CHAPTER EIGHT

Alternatives and Inferences in the Communication of Meaning

Hannah Rohde[*,1] **and Chigusa Kurumada**[§]

[*]Linguistics and English Language, University of Edinburgh, Edinburgh, United Kingdom
[§]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, United States
[1]Corresponding author: Email: hannah.rohde@ed.ac.uk

Contents

Abstract

Communicating meaning is a primary goal of everyday language use. One significant puzzle in the study of linguistic comprehension is that the meaning communicated via language often goes beyond an aggregate of word meaning; comprehenders systematically derive pragmatic—as opposed to lexical or semantic—interpretations by leveraging their extralinguistic contextual knowledge. Key to this process is the insight that comprehenders compare *what was said* against *what could have been said*—the range of alternative messages and alternative utterances—to enrich their inferences about the speaker's intention behind a given choice of the linguistic signal. In this article, we offer a conceptual framework for investigating the roles of alternatives and how they are used in comprehension. The case studies we present span multiple levels, with the common thread that their surface forms fail to uniquely disambiguate the underlying meaning: turn-taking, disfluencies, intonational phonology, and pronoun resolution. We discuss Bayesian statistical inference as a possible computational-level approach to rationally combining assumptions about alternative forms and alternative meanings for pragmatic comprehension.

Psychology of Learning and Motivation, Volume 68
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2018.08.012

Human language permits speakers and listeners to communicate information about the state of the world, which includes not only what is believed to be true on the part of the speaker but also what is desired or intended as well as counterfactually hypothesized. A fundamental problem in language comprehension arises because speakers do not, and cannot, make explicit every detail of their message. Much is left unsaid, and listeners need to use contextual information (e.g., mutual knowledge, visually available information, prior experiences) to infer the speaker's intended meaning, given what was explicitly said.

Past research in linguistics and psycholinguistics offers the insight that listeners make principled inferences about speakers' intentions based on the assumption that a given linguistic output is selected among possible options warranted in context. Let's take an example (Grice, 1975; Recanati, 1989).

(1) Sam: How many kids do you have?
Nick: Two

In (1), Nick's answer "Two" usually implies that he has no more or fewer than two children, even though his answer is technically true if he has more than two children.[1] This is explained by positing that listeners expect speakers to be generally truthful and cooperative, providing only necessary and sufficient linguistic information given an intended meaning (Grice, 1975). If instead Nick needs to communicate that he has three children, "Three" would have been more informative, better allowing Sam to infer the intended meaning. In other words, the selection of the word "Two" over "None", "One", or "Three" strengthens the interpretation that Nick has *exactly* two children. In fact, such strengthening of meaning via consideration of the context of alternatives occurs regularly. Consider another example.

(2) Neel Kantha – This is a temple which is dedicated to Lord Shiva. Situated at a height of 1700 metres, it takes about 4 hours on foot to reach from Rishikesh. (http://www.himalaya2000.com/)

To many of us, this description implies that walking is the only means to access Neel Kantha. You would find the description to be inaccurate or at

[1] In a semantic analysis, "Nick has three children" entails "Nick has two children" because in all possible worlds in which the former is true, the latter is also true.

least misleading if you later learned that buses and cable cars were also available (even though the 4-hour estimate for walking will remain valid). This is because, as in the case of (1), we expect speakers and writers to be appropriately informative and not to arbitrarily leave out relevant facts. If there were in fact buses and cable cars, the description should have mentioned the relevant alternative options.

What could have been said, as opposed to *what was said* is often called an "alternative". A number of studies have portrayed the ability to derive and enumerate the right set of alternatives as a critical step towards arriving at an intended meaning that is never directly observed (e.g., Degen, 2013; Grice, 1975, 1989; Hirschberg, 1985; Horn, 1972; Jäger & Franke, 2014; Levinson, 2000; Potts, 2005; Sauerland, 2012). A question remains, however, with respect to how various alternatives are evoked and combined to constrain listeners' inferences. In particular, as we discuss below, previous research does not always make a clear distinction between listeners' expectations about two levels of alternatives: a) **alternative linguistic forms** that can be defined at multiple levels of a language's inventory of phonemes, morphemes, lexical items, and syntactic structures (e.g., using the lexical item "Two" instead of "One" or "Three" in (1)); and b) **alternative intended meanings** given a context (e.g., Nick having {ZERO, ONE, TWO, THREE, …} children). In the case of the Neel Kantha example in (2), how is it that listeners compute from the form of the passage (i.e., the mention of how long the journey would take on foot) that there is only one way to the temple (among the context of possible speaker intentions {FOOTPATH ONLY, FOOTPATH OR BUS, FOOTPATH OR BUS OR CABLE CAR, …})? Here we argue that, to adequately account for listeners' understanding of speaker intentions, one must avoid conflating these two types of alternatives and instead specify how linguistic comprehension is best understood as a formal inferencing problem.

It turns out to be useful to explain listeners' inferential decoding in terms of their model of a speaker's encoding process. In this way, production and comprehension are inherently linked as two interdependent sides of the communication exchange. This approach contrasts with many models in psycholinguistics, where the tendency is to target one phenomenon or one side of the production/comprehension divide—for example, production models (e.g., Bock & Levelt, 1994) may specify sophisticated speech encoding processes whereby speakers select their messages and choose appropriate linguistic forms, but no claims are made about whether comprehenders capitalize on their knowledge of this system in order to decode the linguistic forms they hear; likewise, comprehension models

that rely on competition/integration (MacDonald, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey & Tanenhaus, 1998) characterize a listener who can integrate a wide variety of cues from context, but the enumeration of that weighted list of cues does not satisfactorily explain how such cues reflect a speaker's choices of particular forms and meanings. In what follows, we elaborate on the notions of alternative forms and meanings, and we show how the two combine in a structured inferential process that serves to link a number of different phenomena.

1. TWO TYPES OF ALTERNATIVES

To convey meaning in a communicative act is to select, as a speaker, an appropriate form to use. The basic assumption common among many theories of language production is that speakers begin this process by formulating their messages (intended meanings) and then selecting linguistic encoding options offered by the grammar of a given language (surface forms; Bock & Levelt, 1994). They assemble conceptual elements (e.g., lemmas) and build structured representations based on syntactic and phonological rules. The possible phonemes, lexical items, and part-of-speech categories are taken to be largely fixed for a given language, and the speaker's choice of an appropriate formulation of their intended meaning is a selection out of the set of available forms. On the other hand, the listeners' job is often thought of as reversing this mapping process: They reconstruct the intended meaning by mapping the observed signal onto increasingly more abstract linguistic representations.

However, the mapping between meanings and forms is not one-to-one; ambiguity is rife, both from the reuse of certain forms within the linguistic system and from the noisy environment in which language is produced and interpreted. What this means is that a listener who encounters an ambiguous form can use what they know of the set of possible alternatives to help recover the speaker's intention. An ambiguous acoustic element can be mapped to a nearby phoneme category; a word is expected to map to one of the available entries in the lexicon; a string of words can be parsed to fit one or more structures that are licensed by the grammar. At the pragmatic level, this context of alternatives has been invoked for analyzing words that participate in scales, along which they are ordered in terms of their evaluative meanings, likelihoods of occurring, and specificities of descriptions (Horn, 1972). It has been hypothesized that listeners recover the intended meaning of a word like "good" by calling to mind a relevant scale (see (3–4)).

(3) Sam: How was your trip to NYC?
Nick: It was good.

(4) BAD < OKAY < GOOD < GREAT < EXCELLENT

A word like "good" is associated with some meaning that is present whenever it is used, namely the meaning NOT BAD and perhaps ABOVE AVERAGE, but it also yields some ambiguity: Its meaning can be associated with any degree of goodness that is not bad (GOOD AND POSSIBLY GREAT AND POSSIBLY EVEN EXCELLENT), or else it can be used to signal a degree of goodness that is not bad but is also no more than good (GOOD BUT NOT GREAT). Given the existence of a scale of alternative meanings like (4) and an ability to reason about cooperative communication (Grice, 1975) (as illustrated in (1) and (2)), a listener who hears something described as "good" is permitted to draw an inference as follows: Had the speaker intended a meaning of GREAT, they would have used that stronger word "great", but they didn't, so the speaker's use of "good" can therefore be enriched to mean GOOD BUT NOT GREAT.

This long-standing approach to the derivation of scalar implicatures has been influential and productive in motivating a great deal of theoretical and empirical work on pragmatic language comprehension. However, it typically sidesteps the questions of where the set of alternative forms comes from and how many there are, problems that have an analogue when one asks what set of candidate meanings a listener must consider. For the interpretation of "good", the two meanings GOOD AND POSSIBLY GREAT and GOOD BUT NOT GREAT are taken to be equally relevant meanings, and it is a computation involving the context of available forms that allows the listener to decide between them.

But not all intended meanings are equally likely. The activation of, and selection among, alternative meanings has been the purview of an increasing number of psycholinguistic studies that try to pin down how listeners use available cues from the context of the utterance to estimate what meaning the speaker likely intends. Candidate lexical items (and the meanings presumably associated with them) are activated based, for example, on the available objects in a visual scene (Altmann & Kamide, 2007; Spivey, Grosjean, & Knoblich, 2005), by a primed concept (Neely, 1977), or more generally by appeal to coherence-driven reasoning and real-world knowledge (Hagoort & van Berkum, 2007; Köhne & Demberg, 2013; Xiang & Kuperberg, 2015). From the perspective of incremental and predictive language processing in general, listeners are posited to assign

different weights to candidate lexical items either at the point when they hear a word or even before the word is encountered if they are anticipating how a sentence will continue (Altmann & Kamide, 1999; Brown-Schmidt & Tanenhaus, 2008; Federmeier, 2007; Kuperberg & Jaeger, 2016; Levy, 2008). Much is still unknown about how such parsing and processing behavior might interact with the time-course of derivation of alternatives and pragmatic comprehension based on them.

