Integration of Pragmatic and Phonetic Cues in Spoken Word Recognition

Hannah Rohde

University of Edinburgh

Marc Ettlinger

VA Northern California Health Care System

Author note

Hannah Rohde, Department of Linguistics and English Language, University of Edinburgh; Marc Ettlinger, Human Cognitive Neurophysiology Lab, Veteran's Administration Research Service, VA Northern California Health Care System.

This research was supported by an Andrew W. Mellon postdoctoral fellowship to Hannah Rohde and by NIH grant T32 NS047987 to Marc Ettlinger. We thank Ann Bradlow and Matt Goldrick for helpful discussion and research assistant Ronen Bay for his help during data collection. We also thank Randi Martin, Bob McMurray, and three anonymous reviewers for many helpful suggestions that led to substantial improvements over earlier drafts. Some portions of this work were first presented in Rohde and Ettlinger (2010).

Correspondence concerning this article should be addressed to Hannah Rohde, Department of Linguistics and English Language, University of Edinburgh, 3 Charles Street, Edinburgh, EH8 9AD, United Kingdom. Email: hannah.rohde@ed.ac.uk

Abstract

Although previous research has established that multiple top-down factors guide the identification of words during speech processing, the ultimate range of information sources that listeners integrate from different levels of linguistic structure is still unknown. In a set of experiments, we investigate whether comprehenders can integrate information from the two most disparate domains: pragmatic inference and phonetic perception. Using contexts that trigger pragmatic expectations regarding upcoming coreference (expectations for either *he* or *she*), we test listeners' identification of phonetic category boundaries (using acoustically ambiguous words on the /hi/~/ʃi/ continuum). The results indicate that, in addition to phonetic cues, word recognition also reflects pragmatic inference. These findings are consistent with evidence for top-down contextual effects from lexical, syntactic, and semantic cues, but they extend this previous work by testing cues at the pragmatic level and by eliminating a statistical-frequency confound that might otherwise explain the previously reported results. We conclude by exploring the time-course of this interaction and discussing how different models of cue integration could be adapted to account for our results.

Keywords: Phonetics, pragmatics, word recognition, coreference, implicit causality

Integration of Pragmatic and Phonetic Cues in Spoken Word Recognition

There is a large body of evidence suggesting that language processing requires the integration of multiple sources of linguistic knowledge across multiple levels of linguistic structure. These relevant knowledge sources range from low-level properties of the acoustic signal, through lexical and morpho-syntactic properties of words and phrases, up to higher-level semantic and pragmatic inferences about the speaker's intended message. Occupying the far ends of this spectrum are phonetics and pragmatics. Therefore, identifying contexts in which comprehenders bring together cues from these two disparate domains would provide a demonstration of the maximum extent of linguistic cue integration.

In the experiments presented here, we test for integrative effects at the pragmatic-phonetic interface using contexts in which listeners' comprehension of acoustically ambiguous words is posited to reflect pragmatic biases in the discourse context. To do this, we use words whose interpretation is inherently discourse dependent—namely, personal pronouns. Based on existing work on pronoun interpretation, we use contexts in which listeners have been shown to anticipate subsequent mention of a particular referent. We capitalize on the fact that the English third person pronouns *he* and *she* constitute phonological minimal pairs and construct acoustically ambiguous pronouns that vary along a *h~sh* continuum. We test listeners' interpretation of acoustically ambiguous pronouns in pragmatically biasing sentences. We also investigate the time course of the pragmatic~phonetic integration in order to test whether these effects depend on the availability of whole words or whether sublexical material triggers similar effects. The results provide a demonstration of the maximum extent of linguistic-cue integration that any successful language processing model must capture. As such, this paper contributes to

the well-established literature on word recognition by broadening the set of known top-down factors known to influence processing and by informing the types of processing models that can capture such effects.

Background

Existing work on the factors that influence word recognition has identified effects from lexical status, syntactic category, and semantic congruity. As we will discuss below, these results have been analyzed both in terms of models of interaction and models of post-perceptual processing, with a long tradition of work attempting to distinguish between these two types of models.

The effects of the lexicon on speech perception have been demonstrated in experiments that show that an ambiguous sound is interpreted differently depending on whether or not its interpretation yields a valid lexical item. For example, Ganong (1980) found that ambiguous sounds along a /t/~/d/ continuum were more likely to be reported as /t/ in contexts in which *t* supports a valid lexical item (e.g. *task* vs. **dask*) and as /d/ in contexts in which *d* supports a valid lexical item (e.g. **tash vs. dash*) (see also Connine & Clifton, 1987; Fox, 1984; McQueen, 1991; Pitt, 1995). Similarly, when a phoneme has been replaced with noise, listeners are more likely to report hearing the missing phoneme in words than non-words (Pitt & Samuel, 1995; Samuel, 1981, 1996, 1997, 2001; Warren, 1970; Warren & Warren, 1970).

When both interpretations of an ambiguous sound yield valid lexical items, listeners also show sensitivity to syntactic and semantic context. The identification of acoustically ambiguous words along a *to~the* continuum has been shown to reflect context-driven part-of-speech constraints: Listeners are more likely to report hearing *to* in contexts with a proceeding verb, as in *We tried ... go*, than in contexts with a proceeding noun, as in *We tried ... gold* (Isenberg,

Walker, & Ryder, 1980; see also van Alphen & McQueen, 2001). The meaning evoked by the sentence also plays a role: Ambiguous words along a *path~bath* continuum are more likely to be reported as /p/ in the context *She likes to jog along the…*, whereas they are more likely to be reported as /b/ in the context *She needs hot water for the…* (Miller, Green, & Schermer, 1984; see also Connine, 1987 and Borsky, Tuller, & Shapiro, 1998). Miller et al. report, however, that semantic congruity effects disappear when the task requires listeners to focus only on the target word, rather than on the full sentence frame. These studies point to the dynamic integration of information sources that range from a sentence's hierarchical syntactic structure (*go* and *gold* constrain the phrase type in which they appear) to real-world event knowledge (contexts that mention *jogging* evoke situations with *paths*, whereas contexts that mention *hot water* evoke *baths*).

However, one possible criticism of this earlier work on syntactic and semantic context is that simple co-occurrence frequencies are themselves sufficient to explain the observed effects. That is, the results may not reflect listeners' deeper understanding or parsing of the sentence and its meaning, but rather reflect statistical frequencies over adjacent words (see Willits, Sussman, & Amato, 2008). In Isenberg et al.'s study, "to go" may simply be a more frequent word pair than "to gold"; in Miller et al.'s study, the probability of seeing the word "bath" within a small window of "water" may be higher than that of seeing the word "path". This co-occurrence-based explanation has been proposed as an alternative explanation for a set of semantic priming results which have typically been attributed to deeper semantic processing and event representations (Ferretti, Kutas, & McRae, 2007; McRae, Spivey-Knowlton, & Tanenhaus, 1998): For example, the word "cooking" may prime "kitchen" either because of listeners' mental models of typical events or because of listeners' knowledge of statistically frequent collocations in language. An

explanation based on collocational frequency as opposed to higher-level event knowledge echoes Pitt & McQueen's (1998) argument that lexical effects may be attributable to transition probabilities between phonemes as opposed to lexical feedback. For both semantic congruity effects and lexical effects, the concern is that what looks to be evidence of higher-level feedback may arise either from higher-level representations or from local statistics. Either way, because the semantic congruity effects are driven by the presence of particular words in the preceding context that make a subsequent word more predictable, we will refer to these results as *collocational effects*. In the work presented here, we replicate the collocational effect and then test for the integration of higher-level linguistic cues in contexts in which co-occurrence frequencies are insufficient to explain listener bias.

Top-down contextual effects like the ones described above have been incorporated into models of word recognition in two different ways (see reviews in McClelland, Mirman, & Holt, 2006; McQueen, Norris, & Cutler, 2006; Mirman, McClelland, & Holt, 2006). Some models allow for top-down contextual information to impact sound perception directly (e.g., TRACE, McClelland & Elman, 1986) or through emergent activation based on bottom-up information and top-down support (e.g., ART, Grossberg & Myers, 2000), whereas others argue for an encapsulated perceptual system that operates independently of other levels of language processing, such that top-down factors exert an influence post-perceptually (e.g., Merge, Norris, McQueen, & Cutler, 2000). This interaction/post-perceptual distinction has also been conceptualized as a difference between ‘interactive’ and ‘modular’ theories of cognition (Bowers & Davis, 2004; see also Baese-Berk & Goldrick, 2009; Magnuson, McMurray, Tanenhaus, & Aslin, 2003; Samuel & Pitt, 2003).

Our primary goal in this work is to extend the observed range of top-down effects beyond the previously reported lexical, syntactic, and semantic levels; evidence for pragmatic~phonetic integration would be compatible with both interaction-based and post-perceptual models. In our last experiment, we ask whether sub-lexical acoustic cues are sufficient to trigger top-down pragmatic effects and consider how different models can account for such effects.

Pragmatic Manipulation

Although a variety of linguistic and extralinguistic cues are often studied under the label of "pragmatics", for the purposes of this paper, we intend pragmatics to refer to the linguistic notion of "what is meant beyond what is said" (Bach, 1994; Grice, 1975), i.e., the information that must be inferred in order for a sentence to stand in a coherent relationship with the linguistic material in the surrounding discourse context. This excludes extralinguistic cues (e.g., physical properties of the speaker; Kraljic, Brennan, & Samuel 2008). In the present study, we focus on the pragmatics of coreference, a phenomenon that underlies listeners' ability to track who and what is being talked about across clauses in a discourse and to thereby infer how a series of utterances come together to convey a meaningful and coherent message (Levinson, 1987). In the remainder of this section we review a set of coreference results that motivate the logic of our experimental design.

In order to test whether pragmatic biases yield top-down effects, we manipulate a property of the discourse context that is known to guide listeners' expectations about upcoming patterns of coreference: the presence of implicit causality (IC) verbs (e.g., *annoy, hate, admire, impress,* etc.). This well-studied class of verbs has been shown to influence listeners' expectations about who will be mentioned next as the discourse proceeds based on listeners' real-world knowledge about events and the inferences they make about the typical causes of

events (Arnold, 2001; Au, 1986; Caramazza, Grober, Garvey, & Yates, 1977; Garvey & Caramazza, 1974; Koornneef & van Berkum, 2006; McDonald & MacWhinney, 1995; McKoon, Greene, & Ratcliff, 1993; Kehler, Kertz, Rohde, & Elman, 2008; Pyykkönen & Järvikivi, 2010; Stevenson, Crawley, & Kleinman, 1994; Stevenson, Knott, Oberlander, & McDonald, 2000; Stewart, Pickering, & Sanford, 2000). These verbs appear to guide listeners' coreference biases because they describe events in which one participant is implicated as central to the event's cause and is thus likely to be re-mentioned in a subsequent clause explaining the event in question.

Early coreference experiments involving IC verbs asked listeners to interpret ambiguous pronouns in sentence-completion tasks (Garvey & Caramazza, 1974; Caramazza, Grober, Garvey, & Yates, 1977). The results showed that IC verbs are characterized by strongly divergent coreference biases in causal contexts: For example, certain verbs like *annoy* in a context like *John annoys Tom because he...* strongly favor a subject interpretation of the ambiguous pronoun (*...because* he_{JOHN} *always tries to outdo Tom*), whereas verbs like *hate* in a context like *John hates Tom because he…* favor an object interpretation (*...because* he_{TOM} *once humiliated John in public*). With subject-biased IC verbs, the cause is typically attributed to the individual mentioned in subject position (e.g., *annoy, impress, amaze, bore, disappoint*), whereas, with object-biased verbs, the cause is attributed to the individual mentioned in object position (e.g., *hate, scold, congratulate, admire, fear*). The results from off-line sentence-completion studies have been confirmed in on-line reading studies using contexts with two opposite-gendered referents; these studies reveal comprehension difficulty when the pronoun does not match the gender of the causally implicated referent (e.g., *John annoys Mary because* <u>*she*</u> *...*; Koornneef & van Berkum, 2006). Furthermore, recent visual-world eye-tracking

experiments show that comprehenders make anticipatory looks to the causally implicated referent even before listeners encounter the causal connective or the pronoun, suggesting that IC biases are expectation-driven (Pyykkönen & Järvikivi, 2010).

