnold, J. E. (2008). *THE BACON* not *the bacon*: How children and adults understand accented and unaccented noun phrases. *Cognition, 108*, 69-99.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*, 25-36.

Arnold, J. E., Hudson Kam, C. L., & 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,* 914-930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science, 15,* 578–581.

Baayen, R. H., Davidson, D. H., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory & Language*, *49*, 183-200.

Bailey, K. G. D., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 485-500). Oxford, UK: Elsevier.

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*, 255-278.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes, 25*, 441-455.

Bates, D., Maechler, M., & Dai, B. (2012). lme4: Linear mixed-effects models using S4 classes. Available at http://lme4.r-forge.r-project/org.

Benatar, A., & Clifton, C., Jr. (2014). Newness, givenness and discourse updating: Evidence from eye movements. *Journal of Memory and Language, 71*, 1-16.

Birch, S., & Clifton, C., Jr. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38*, 365-391.

Birch, S., & Clifton, C., Jr. (2002). Effects of varying focus and accenting adjuncts on the comprehension of utterances. *Journal of Memory and Language, 47,* 571-588.

Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *25*, 653-660.

Birch, S., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition, 38*, 740-752.

Bortfeld, H., Leon, S., Bloom, J., Schober, M., & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123-147.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition, 136*, 135-149.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368-407.

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*, 696-702.

Corley, M. (2010). Making predictions from speech with repairs: Evidence from eye movements. *Language and Cognitive Processes, 25*, 706-727.

Corley, M., & Hartsuiker, R. J. (2011). Why *um* helps auditory word recognition: The temporal delay hypothesis. *PLoS One, 6*, e19792.

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*, 658-668.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7*, 49-59.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20*, 1-10.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*, 292-314.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117-1121.

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science, 21*, 629-634.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*, 12203-12208.

Ferreira, F., & Bailey, K. G. D. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Sciences, 8*, 231-237.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11-15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*, 3-20.

Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science, 28*, 721-749.

Ferreira, F., & Lowder, M. W. (in press). Prediction, information structure, and good enough language processing. *Psychology of Learning and Motivation.*

Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass, 1*, 71-83.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709-738.

Gibson, E., Bergen, L. & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences, 110,* 8051-8056.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science, 24*, 1079-1088.

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar.* Cambridge, MA: MIT Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446.

Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The tine-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133-156.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science, 14*, 925-934.

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience, 24*, 1104-1112.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*, 481-529.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57*, 519-543.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*, 32-59.

Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes, 20*, 633-666.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41-104.

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies. *Language, Cognition and Neuroscience, 31*, 19-31.

Lowder, M. W., & Gordon, P. C. (2015). Focus takes time: Structural effects on reading. *Psychonomic Bulletin & Review, 22*, 1733-1738.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 913-934.

Maxfield, N. D., Lyon, J. M., & Silliman, E. R. (2009). Disfluencies along the garden path: Brain electrophysiological evidence of disrupted sentence processing. *Brain & Language, 111*, 86-100.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs and nouns. *Memory & Cognition, 33*, 1174-1184.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language, 66*, 545-567.

Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition, 26*, 1313-1322.

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language, 74*, 36-45.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language, 67*, 426-448.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 514-528.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, 75-116.

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review, 11*, 882-888.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*, 176-190.

Appendix

The experimental items from Experiments 1 and 2 are presented below as they were spoken in the Disfluency condition. Each item was also presented in the various other conditions described in the text. The visual display included pictures corresponding to NP1 and NP2 (presented within each sentence), as well as pictures corresponding to a critical distractor and random distractor (presented in parentheses following each item).