Another question concerns what representations must be included as members of alternatives. Extant theories often take lexical items as primitives of alternatives. In our daily language use, however, a wider range of elements are often considered and compared against each other. To illustrate, let's think about another example. Imagine Bob and Sarah are sitting together, and Bob says "I love you". Now, in a canonical conversational context, Sarah is obliged to respond, indicating whether she reciprocates the sentiment or not. In other words, the two relevant alternative meanings on the table are LOVE and ¬LOVE ("not" Love). (There are probably more subtle variants in between, but let's stick to those as the two possible states of Sarah's intentions for the purpose of this discussion.)

As for the relevant forms that Sarah's response can take, it is easy for her to choose a response if the intention is LOVE: She can simply reuse Bob's words to utter "I love you too". It is much harder to select an appropriate response if Bob's affection is not reciprocated. If Sarah needs to convey ¬LOVE, there are in fact many things besides "I don't love you" such as "I'm sorry", "well, actually I was meaning to say …", "it's complicated", etc., some of which could potentially be ambiguous. One can even imagine a situation in which a slight (e.g., less than a second) delay in Sarah's response would already convey some of the answer, via an inference that if someone's intended meaning were LOVE, they would probably be less likely to hesitate in generating their response. If alternatives are defined as possible options that the speaker could have produced, these linguistic and non-linguistic options beyond word choice may need to be considered as relevant alternatives.

A main goal of this paper is to propose a conceptual framework of alternatives that can encompass various aspects of language use that contribute to the pragmatic communication of meaning. Here let us introduce our general approach by elaborating on this example of conveying LOVE and ¬LOVE. Fig. 1 shows a schematic representation of this generative process—one that will be expanded on for the case studies in the upcoming sections.

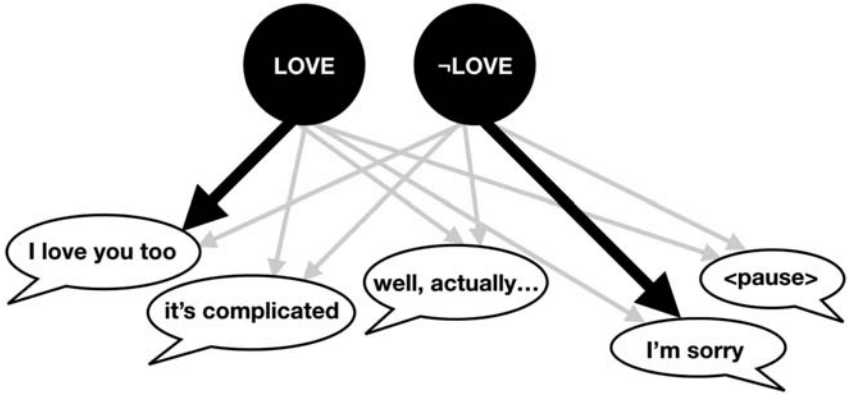

Figure 1 Graphical model of utterance production, showing two underlying meaning states LOVE and ¬LOVE and a sample of possible output forms.

The intended meaning states LOVE and ¬LOVE are shown to be linked to a set of possible surface forms, including the aforementioned hesitation (indicated as <pause>). In the figure, the strength of the form ~ meaning links is indicated by the darkness and width of the arrows, so that, for example, "I love you too" is the most likely utterance given that Sarah's intention is LOVE, whereas "it's complicated" is much less likely. In the model, every form is potentially associated with both underlying meaning states. This allows some probability (albeit near-zero in some cases) to be assigned to the generation of each of the outputs—speakers in LOVE may sometimes fail to articulate that meaning with the best option, just as speakers in ¬LOVE have been known to say all kinds of misleading or awkward things.

Fig. 1 drives home the point that there are two distinct levels of alternatives. The first is concerned with alternative intended **meanings**, represented here with the two black circles. One can think of them as alternatives at the mental state or proposition level, which are abstracted out from any specific linguistic encoding options. The speaker is postulated to be in one of those meaning states amongst the set of alternatives and to then encode that state linguistically. The other level pertains to alternative **forms**, illustrated here as utterances in speech bubbles. Speakers are taken to select linguistic encoding options (e.g., lexical lemmas, grammatical functions, syntactic constructions, phonological forms, phonetic realizations) that best represent their intended meanings. The listener is in turn tasked with leveraging the produced form to make an inference about the intended

meaning. In this sense, the process is generative.[2] That is, listeners have a mental model of the process that *generates* the observed linguistic signal. We consider alternative meanings and alternative forms to be variables in this mental model at two different levels: *1) possible meaning options that generate the linguistic signal* and *2) possible signals that can be generated by these meanings*.

This distinction between alternative intended meanings and alternative forms is not always made salient in a discussion about a conventional scale such as GOOD < GREAT < EXCELLENT. This is because the meaning-level alternatives (the concepts of GOOD < GREAT < EXCELLENT) and the linguistic level alternatives (the words "good", "great", "excellent") appear to exhibit close correspondences, at least semantically. However, it is more the norm than the exception in language comprehension that meaning-level alternatives and form-level alternatives yield many-to-many relationships, like the example illustrated in Fig. 1. That is, one meaning can be encoded in multiple alternative forms (including an absence of a form, i.e., pause), and one form can be associated with multiple meanings. The listener therefore needs to navigate the resulting uncertainty to arrive at the most likely intended meaning given an observed form.

In what follows, we explore how the listener may infer a speaker's intended meaning by combining their assumptions about alternative meanings and alternative forms. To do so, we link multiple phenomena not typically analyzed together: filled and unfilled pauses (Section 2), intonational phonology (Section 3), and finally pronoun interpretation (Section 4). Seeing these disparate phenomena as similar requires a new lens, one which formally combines expectations at two different levels of representation. As we mentioned above, much of the previous work that discusses roles of alternatives focuses on those at the level of lexical items (e.g., "good", "great" vs. "excellent", or "some" vs. "all"). By departing from these common examples, we aim to illuminate the utility of postulating two levels of alternatives. Unlike the lexical alternatives that have traditionally been examined, some of the alternative forms we discuss (e.g., silence, disfluency) do not have commonly accepted,

[2] Note that the term 'generative model' has had a long and multi-sided history in the field. We intend it here in the sense of a mental model of the speaker's production process that defines how different possible speaker outputs can be *generated*. Generative Grammar as a conceptual framework was built on similar principles, but the specifics of that framework extend far beyond the underlying sense of 'generative' intended here.

conventionalized, meanings. Rather, listeners actively need to seek possible and likely meanings based on their contextual knowledge and infer how a perceptible form (or a lack thereof) might have been produced (generated) among other alternative forms for a particular intended meaning. Most researchers would acknowledge that context has a role to play, but what is new in this approach is the specification of which contextual features affect which linguistic calculations. Our approach uses an inferential architecture that makes specific predictions about what features can be manipulated with what systematic effects on interpretation. Finally, in Section 5, we lay out a computational-level inference framework for combining the listener's assumptions about alternative meanings and alternative forms.

2. WHEN NOTHING SAYS SOMETHING: ROLES OF ALTERNATIVE FORMS

Let's go back to the LOVE/¬LOVE example discussed above. We mentioned the intuition that a delayed onset of a response can bias the listener's interpretation of the speaker's state to ¬LOVE. How does this interpretation come about?

One important fact about conversation is that contributions from multiple parties are usually very tightly connected, not only in their contents but also in their timing. In a face-to-face conversation between adults, the average gap between two parties' turns is often 200 ms or shorter, with only rare overlap between turns (Casillas, 2013; Levinson & Torreira, 2015; Stivers et al., 2009; de Ruiter, Mitterer, & Enfield, 2006). A deviation from this can be registered as too early or too late, as seen in N400 effects, which are normally associated with contextual expectation violation but disappear with abnormally long pauses (Bögels, Kendrick, & Levinson, 2015). This strong expectation about when to respond creates an arena for an elaborate dance between conversational partners. The speaker must hold the floor while it is still her turn and signal whether and when to end her turn (Sacks, Schegloff, & Jefferson, 1974). The listener, on the other hand, needs to plan her own response so as to jump in as soon a transition of turns is indicated. A skilled speaker can take advantage of this strict temporal coordination to achieve discourse functions. A pregnant pause builds up suspense because the longer-than-expected silence (e.g., "Well, I meant to … tell you … ') forces the listener to wait while feeling compelled to pick up a turn and start talking.

This strong pressure for temporal coordination can turn very subtle signs — including ones produced inadvertently — into alternative forms to communicate pragmatically enriched meanings. Studies on speech disfluency have elegantly explained how this likely occurs. When a speaker experiences difficulty in planning or producing an utterance, her speech may become less fluent. Two main types of disfluencies that have been investigated so far are unfilled pauses (i.e., silence) and filled pauses such as "uh" and "um". Clark and Fox Tree (2002) were among the first to point out that these disfluencies can be produced and employed by the speaker in a systematic manner, just like any other lexical items. In their account, the speaker produces "uh"s and "um"s before items that are difficult to retrieve or produce such as words that are rarely used or those that have not been introduced into discourse yet (see also Brennan & Schober, 2001; Brennan & Williams, 1995; Clark, 1994; Clark & Wasow, 1998; Fox Tree, 1995; Fox Tree & Clark, 1997; Levelt, 1989; Shriberg, 1996; Smith & Clark, 1993). (For other researchers, the production of unfilled pauses is less strategic and need not be based on the speaker's conscious decision to signal their production difficulty e.g., Bock, 1987; Barr, 2001; Bortfield, Leon, Bloom, Schober, & Brennan, 2001; Lake, Humphreys, & Cardy, 2011; Oviatt, 1995).

Listeners, on the other hand, evaluate these disfluencies against their expectations about other (alternative) forms of utterances the speaker could produce. Let's think about the case in which Sarah fails to respond to Bob's declaration of his love within an expected time window (i.e., <pause>). In this case, the delayed onset is compared against possible alternative forms of production such as those in the speech bubbles in Fig. 1. Unlike the case of evaluative adjectives (e.g., "good" vs. "great") that comprise a lexically defined scale, interpretations of a delayed onset do not depend on a clearly ordered set of alternative forms and their associated semantic meanings. Rather, these alternative meanings are socially and situationally defined: Responses are ordered as positive (affirmative, preferred) and negative (contradictory, dispreferred) in terms of the expectation displayed in the initiating action (Kendrick & Torreira, 2015; Levinson, 1983). An optimally-timed response onset is normally associated with a positive response (e.g., LOVE), and an alternative form is associated with a negative response, one that is socially more costly and therefore potentially more difficult to produce (e.g., ¬LOVE). These situationally evoked and constrained ordering of alternatives (both of forms and meanings) are often discussed as ordered on an "ad-hoc" scale as opposed to a lexically defined scale (Stiller, Goodman, & Frank, 2015).