For our target manipulation, we use discourse contexts with two opposite-gender names in the first clause and an acoustically ambiguous *he/she* pronoun in a subsequent *because* clause. If listeners rely entirely on bottom-up acoustic cues, their responses are predicted to only reflect acoustic properties of the particular pronoun stimulus along the *he/she* acoustic continuum. Alternatively, if listeners rely only on top-down pragmatic cues, their responses are predicted to vary only with the verb type, not with the acoustic cues. However, if listeners are combining bottom-up and top-down cues, the presence of an IC verb is predicted to overlay an additional bias alongside the acoustic cues in favor of *he* in contexts with a causally implicated male (1a-b) and in favor of *she* in contexts with a causally implicated female (2a-b).

(1) *he*-biasing contexts

a. subject-biased verb

Tyler deceived Sue because ☐ couldn't handle a conversation about adultery.

b. object-biased verb

Joyce helped Steve because ☐ was working on the same project.

(2) *she*-biasing contexts

a. subject-biased verb

Abigail annoyed Bruce because ☐ was in a bad mood.

b. object-biased verb

Luis reproached Heidi because ☐ was getting grouchy.

Our predictions are based on the following reasoning: The verbs in (1-2) are presumed to drive listeners' pragmatic inferences about which referent is causally implicated; this causal inferencing determines which referent listeners expect to hear mentioned first in a subsequent *because* clause; these coreference expectations in turn yield a lexical expectation for either *he* or *she*, which serves to influence the interpretation of the acoustically ambiguous pronoun. The target passages were constructed with the intention of permitting either interpretation of the acoustically ambiguous pronoun (marked with □ in (1-2)). This is essential if we are to see evidence of bottom-up cues—both a *he* and *she* interpretation should be pragmatically plausible in order to avoid listeners categorically ignoring bottom-up acoustic cues. We therefore capitalize on the fact that IC biases are probabilistic, and not categorical; a reference to either individual is technically permitted, even if one is strongly dispreferred (the Methods section of Experiment 3 presents plausibility ratings of the *he* and *she* versions of such passages).

The presence of both male and female names in the preceding context also avoids the statistical co-occurrence confound at issue in the earlier semantic congruity studies: Unlike the collocational predictability of the word *path* in a context that contains the word *jog*, neither *he* nor *she* is inherently more predictable in a context that mentions both a female and a male, at least not without appeal to deeper event-level knowledge and causal reasoning. By using both subject-biased and object-biased verbs which vary the position of the preferred referent, we also control for any recency effect: A *he* interpretation is predicted to be preferred both in contexts where the male is the more recent referent (1b) and in ones where the male is the more distant referent (1a); likewise for *she*. The only difference between (1) and (2) is the verb-driven causal inference that implicates a male referent or a female referent.[1]

In order to first establish the validity of our *h~sh* continuum, Experiment 1 replicates the Ganong effect for words and non-words on that continuum. Experiments 2 replicates the collocational effect by eliciting *he/she* judgments in contexts with only male or only female referents as a measure of listeners' sensitivity to the availability of particular gendered referents in the sentential context. This also serves to validate the use of pronouns as potential targets of top-down and bottom-up biases. Experiment 3 tests our primary pragmatic manipulation using contexts with opposite-gender referents in causally biasing contexts. Experiment 4 uses a gating task to establish the import of full words on phoneme decisions.

Experiment 1: Lexical Effects

It is first necessary to establish that a /h/~/ʃ/ continuum is a valid one for assessing bottom-up and top-down influences in phoneme identification. Unlike the oft-used *t~d* or *s~sh* continua, for which single continuous variables can be manipulated to create a continuum (voice onset time and spectral mean, respectively), no simple acoustic feature distinguishes *sh~h*. Therefore, we verified the effect of lexical status for the continuum generated using the procedure described below. We tested whether listeners would judge the ambiguous onset of a monosyllabic item (e.g., /□ik/) as more /ʃ/-like if the English lexicon contains a word with a /ʃ/- onset and lacks a corresponding word with a /h/- onset (e.g. *sheik/*heik*) and as more /h/-like in the reverse condition (e.g. *heave/*sheeve*). If the /h/~/ʃ/ continuum follows other acoustic continua that have been tested (Connine & Clifton, 1987; Fox, 1984; Ganong 1980; McQueen, 1991; Pitt, 1995), we predict that listeners will be sensitive both to the step along the continuum (bottom-up cues) and to lexical status (a top-down cue).

Method

Participants. 35 native English-speaking Northwestern University undergraduates received either $6 or course credit for their participation in the study.

Materials. Six pairs of items consisted of a word and a non-word. The pairs *sheik/*heik*, *sheen/*heen*, and *sheaf/*heaf* were /ʃ/-biasing in that the /ʃ/- onset constitutes a valid word. The pairs *heeds/*sheeds*, *heels/*sheels*, and *heave/*sheave* were /h/-biasing in that the /h/- onset constitutes a valid word. Onsets ranged from /h/ to /ʃ/ along a 20-step acoustic continuum. A token of a male speaker saying each word and non-word was recorded using Praat (Boersma & Weenink, 2001). To construct the steps we combined two naturally produced tokens of *he* and *she* at varying intensities (Cristia, McGuire, Seidl, & Francis, *in press*; Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McGuire, 2007; Pitt & McQueen, 1998). We opted for this method of stimulus construction (versus synthetic generation of stimulus) because it does not artificially minimize stimulus complexity and, more importantly, does not require the insertion of an artificially generated stimulus into the context of a naturally spoken sentence, which would otherwise have been the case for Experiments 2-4. This method allows us to maintain all relevant speaker/voice characteristics throughout the sentences, an important factor for speech recognition (Goldinger, 1996). The method is further supported by studies showing a center-of-gravity effect in perceiving stimulus generated in this manner, whereby listeners perceive a weighted average of merged stimuli that are similar sounding (e.g. Chistovich & Lublinskaya, 1979; Delattre, Liberman, Cooper, & Gerstman, 1952; Xu, Jacewicz, Feth, & Krishnamurthy, 2004). In addition, Cristia et al. (in press) showed high naturalness ratings of fricatives generated in this way, as compared to manually manipulated natural stimuli. Since the duration of the fricative portion may also serve as a cue to differentiate these items, the duration was the average of the /hi/ and /ʃi/ tokens. Items were constructed such that each of the 6 pairs appeared

with each of the 20 steps along the /hi/~/ʃi/ continuum. The waveforms in Figure 1 show sample steps.

[Figure 1]

Procedure. Participants listened to the items through headphones while sitting in a sound-attenuated booth. For each item, they were asked to indicate using a button box whether the onset of the item sounded more *h*-like or more *sh*-like on a 4-point scale (1="definitely h", 2= "probably h", 3="probably sh", 4="definitely sh"). Participants heard all items twice. A subset of participants also completed Experiments 2 and 3. In experiment sessions that included multiple tasks, this lexical status experiment was always completed as the last part of the session.

Results and Discussion

We submitted the data to two-way lexical status × acoustic step ANOVAs, by subjects (F_1) and by items (F_2). As predicted, there was a main effect of lexical status ($F_1(1,34)=198.15$, $p<0.001$; $F_2(1,4)=11.56$, $p<0.03$), whereby listeners reported hearing more initial /ʃ/ sounds for /ʃi/-biasing items (items on the *sheik~heik, sheen~heen*, and *sheaf~heaf* continua; mean score=2.9, where 1 is /h/ and 4 is /ʃ/) than for /hi/-biasing items (items on the *heeds~sheeds, heels~sheels*, and *heave~sheave* continua; mean score=2.4). Also as predicted, we found a main effect of step ($F_1(1,34)=1647.70$, $p<0.001$; $F_2(1,112)=125.48$, $p<0.001$), whereby items whose onsets were acoustically more /ʃi/-like yielded higher /ʃ/ ratings than items whose onsets were acoustically more /hi/-like. There was no reliable interaction between the two effects ($F_1(1,34)=2.67$, $p=0.11$; $F_2<1$). The lexical status results are shown in Table 1 collapsed across steps and in Figure 2 with means for each step. Proportions throughout the paper represent subject means and are shown in figures with error bars for standard error of the mean.

[Table 1]

[Figure 2]

The main effect of lexical status replicates the top-down lexical effect originally reported by Ganong with a /t/~/d/ continuum for this new /h/~/ʃ/ continuum. The main effect of step confirms that participants were not relying only on lexical status in assigning their ratings, but were using bottom-up acoustic cues as well. Because a subset of the participants had already participated in Experiment 2 and 3 during the experiment session, we also compared performance based on prior experiment participation. There was no difference in phoneme ratings between participants who had only participated in Experiment 1 and those that had participated in multiple experiments (Fs<1).

Experiment 2: Collocational Effects

Borsky et al. (1998), Connine (1987), Isenberg et al. (1980), and Miller et al. (1984), investigated top-down effects of sentence context by asking whether listeners are sensitive to the predictability of a word in contexts which evoke a related concept (e.g., *path* or *bath* in contexts about jogging or hot water). In keeping with those previous investigations, we ask whether listeners are sensitive to the predictability of a gendered pronoun in a context that evokes only one gender. If so, listeners are expected to report hearing more *he* pronouns in contexts that describe events with only male referents and more *she* pronouns in contexts that describe events with only female referents. As in Experiment 1, listeners are expected to show sensitivity to the acoustic step along the continuum as well. This experiment is similar to the earlier studies of collocational effects in that an observed effect can be accounted for both in terms of real-world event knowledge and through collocational frequencies of related words. This experiment serves to confirm that listeners are able to use gender cues in determining the predictability of an

upcoming pronoun and that gendered pronouns are a valid target for assessing top-down effects in spoken word recognition.

For this experiment as well as the following ones, it is important to consider how listeners' expectations about pronominalization (their bias to hear a pronoun instead of a name) may influence the results. This investigation of the effect of gender congruity measures listeners' expectations regarding the re-mention of a salient individual already introduced in the preceding context. The effect is therefore predicted to be modulated by listeners' expectations regarding the *form* of reference that the speaker will use in re-mentioning a particular individual. Choice of referring expression in English (pronoun vs. proper noun) is known to be guided by the syntactic position of the previous reference to that individual. In story completion tasks (Stevenson, Crawley, & Kleinman, 1994; Arnold, 2001; Rohde & Kehler, 2009), speakers typically produce pronouns when referring to individuals that were last mentioned in subject position, even in contexts in which a pronoun will be ambiguous (e.g. *Abigail annoyed Dorothy because* <u>*she*</u>$_{\text{ABIGAIL}}$ *talked nonstop*); in contrast, they tend to use proper names (or other non-pronominal forms) when re-mentioning non-subjects, even when a pronoun would be unambiguous (e.g. *Abigail annoyed Bob because* <u>*Bob*</u> *was in a bad mood*). This has been attributed to speakers' bias to mention topical information in subject position (Lambrecht, 1994) and to realize more topical information with reduced referring expressions such as pronouns (Ariel 1990; Gundel, Hedberg, & Zacharski, 1993; Prince 1992). We therefore may find stronger gender congruity effects in contexts which favor re-mention of the subject, due to listeners' increased anticipation for and subsequent discovery of a *pronominal* reference in those contexts.

Method

Participants. 27 native English-speaking Northwestern undergraduates participated for $6 or course credit.

Materials. 40 sentences were constructed consisting of two clauses connected by the word *because*. The first clause introduced two individuals of the same gender and the second clause contained an acoustically ambiguous pronoun. Half the items contained female referents and half male referents. The items resemble those presented in (1-2) in the introduction and were constructed by minimally changing the passages used in our target manipulation in Experiment 3: The names were changed so that they would be of the same gender, and the post-pronoun sentence continuations were revised to more directly describe the causally implicated referent (in comparison to the Experiment 3 continuations, which were required to be compatible with either referent). Gender bias was manipulated within subjects and between items. Because the items were adapted from the Experiment 3 materials, verb bias (subject-bias versus object-bias) also varied within subjects and between items. The sample items in (3-4) use bold to highlight the gendered names and underlining to mark the referent favored by the IC verb. Based on reported pronominalization biases, the subject-biased verbs (3a,4a) may show stronger effects. The complete set of experimental items can be found in the appendix.