1. The woman next door went to the animal shelter and brought home a dog uh I mean a rabbit, even though her apartment doesn't allow pets. (critical distractor: cat; random distractor: plant)
2. Susan wanted to make some toast, but she realized she was all out of bread uh I mean honey, so she went to the store. (critical distractor: butter; random distractor: carrots)
3. Frank had parked illegally, and he discovered that someone had towed his car uh I mean his bike, and he was very angry. (critical distractor: truck; random distractor: tree)
4. The wedding is just about to start, but no one can find the bride uh I mean the priest, and so all the guests are waiting around. (critical distractor: groom; random distractor: nun)
5. When Matt cooked breakfast this morning he overcooked the bacon uh I mean the potatoes, which was a bit disappointing. (critical distractor: eggs; random distractor: towel)
6. When Mark went on vacation to the jungle he saw a ferocious lion uh I mean a monkey, and he was a bit scared. (critical distractor: tiger; random distractor: jeep)
7. Mary loves fruit, so every day she eats at least one apple uh I mean one kiwi, and she's very healthy. (critical distractor: orange; random distractor: teapot)
8. After the rock concert, the boys decided they wanted to play the drums uh I mean the keyboard, so we signed them up for lessons. (critical distractor: guitar; random distractor: speakers)
9. The writer suddenly had some inspiration, so he grabbed his pen uh I mean his folder and got to work. (critical distractor: pencil; random distractor: plate)
10. Before Nancy can begin cooking dinner, she needs a large pot uh I mean a bowl, and then she'll have everything she needs. (critical distractor: pan; random distractor: desk)
11. Jill poured herself a cup of coffee, but she was unable to find the creamer uh I mean the napkins, so she went back to her office. (critical distractor: sugar; random distractor: briefcase)
12. The meat was pretty bland, so the chef reached for some salt uh I mean some ketchup, which made it much more flavorful. (critical distractor: pepper; random distractor: milk)
13. As Sarah was getting dressed, she noticed that there was a hole in her shoe uh I mean her skirt, which was very upsetting. (critical distractor: sock; random distractor: soap)
14. Eric is trying to fix the sink, and right now he really needs a wrench uh I mean a hammer before he can finish. (critical distractor: plunger; random distractor: barbell)
15. Emma needed something to sit on in her new apartment so she bought a sofa uh I mean a stool, and it looks very nice. (critical distractor: chair; random distractor: frame)

16. Thomas saw some animals in the pasture and he thinks they were cows uh I mean deer, but it was hard to tell. (critical distractor: sheep; random distractor: rocks)
17. Tim had a plate in front of him but he was still missing a fork uh I mean a glass, so he couldn’t enjoy his dinner. (critical distractor: knife; random distractor: book)
18. Entering the royal chamber, the servant hurried toward the king uh I mean the jester with an urgent message. (critical distractor: queen; random distractor: fountain)
19. Cathy went to the jewelry store and brought home a necklace uh I mean a keychain, and she has plenty of money left over. (critical distractor: bracelet; random distractor: mirror)
20. For dessert, everyone at the table had a piece of cake uh I mean cheese, which was a nice end to the meal. (critical distractor: pie; random distractor: ice)
21. The drink was a bit sour because Rachael had added some lemons uh I mean some cherries, and there wasn’t enough sugar. (critical distractor: limes; random distractor: tongs)
22. Every Sunday, the old man sits on the park bench and feeds the ducks uh I mean the fish, and then he goes home. (critical distractor: geese; random distractor: streetlamp)
23. For Christmas, Billy got a coloring book and some crayons uh I mean some puzzles, which he absolutely loves. (critical distractor: markers; random distractor: windows)
24. Michael likes his cheese steak sandwich to be topped with grilled peppers uh I mean tomatoes, and he has one every week. (critical distractor: onions; random distractor: pigeons)
25. During the blackout, Robert searched for a flashlight uh I mean his glasses because he couldn’t see anything. (critical distractor: candle; random distractor: pushpin)
26. After Charlie finished mowing the lawn, he opened up a cold beer uh I mean a water, and then he watched TV. (critical distractor: soda; random distractor: table)
27. Whenever Kelly goes to the movies, she always gets some popcorn uh I mean some coffee, and she sits in the back. (critical distractor: candy; random distractor: sun)
28. Carla went down to the nail salon to treat herself to a manicure uh I mean a massage, and she had a great afternoon. (critical distractor: pedicure; random distractor: workout)
29. Last summer, Betty got a sunburn all over her arms uh I mean her nose, and it was just awful. (critical distractor: legs; random distractor: seashell)
30. The thief spotted a victim and quickly snatched the wallet uh I mean the phone, and he was never caught. (critical distractor: purse; random distractor: jar)
31. The secretary needed to keep her papers together, so she reached for a stapler uh I mean a rubber band, and then she filed them away. (critical distractor: paperclip; random distractor: computer)
32. The bake sale offered a wide selection of cookies uh I mean doughnuts, and the money that was raised went to the school. (critical distractor: cupcakes; random distractor: jacket)
33. The vaccine was administered to the child by the doctor uh I mean the surgeon, and the child didn’t cry at all. (critical distractor: nurse; random distractor: clipboard)
34. While in the waiting room, Vince read a book uh I mean a map, and finally his name was called. (critical distractor: magazine; random distractor: umbrella)