Bob's utterance,"I love you", thus creates an expectation that the LOVE meaning is the probable and expected meaning for Sarah as well, and if Sarah is to convey that meaning, she should be experiencing no difficulties in doing so. A delayed onset contrasts with that scenario, and thereby triggers an inference that Sarah must be conveying an alternative meaning that causes production difficulties. The basic machinery of that inference can be explained in a similar manner to the case of GOOD < GREAT < EXCELLENT. Observing a given form (e.g., "good") raises a question about why the speaker did not produce other options (e.g., "great"), which leads to an inference that the meaning of GREAT was probably not warranted.

Now let's consider the case where Sarah says "well …" or "uh", which can be taken as her attempt to begin her response (perhaps to avoid inferences associated with a long silence) but to not yet go on record with a committed answer (Fox Tree & Clark, 1997). In a manner similar to what is discussed above for unfilled pauses, these forms are compared against other linguistic items that could be produced in this context. If Sarah were trying to convey the positive, socially more expected meaning (LOVE), there is a form that is readily available (i.e., "I love you, too"). Filled pauses again indicate a deviation from these expected forms, which triggers the inference that Sarah's attempt to articulate her meaning is causing production difficulty, and such difficulty is often attributed to production of linguistic elements that are new, infrequent, or difficult to articulate (e.g., Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Arnold, Hudson Kam, & Tanenhaus, 2007; Barr & Seyfeddinipur, 2010; Loy, Rohde, & Corley, 2017). In this case, it likely leads to an inference that Sarah is in a dispreferred meaning state, i.e., ¬LOVE.

Existing studies on comprehension of disfluencies demonstrate that listeners routinely carry out these inferences in their real-time language comprehension (Arnold et al., 2007, 2004; Bögels et al., 2015; Corley, MacGregor, & Donaldson, 2007; Heller, Arnold, Klein, & Tanenhaus, 2015; Loy et al., 2017). Perception of disfluencies can immediately impact their predictive processing, evoking otherwise less probable or expected elements in a current context. These inferences offer an important insight into theories of alternatives in communication of meanings. That is, alternative meanings and forms are not necessarily static representations stored as part of our linguistic knowledge. Rather, they are emergent constructs in context. Listeners derive expectations about possible and likely meanings according to their assumptions about the context, conversational partners and common ground shared with them (Clark, 1996, 2015; Clark, 2015), prior linguistic experiences, and what is socially and communicatively licensed and preferred. Similarly,

the linguistic signal (as well as an absence of it) is evaluated against contextually-constrained expectations about what linguistic encoding options are more or less likely in the context at hand. This resonates with the past 20 years of psycholinguistic research, suggesting that listeners are constantly predicting upcoming input so as to process the linguistic signal despite its fleeting and ambiguous nature (Altmann & Kamide, 1999; Brown-Schmidt & Tanenhaus, 2008; Federmeier, 2007; Kuperberg & Jaeger, 2016; Levy, 2008).

Recent investigations into real-time processing of filled pauses show that listeners are capable of calibrating their expectations at an incredibly fine-grained scale. As we discussed above, an instance of a disfluency such as "uh" or "um" triggers anticipation for a referent that is new in discourse or hard to label for other reasons. However, when listeners are told that the speaker has object agnosia and thus experiences difficulty naming ordinary objects, they do not show a bias toward objects lacking a conventional name when they hear a speaker's disfluency (Arnold et al., 2007). A similar effect of an explicit instruction was observed even with young children, who were told that the speaker is particularly forgetful (Orena & White, 2015). Additionally, listeners show sensitivity to the identity of the speaker in determining which objects have already been referred to, with disfluency biasing toward objects that have not been referred to by the particular speaker (Barr & Seyfeddinipur, 2010; Yoon & Brown-Schmidt, 2016). These results suggest that listeners condition their expectations on factors at the speaker level, fine-tuning their inferences about intended meanings given the specificities of the speaker. This likely includes causal reasoning about why the speaker may be producing disfluencies (Heller et al., 2015). An instance of disfluency produced by a non-native speaker does not trigger an inference about the referent's discourse status or lexical frequency in the same way as when it is produced by a native speaker, as the same assumptions do not hold (Bosker, Quené, Sanders, & de Jong, 2014; see also King, Loy, & Corley, 2018 for other context-driven adjustments given an alternative explanation for observed disfluency). Similar types of speaker-based calibration are possible for pragmatic interpretations of unfilled pauses. For instance, silence of a given duration (e.g., 600 ms) can be considered long for adults but short for young children (Casillas, 2013; Casillas, Bobb, & Clark, 2016; Stivers, Sidnell, & Bergen, 2018) or long in a casual, face-to-face conversation but short in a long-distance phone call. Thus deriving alternative meanings and alternative forms often requires situationally calibrated expectations about interlocutors and their language use, as well as general knowledge about how various meanings are encoded by linguistic alternatives.

3. INTONATIONAL PHONOLOGY: ROLE OF ALTERNATIVE MEANINGS

From the viewpoint of information processing, listeners' expectations for alternative forms play an important role in most domains of speech comprehension. The speech signal is produced and perceived in the presence of noise, and the same physical signal can be mapped onto different linguistic representations (e.g., phonemes) depending on speakers and situations (Liberman, 1960). Listeners therefore need to integrate multiple sources of information to anticipate which forms are more or less likely to be observed in a given context (Cutler, 2015). For instance, perception of a phoneme (e.g., /b/) occurs through integration of information about other, similar-sounding, phonemes (e.g., /p/, /d/, /v/), general knowledge of phonology and phonotactics (e.g., /b/ is more frequent than /v/ in English), distributional patterns of acoustic cues over phoneme categories (e.g., voice onset time), as well as contextual information (e.g., the target sound appears in a word "bathtu__", which makes /b/ more likely than other alternative sounds) (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Ganong, 1980; Kleinschmidt & Jaeger, 2015; Kraljic, Brennan, & Samuel, 2008; McMurray & Jongman, 2011; Norris, McQueen, & Cutler, 2003, 2000; Pisoni & Levi, 2007; Samuel, 2001). A number of knowledge sources can thus guide listeners' expectations about possible and likely alternative forms.

Listeners' knowledge and expectations for alternative meanings are often discussed as top-down information that constrain their hypotheses about the speaker's intended forms. The significance of top-down information is best appreciated when alternative forms do not form a closed, easily enumerable class (cf. phonemes). Speech prosody (intonation, in particular) can provide one such case. In English and other languages, stress, accents, and other forms of prosodic emphasis can highlight a linguistic element of importance in semantic and pragmatic communication (e.g., Cruttenden, 1997; Dahan, 2015; Ladd, 2008; Selkirk, 1996). For instance, the utterance "she LOVES you" can be used to convey "she LOVES you, not just LIKES you" while "she loves YOU" can convey "she loves YOU, not SOMEONE ELSE"(Pierrehumbert & Hirschberg, 1990). The speaker can modulate various prosodic cues (e.g., pitch, duration, intensity, loudness) to highlight elements in focus while backgrounding other elements in an utterance. The domain of research that concerns ways in which these different cues are combined to mark different informational statuses is called intonational phonology.

A perennial question in the domain of intonational phonology pertains to mappings between prosodic cues such as perceived pitch contour and intended meanings. Many researchers have largely agreed that speakers make systematic uses of prosodic cues to mark intonational categories, and these categories indicate how a given linguistic element (e.g., a word) should be interpreted in relation to information and assumptions in context (e.g., Beckman & Pierrehumbert, 1986; Calhoun, 2006; Cutler, Dahan, & Donselaar, 1997; Ladd & Morton, 1997; Ladd & Schepman, 2003; Pierrehumbert & Hirschberg, 1990; Watson, Tanenhaus, & Gunlogson, 2008). For instance, the type of emphasis used to convey the meaning [X, not Y] (e.g., LOVE, not LIKE) is often called a contrastive pitch accent and is analyzed separately from an emphasis that can be used for items that are simply new in discourse (Calhoun, 2004; Dilley, Ladd, & Schepman, 2005; Ito & Speer, 2008; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990; Silverman & Pierrehumbert, 1990; Watson et al., 2008). New information is often introduced with a heightened pitch peak aligned with the stressed syllable of a word (e.g., "Who does she love?" … "she loves BOB") as schematically illustrated on the left in Fig. 2. The annotation conventions of the ToBI system (Silverman et al., 1992) identifies it as an H* (a high tone aligned with a stressed syllable).[3] On the other hand, the contrastive interpretation (e.g., "Does she love Ted?" … "No, she loves BOB") is often conveyed by a pitch accent that includes a low tone and a high tone aligned with the stressed syllable. It is often annotated as an L + H*, as illustrated in the right figure in 2.

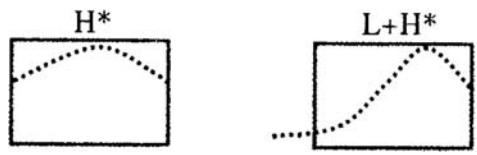

Figure 2 New accent vs. contrastive accent (Adapted from Arvaniti, A., & Garding, G (2007). Dialectal variation in the rising accents of American English, *Papers in Laboratory Phonology, 9.*).

[3] We must note that even a cursory glance at the literature makes it clear that these categories are far from self-evident. Some frameworks of intonational phonology reject the assumption that there are categories that can be determined on the basis of acoustic properties of speech e.g., (e.g., Xu, 2005). Substantial work has been done to explore efficacies and limitations of proposing a particular inventory of intonational categories (e.g. Dilley, Breen, Gibson, Bolivar, & Kraemer, 2012, 2006; Wightman, 2002). An exhaustive discussion of such attempts is beyond the scope of the current paper, but we acknowledge the richness of the literature and the fundamental complexity of the topic.