(3) *he*-biasing contexts

a. subject-biased verb

Tyler deceived **Gabe** because ☐ didn't want anyone to know the truth.

b. object-biased verb

Luis reproached **Joe** because ☐ hadn't done the work.

(4) *she*-biasing contexts

a. subject-biased verb

Abigail annoyed **Dorothy** because ☐ talked nonstop.

b. object-biased verb

Joyce helped **Sue** because ☐ was up against a deadline.

In order to increase the number of trials at each data point without repeating items, we selected a subset of 5 steps from the 20 in Experiment 1. Additional pilot experiments using the full-sentence contexts helped identify the subset of steps that were centered around the point of maximum ambiguity for listeners. Figure 1 suggests that steps 11/12 would be the point of maximum ambiguity, but pilot studies showed that stimuli centered around step 14 displayed the maximum divergence between *he* and *she* identification functions, perhaps because of the effects of context and speech rate on perception (recall that typical duration for *h* and *sh* are different). In order to maintain an even distribution of stimulus, steps 12-16 were used instead of a set incorporating the endpoints (e.g., 1, 13, 14, 15, 20). This choice of 5 steps served to maximize the number of trials that were ambiguous and might therefore show a top-down effect. Furthermore, as the results below show, the 5 steps we used were adequately distributed such that a bottom-up effect of acoustic step could be established, obviating the necessity of including end-points.

Each of the frame sentences was recorded by the same male native English speaker as above. These original sentences contained *he* as the pronoun, and the speaker recorded the sentences with a pause before the pronoun to minimize co-articulatory cues from the preceding word *because*. The pronoun was then replaced with one of the 5 ambiguous pronouns from the /hi/~/ʃi/ continuum above using Praat. The pause before the pronoun was then standardized to 70 milliseconds between the offset of frication in *because* and the onset of frication for the pronoun for all items. This pause served to minimize any listener compensation for co-articulation due to

because. As described in the Methods section of Experiment 1, those pronouns were created by combining two tokens of naturally produced *he* and *she* pronouns at varying intensities. By using naturally produced pronouns instead of synthesized speech, we avoided inserting synthetic speech into otherwise naturally recorded sentences. The spectrograms and waveforms in Figure 3 show the most /ʃi/-like and most /hi/-like steps among the 5 pronouns we used.

[Figure 3]

Procedure. Participants listened to the sentences through headphones while sitting in a sound-attenuated booth. For each item, they were asked to indicate on a button box whether the sentence contained the word *he* or *she*, using a 4-point scale (1="definitely he", 2= "probably he", 3="probably she", 4="definitely she"). Participants were not able to respond until the end of the second clause. After each sentence, participants were asked a yes/no comprehension question based on the sentence's meaning (but not the interpretation of the pronoun) to ensure they were focused on understanding the sentence and not focused exclusively on the ambiguous phoneme (see Miller et al., 1984). Participants heard all items once so as to avoid gender expectations for the second clause based on repeated mention of particular verbs or names in the first clause and their prior resolution of the pronoun in that context.

Participants completed this experiment along with Experiment 3 in the same session. This task was always completed after Experiment 3 (our key pragmatic manipulation) to ensure that any measured effect in Experiment 3 could not be attributed to verb-repetition effects that might arise due to item similarity across tasks.

Results and Discussion

Only trials with correctly answered comprehension questions were included in the results; this excluded 13.1% of the responses. We submitted the data to a three-way gender bias

× acoustic step × verb bias ANOVA, by subjects and by items. As predicted, there was a main effect of gender bias ($F_1(1,23)=34.58$, $p<0.001$; $F_2(1,30)=58.26$, $p<0.001$), whereby listeners assigned higher *she* ratings to items containing two female referents than to items containing two male referents, and a main effect of acoustic step ($F_1(1,23)=32.06$, $p<0.001$; $F_2(1,30)=19.78$, $p<0.001$), whereby more /ʃi/-like steps yielded higher *she* ratings than more /hi/-like steps. We also found a main effect of verb bias ($F_1(1,23)=19.36$, $p<0.001$; $F_2(1,30)=6.54$, $p<0.02$), whereby items containing subject-biased verbs yielded lower *she* ratings than items containing object-biased verbs. The effect of verb bias was driven by a two-way interaction between gender bias and verb bias ($F_1(1,24)=18.95$, $p<0.001$; $F_2(1,30)=9.85$, $p<0.004$), whereby items containing subject-biased verbs yielded a larger difference between *she*-biasing and *he*-biasing contexts than items with object-biased verbs. This interaction is in keeping with the claim that subject-biased verbs may yield stronger pragmatic effects due to their bias for an upcoming *pronominal* referring expression. Post-hoc analyses of the two verb types revealed significant effects of gender bias in both the subject-biased verbs ($F_1(1,25)=55.89$, $p<0.001$; $F_2(1,16)=84.60$, $p<0.001$) and the object-biased verbs ($F_1(1,26)=17.47$, $p<0.001$; $F_2(1,17)=4.53$, $p<0.05$). There were no other two way interactions (gender bias × step: $F_1(1,24)=2.56$, $p=0.12$; $F_2(1,30)=2.95$, $p=0.10$; verb bias × step: $Fs<1$) and no three-way interaction ($F_1(1,25)=1.23$, $p=0.28$; $F_2(1,30)=1.51$, $p=0.23$).[2] The gender bias and verb bias results are shown in Table 2 collapsed across steps and in Figure 4 with means for each step. Note that the full 'S' shaped curve found in Experiment 1 is not visible in Figure 4 because the steps represent only a subset of the curve (steps 12-16).

[Table 2]

[Figure 4]

The results from Experiment 2 are in keeping with the previously reported collocational results—namely, that word recognition depends on a combination of bottom-up cues from the acoustic signal and top-down cues from sentential context. However, the same question that can be raised for the earlier sentential context effects applies here—do these contextual manipulations specifically test listeners' pragmatic biases or could the observed effects also be attributed to semantic neighborhood or co-occurrence effects? The word *she* may simply occur more frequently within a small window of female names; the word *he* may occur more frequently near male names. Given this concern, the pragmatic bias manipulation in Experiment 3 uses contexts with both a female name and a male name and the distance between the gendered names and the pronoun is balanced across items, such that a co-occurrence-based account is insufficient.

Experiment 3: Pragmatic Bias

In the introduction, we laid out the principle design of our pragmatic manipulation, which relies on the causal inferencing that is induced in contexts with IC verbs. We hypothesized that listeners would report hearing more *she* pronouns in contexts in which the verb creates a bias for upcoming reference to a causally-implicated female and more *he* pronouns in contexts with a causally-implicated male. Because we use contexts that mention both a female and a male referent and balanced the position of the names, the mere presence of a female or male name cannot, by itself, explain an observed effect on word recognition, as was a possibility in Experiment 2. As in Experiment 2, stronger pragmatic effects may emerge in contexts which favor re-mention of the subject, due to listeners' increased anticipation of a pronominal reference in those contexts.

Method

Participants. 27 native English-speaking Northwestern undergraduates participated for $6 or course credit.

Materials. 40 sentences were constructed consisting of two clauses connected by the word *because*. The first clause introduced two individuals of opposite gender and contained an IC verb; the second clause contained an acoustically ambiguous pronoun and a sentence completion that provided an explanation that could plausibly be attributed to either individual. Items were balanced for verb bias (subject-biased vs. object-biased) and the position of the male and female names (subject vs. object), as shown in examples (1-2), repeated here as (5-6), with bold to highlight the causally implicated referent. All other aspects of the materials were the same as in Experiment 2. The complete set of experimental items can be found in the appendix.

(5) he-biasing contexts

a. subject-biased verb

Tyler deceived Sue because ☐ couldn't handle a conversation about adultery.

b. object-biased verb

Joyce helped **Steve** because ☐ was working on the same project.

(6) she-biasing contexts

a. subject-biased verb

Abigail annoyed Bruce because ☐ was in a bad mood.

b. object-biased verb

Luis reproached **Heidi** because ☐ was getting grouchy.

In order to ensure that both a subject-referring and an object-referring pronoun yielded a plausible continuation (e.g., *...she/he was in a bad mood* for sentence (6a) above), we conducted a norming study with 12 participants who did not participate in any of the other experiments

presented here. The participants listened to both acoustically unambiguous *he* and *she* versions of sentences like the items in (5-6) and rated the plausibility of each sentence on a scale of 1 to 4. Over the course of the norming study, participants heard each item twice, once with a subject-referring pronoun and once with an object-referring pronoun. A set of plausible and implausible fillers were included to ensure that participants were using the 4-point scale correctly and to provide points of comparison for the elicited plausibility judgments. The fillers were adapted from items in a reading study about possible and impossible events (Warren, McConnell, & Rayner 2008). The norming results confirmed that participants rated all conditions of the experimental items as significantly more plausible than implausible fillers: subject-biased verbs with subject-referring pronouns ($F_1(1,11)=776.67$, $p<0.001$; $F_2(1,78)=626.85$, $p<0.001$), subject-biased verbs with object-referring pronouns ($F_1(1,11)=99.06$, $p<0.001$; $F_2(1,78)=365.20$, $p<0.001$), object-biased verbs with subject-referring pronouns ($F_1(1,11)=318.63$, $p<0.001$; $F_2(1,78)=656.53$, $p<0.001$), and object-biased verbs with object-referring pronouns ($F_1(1,11)=1095.10$, $p<0.001$; $F_2(1,78)=500.23$, $p<0.001$). Within the experimental items alone, however, the results revealed comparatively lower mean scores for subject-biasing verbs paired with object-referring pronouns (main effect of verb bias: $F_1(1,11)=20.98$, $p<0.001$; $F_2(1,27)=22.97$, $p<0.001$; main effect of pronoun reference significant only by subjects: $F_1(1,11)=7.54$, $p<0.05$; $F_2(1,27)=1.54$, $p=0.22$; verb bias × reference interaction: $F_1(1,11)=7.53$, $p<0.05$; $F_2(1,27)=9.43$, $p<0.005$). We discuss possible ramifications of the norming results in the Results section below. The norming results are shown in Figure 5.

[Figure 5]

Procedure. The experimental procedure was the same as in Experiment 2. Participants heard all items once so as to avoid expectations based on repeated verb-name combinations. Of

the three experiments conducted in multi-experiment sessions, this task was always presented first.

Results and Discussion

Only trials with correctly answered comprehension questions were included in the results; this excluded 6.2% of the responses. We submitted the data to a three-way pragmatic bias × acoustic step × verb bias ANOVA, by subjects and by items. As predicted, there was a main effect of pragmatic bias ($F_1(1,22)=39.83$, $p<0.001$; $F_2(1,31)=16.80$, $p<0.001$), whereby listeners assigned higher *she* ratings to items with a causally implicated female referent than to items with a causally implicated male referent, and a main effect of acoustic step ($F_1(1,22)=110.09$, $p<0.001$; $F_2(1,31)=39.78$, $p<0.001$), whereby more /ʃi/-like steps yielded higher *she* ratings than more /hi/-like steps. There was no main effect of verb bias ($F_1<1$; $F_2(1,31)=2.61$, $p=0.12$). We found two-way interactions between pragmatic bias and verb bias ($F_1(1,23)=56.89$, $p<0.001$; $F_2(1,31)=15.76$, $p<0.001$) and between pragmatic bias and acoustic step (marginal by items: $F_1(1,23)=9.64$, $p<0.005$; $F_2(1,31)=3.34$, $p=0.08$). As noted in the discussion of Experiment 2, the pragmatic bias × verb bias interaction is in keeping with the claim that subject-biased verbs may yield stronger effects due to their bias for an upcoming pronominal referring expression. Post-hoc analyses of the two verb types revealed significant effects of pragmatic bias for the subject-biased verbs ($F_1(1,25)=117.31$, $p<0.001$; $F_2(1,17)=21.84$, $p<0.001$) but not the object-biased verbs (Fs<1). The pragmatic bias × acoustic step interaction can be attributed to the stronger pragmatic effect at certain steps (notably step 13). There was no verb bias × acoustic step interaction ($F_1(1,24)=1.14$, $p=0.30$; $F_2<1$). The three-way interaction was only significant by subjects ($F_1(1,25)=4.52$, $p<0.05$; $F_2<1$) and could be attributed to the fact that the stronger pragmatic effect at certain steps is apparent only with

subject-biased verbs.[3] The results are shown in Table 3 collapsed across steps and in Figure 6 with means for each step. As in Experiment 2, the full 'S' shaped curve found in Experiment 1 is not visible in Figure 6 because the steps represent only a subset of the curve (steps 12-16).