35. Oscar’s cat always liked to chase the mice uh I mean the ants while everyone else was asleep. (critical distractor: rats; random distractor: keys)
36. Max’s dog always loved to jump up and catch the ball uh I mean the bone every Saturday morning. (critical distractor: dog toy; random distractor: watch)
37. After the meal, Jenn left a big tip on the table for the waiter uh I mean the busboy, and then she went home. (critical distractor: waitress; random distractor: fireman)
38. Before making the salad, Laura had to wash the lettuce uh I mean the grapes, which took longer than she thought. (critical distractor: tomatoes; random distractor: blendor)
39. Steve likes to cover his hot dog with ketchup uh I mean mayonnaise before he digs in. (critical distractor: mustard; random distractor: lemonade)
40. Bobby asked for some maple syrup to put on his pancakes uh I mean his oatmeal, and then he finished his breakfast. (critical distractor: waffles; random distractor: toaster)

Author Note

This research was supported by R56 DC013545 from NIDCD. We thank Tess Dawkins, Elena Galkina, and John Lim for assistance in conducting the experiments, and Jimmy Nye for assistance in creating and pretesting the stimuli.

.

Table 1

Experiment 1. Results of mixed effects analyses.

	NP1				NP2				Critical Distractor			
	Estimate	*SE*	*z*	*p*	Estimate	*SE*	*z*	*p*	Estimate	*SE*	*z*	*p*
Window 1												
Intercept	2.127	0.417	5.096	< .001	-1.267	0.203	-6.243	< .001	-0.743	0.221	-3.365	< .001
Disfluency vs. Coordination	-0.261	0.415	-0.630	*n.s.*	-0.281	0.275	-1.023	*n.s.*	0.178	0.209	0.852	*n.s.*
Random effects structure	(0+Condition|Subject) + (1|Item)				(1+Condition|Subject) + (1+Condition|Item)				(1+Condition|Subject) + (1+Condition|Item)			
Window 2												
Intercept	1.373	0.342	4.020	< .001	-0.932	0.190	-4.871	< .001	0.286	0.259	1.103	*n.s.*
Disfluency vs. Coordination	0.321	0.279	1.149	*n.s.*	-0.657	0.292	-2.247	< .05	-1.331	0.245	-5.429	< .001
Random effects structure	(1+Condition|Subject) + (1+Condition|Item)				(1+Condition|Subject) + (0+Condition|Item)				(1+Condition|Subject) + (0+Condition|Item)			
Window 3												
Intercept	-1.251	0.226	-5.541	< .001	2.655	0.511	5.201	< .001	-0.327	0.161	-2.035	< .05
Disfluency vs. Coordination	0.973	0.220	4.433	< .001	-0.796	0.438	-1.819	< .10	-0.636	0.212	-3.006	< .005
Random effects structure	(1+Condition|Subject) + (1+Condition|Item)				(1+Condition|Subject) + (1+Condition|Item)				(1+Condition|Subject) + (1+Condition|Item)			

Table 2

Experiment 2. Results of mixed effects analyses.