The ease with which a native listener distinguishes these patterns in their acoustic realizations and their associated meanings belies the complexity of the actual processes involved in comprehending intonational meanings. To date, there is little consensus with respect to how listeners distinguish alternative forms, such as those illustrated in Figure (2) as postulated in theories of intonational phonology. A major source of the difficulty is the fact that actual realizations of the forms (i.e., perceptible acoustic cue distributions) are influenced by many other factors (e.g., phonological properties of words, prosodic features of an utterance, speech rate, emotional states of the speaker, individual differences in pitch height and speech rate, dialectal and sociolectal variations), and they are not deterministically mapped onto the theoretically posited intonational categories. Fig. 3 plots the distributions of fundamental frequency (F0) and segment duration of items that are either contrastive or new, produced in a naturalistic task environment (Buxó-Lugo, Toscano, & Watson, 2018). These two cues are chosen because they are often treated as main sources of information used by listeners to distinguish the two meaning categories (Pierrehumbert & Hirschberg, 1990). Actual distributions of these prosodic cues are, however, completely overlapping with each other, making it unlikely that

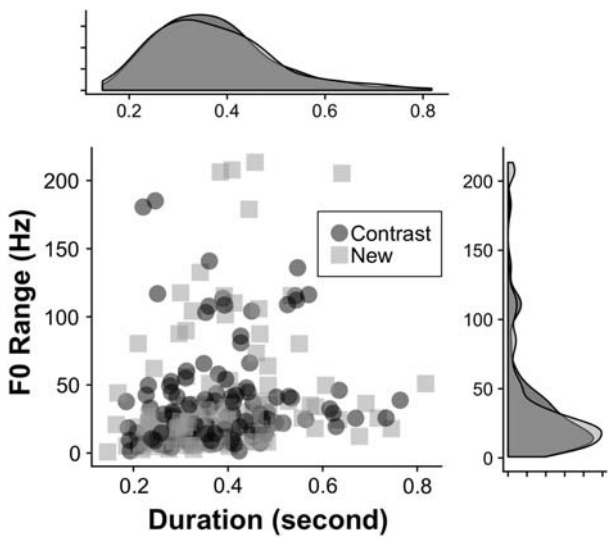

Figure 3 Mean word duration and F0 range of items produced in two different discourse statuses (Contrast (dark gray) vs. New (light gray)), showing large empirical overlap. This data was collected for a recent related study (Buxó-Lugo et al., 2018), re-visualized here courtesy of Andrés Buxó-Lugo.

listeners would arrive at one of these two alternative meanings through the acoustic information alone (without any other normalization schemes).

One potential source of information that can be useful is listeners' expectations about the alternative meanings (Cole, Mo, & Hasegawa-Johnson, 2010; Kurumada, Brown, & Tanenhaus, 2012; Pierrehumbert & Hirschberg, 1990). In a given conversational context, listeners may have a strong expectation for what types of meanings (e.g., new vs. contrast) will likely be conveyed by the speaker. In the example above, the question "Who does she love?" sets up an expectation that a referent introduced in a following utterance will be new information, whereas "she loves BOB" in response to "does she love Ted?" will likely be interpreted as contrastive. That is, listeners can estimate the relative likelihoods with which various alternative meanings may be expressed in a given discourse context. Such expectations for the meanings may, in turn, shape their expectations as to which alternative forms will be used by the speaker.

Kurumada, Brown, Bibyk, and Tanenhaus (2018) tested this prediction, asking how context-driven expectations regarding possible and likely alternative meanings may guide listeners' pragmatic interpretations of speech. The study used the looks-like-an-X construction, as in (5), which can support two interpretations depending on the intonation contour used. Noun-focus intonation puts the emphasis on the final noun (zebra) in (5a) and favors a meaning in which the speaker proffers a guess about the identity of the animal, because the accent placement on the noun works to highlight the relevant set of alternative animals (e.g., "a ZEBRA, not a horse"). On the other hand, contrastive accent on the verb "looks" and the rising intonation at the end of an utterance, as in (5b), can be taken to mean that "it resembles a zebra but it is not one" (Kurumada & Clark, 2016; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014). It has been observed that adult speakers often use such constructions to introduce a new, or less familiar, lexical item in child-directed speech (Clark & Wong, 2002). Perhaps, in the case of (5b), the speaker may be looking at an animal with a zebra-like appearance (Fig. 4, right).

(5) a. "it looks like a ZEBRA!" [Noun focus, proffered guess meaning = ZEBRA]

b. "it LOOKS like a zebra …" [Verb focus, correction meaning = ¬ZEBRA]

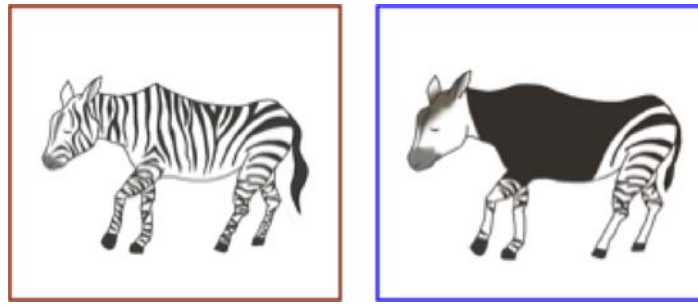

Figure 4 Visual illustrations of intended alternative meanings of (5): a zebra on the left (intended by (5a) "It looks like a ZEBRA") and an okapi on the right (intended by (5b) "It LOOKS like a zebra …").

In a series of experiments, Kurumada et al. (under review) manipulated the intonation contour of looks-like-an-X utterances (noun-focus vs. verb-focus) and the discourse context in which those utterances were embedded (bias to GUESS vs. bias to CORRECTION). In one experiment, they manipulated factors that would affect listeners' expectations for alternative meanings only. Participants were randomly assigned to an Expert condition or a Non-expert condition. In the Expert condition, participants were presented with a cover story in which the speaker is a kindergarten teacher talking to children (an expert who is in a position to plausibly correct mistaken evaluations). In the Non-expert condition, the speaker was introduced as a 14 year-old girl talking to her friend in a science museum (a non-expert who is in a position to plausibly be proffering a guess). Each condition was accompanied by pictures and a short text describing the situation, as in (6–7).

(6) Expert condition

Today a kindergarten teacher Ms. Jones had a story time with her students. She brought a book with a lot of pictures of different kinds of objects and animals. The children are very excited about this book. Every time she turns a page, somebody asks, "What's that?"

(7) Non-Expert condition

Today a group of eighth graders went to a natural science museum. Their assignment was to go through different exhibitions and write a journal about what they saw. Catie and Lauren decided to do the assignments together. There were a lot of strange-looking animals and artifacts they had never seen before, and they kept asking each other, "What's that?"

Participants then heard an identical set of 14 instances of "it looks like an X!" vs. "it LOOKS like an x …" (seven instances each) and provided their interpretation of the speaker's intention by selecting one of the following disambiguated statements regarding the speaker's GUESS, CORRECTION, or uncertainty: a) The speaker is implying "It is an X", b) The speaker is implying "It is not an X", and c) The speaker is implying "I'm not sure". These options were printed in text, in this order, with clickable radio buttons.

As expected, there was a main effect of intonation type such that "it LOOKS like an x …" yielded significantly more CORRECTION [It is not an X] interpretations. There was no significant effect of the Expert conditions. Crucially, however, the interaction between the intonation type and Expert condition was significant (see Fig. 5). Participants were more likely to interpret the verb-focus intonation as a CORRECTION [It is not an X] when the speaker was presented as an expert (i.e., the kindergarten teacher talking to children), who is likely to know an identity of a referent and correct the listener's assumption ("it LOOKS like a zebra, but it is not."). Remember that the audio stimuli presented to participants in the Expert and in the Non-Expert conditions were identical and hence the differences in the interpretations cannot be attributed to the perception of intonational contours (i.e., alternative forms) per se. It is more likely that the preamble and expectations of the speaker's knowledge states made the meaning of

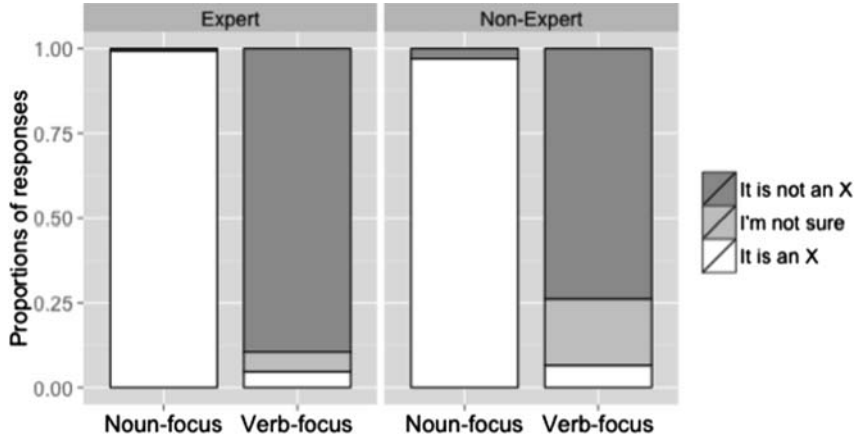

Figure 5 Meaning interpretations across focus and expertise conditions, showing the way context-driven expertise bias modulates the effect of form-driven differences between "It looks like an X" and "It LOOKS like an x" (Adapted from Kurumada, C., Brown, M., Bibyk, S., & Tanenhaus, M.K. (under review). Probabilistic inferences and adaptation in pragmatic interpretation of contrastive prosody.).