[Table 3]

[Figure 6]

These results suggest that word recognition depends on both bottom-up and top-down information. Bottom-up acoustic cues yielded differences across steps; top-down cues yielded differences based on pragmatic bias. Unlike the results in Experiment 2, however, these effects cannot be reduced to a lexical co-occurrence effect because these materials contained both a female name and a male name in the same window. In fact, we found stronger effects in contexts with subject-biasing verbs where the causally implicated referent is linearly more distant from the target pronoun than in contexts with object-biasing verbs.

As in Experiment 2, the verb bias differences seen in Figure 6 may be attributable to differences in listeners' pronominalization expectations. Alternatively, the verb-bias differences may arise from differences in the plausibility of the sentence completions. Neither explanation contradicts the principle finding regarding pragmatic bias, but we explore the plausibility account here because it pertains to the constraints that our results place on models of word recognition.

Recall that the experimental items which received the lowest scores in the norming study were the sentences that contained subject-biased verbs followed by an explanation about the *object* referent (e.g., *Abigail annoyed Bruce because <u>he</u> was in a bad mood*, as in (6a)). One interpretation of the pattern of results in Experiment 3 is that listeners evaluated the acoustically ambiguous pronoun against the plausibility of the sentence completion and not solely on their

verb-driven expectations about which referent would be mentioned next. In other words, for subject-biased verbs, they may have found the sentence completion to be implausible for the pronoun they initially identified, leading to post-hoc revision. To show how such revision could generate the observed effects, consider examples (7) and (8). Both contexts are *he*-biasing and appeared in Experiment 3 with the most *she*-like pronoun (step 16).

(7) subject-biased verb, bias to *he*

Mark exasperated Ilana because ☐ was running late.

(8) object-biased verb, bias to *he*

Jill detested **Peter** because ☐ was a malicious person.

For (7), the norming results show that the subject-referring version was judged to be more plausible than the object-referring version (plausibility score of 3.6 for *he* and 2.4 for *she*). In contrast, for (8), the norming results show that the object-referring and subject-referring versions were both judged to be fairly plausible (3.8 for *he*; 3.3 for *she*). Despite the fact that both items contained the same *she*-like pronoun in *he*-biasing contexts, (7) received lower *she* ratings than (8) (mean ratings of 1.8 for (9) and 2.3 for (10), where 1 is *he* and 4 is *she*). This suggests that the degraded plausibility of the reference to the non-causally-implicated individual in (7) (*Mark exasperated Ilana because <u>she</u> was running late*) may have lowered the *she* ratings; for both items, lower *she* ratings match the verbs' reported pragmatic biases, so the pragmatic effect we observed may have been driven in part by sentence plausibility. Listeners may have revised their judgment of the acoustically ambiguous pronoun when they applied their causal inferencing to integrate the pronoun with the rest of the sentence. However, even if we remove data from items that received plausibility scores below 3 (20% of the total data), the results still show a main effect of pragmatic bias ($F_1(1,24)=35.33$, $p<0.001$; $F_2(1,23)=10.63$, $p<0.004$), along

with the same pattern of other main effects and interactions observed in the original analysis (step: $F_1(1,24)=122.12$, $p<0.001$; $F_2(1,23)=33.69$, $p<0.001$; verb bias: $Fs<1$; pragmatic bias × verb bias: $F_1(1,24)=29.06$, $p<0.001$; $F_2(1,23)=9.58$, $p<0.006$; pragmatic bias × step: $F_1(1,24)=9.85$, $p<0.005$; $F_2(1,23)=3.30$, $p=0.08$; verb bias × step: $Fs<1$; pragmatic bias × step × verb bias: $F_1(1,25)=2.10$, $p=0.16$; $F_2<1$).

Given the pattern in the norming and the possibility of listeners' post-hoc revision, the results here point to models in which listeners integrate top-down information sources late in processing. On one hand, this may be taken as support for models in which top-down effects emerge only *after* bottom-up perceptual processes are complete (e.g., Norris, McQueen & Cutler, 2000). On the other hand, these results are also compatible with an interaction-based model, so long as late-arriving information can be subsequently incorporated at an additional post-perceptual stage. In support of the latter account, there is evidence that listeners integrate sub-phonetic detail with late-occurring top-down lexical cues—though late-occurring in such data is on the order of syllables, not words as in our data—and such evidence supports interactive models in which information from the perceptual system is available for later stages of processing (McMurray, Tanenhaus, & Aslin, 2009; see also Connine, Blasko & Hall, 1991). An interactive account is further supported by the alternative account of the verb-bias differences which attributes the differences to variation in the likelihood of an upcoming pronominal form (see discussion of Experiment 2). The role of late-occurring sentence-plausibility cues in an interactive model is compatible with the time course of effects laid out in ART (Grossberg & Myers, 2000), in which interacting bottom-up and top-down processes require sufficient time to achieve resonance. Experiment 4 below considers the time course of the pragmatic effect, asking whether sublexical acoustic cues are sufficient to trigger integration with top-down cues.

Experiment 4: Time course of pragmatic biases

In order to test whether pragmatic factors influence listeners' phoneme decisions for whole words or whether word subparts are similarly affected, we adapted the materials from Experiment 3 for a gating judgment task (see Grosjean, 1980). For this task, listeners were asked to make *he/she* judgments after hearing truncated portions of a single acoustically ambiguous pronoun and prior to hearing the rest of the sentence. This task allows us to test several aspects of listeners' phoneme decisions.

First, the gating task allows us to compare listeners' treatment of the fricative portion of the pronoun with their treatment of the whole word and thereby to test whether the pragmatic effect observed in Experiment 3 interacts with time. Second, this task allows us to clarify the interactions with verb bias that were observed in Experiments 2 and 3. In Experiment 3, we considered explanations of the differences by verb type based both on the norming results (the plausibility of the continuations) and on properties of the verbs themselves (their pronominalization preferences). To distinguish between these two possibilities, we can now examine listeners' *she/he* judgments at the completion of the pronoun, prior to the rest of the second clause. Presence of an interaction before listeners encounter the rest of the second clause would point to differences in the verbs themselves, whereas the disappearance of the interaction would indicate that sentence plausibility likely drove the verb bias differences observed in Experiment 3. Lastly, in order to see if listeners have expectations about the upcoming pronoun, we also ask listeners to make *he/she* judgments when acoustic cues are still quite limited. Based on recent research demonstrating listeners' verb-driven expectations about upcoming patterns of coreference (Kehler, Kertz, Rohde, & Elman, 2008; Pyykkönen & Järvikivi, 2010), we predict listeners will show a pragmatic effect prior to hearing the acoustic material.

Method

Participants. 14 native English-speaking Northwestern undergraduates participated for either $6 or course credit. None had participated in the other experiments reported here.

Materials. The materials consisted of the same sentences from Experiment 3 (e.g. *Abigail annoyed Bruce because* ☐ *was in a bad mood*), but only one acoustically ambiguous pronoun was used (step 13). Each sentence was spliced into two parts at one of five different time points (gates) measured from the offset of *because*: 50ms, 100ms, 250ms, 300ms, and 450ms. Due to the 70ms gap between *because* and the pronoun, at 0ms, there was no audible portion of the pronoun; at 100ms, the beginning of the onset consonant (~20ms) is audible; at 250ms the entire onset consonant is perceivable; at 300ms a portion of the vowel is perceptible, providing onset consonant information in the form of its length and formant transitions; and the full pronoun is heard at 450ms (Figure 7).

[Figure 7]

Each sentence appeared with the same ambiguous token—step 13 on the continuum, which had the largest pragmatic effect in Experiment 3—in order to increase the number of observations from each participant at each time point. We manipulated pragmatic bias, verb bias, and gate within subjects and between items.

Procedure. The procedure was a variation of that used in Experiments 2 and 3. Participants heard the first part of each item, up to and including a portion of the pronoun. They then made their *she/he* decision, listened to the rest of the sentence, and answered a comprehension question. Participants heard all items once.

Results and Discussion

Only trials with correctly answered comprehension questions were included in the results; this excluded 5.6% of responses. Table 4 shows the means for each gate.

[Table 4]

We submitted the data to a three-way pragmatic bias × verb bias × gate ANOVA, by subjects and by items. Regarding the time course of the pragmatic effects, we find a main effect of pragmatic bias ($F_1(1,11)=79.13$, $p<0.001$; $F_2(1,20)=23.88$, $p<0.001$) and a pragmatic bias × gate interaction ($F_1(4,50)=17.22$, $p<0.001$; $F_2(4,20)=5.09$, $p<0.006$) driven in part by the disappearance of the pragmatic effect at 300ms, the gate which represents the full fricative portion of the pronoun (Table 5). There were also main effects of verb bias ($F_1(1,11)=17.36$, $p<0.002$; $F_2(1,20)=5.94$, $p<0.03$), whereby subject-biased verbs yielded higher *she* ratings than object-biased verbs, and gate ($F_1(4,50)=6.46$, $p<0.001$; $F_2(1,20)=2.87$, $p=0.05$), whereby the second gate yielded the highest *she* ratings. The remaining interactions failed to reach significance in one or both of the subjects and items analyses (pragmatic bias × verb bias: $F_1(1,11)=3.95$, $p=0.07$; $F_2<1$; verb bias × gate: $F_1(4,50)=5.85$, $p<0.001$; $F_2(4,20)=1.11$, $p=0.38$; pragmatic bias × verb bias × gate: $F_1(4,50)=4.50$, $p<0.004$; $F_2<1$).[4]

The main effect of pragmatic bias is consistent with the results from Experiment 3: Contexts with a causally implicated female yield higher *she* ratings than contexts with a causally implicated male. Regarding the time course question, the results (Figure 8; Table 5) show that the pragmatic effect appears not only with the full lexical item, but also when lexical material is minimal or even absent: The pragmatic effect is apparent at the first gate, when no acoustic stimulus is present; the effect is also present at the second gate, with limited acoustic cues for the fricative, and at the third gate, the end of the fricative but before any vowel formant transition cues or cues signaling the end of the fricative; the effect then disappears at the 300 ms gate,

which contains the fricative plus the vowel formant transitions; the effect reappears at the final gate, which contains the full pronoun. Post-hoc analyses of the results at each gate are shown in Table 5 in rows labeled Pragmatic bias.

[Figure 8]

[Table 5]

Regarding the gender × verb bias interaction observed in Experiment 3, we find here that the results at the last gate (the full pronoun) show no such interaction, only a main effect of pragmatic bias (see Table 5, specifically the 450ms-gate rows for Pragmatic bias and Pragmatic × Verb as Sources of Variance). This suggests that the pragmatic bias × verb bias interaction in Experiment 3 was the result of sentence completion plausibility. The principle pragmatic effect is maintained when listeners hear only the first clause and the pronoun from the second clause.

Regarding claims in the IC literature about the anticipatory nature of IC biases, the results at the first two gates (the gates with no/limited acoustic cues) are in keeping with expectation-driven accounts of IC processing: The effect of pragmatic bias is apparent before the listener encounters the acoustic material of the pronoun (significant by subjects, marginal by items, see above). We also conducted an analysis of responses collapsed across the first two gates in order to increase power, and the results confirm the anticipatory bias (main effect of pragmatic bias: $F_1(1,13)=64.37$, $p<0.001$, $F_2(1,12)=7.67$, $p<0.02$, effect of verb bias: $F_1(1,13)=18.11$, $p<0.001$; $F_2<1$; no interaction: $F_1(1,13)=2.56$, $p=0.13$; $F_2<1$).[5]

The collapse and disappearance of the pragmatic effect at the fourth gate is challenging to explain purely within the context of available models of cue integration. A fully interactive model permits top-down factors to be combined at any stage of acoustic processing, in keeping with the observed pragmatic effect at sublexical gates, but such a model would not necessarily

predict the observed change at the fourth gate. However, an anonymous reviewer points out that the gating paradigm is known to emphasize the salience of word-initial information (see Allopenna, Magnuson, & Tanenhaus, 1998), such that the top-down cues in our experiment may have been overshadowed at some gates by the available but ambiguous bottom-up cues. This biased competition between bottom-up and top-down cues is consistent with interactive models that permit direct interaction between available cues (Mirman, McClelland, Holt, & Magnuson, 2008).