	NP1				NP2				Critical Distractor			
	Estimate	*SE*	*z*	*p*	Estimate	*SE*	*z*	*p*	Estimate	*SE*	*z*	*p*
Window 1												
Intercept	1.794	0.236	7.595	< .001	-1.259	0.186	-6.785	< .001	-0.522	0.151	-3.466	< .001
Disfluency vs. Coordination	0.085	0.238	0.358	*n.s.*	0.328	0.198	1.659	*n.s.*	0.117	0.214	0.547	*n.s.*
Disfluency vs. Focus	-1.111	0.217	-5.130	< .001	0.553	0.195	2.832	< .005	0.936	0.188	4.987	< .001
Random effects structure	(1|Subject) + (1|Item)				(1|Subject) + (1|Item)				(0+Condition|Subject) + (1|Item)			
Window 2												
Intercept	1.517	0.206	7.380	< .001	-0.826	0.199	-4.159	< .001	0.143	0.219	0.653	*n.s.*
Disfluency vs. Coordination	-0.123	0.216	-0.567	*n.s.*	0.023	0.191	0.123	*n.s.*	-0.734	0.237	-3.094	< .005
Disfluency vs. Focus	-2.046	0.210	-9.769	< .001	0.772	0.187	4.130	< .001	0.573	0.247	2.319	< .05
Random effects structure	(1|Subject) + (1|Item)				(1|Subject) + (1|Item)				(0+Condition|Subject) + (1|Item)			
Window 3												
Intercept	-0.931	0.186	-5.013	< .001	2.428	0.307	7.901	< .001	-0.437	0.163	-2.684	< .01
Disfluency vs. Coordination	0.666	0.186	3.582	< .001	-0.214	0.263	-0.814	*n.s.*	-0.524	0.196	-2.678	< .01
Disfluency vs. Focus	-1.000	0.222	-4.516	< .001	0.300	0.286	1.049	*n.s.*	0.152	0.185	0.819	*n.s.*
Random effects structure	(1|Subject) + (1|Item)				(1|Subject) + (1|Item)				(0+Condition|Subject) + (1|Item)			

Figure 1. Example visual display for Experiments 1 and 2.

Figure 2. Results of Experiment 1. Proportion of fixations to each picture type across the four conditions.