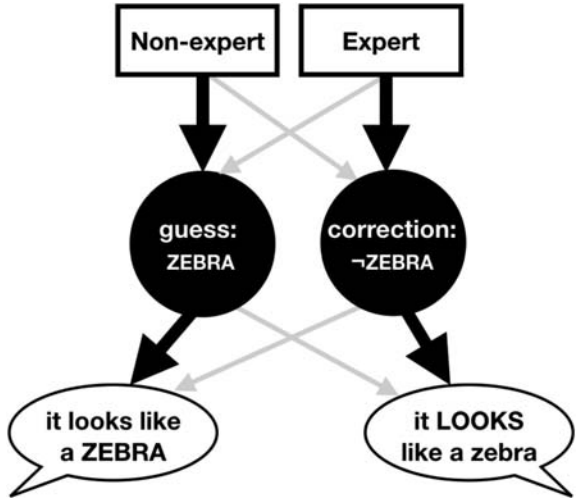

Figure 6 Graphical model of intonation production for "it looks like an X", showing alternative output forms and alternative underlying meaning states with two conditioning contexts.

[it is not an X] more or less likely, which resulted in different interpretations of the physically identical input.

Fig. 6 shows the preferences among possible meanings (conditioned on discourse context) and the preferences among forms (conditioned on intended meaning). Preferences — i.e., favored meanings and their associated forms — are indicated by thicker arrows. These are intended to signal production preferences whereby a teacher is being characterized as more likely to produce the CORRECTION meaning when speaking to students, and a teenager is more likely to produce the GUESS meaning when speaking to a peer. For the choice of forms, we expect the two speakers to possess an equivalent grammar of prosody: A speaker who intends the CORRECTION meaning is posited to favor the "it LOOKS like a zebra" contour, whereas a speaker who intends the GUESS meaning is posited to favor the "it looks like a ZEBRA" tune.[4]

For the listener, the puzzle is how to reverse engineer the likely meaning from the potentially ambiguous tune. The listener encounters, say, an instance of a kindergarten teacher producing "it LOOKS like a zebra".

[4] It is possible that speakers also differ in their preferences for forms in addition to their preferences for meanings. That is, a school teacher, as compared to a teenager, might be more likely to produce the verb-focus contour ("it LOOKS like a zebra") when her intended meaning is CORRECTION.

If the listener's generative model of the speaker's production process looks like Fig. 6, they are aware that this expert speaker is more likely to intend the CORRECTION meaning and that the CORRECTION meaning is more likely to be realized with verb focus (for similar effects of speaker knowledge, see Bergen and Grodner (2012)). In contrast, if the listener believes that a non-expert speaker is less likely to intend the CORRECTION meaning, an outcome of their inference is biased against this meaning even with the same preference at the level of choosing the verb-focus contour given the CORRECTION meaning (see the final column of Fig. 5). The current framework, as graphically illustrated in Fig. 6, can thus capture the long-standing intuition in the field that some of the meanings are *a priori* more or less probable given the listeners' knowledge about the context and that an appropriate integration of such expectations sharpens interpretations of intonation contours (Cole, 2015; Cole et al., 2010; Turnbull, Royer, Ito, Speer, & Turnbull, 2017). Contextually-sensitive weighting of alternative meanings is likely a key to the clear and reliable interpretations listeners can draw from the complex and often ambiguous intonational contours.[5]

Further, when combined with additional computational assumptions, the current approach makes a novel and quantitative prediction as to how these context-based expectations about alternative meanings would influence outcomes of pragmatic interpretations of intonation. Crucially, the current approach posits that listeners' expectations about possible meanings and possible forms are computed as distinct sets of alternatives, and contextual factors can affect them separately. That is, being in a particular context (e.g., listening to a kindergarten teacher who is talking about animals) affects both listeners' expectations about what meanings are more or less likely to be conveyed (alternative meanings) and how each of the meanings might be linguistically encoded (alternative forms). Previous approaches showed in broad strokes that both of these expectations matter in intonational interpretations, but few could articulate a prediction about how they should be integrated with each other

[5] Here we assume at least two types of ambiguity. One pertains to purely perceptual ambiguity resulting from noise in the production and comprehension process. The other pertains to ambiguity in mappings between a perceived intonational contour and a more abstract intonational category: As illustrated in Fig. 3, the same physical contour can be mapped onto two intonational categories (and their associated meanings). In both cases, we believe, listeners may be navigating the ambiguity by adopting systematic inferences along the lines of what we will propose in Section 5. We thank Alice Turk for pointing out these multi-layered sources of ambiguity.

(e.g., Is the relationship additive or interactive?). We come back to this discussion in Section 5 below.

4. COREFERENCE: COMBINING EXPECTATIONS FOR MEANINGS AND FORMS

Similar to the discussions of disfluencies and intonation, the interpretation and production of referring expressions offers a parallel window into the two different types of alternatives discussed in this paper. Let us consider the simple example below.

(8) Do you see the guy sitting by the window? He is Bob's high school friend, Nick.

By *referring expression*, we mean the speaker's choice of a nominal expression to linguistically encode a reference to a particular entity (an object or individual or concept or event). For instance, "the guy (sitting by the window)", "he", "Bob's high school friend", and "Nick" are all referring expressions (alternative forms) that can be used to talk about the same person (an intended meaning). The speaker needs to select an appropriate form depending on linguistic and non-linguistic information available in context to refer to the person. By *coreference*, we mean instances of two or more referring expressions picking out the same entity within or across utterances. For instance, the speaker of (8) needed to choose a referential form at the outset of the second utterance ("Do you see the guy sitting by the window? □ is …") to clearly indicate the overlapping identity of the person being talked about. Unlike disfluency or intonational phonology, where the alternatives of form can vary continuously, the choices among referring expressions are discrete. A speaker must choose a form out of the (potentially very large) set of possible referring expressions; for our purposes here we will contrast the behavior of pronouns and names (but see Hemforth et al. (2010) for discussion of reference alternatives at the level of syntactic formulation).

Now let us consider how a listener might process various cases of coreference to identify who the speaker intended to refer to. This might appear trivial when looking only at cases such as (8), where there is only one possibility. However, resolving coreference can be a challenge for a listener when there are two possible candidates that cannot be distinguished by gender or animacy. Compare the minimal pair (9).

(9) **a.** John congratulated Bob because he won the race. [he ≈ BOB]
b. John impressed Bob because he won the race. [he ≈ JOHN]

The sentences in (9a-b) are identical except for the verb in the first clause, yet the preferred interpretation of the pronoun (as indicated in brackets) appears to flip from being coreferent with the OBJECT REFERENT BOB in (9a) to the SUBJECT REFERENT JOHN in (9b). The verbs "congratulate" and "impress" are well-studied members of a set of so-called implicit causality (IC) verbs, which are notable for inducing strong biases regarding the likely individual to whom cause can be attributed and thereby influencing the interpretation of ambiguous pronouns in contexts like (9) (Au, 1986; Brown & Fish, 1983; Caramazza, Grober, Garvey, & Yates, 1977; Garvey & Caramazza, 1974; Kehler, Kertz, Rohde, & Elman, 2008; McKoon, Greene, & Ratcliff, 1993).

Subject-biased IC verbs (e.g., "impress", "amaze", "disappoint") describe events in which the cause is typically attributed to the referent in subject position; object-biased IC verbs (e.g., "congratulate", "scold", "thank") typically attribute cause to the object. These are not categorical distinctions; it is possible for the subject to be the causally-implicated referent of an object-biased verb and vice versa. Examples (10a-b) show that the cause of a scolding event can be attributed unambiguously either to the object Bob (who did something undesirable) or to the subject John (for his desire to avoid blame). Examples (10c-d) show that both meanings can be formulated using a pronoun.

(10) John scolded Bob because …
a. … Bob had been boastful.
b. … John was trying to deflect blame from himself.
c. … he had been boastful. [he ≈ BOB]
d. … he was trying to deflect blame from himself. [he ≈ JOHN]

A speaker who intends to describe a situation in which it is Bob's boastfulness that led to the scolding can choose to articulate that meaning with (10a) or (10c); for a situation in which the reason behind the scolding is John's blame deflection, the choices include (10b) and (10d). A listener who encounters a pronoun ("he") and wants to recover the intended meaning can use their generative model of the speaker's production process to weigh the probability of different alternative coreference meanings ({JOHN, BOB})

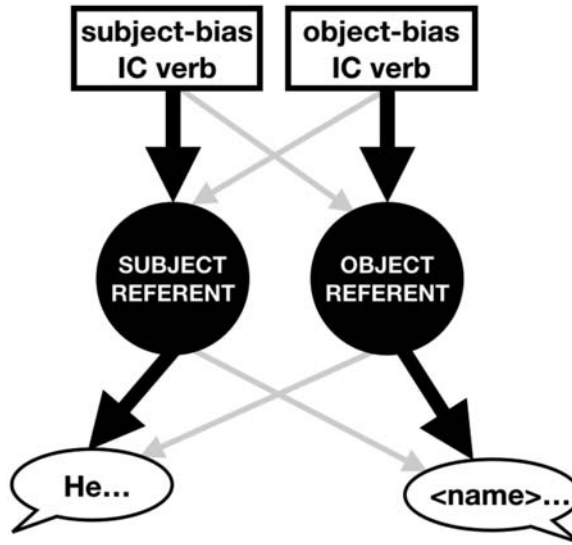

Figure 7 Graphical model of referring-expression production, showing alternative output forms and alternative underlying meaning states with two conditioning contexts.

and the probability of those meanings being articulated with different alternative forms ({"he", "John", "Bob", …}).

In this way, parallels can be drawn with the looks-like-an-X example from Section 3. The possible meanings and the possible forms are different, but the inferential structure is the same. Fig. 7 shows a model of a speaker's production choices. Again there is a contextual factor (here, verb type) which influences which meaning a speaker is more likely to produce (here, re-mention of the SUBJECT vs. OBJECT referent). The speaker conveys such meanings via a choice among various forms (here, pronominal forms vs. names). In Fig. 7, the thickness of the arrows again indicates production preferences. IC verbs are depicted as influencing the preference over the re-mention of the SUBJECT and OBJECT referents, and the referent's SUBJECT/OBJECT status is depicted as influencing the eventual preferred form of reference.