On the other hand, a post-perceptual model might be expected to constrain the sources of information that are available when making a pragmatic decision (due to the encapsulation of the acoustic perception system) and thus limit the influence of top-down effects during sublexical processing. However, such models have been used successfully to account for lexical effects in non-words (Connine, Titone, Deelman, & Blasko, 1997; Newman, Sawusch, & Luce, 1997). Furthermore, if pragmatic cues guide lexical expectations and if a post-perceptual model permits lexical processes to start early during processing, then the acoustic and lexical information could plausibly be integrated at a later decision layer, rendering our results compatible with a post-perceptual model as well.[6]

General Discussion

The results presented here are in keeping with a body of accumulating evidence in the psycholinguistic literature that points to multiple information sources that are integrated during language processing. These results suggest that the range of integrated cues spans the conceivable range of linguistic information sources: Bottom-up phonetic information is integrated with high-level causal inferencing about events, event participants, and the likelihood of co-reference across clauses in a discourse. We find that listeners' interpretation of sounds

along the *h~sh* continuum reflects top-down cues including lexical status (Experiment 1), collocational information (Experiment 2), and pragmatic inference (Experiments 3 & 4). Listeners' interpretation also reflects bottom-up acoustic cues, as evident in the different ratings assigned to tokens that occupy non-adjacent positions along the continuum (effect of acoustic step in Experiments 1, 2, & 3) and the different treatment of different subcomponents of the acoustically ambiguous pronoun (Experiment 4). Together, the results from Experiments 3 and 4 suggest that pragmatic biases play a role both in listeners' anticipation of upcoming words as well as in their integration of acoustically ambiguous words into a larger discourse context. This is consistent with the growing body of evidence showing that listeners are sensitive to real-world cues such as information about the speaker or other properties of the discourse context (Massaro, 1998; Staum Casasanto, 2008; Kraljic & Samuel, 2005; Hay & Drager, 2010). In our case, listeners' word recognition was guided by linguistic properties of the sentence itself but depends crucially on higher-level inferences about causality and the real world.

As described above, existing models of word recognition currently account for contextual effects in one of two ways. Highly interactive models permit direct interaction between acoustic cues, the lexicon, and contextual cues (contextual cues broadly construed, e.g. visual cues, speaker information, acoustic context) such that top-down biases can influence the perceptual system itself (Grossberg & Myers, 2000; Johnson & Mullennix, 1997; McClelland & Elman, 1986). Such models are supported by recent evidence on the neural bases of lexical effects on phonetic perception (Myers & Blumstein, 2008). On the other hand, post-perceptual models have been proposed that specify a separate phoneme decision layer as the stage at which listeners combine higher-level lexical information sources with lower-level phonetic cues (Norris, McQueen & Cutler, 2000). Both types of models could in principle be adapted to account for the

pragmatic effects observed here, so long as the range of contextual cues is not restricted to lexical or co-occurrence-based input.

For interactive models, an important question is whether pragmatic information is directly available during the speech perception process, adding an additional set of non-acoustic cues into the perceptual process, or whether pragmatic context yields an expectation for a particular word, which in turn makes the perceptual process more sensitive to certain acoustic cues. Both may be at play, since the results from Experiment 4 point to expectations, whereas the results from Experiment 3 point to additional non-acoustic post-hoc constraints such as sentence plausibility.

For models that rely on post-perceptual integration of information, context serves as a check on an encapsulated perceptual process. To account for our results within a post-perceptual model, pragmatic biases must be permitted to act as even higher-level top-down constraints, in addition to other biases that are introduced by lexical status, syntax, and semantic congruity. In the Merge model, however, this top-down check is attributed to a postlexical processing stage reserved for metalinguistic judgments (McQueen, Cutler, & Norris, 2006). To measure effects in non-metalinguistic word recognition, researchers have turned to adaptation paradigms in which listeners' phonetic category boundaries are shifted following exposure to ambiguous sounds in biasing contexts (e.g., Kraljic & Samuel, 2005; Maye, Aslin, & Tanenhaus, 2008, though see Norris, McQueen, & Cutler, 2003, for a critique of such results as merely evidence of shifts in learning not shifts in on-line processing). Although the task presented here required listeners to make a metalinguistic judgment, we used full-sentence contexts and comprehension questions for each stimulus item, in order to encourage participants to rely on more naturalistic processing. It remains an open question how best to engage non-metalinguistic word recognition. An

alternative approach is the indirect evaluation of top-down effects on a secondary phonetic effect (Elman & McClelland, 1988), though the interpretation of such effects remains controversial (Pitt & McQueen, 1998; Magnuson, McMurray, Tanenhaus & Aslin, 2003; Samuel & Pitt, 2003).

Just as existing models of word recognition could in principle be extended to include higher-level top-down biases, another option for modeling our results would be to adapt existing sentence processing models to capture effects at lower levels of processing. Existing constraint-based sentence processing models have up until now primarily targeted syntactic processes not phoneme decisions (MacDonald 1994; Jurafsky 1996; Spivey & Tanenhaus 1998; McRae, Spivey-Knowlton, & Tanenhaus 1998; Levy 2008, among others). These models—crucially their architectures for integrating multiple cues—could be adapted to fit our data by incorporating discourse-based constraints that interact fully with other processing biases, including those generated at the phonetic level. The work described in this paper attests to the importance of a unified approach that models a range of information sources and their combined impact on processing.

Existing processing models have thus not fully addressed the question of precisely which information sources at which linguistic levels are integrated and what mechanism would allow phonetic and pragmatic information to be combined. Our results, which present a new type of integrative effect, help establish the extent of possible integration that must be accounted for, though the results also raise questions regarding the exact nature of these effects. Our findings leave open the possibility that some contextual effects require a post-perceptual approach, while others, perhaps those at lower levels of linguistic structure, can be captured with an interaction-based approach. The paradigm we have introduced here provides useful contexts for such work

precisely because these contexts permit the manipulation of biases that may be active when listeners are interpreting sounds in rich discourse contexts.

References

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

Ariel, M. (1990). Accessing Noun-Phrase Antecedents. London: Routledge.

Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes*, 31, 137-162.

Au, T. K. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, 25, 104-122.

Bach, K. (1994). Conversational impliciture. *Mind and Language*, 9, 124-62.

Baese-Berk, M. & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527-554.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustic Society of America*, 103(5), 2670-2676.

Bowers, J. S. & Davis, C. J. (2004). Is speech perception modular or interactive? *TRENDS in Cognitive Sciences*, 8(1), 3-5.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behaviour*, 16, 601-609.

Chistovich, L. A., & Lublinskaya, V. V. (1979). The 'Center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research*, 1, 185-195.

Cristia, A., McGuire, G. L., Seidl, A., & Francis, A. L. (in press). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*.

Connine, C. M. (1987). Constraints on Interactive Processes in Auditory Word Recognition: The Role of Sentence Context. *Journal of Memory and Language*, 26, 527-538.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, 30, 234-250.

Connine, C. M. & Clifton, C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 291-299.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, 37, 463-480.

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 8, 195-210.

Elman, J. L. & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143-165.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb Aspect and the Activation of Event Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 182-196.

Ferstl, E.C., Garnham, A., & Manouilidou, C. (2011). Implicit Causality Bias in English: A corpus of 300 verbs. *Behavior Research Methods*, 43(1), 124-135.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 526–540.

Ganong, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–125.

Garvey, C. & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5, 459-464.

Goldinger, S. D. (1996). Words and Voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 1166-1183.

Grice, H. P. (1975). Logic and conversation, in P. Cole & J. Morgan (ed.), Syntax and Semantics, 3: Speech Acts, pp. 41–58, New York: Academic Press.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267-283.

Grossberg, S. & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, 107, 735-767.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203-225.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274-307.

Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4), 865-892.

Isenberg, D., Walker, E. C. T., & Ryder, J. M. (1980). A top-down effect in the identification of function words. Paper presented at the Acoustical Society of America, Los Angeles, CA.

Johnson, K. & Mullennix, J. (1997). Talker variability in speech processing. San Diego: Academic Press.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20(2), 137–194.

Kehler, A. (2002). Coherence, reference, and the theory of grammar. Stanford, CA: CSLI Publications.

Kehler, A., Kertz, L., Rohde, H. & Elman, J. L. (2008). Coherence and Coreference Revisited. *Journal of Semantics Special Issue on Processing Meaning*, 25, 1-44.

Koornneef, A. W., & van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54, 445-465.

Kraljic, T. & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141-178.

Kraljic, T, Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 19, 332-338.

Lambrecht, K. (1994). Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Cambridge: Cambridge University Press.

Levinson, S. C. (1987). Pragmatics and the grammar of anaphora. *Journal of Linguistics*, 23, 379-434.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9(2), 157–201.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical Effects on Compensation for Coarticulation: The Ghost of Christmas Past. *Cognitive Science*, 27, 285-298.

Massaro, D. W. (1998). Perceiving talking faces: From speech perception to a behavioral principle. MIT Press: Cambridge, MA.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32, 543-562.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(8), 363-369.

McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphora resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, 34, 543-566.

McGuire, G. (2007). Phonetic Category Learning. Ph.D. Dissertation, The Ohio State University.

McKoon, G., Greene, S., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology*, 19, 1040-1052.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60(1), 65-91

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 433–443.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30, 1113-1126.

McQueen, J. M., Norris, D., & Cutler, A. (2006). Are there really interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(12), 533.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics*, 36(4), 329-337.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). Reply to McQueen et al.: Theoretical and empirical arguments support interactive processing. *Trends in Cognitive Sciences*, 10(12), 534.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J.S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science*, 32, 398-417.

Munson, B., & Coyne, A. C. (in press). The Influence of Apparent Vocal-Tract Size, Contrast Type, and Implied Sources of Variation on the Perception of American English Voiceless Lingual Fricatives. *Journal of the Phonetic Society of Japan*.

Myers, E. B. & Blumstein, S. E. (2008). The neural bases of the lexical effect: An fMRI investigation. *Cerebral Cortex*, 18(2), 278-288.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 873-889.

Norris, D. G., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299-370.

Norris, D. G., McQueen, J. M., & Cutler, E. A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204-238.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 1037–1052.

Pitt, M. A. & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347-370.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, 29, 149–188.

Prince, E. F. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thompson and W. Mann, eds., Discourse Description: Diverse Analyses of a Fundraising Text. Amsterdam/Philadelphia: John Benjamins. 295-325.

Pyykkönen, P. & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, 57, 5-16.

Rohde, H. & Ettlinger, M. (2010). Effects of Pragmatic Inference on Phonetic Category Perception. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 381-386.

Rohde, H. & Kehler, A. (2009). Grammatical and Coherence-Driven Biases in Pronoun Interpretation. Talk presented at the 83rd Annual LSA Meeting. San Francisco, CA, January 8-11, 2009.

Rohde, H., Kehler, A., & Elman, J. L. (2007). Pronoun Interpretation as a Side Effect of Discourse Coherence. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 617-622.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474–494.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125, 28–51.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, 32, 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, 12, 348–351.

Samuel, A. G. & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48, 416-434.

Spivey, M. J. & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential content and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6),1521-1543.

Staum Casasanto, L. (2008). Does Social Information Influence Sentence Processing? In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, 799-804.

Stevenson, R., Crawley, R., & Kleinman, D. (1994). Thematic roles, focusing and the representation of events. *Language and Cognitive Processes*, 9, 519-548.

Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles, and coherence relations. *Language and Cognitive Processes*, 15, 225-262.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, 42, 423-443.

van Alphen, P. & McQueen, J. M., (2001). The time-limited influence of sentential context of function word identification. *Journal of Experimental Psychology: Human Perception and Performance*, 27(5), 96-1523.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.

Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American*, 223, 30–36.

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of Context on Eye Movements When Reading About Possible and Impossible Events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 1001–1010.