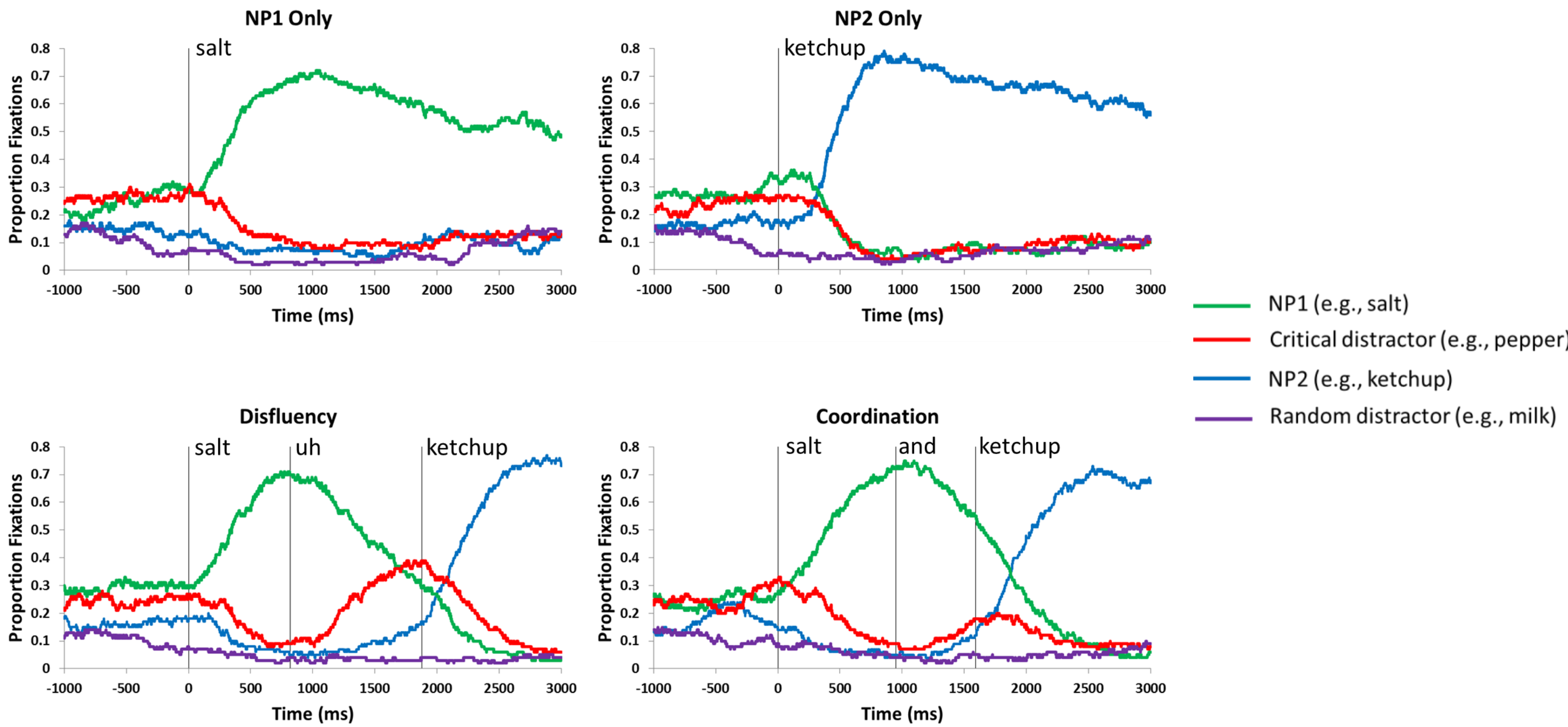

Figure 3. Predictable filler items from Experiment 1. Proportion of fixations to each picture type.

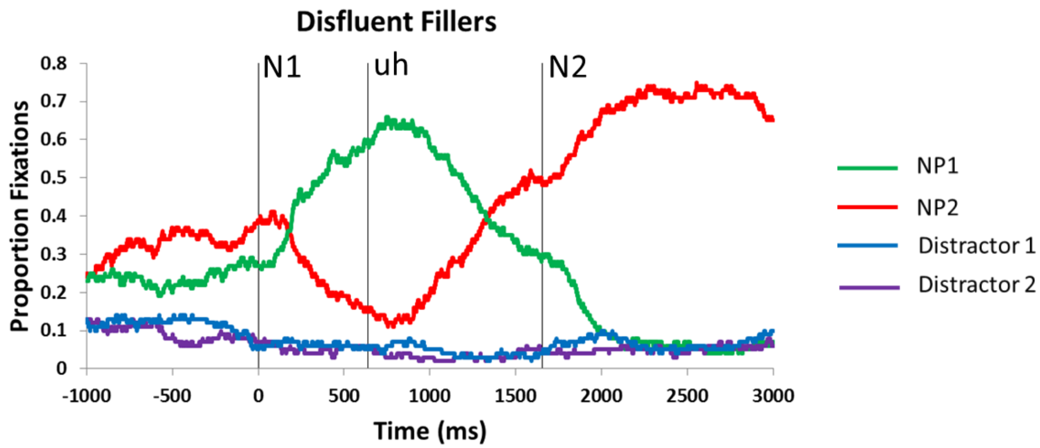

Figure 4. Results of Experiment 2. Proportion of fixations to each picture type across the four conditions.