This latter pronominalization preference is important because it has the power to influence listeners' interpretations even in contexts where a pronoun is fully ambiguous (i.e., "He" where the available referents are JOHN and BOB). The preference for speakers to produce pronouns when re-mentioning a subject referent is often taken to reflect, for English, the presumed topicality of the subject (Givón, 1983) combined with an understanding that continued reference to the same topic is best achieved with a pronoun (Ariel, 1990; Grosz, Joshi, & Weinstein, 1995; Gundel, Hedberg,

& Zacharski, 1993; Lambrecht, 1994). Evidence for this preference comes generally from observations of a "subject preference" for pronoun coreference (Crawley, Stevenson, & Kleinman, 1990; Frederiksen, 1981; Gernsbacher & Hargreaves, 1988; Gernsbacher, Hargreaves, & Beeman, 1989; Järvikivi, van Gompel, Hyönä, & Bertram, 2005), and more specifically from production tasks that measure participants' pronominalization rates (Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). Such production tasks have used IC prompts as in (11) to measure both who participants favor for next mention and what form they prefer to use.

(11) Story-continuation stimuli
- **a.** [obj-bias IC] John congratulated Bob. _____
- **b.** [subj-bias IC] John impressed Bob. _____

For prompts like (11), participants produce more pronouns when the referent is the SUBJECT than when the referent is the OBJECT. Intriguingly, this pattern whereby more pronouns are produced for the subject referent holds even when the subject referent isn't favored for next mention, as is the case with object-biased IC verbs (Fukumura & van Gompel, 2010; Rohde & Kehler, 2014; Kehler & Rohde, 2018; cf. Rosa & Arnold, 2017, for a comparison with transfer-of-possession verbs). In Rohde and Kehler's results for subject-biased IC verbs, subject referents were re-mentioned with pronouns quite often (77.5% of the time, compared to a pronominalization rate of 26.6% when the object referent was re-mentioned). What is interesting is that for object-biased IC verbs, it is still the *subject* referent that is pronominalized at the higher rate (80.8% of the time, compared to a rate of 21.7% for objects). In those cases, subject referents are rarely mentioned in participants' continuations (due to the semantic bias of the object-biased IC verb), but when they are, they are mentioned with a pronominal referring expression.

Based on this, a listener who encounters an ambiguous pronoun is in a position to consider two sources of biases that are at work in the speaker's generative production process. First are the biases about who will be mentioned next (i.e., meaning-driven expectations). Subject-biased IC verbs favor upcoming mention of the SUBJECT; object-biased IC verbs favor the OBJECT. The other bias source pertains to syntactic position. Regardless of the IC verb types, the previous subject is more likely to be mentioned via a pronoun compared to the previous object (i.e., form-driven expectations). The upshot is that comprehenders should use IC cues to inform their

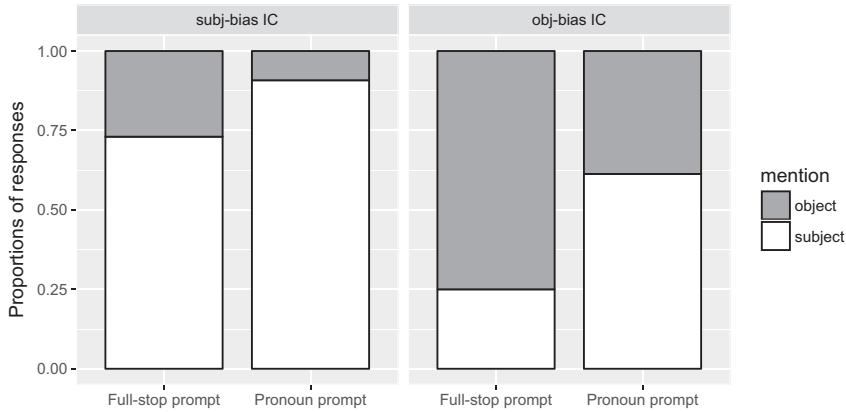

Figure 8 Coreference choices across verb and prompt types, showing the combination of context-driven IC biases and form-driven prompt differences (Adapted from Rohde, H., Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience, Special Issue on Production of Referring Expressions: Models and Empirical Data,* 912–927.).

meaning-driven expectations about who will be mentioned next before they have encountered any referential form. Then, if they encounter a pronoun, they should update their next-mention expectations in favor of the subject.

This prediction was tested by Rohde and Kehler (2014). They used a story-continuation task that manipulated a contextual cue (subject vs. object biased IC verbs) and the form of the prompt, so that participants saw half the items in a full-stop condition as in (11) and half with a pronoun prompt (as in "John congratulated Bob. He _____"). The results, shown in Fig. 8, confirm the predicted patterns: The contextual cue guided participants' preferences about which meaning should hold (more continuations that mentioned the subject following subject-biased verbs than object-biased verbs), and the form-based cue encouraged participants to update their preferences (more subject continuations when the prompt contained a pronoun than when it was just a full stop). There was no interaction between verb bias and prompt type.

Note that, for Rohde and Kehler's materials, the pronoun itself is fully ambiguous — if the participant wanted to use the pronoun prompt to corefer with, for example, Bob, the causally implicated referent for the object-biased prompt in (11a), they certainly could (and many of them do). However, the mere presence of a pronoun is enough to yield an increase in subject coreference when compared with the rate of subject

coreference in the full-stop condition. As Rohde and Kehler note, the results therefore suggest that comprehenders are taking into account both the context-driven expectations over probable meanings as well as their expectations for different forms given those possible meanings, as captured by the causal structure in Fig. 7.

Why does this matter? The results that Rohde and Kehler (2014) report matter because they distinguish coreference models that rely on a general notion of prominence from models based on the framework advocated here (i.e., differentiating between prominent meanings and preferred forms). A number of coreference models posit a link between a pronoun and the referent who is most prominent, salient, accessible, in-focus, activated, etc. in a context (e.g., Gundel et al.'s (1993), accessibility hierarchy, Arnold's (2001), Expectancy Hypothesis, Grosz et al.'s (1995), Centering Theory), but such models cannot explain the difference shown in Fig. 8 between the full-stop prompt and the pronoun-prompt. Why would participants not simply use the pronoun in the pronoun prompt to refer to the most prominent referent? In the current framework, this difference can be explained by the assumption that the speaker's (or writer's) choice of a referential form out of possible alternatives carries useful information about the intended meaning. Participants' continuations in the full-stop condition reveal which referent they are treating *a priori* as more prominent (a reflection likely of a combination of biases dictated by the IC verb or the prominence of the subject as topic or any number of other factors). In contrast, their continuations in the pronoun-prompt condition reveal their expectations as updated by observing the speaker's choice of a pronoun, which permits them to draw an inference: If the speaker had intended an OBJECT continuation, she would have likely selected an alternative referential form like a name, but she used a pronoun instead and pronouns are commonly used for SUBJECT mentions, so a SUBJECT mention is just that much more likely here. In other words, participants are combining these two sources of biases to reverse-engineer a possible meaning that is most likely to have generated the observed referential form. In the next section we discuss this inference process in detail.

5. INFERENCES BASED ON ALTERNATIVE FORMS AND MEANINGS: A BAYESIAN APPROACH

In the various examples discussed in Sections 1–4, we have consistently identified two levels of alternatives: alternative meanings that speakers

may be intending to convey and alternative forms that they can select to convey an intended meaning. An important insight to be reiterated here is that listeners' expectations about these two levels of alternatives can be derived independently. Taking a case of coreference as an example, a listener can *a priori* assign different probability estimates to different meanings (i.e., discourse referents such as BOB, JOHN, TOM, SARAH, …) in terms of how likely each of them is to be talked about in a current utterance. These estimates are independent of those the listener can derive for different linguistic expressions (e.g., When the intended meaning is BOB, it can be encoded as "Bob", "he", "that guy", "my neighbor"). In this section we discuss how these two levels of probability estimates might be combined to guide the listener's inferences.

Among different approaches, here we focus on those that use Bayesian statistics to tackle this problem. Bayesian statistics are used widely to model subjective probability values (or beliefs) accrued and projected by human agents as opposed to objective probabilities of events (Anderson, 1990). The basic idea is that humans have hypotheses (Hs) about true states of the world that are often not directly observable. These are often called *prior beliefs*. When they observe evidence/data (E), humans evaluate it according to their estimates of the probability of observing E given a particular H. These estimates are called *likelihood estimates*. For example, continuing with the coreference example, listeners can have multiple hypotheses (Hs) about what an intended referent of a pronoun must be (e.g., BOB, JOHN, TOM, SARAH …). Different Hs are assigned different prior probabilities. When the listener encounters a referential form (e.g., "he"), that constitutes evidence, E. Listeners can derive a likelihood estimate for the piece of evidence given a particular hypothesis (e.g., What's the likelihood of the speaker producing "he" if the intended meaning was BOB?) Using these two parameters, Bayesian inference computes the *posterior probability* of the hypothesis given the evidence.

$$\textbf{(12)}\quad P(H|E) = \frac{P(E|H)\cdot P(H)}{P(E)}$$

In this notation, P stands for probability. P(H), for example, means the probability assigned to a particular hypothesis. P(E|H) is the probably of a piece of evidence given a hypothesis. The right side of the equation represents a product of the likelihood estimate (i.e., P(E|H)) and the prior probability divided by the normalizing constant, P(E). The denominator P(E) is

included to make sure that the outcome of the inference is proportional to the probability of observing the given evidence. The left hand side of the equation represents the *posterior probability* of the hypothesis given the data. In the case of the coreference example mentioned above, this represents how likely it is that a given hypothesized meaning (e.g., BOB) is chosen given an observed referential form (e.g., "he"). This Bayesian approach is precisely that outlined in Kehler et al. (2008) for the coreference patterns as presented in Section 4.