Willits, J. A., Sussman, R. S., & Amato, M. S. (2008). Event knowledge vs. verb knowledge. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 2227–2232.

Xu, Q., Jacewicz, E., Feth L. L., & Krishnamurthy, A. K. (2004). Bandwidth of spectral resolution for two-formant synthetic vowels and two-tone complex signals. *Journal of the Acoustic Society of America*, *115*, 1653–1664.

Appendix

Target items for collocation manipulation (Experiment 2)

Subject-biased verbs

Malcolm aggravated Brett because ☐ was arrogant. (N Did someone aggravate Malcom?)

Eliza amazed Natalie because ☐ was incredibly strong. (Y Was Natalie very strong?)

Ronald amused Bruce because ☐ stood on his head. (N Did someone stand on the table?)

Abigail annoyed Dorothy because ☐ talked nonstop.(Y Was it Dorothy who was annoyed?)

Nathan apologized to Owen because ☐ was late. (N Was it Owen who apologized?)

Ethel bored Jasmine because ☐ never left the house. (Y Was someone a bore?)

Tony charmed Dennis because ☐ baked muffins for breakfast. (N Did Tony bake brownies for breakfast?)

Valerie confessed to Ella because ☐ had forged the check. Y Did someone fake a signature on a check?)

Tyler deceived Gabe because ☐ didn't want anyone to know the truth. (N Was it Tyler who was deceived?)

Bethany disappointed Naomi because ☐ failed the test. (Y Was it Bethany who was a disappointment?)

Mark exasperated Tom because ☐ forgot their meeting. (N Did someone please Tom?)

Lucy fascinated Ilana because ☐ could ride a unicycle. (Y Could someone ride with only one wheel?)

Noah frightened Ian because ☐ drove over 100 miles per hour. (N Was it Noah who was frightened?)

Andrea humiliated Lillian because ☐ brought out the family photo album. (Y Was it Andrea who humiliated someone?)

Hal infuriated Paul because ☐ flirted with everyone. (N Did Paul infuriate someone?)

Ann inspired Gloria because ☐ studied every night. (Y Was someone a hard working student?)

Dwayne intimidated Curtis because ☐ had an expensive car. (N Did someone have an expensive yacht?)

Beth offended Jacqueline because ☐ made fun of cat owners. (Y Was it Jacqueline who was offended?)

Greg scared Dustin because ☐ bared his teeth and growled. (N Was it Dustin who scared Greg?)

Cecelia surprised Tracy because ☐ baked a birthday cake. (Y Did someone surprise Tracy?)

Object-biased verbs

Doug assisted Bob because ☐ needed to pass the exam. (N Was the exam irrelevant?)

Katherine blamed Ebony because ☐ didn't read the directions. (Y Should someone have read the directions?)

Timothy comforted Carl because ☐ was nervous. (N Was it Timothy who was comforted?)

Kristen congratulated Stephanie because ☐ had gotten a new job. (Y Was it Kristen who congratulated Stephanie?)

Grant corrected Peter because ☐ had added two numbers wrong. (N Did someone subtract two numbers incorrectly?)

Jill detested Susan because ☐ was so unsympathetic. (Y Was someone unsympathetic?)

John envied Christopher because ☐ went skiing in Aspen every year. (N Was it John who was envied?)

Kara feared Claire because ☐ enjoyed watching people suffer. (Y Did Kara fear someone?)

Justin hated Steve because ☐ ruined the party. (N Was someone the subject of adoration?)

Joyce helped Sue because ☐ was up against a deadline. (Y Was time running out before the deadline?)

Frank mocked Steward because ☐ forgot everyone's name. (N Did someone remember everyone's name?)

Brooke noticed Eileen because ☐ was taller than everyone else. (Y Was it Eileen who was noticed?)

Austin pacified Burt because ☐ was throwing a tantrum. (N Was it Austin who was pacified?)

Tina praised Eleanor because ☐ had worked hard and improved a lot. (Y Did someone offer praise?)

Luis reproached Joe because ☐ hadn't done the work. (N Had everyone done their work?)

Theresa scolded Heidi because ☐ was fidgeting. (Y Was someone unable to sit still?)

Rob stared at Lance because ☐ was really cute. (N Was it Rob who was stared at?)

Rachel thanked Elizabeth because ☐ had offered to help. (Y Was it Rachel who said thanks?)

Charles trusted Josh because ☐ was very reliable. (N Was someone distrustful?)

Kate valued Eve because ☐ was honest. (Y Was someone honest?)

Target items for pragmatic manipulation (Experiment 3)

Subject-biased verbs

Malcom aggravated Natalie because ☐ was too impatient. (Y Was it Malcom who aggravated someone?)

Eliza amazed Brett because ☐ was so gullible. (N Was it Eliza who was amazed by someone?)

Ronald amused Dorothy because ☐ had a good sense of humor. (Y Did someone have a good sense of humor?)

Abigail annoyed Bruce because ☐ was in a bad mood. (N Was Bruce in a good mood?)

Nathan apologized to Jasmine because ☐ is honorable. (Y Was it Jasmine who received an apology?)

Ethel bored Owen because ☐ was pre-occupied with work. (N Was it Ethel who bored someone?)

Tony charmed Ella because ☐ was eager to hear about life at college. (Y Was it Tony who charmed someone?)

Valerie confessed to Dennis because ☐ believed in honesty at all times. (N Was it Valerie who listened to a confession?)

Tyler deceived Naomi because ☐ couldn't handle a conversation about adultery. (Y Was someone unable to talk about cheating?)

Bethany disappointed Gabe because ☐ couldn't accept mistakes. (N Was someone forgiving of mistakes?)

Mark exasperated Ilana because ☐ was running late. (Y Did Mark exasperate someone?)

Lucy fascinated Tom because ☐ loves all people. (N Was Lucy fascinated by Tom?)

Noah frightened Lillian because ☐ believes in vampires. (Y Does someone believe in vampires?)

Andrea humiliated Ian because ☐ opened the bathroom door. (N Did someone close the bathroom door?)

Hal infuriated Gloria because ☐ is a very negative person. (Y Did Hal infuriate someone?)

Ann inspired Paul because ☐ was working to become an artist. (N Was someone working to become a lawyer?)

Dwayne intimidated Jacqueline because ☐ takes things too seriously. (Y Was it Jacqueline who was intimidated?)

Beth offended Curtis because ☐ was too worried about hygiene. (N Did Curtis offend someone?)

Greg scared Tracy because ☐ was standing alone in the attic. (Y Did Greg scare someone?)

Cecelia surprised Dustin because ☐ was home sick. (N Cecelia taken by surprise by something?)

Object-biased verbs

Doug assisted Ebony because he wanted the job done by 5pm. (Y Did someone want to get the work done by the end of the day?)

Katherine blamed Bob because he couldn't see that some things take time. (N Did someone explain that some things take time?)

Timothy comforted Stephanie because ☐ knew it was time to say goodbye. (Y Was it time to say goodbye?)

Kristen congratulated Carl because ☐ had come so close to winning. (N Did Carl congratulate someone?)

Grant corrected Susan because ☐ hates mistakes. (Y Did Grant correct someone?)

Jill detested Peter because ☐ was a malicious person. (N Was it Jill who was detested?)

John envied Claire because ☐ was studying at a college close to home. (Y Was it Claire who was envied?)

Kara feared Christopher because ☐ was going to get caught sooner or later. (N Was it Kara was feared?)

Justin hated Rebecca because ☐ is hostile. (Y Was someone hostile?)

Joyce helped Steve because ☐ was working on the same project. (N Was everyone working on a different project?)

Frank mocked Eileen because ☐ didn't understand why studying endocrinology is important. (Y Was someone studying endocrinology?)

Brooke noticed Stewart because ☐ happened to be standing on the same train platform. (N Were Brooke and Stewart in line at the same airport?)

Austin pacified Eleanor because ☐ was hoping to get out of the house. (Y Was someone trying to leave the house?)

Tina praised Burt because ☐ was being nice. (N Was it Tina who got praised?)

Luis reproached Heidi because ☐ was getting grouchy. (Y Was it Luis who reproached someone?

Theresa scolded Damien because ☐ had no patience. (N Was it Damien who scolded someone?)

Rob stared at Elizabeth because ☐ was sitting across the table. (Y Was someone sitting at a table?)

Rachel thanked Lance because ☐ is kind. (N Was someone unkind?)

Charles trusted Eve because ☐ knew the importance of family. (Y Did someone put a high value on family?)

Kate valued Josh because ☐ was in the same situation. (N Were Kate and Josh in different situations?)

Footnotes

[1] This treatment of IC effects as evidence of listeners' causal inference is echoed in the introduction to Fersl, Garnham, and Manouilidou's recent corpus study of IC verbs:

> The effects of implicit causality in sentence comprehension and production have been manifested with great regularity across different research paradigms, across different languages and cultures, and for children as well as adults (for a review, see Rudolph & Försterling, 1997). For instance, in psycholinguistics, implicit causality is known to play a role in the comprehension of discourse, since the causal inferences reflect part of the general knowledge one must have access to in order to grasp the meaning of the text. (Ferstl et al, in press)

[2] Including responses for items with incorrectly answered comprehension questions yields the same reliable main effects of gender bias, acoustic step, and verb bias, as well as the interaction between gender bias and verb bias. The only difference is an additional gender × step interaction, significant only by subjects and driven by the reduction of the gender bias in the most /hi/-like step.

[3] Including responses for items with incorrectly answered comprehension questions yields the same reliable main effects of pragmatic bias and verb bias, as well as the two-way interactions between pragmatic bias and verb bias and between pragmatic bias and acoustic step, though the latter is marginal by items. The three-way interaction is not significant by subjects or by items.

[4] Including responses for items with incorrectly answered comprehension questions yields the same main effects of pragmatic bias, verb bias, and gate (though gate is marginal by

items), as well as the two-way interaction between pragmatic bias and gate. No additional interactions were significant by subjects and by items.

[5] Increasing power by collapsing responses across the 250ms and 300ms gates does not result in a main effect of pragmatic bias appearing for that pre-final region.

[6] We thank Bob McMurray and an anonymous reviewer who pointed out the compatibility of our results with both interactive and post-perceptual models in their reviews of an earlier version of this paper. The anonymous reviewer noted that lexical and acoustic processing can be cascaded and can thereby both start early even if the processing mechanisms themselves are not interacting.

Table 1

Phoneme Category Ratings Collapsed Across Steps (1=h / 4=sh)

	Mean *sh*-rating
***h*-biasing contexts**	2.44 $\pm$ 0.04
***sh*-biasing contexts**	2.91 $\pm$ 0.04

Table 2

He/She Ratings Collapsed Across Steps (1=he / 4=she)

	Subject-biased verbs	**Object-biased verbs**
***he*-biasing contexts**	1.36 $\pm$ 0.07	1.75 $\pm$ 0.09
***she*-biasing contexts**	2.26 $\pm$ 0.11	2.31 $\pm$ 0.11

Table 3

Phoneme category ratings collapsed across steps (1=he / 4=she)

	Subject-biased verbs	**Object-biased verbs**
***he*-biasing contexts**	1.77 $\pm$ 0.09	2.22 $\pm$ 0.11
***she*-biasing contexts**	2.26 $\pm$ 0.10	2.31 $\pm$ 0.12

Table 4

He/she ratings by gate and by condition (1=he / 4=she)

	Gate (ms)	Subject-biased verbs	Object-biased verbs
he-biasing contexts	50	1.85 ± 0.16	1.79 ± 0.16
she-biasing contexts	50	3.11 ± 0.12	2.28 ± 0.10
he-biasing contexts	100	2.57 ± 0.15	2.39 ± 0.11
she-biasing contexts	100	3.27 ± 0.12	3.32 ± 0.10
he-biasing contexts	250	2.79 ± 0.13	2.04 ± 0.16
she-biasing contexts	250	3.11 ± 0.18	2.54 ± 0.15
he-biasing contexts	300	3.07 ± 0.16	2.25 ± 0.19
she-biasing contexts	300	2.39 ± 0.22	2.21 ± 0.23
he-biasing contexts	450	2.00 ± 0.14	1.65 ± 0.17
she-biasing contexts	450	3.04 ± 0.21	3.39 ± 0.22

Table 5

Mean Square Errors and F-test Results for Experiment 4 Post-Hoc Analyses

		Subjects		Items	
Gate (ms)	**Source of Variance**	**MSE**	**F_1**	**MSE**	**F_2**
50	Pragmatic bias	.219	49.00**	.208	6.29†
	Verb bias	.117	23.83**	.208	1.37
	Pragmatic × Verb	.392	5.02*	.208	< 1
100	Pragmatic bias	.265	34.40**	.268	4.65†
	Verb bias	.079	1.46	.268	< 1
	Pragmatic × Verb	.101	1.73	.268	< 1
250	Pragmatic bias	.237	9.98**	.047	7.30*
	Verb bias	.237	25.83*	.047	18.77*
	Pragmatic × Verb	.160	< 1	.047	< 1
300	Pragmatic bias	.497	3.59†	.138	1.54
	Verb bias	.577	6.07*	.138	3.68
	Pragmatic × Verb	.158	9.16**	.138	1.12
450	Pragmatic bias	.402	63.38**	.149	27.21**
	Verb bias	.375	< 1	.149	< 1
	Pragmatic × Verb	.439	2.80	.149	1.99

Note: Degrees of freedom (*df*) for all effects = 1. Error *df* for all Subjects analyses = 13; Error *df* for all Items analyses = 4. '*' indicates significance at or below 0.05, '**' significance at or below 0.01, '†' marginal at or below 0.10.