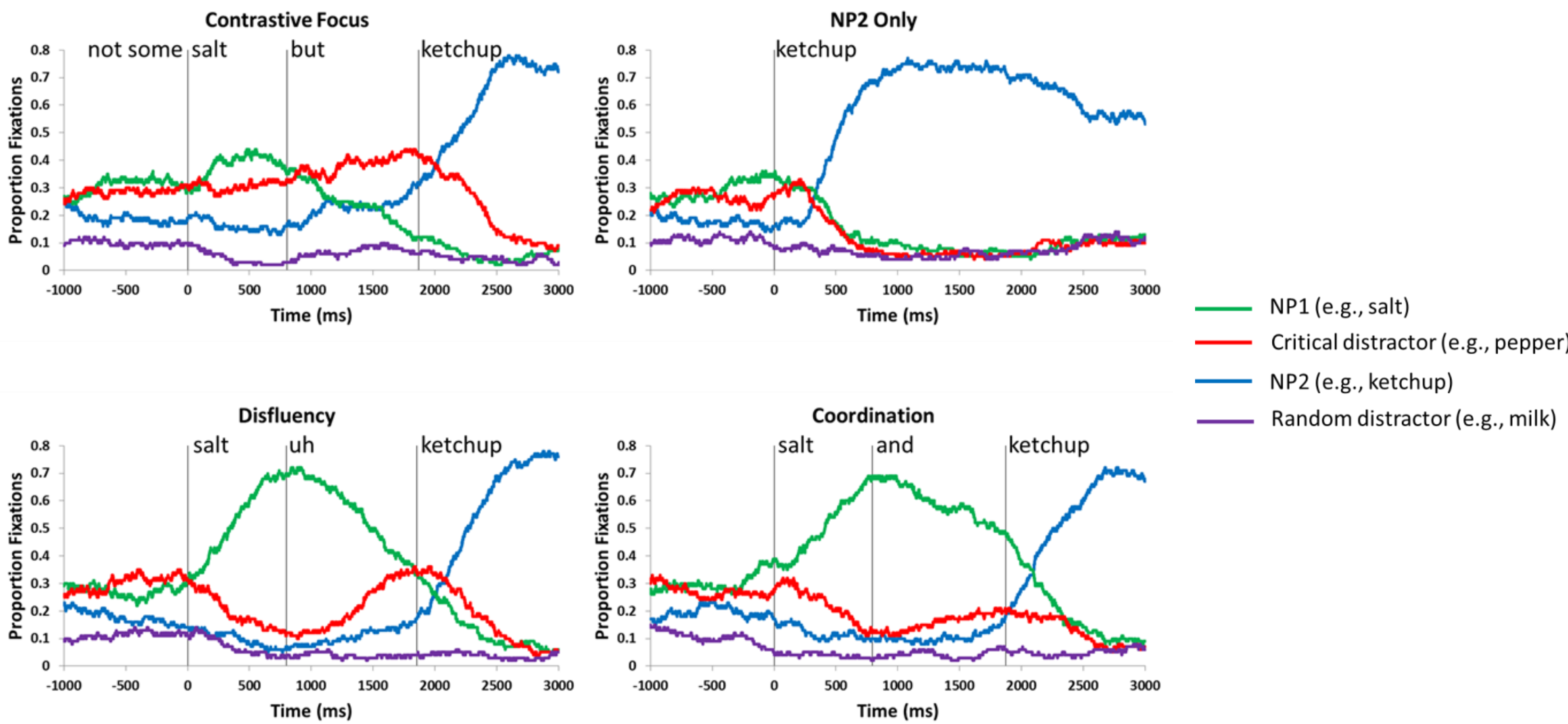

Figure 5. Predictable filler items from Experiment 2. Proportion of fixations to each picture type across the three conditions.