A number of recent studies have applied this basic inference model to language comprehension (Phoneme perception and learning: Feldman & Griffiths, 2009; Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008; Sonderegger & Yu, 2010; Syntactic parsing: Hale, 2006; Levy, 2008; Pragmatic reasoning: Bergen & Grodner, 2012; de Ruiter & Cummins, 2012; Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013; Jurafsky, 2008; Kehler et al., 2008). In the domain of pragmatic inferences and interpretation, one approach has been called "Rational Speech Act (RSA)" model, in which hypothesized meanings can be computed given an instance of observed form (Frank & Goodman, 2012, 2014).[6] It is indeed a powerful reasoning framework that yields a wide range of predictions about how listeners derive their pragmatic interpretations of linguistic input. Let us go back to the earlier example where Bob says "I love you" to Sarah and Sarah responds. In this scenario, Bob can posit alternative meanings, which constitute his hypotheses about Sarah's love (i.e., the distribution P(H)). He has a model of language production that assigns different probability values to a particular linguistic form given the hypothesis (i.e., P(E|H)), e.g., What would Sarah say if her meaning state was one of reciprocated LOVE? When Bob observes the linguistic evidence (E), he can compute the probability to be assigned to his hypothesis given the observed evidence (i.e., P(H|E)).

One major strength of this approach lies in its being a principled, mathematical framework that allows researchers to make quantitative predictions over outcomes of inferences. Specifically, it can capture two key determining components in the communication of meaning: our beliefs about meanings (e.g., what kinds of intentions/meanings are more likely in a given

[6] For those who are interested in this framework, there are a number of informative and accessible papers written from a standpoint of introducing the general reasoning and inference framework to problems related to language processing and acquisition (e.g., Perfors et al., 2011; Sobel, Tenenbaum, Gopnik, 2010; Tenenbaum & Griffiths, 2001).

context) and our model of language (e.g., how a given meaning is mapped onto linguistic encoding options). The Bayesian inference framework provides a systematic approach for how these two knowledge sources might be combined under uncertainty.[7] An advantage of this architecture is that it captures the structured influence of discourse context on pragmatic reasoning about meanings by integrating prior beliefs that are relevant and yet independent of meanings encoded in the linguistic signal. For instance, in the case of the LOVE vs. ¬LOVE example, Bob might *a priori* believe that Sarah is no longer in love with him based on her demeanor or some information he has received from a friend. If the probability of one hypothesis is very high, that prior probability (of, say, ¬LOVE) dominates the outcome of the inference. The prior probability thus represents how likely Bob will infer LOVE or ¬LOVE as an intended meaning in the absence of any linguistic signal from Sarah.

As simple as it may seem, this effect of context on inferred meaning *irrespective of linguistic forms used* is hard to explain in other traditional accounts of linguistic comprehension. This is because many accounts generally assume that listeners first construct a semantic (or logical) meaning of an utterance from semantic and syntactic elements of an utterance. They then enrich it in consideration of relevant components of context. In the current account, the assumption is reversed. Listeners begin to process the linguistic input based on their expectations about possible and likely meanings in an immediate environment as well as on their life-long experiences with linguistic communication. They expect the speaker to produce a linguistic signal that can meaningfully alter (update) the existing hypotheses about intended meanings. As a result, the same physical linguistic evidence can lead to distinct meanings, depending on which meanings were expected. As was illustrated in the example of the assumed level of expertise in the interpretation of "it looks like an X" (Section 3), inferences about the speaker's intended meaning are contingent on a wide variety of contextually-supported assumptions about the speaker (e.g., an expected level of expertise) as well as non-linguistic features of referents (Frank & Goodman, 2012).

[7] We are not proposing that the human brain is necessarily processing information according to the Bayes Rule. Bayesian statistics make no assumptions about memory capacity and metabolic costs for processing information, and often rendered inadequate to capture neural activities. Such an approach, however, offers powerful machinery to make computational level predictions for neural behaviors, and our discussion here also primarily aims to make a computational level prediction for pragmatic inferencing behaviors.

The current approach provides a conceptual framework for understanding how such meaning-related expectations might affect outcomes of pragmatic inferencing.

This way of reasoning itself is, of course, not at all new. Within computational linguistics, Bayesian approaches have long provided the inferential underpinnings for many Natural Language Processing systems ranging from machine translation (Blunsom, Cohn, Dyer, & Osborne, 2009) to word segmentation (Goldwater, Griffiths, & Johnson, 2009) and other tasks (see reviews in Jurafsky & Martin, 2009, and Cohen, 2016). At its core, a Bayesian NLP system builds a generative model of the observed signal, representing the uncertainty about some underlying parameter. For speech recognition, for example, the observed signal is the acoustic signal. The system works by modeling the set of underlying word sequences that could plausibly yield the encountered speech, estimating the prior probability of those sequences of words and the likelihood that such words would be realized with the acoustic forms encountered.

Closer to home, in the domain of pragmatics, a number of observations have been made about the critical role of conceptual statuses of intended meanings under the terms such as *accessibility*, *salience*, *topic/focus-hood*, or *prominence* for coreference (Ariel, 1990; Chafe, 1974; Gundel et al., 1993). For instance, it has been demonstrated that a referent's established salience in a discourse (the probability of a speaker choosing to re-mention it in a subsequent sentence) influences the interpretation of an upcoming pronoun (Arnold, 2001). The current inference framework can correctly predict such "salience" effects by assuming that the semantic factors that influence salience are what lead to a higher prior probability of choosing a given referent for re-mention. It has also been demonstrated that pronominal forms are unique in the frequency of their use when a prominent (subject) entity is being re-mentioned (Gordon, Grosz, & Gilliom, 1993; Grosz et al., 1995). The current framework likewise captures this bias in production choices for prominent subjects. The added value of applying the current framework is in being able to pinpoint which factors influence which choice and to thereby move away from underspecified notions of salience and prominence. The factors that influence a speaker's choice among prominent alternative meanings need not be the same as those that determine prominence for the selection among alternative forms (see Kehler and Rohde (2013) for this argument and the explanation of how salience effects and pronominalization biases can be accounted for within a Bayesian model). Without this model, the appeal to salience and prominence can become

circular: The typical story goes that a speaker produces a pronoun to refer to something salient/prominent; and how do we know that the referent is salient/prominent? Well, we know because that referent can be pronominalized! Therefore, in addition to making specific predictions and being quantitative, the current approach avoids this circularity and avoids the simple descriptive enumeration of referent properties that might contribute to salience and prominence.

Another reason that this inferential approach is powerful is that both the prior probabilities and likelihood estimates can be learned and fine-tuned with respect to various conditioning factors. Intuitively, our expectations about possible intended meanings and possible linguistic forms used to convey these meanings vary across contexts. Imagine that Sarah is a generally very expressive and affectionate person, who says "I love you" to all of her friends. In that case, her saying "I love you" might not qualitatively differ from (others') saying "I like you". On the other hand, if Sarah is generally very shy and reserved in the way she expresses her emotions, "I love you" might signal a true declaration of love. In other words, one's expectations for P(E|H) (e.g., What would the speaker say if the intended meaning is LOVE vs. LIKE?) can be conditioned on the identity of the speaker (i.e., P(E|H, speaker)). As a result, the estimates about P(H|E) will also reflect the speaker, where the same linguistic evidence can support different inferences depending who said it. Conversely, the listeners' estimate about the prior (P(H)) may itself vary across listeners since it is not necessarily the case that all individuals across ages and contexts should converge on the same expectations about the messages they hypothesize to be most probable.

The importance of interlocutor identities in linguistic comprehension has been discussed in the long tradition of socio-phonetics (e.g., Drager, 2010b; Hay & Drager, 2007; Labov, 2006). Listeners are sensitive to subtle variations of language use, finding traces of the speaker's linguistic background and social membership. The knowledge about speakers, in turn, changes the way the linguistic signal is processed. It has been shown, for example, that native listeners' perception of vowels (diphthongs) can be influenced by the purported origin of the speaker (Canada vs. Detroit, MI; Niedzielski, 1999). Listeners' perception of /aw/ (e.g., **ou**t) or /ay/ (e.g., n**igh**t) reflects what they expected to hear from a speaker from a particular region. One way to explain this phenomenon is that listeners have different expectations for acoustic values that count as /aw/ and /ay/ for a given dialect, e.g., P(F1 and F2 values | /aw/, Canadian) vs. P(F1 and F2 values | /aw/, Detroiter). The same input then inferentially

supports perception of different sound categories depending on these expectations.

More generally, a growing literature targets the extent to which perceived social group membership affects language comprehension, including in phonological and syntactic processing (Drager, 2010a,b; Hanulíková, van Alphen, van Goch, & Weber, 2012; Hay & Drager, 2010; Hay, Warren, & Drager, 2006; Staum Casasanto, 2008; Warren, 2017). In pragmatic interpretation, too, assumptions about the speaker and her knowledge/intention as well as linguistic traits can have systematic effects on outcomes of inferences. Recently, Kurumada, Brown, and Tanenhaus (2017) have provided experimental evidence that native English listeners can adapt their pragmatic interpretations of intonation contours according to distributional characteristics of a given speaker's production. Based on our approach, we can now envision that similar reasoning might apply to investigations of silence and turn-taking. A pause of 600 ms can be perceived as a delay of speech onset, signaling hesitation for a speaker who is generally very eloquent. In other words, P(pause = 600 ms | hesitation) can be really high for this speaker. A pause spanning the same amount of time, however, may not be particularly meaningful for another speaker who is generally careful in choosing words and regularly produces long pauses. Thus the approach described above can pinpoint possible loci of such speaker-based adjustment of inferences, which can facilitate accurate inferences about what the speaker meant to communicate.

6. WHERE NEXT?

As we chronicled in the previous sections, past research has provided a wide variety of observations regarding alternative forms and alternative meanings evoked in communication of meaning. Bayesian statistics offers a principled computational-level account as to how these assumptions and knowledge at different levels of representations might be combined to guide the listener's inferences about communicated meanings. Challenges remain, however, in predicting and simulating how such inferences may be implemented in real-time language comprehension with a rich set of contextual information. How can speakers and listeners converge on a particular set of alternative forms and meanings among an infinite number of possibilities? Indeed, much is still unknown about how probabilistic knowledge about linguistic forms and meanings is stored and represented. Would it be feasible to assume that human language users, with a finite cognitive and computational

capacity, are engaging in the inferences outlined in Section 5? How do human infants become proficient in the types of inferences discussed above? Resolving these problems clearly goes beyond the scope of the current work. Here we highlight what we take to be promising avenues for future research in which the insights laid out in the current paper could contribute to theoretical and empirical advances in existing scientific endeavors.