Figure Captions

Figure 1. 4 of the 20 steps used in Experiment 1 from most /ʃi/-like (step 1) to most /hi/-like (step 20), constructed by resynthesizing two naturally occurring tokens combined at varying intensities

Figure 2. Effect of lexical status on reported phoneme category (Experiment 1)

Figure 3. First and last steps along the /ʃi/~/hi/ continuum that were used in Experiments 2 and 3

Figure 4. Effect of gender bias on word recognition, broken down by verb bias (Experiment 2)

Figure 5. Norming results for Experiment 3 sentence completions. Bar labels mark verb bias and pronoun referent (e.g., 'subjBias.subj' refers to sentences containing a subject-biased verb in the first clause with a subject-referring pronoun in the second clause).

Figure 6. Effect of pragmatic bias on word recognition, broken down by verb bias (Experiment 3)

Figure 7. Time course of example stimulus showing position of the 5 gates

Figure 8. Effect of pragmatic bias in gating task (Experiment 4)

Figure 1

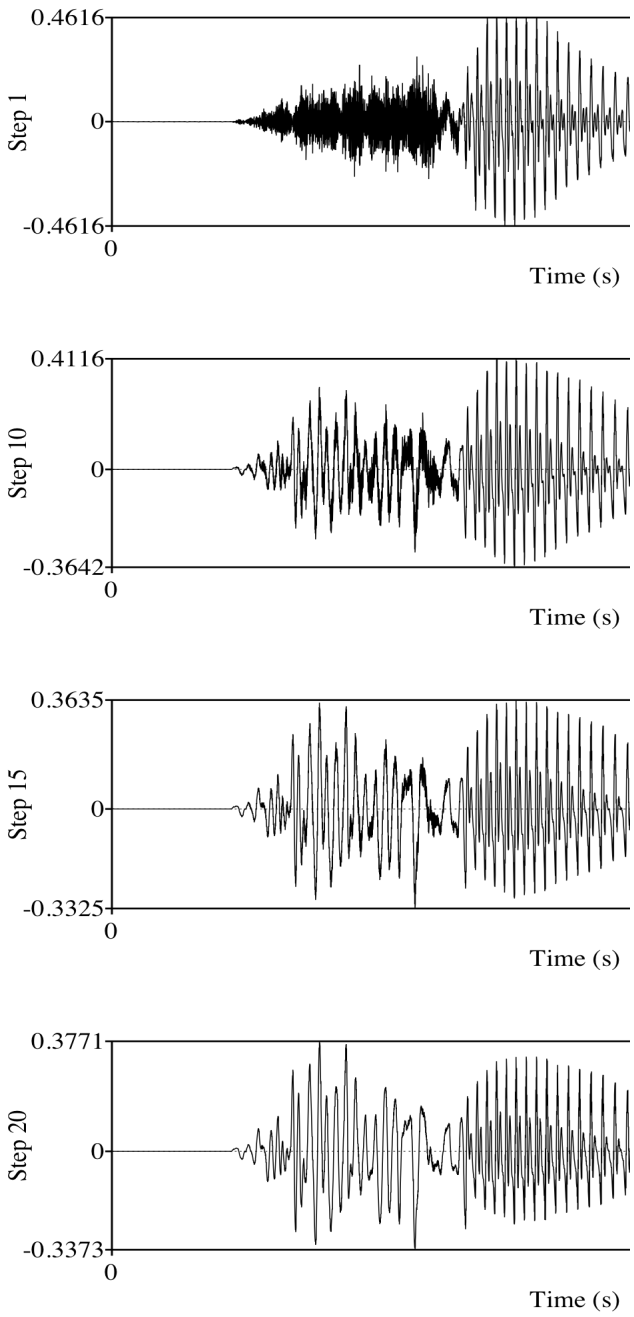

0.4616
Step 1
0
-0.4616
0
Time (s)
0.4116
Step 10
0
-0.3642
0
Time (s)
0.3635
Step 15
0
-0.3325
0
Time (s)
0.3771
Step 20
0
-0.3373
0
Time (s)

Figure 2

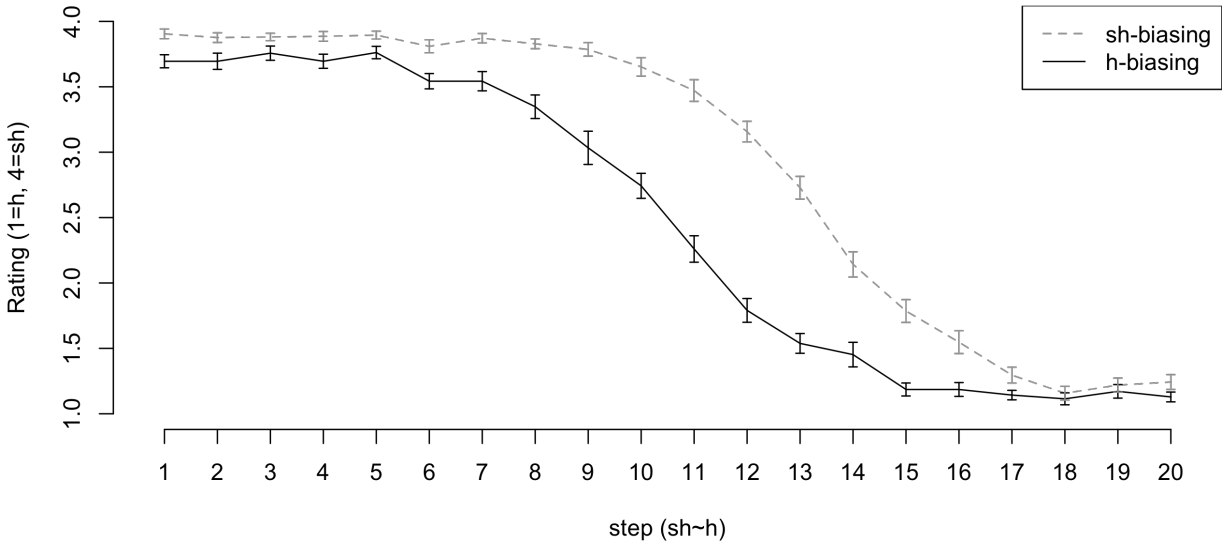

sh-biasing
h-biasing
Rating (1=h, 4=sh)
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
step (sh~h)

Figure 3

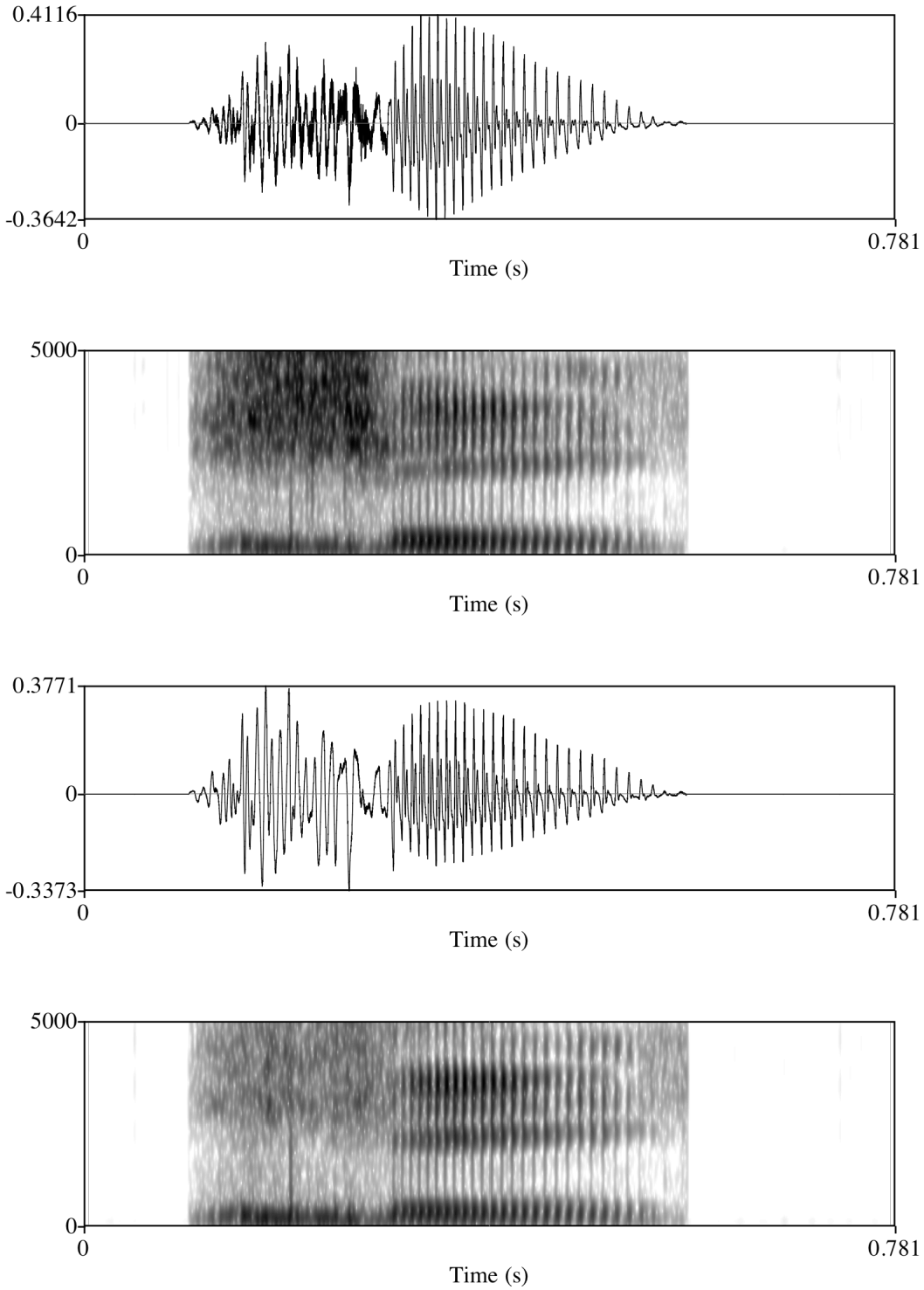

0.4116
0
-0.3642
0
0.781
Time (s)
5000
0
0
0.781
Time (s)
0.3771
0
-0.3373
0
0.781
Time (s)
5000
0
0
0.781
Time (s)

Figure 4

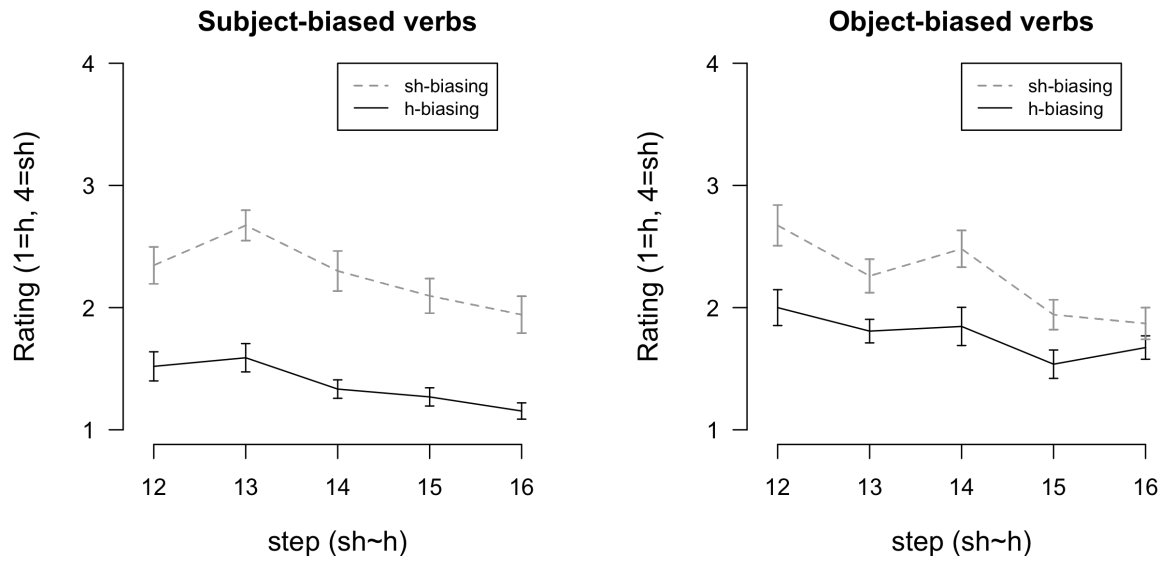

Subject-biased verbs
Object-biased verbs
sh-biasing
h-biasing
Rating (1=h, 4=sh)
step (sh~h)
1
2
3
4
12
13
14
15
16

Figure 5

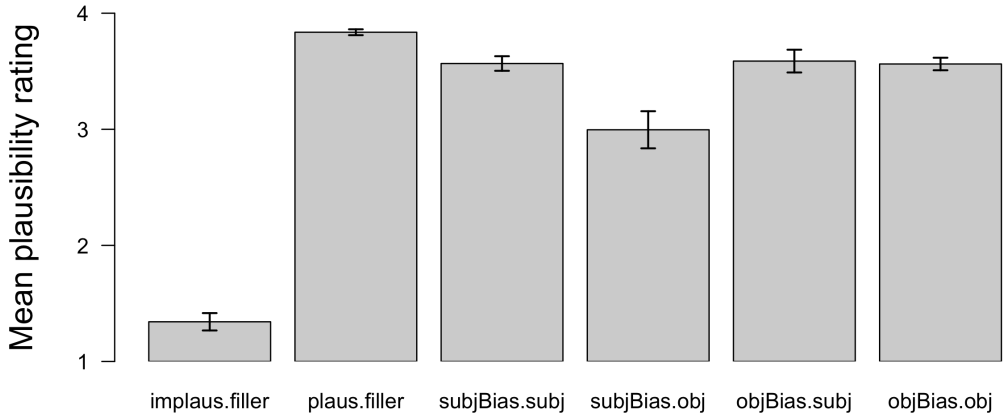

Mean plausibility rating
4
3
2
1
implaus.filler
plaus.filler
subjBias.subj
subjBias.obj
objBias.subj
objBias.obj

Figure 6

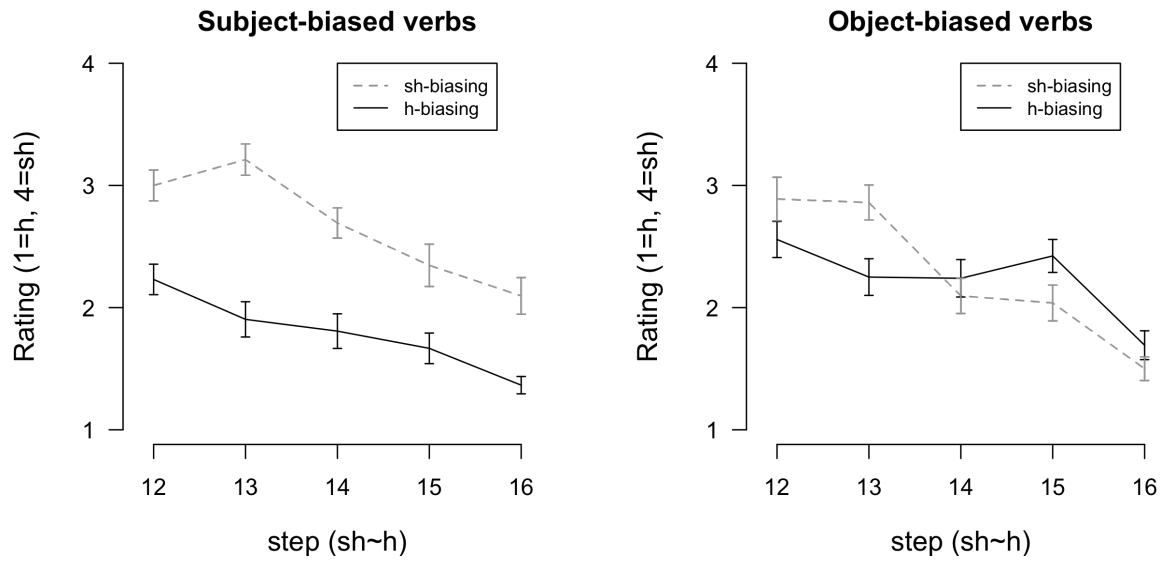

Subject-biased verbs
Object-biased verbs
sh-biasing
h-biasing
Rating (1=h, 4=sh)
step (sh~h)
1
2
3
4
12
13
14
15
16

Figure 7

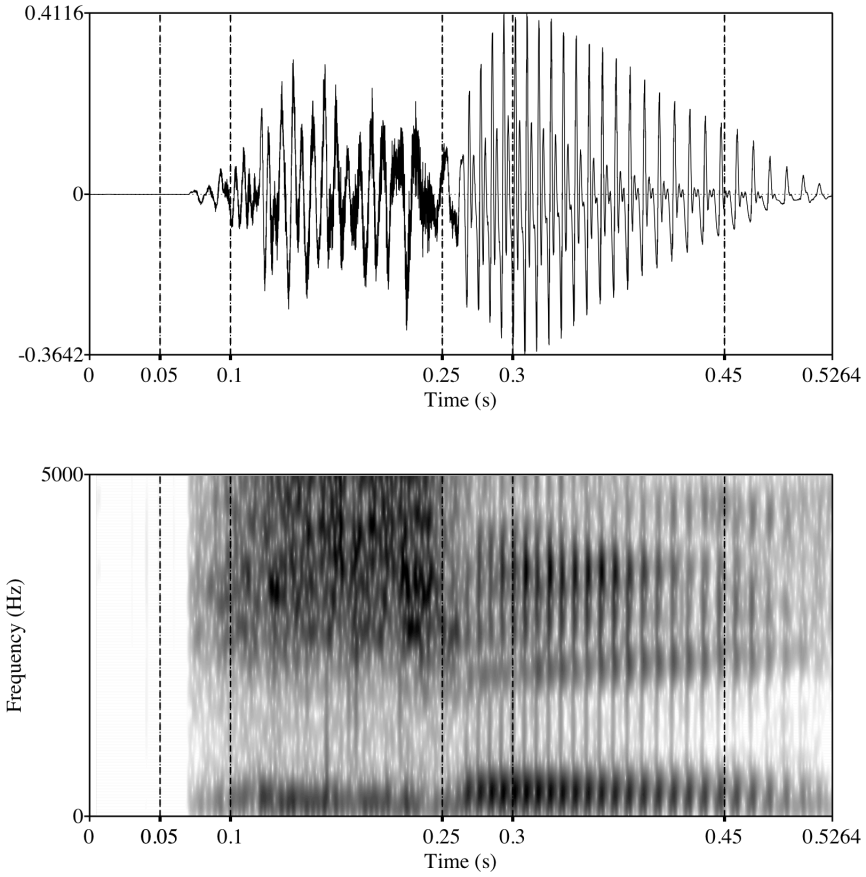
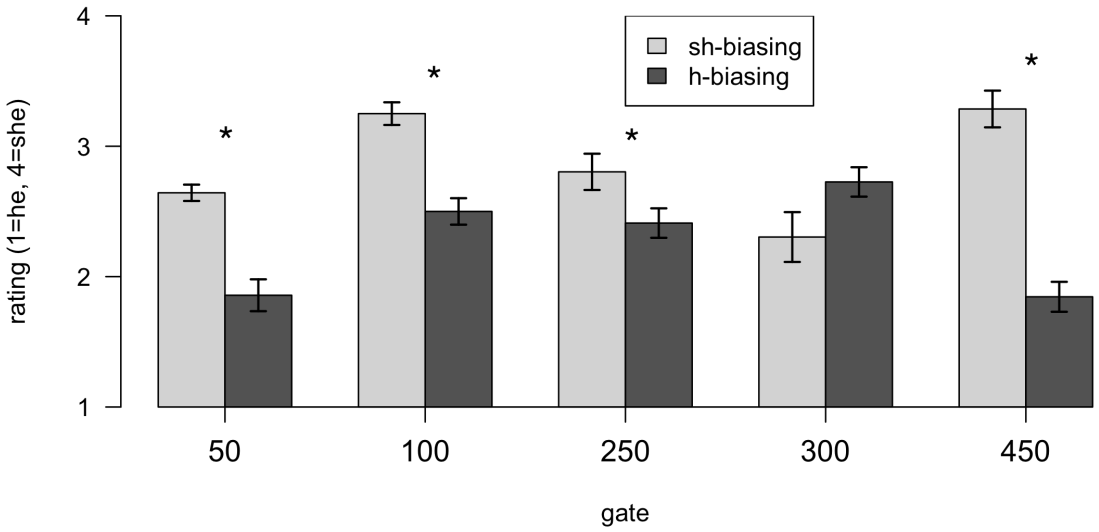
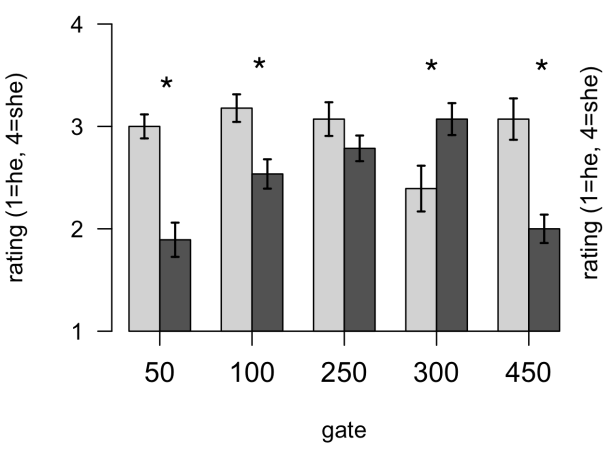

0.4116
0
-0.3642
0
0.05
0.1
0.25
0.3
0.45
0.5264
Time (s)
5000
Frequency (Hz)
0
0
0.05
0.1
0.25
0.3
0.45
0.5264
Time (s)

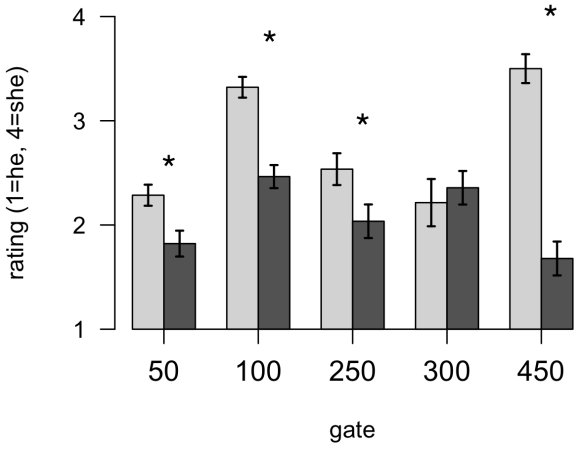
Figure 8



(a) Ratings collapsed across verb type



(b) Ratings for subject-biased verbs



(c) Ratings for object-biased verbs