Real-time Processing: As mentioned above, a number of experiments utilize psycholinguistic measurements that are time-locked to the unfolding linguistic signal (e.g., eye-movements, ERP). One of the goals of such studies is to account for the time course of semantic and pragmatic understanding of speech: How do listeners derive contextually-enriched interpretations based on the incoming linguistic signal (Breheny, Ferguson, & Katsos, 2013; Degen & Tanenhaus, 2016; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2009, 2011; Van Berkum, Brown, & Hagoort, 1999)? Not all of these studies subscribe to the idea that listeners are actively comparing alternative forms and meanings. It is, in fact, actively debated whether listeners' moment-by-moment behavior may be compatible with the type of reverse engineering posited under the Bayesian account (Huang & Snedeker, 2018). Regardless of one's theoretical stance, however, it is important to point out that a typical "Visual World" eye-tracking experiment often presupposes some framework closely related to what we have described here. That is, most eye-tracking studies that use visual referents and linguistic instructions encourage the listener to arrive at alternative (intended) meanings based on visual context as well as to evaluate the incoming linguistic signal with respect to the meaning options.

A classic study conducted by Tanenhaus and his colleagues provides a useful illustration of this (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Two types of visual scenes were created with four objects each to modulate expectations for alternative intended meanings. In one, listeners saw four distinct object kinds. In the other, they saw two instances of the same object kind (e.g., two apples). In the latter case, a speaker is presumed to be more likely to intend to distinguish the target apple from the distractor apple, whereas she is less likely to do so when there is only one object that can be referred to as "the apple". In other words, these two scenes altered P(Intended meaning = DISTINGUISHING TWO APPLES). These differing expectations in meanings are associated with expectations about alternative linguistic forms. Listeners have some likelihood estimates for the probability that the speaker will produce a modifier qualifying the target, such as a post-nominal prepositional phrase i.e., P(a post-nominal

phrase | Intended meaning = DISTINGUISHING TWO APPLES). According to Bayesian inference, the posterior probability of P(Intended meaning = DISTINGUISHING TWO APPLES | a post-nominal phrase) is higher in the scene with two apples due to the increased prior probability of the meaning.

In the behavioral results, Tanenhaus et al. indeed found that a temporarily ambiguous prepositional phrase "Put the apple **on the towel** …" is more likely to be interpreted as a post-nominal modifier (i.e., the apple that is on the towel) in the presence of two apples. The same phrase **on the towel** is more readily interpreted as a destination (i.e., Move the apple so it will be on the towel) when there is only one apple in sight. This study, and a number of others following it, illuminate the possibility that the linguistic input is processed according to the rich backdrop of contextually-constrained expectations about possible and likely forms and meanings (e.g., Brown-Schmidt & Tanenhaus, 2008; Federmeier, 2007; Ni, Crain, & Shankweiler, 1996; Papadopoulou & Clahsen, 2006; Sedivy, 2002; Van Berkum et al., 1999).

An outstanding question remains as to whether such an approach can scale to contexts with increased complexity. The majority of visual world eye-tracking studies employ a simple visual context (cf. Brown-Schmidt and Tanenhaus (2008)), and the listener's goal is to resolve the speaker's referential intention. In a more naturalistic context, speakers often intend to convey a multitude of referential, communicative, and social meanings. Identifying a set of alternative meanings must encompass many other sources of information. Additionally, speakers avail themselves of a wide variety of linguistic and non-linguistic means to encode their intended meanings. For instance, a unique reference in the scene with two apples in Tanenhaus et al. (1995) can be just as easily achieved by directing eye-gaze or pointing in the direction of one of the apples (in lieu of a post-nominal modifier). Likewise, skilled listeners take into account contextual expectations and task-relevant affordances of possible referents in a scene (Chambers et al., 2002, 2004). Listeners' successful computation of likelihood estimates for a given linguistic option would therefore be well-placed to take into account a wide variety of options. Furthermore, each speaker may differ in their uses of these different options and combinations of them (Grodner & Sedivy, 2011). Future studies must explore whether and how listeners condition their expectations for alternative forms and meanings on contextual factors including individual differences across speakers.

As for the question of how to scale processing models to the potentially infinite set of alternatives (of form and of intended meaning), this is likewise not a new problem. Just as our use of a Bayesian approach echoes its long-standing utility in computational linguistics, cognitive modeling may benefit from a consideration of the strategies in Natural Language Processing domains where pruning the search space is likewise necessary (e.g., syntactic parsing, Vieira & Eisner, 2017). With pruning strategies, the full combinatorial complexity is reined in by eliminating low-probability options. Such pruning in turn may account for the garden-path processing difficulty that comprehenders experience when they struggle to recover a valid but low-probability parse Jurafsky (1996).

Acquisition of accurate inferencing: The way of reasoning laid out in the current paper presents a significant question about language learning. How do young children acquire the ability to enumerate appropriate alternative forms and meanings, combining them in a principled manner? In fact, pragmatic comprehension of utterances has often been identified as a domain of significant difficulty for young children (e.g., Braine & Rumain, 1981; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2006; Eskritt, Whalen, & Lee, 2008; Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001; Lidz & Musolino, 2002; Noveck, 2000; Papafragou, 2006; Papafragou & Tantalou, 2004), particularly for speaker intentions (e.g., Ackerman, 1981; Ackerman, Szymanski, & Silver, 1990), information status (e.g., Arnold, 2008; Stephens, 2010) and identity of discourse referents (e.g., Surian, 1991). Well after other areas of linguistic abilities become proximal to those of adults, school-aged children continue to fail in pragamtically-enriched interpretations of the linguistic input (e.g., Children older than eight years of age often endorse sentences like "Some elephants have trunks" while adults reject them because "some" implies that there are elephants with no trunk (Noveck, 2000; Smith, 1980)).[8]

[8] This does not necessarily mean that young children are incapable of making pragmatic inferences. To the contrary, there are a number of descriptive studies that observe that children can take into account the speaker's (and the agent's) goals and intentions (Csibra & Gergely, 2009; Meltzoff, 1995; Vouloumanos, Onishi, & Pogue, 2012) and draw contextually-situated meanings (e.g., Akhtar & Tomasello, 1996; Clark & Grossman, 1998; Ochs & Schieffelin, 1979). However, the studies cited here repeatedly found that children's responses in pragmatically-involved language comprehension tasks deviated from those of adults in experimental environments. This might be due to a mismatch between children's construals of the experimental tasks and what the tasks are designed to test. When task demands were alleviated, children can show behaviors similar to adults (Katsos & Bishop, 2011).

Understanding the basic mechanisms of adult language comprehension will facilitate the investigation of how a similar process may develop in children. A particularly intriguing possibility is that the slow and protracted pragmatic development may at least partially reflect children and adults' different assignment of probabilities to alternative forms and meanings due to their diverging linguistic experiences (or even that children are not monitoring for potential alternatives to the same extent as adults; see eye-tracking evidence on awareness of scene ambiguity reported by Rabagliati & Robertson, 2017). If so, outcomes of children's and adults' inferences will necessarily be different even if children are endowed with fully-fledged cognitive and computational abilities.

Some recent results suggest that even young language learners are equipped with the basic computational ability to compare alternative forms and meanings in a rational manner (Endress, 2013; Frank, Goodman, & Tenenbaum, 2009; Perfors, Tenenbaum, Griffiths, & Xu, 2011, 2010). For instance, the manner in which 3- and 4-year-olds form categories reflect inductive biases based on differing prior probabilities of exemplars (Xu & Tenenbaum, 2007). Similar expectations about underlying generative models have been discussed in word learning (Frank et al., 2009), and in syntactic learning (Perfors et al., 2011). Simply put, the generative framework allows learners to use inductive reasoning that supports learning of representations beyond what have been directly observed in the data. Such reasoning is thereby expected to help them effectively bootstrap their linguistic knowledge.

One of the best known examples of such inference may be what is often called Principle of Contrast (Clark, 1987, 1990, 1999, 2009).[9] The basic observation is that children as young as a two years old readily associate an unfamiliar label (e.g., "whisk") with an unfamiliar object in the presence of two objects differing in familiarity (e.g., a spoon and a whisk). This can be understood as a bias stemming from a competition between alternative form-meaning mappings. In this case, for the child, the word "spoon" is likely generated when the intended referential meaning is a spoon

[9] Principle of Contrast is often contrasted with Mutual Exclusivity (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Markman, Wasow, & Hansen, 2003). Mutual Exclusivity is considered as a normative bias discussed primarily in the domain of novel noun learning. The bias was proposed to explain children's tendency to maintain one-to-one mapping between a noun and an object. We adopt Principle of Contrast here because we assume that the inference process here involves computation of intended meanings and applies more broadly to lexical classes besides nouns, as proposed by Principle of Contrast.

(i.e., high P("spoon" | SPOON)). This proportionally lowers the probability assigned to the speaker's production of other (alternative) word forms (e.g., low P("whisk" | SPOON)), which contributes to a lower posterior probability of P(SPOON | "whisk") and a higher posterior probability of P(WHISK | "whisk"). In other words, even though the learner does not yet have any precise likelihood estimate of P("whisk" | WHISK), their assumption that words (forms) generally compete for the same meaning can allow them to make a new form-meaning mapping in context. If, on the other hand, learners were simply tracking one-to-one mappings between forms and meanings (without considering them as *alternatives*), such mapping would not be readily predictable.

Extending this approach can facilitate our understanding of how children may expand their ability to extract meaning given the linguistic signal. Here we end with an example in the domain of intonational phonology. Preschoolers are generally known to be insensitive to the types of intonational meanings discussed in Section 3, and the ability develops only gradually through their school years (Cruttenden, 1985; Cutler & Swinney, 1987; Gualmini, Maciukaite, & Crain, 2003; Ito, Bibyk, Wagner, & Speer, 2013; Solan, 1980; Speer & Ito, 2009; Wells, Peppe, & Goulandris, 2004). Kurumada and Clark (2016), however, found that preschoolers are capable of interpreting "It LOOKS like a zebra …" as ¬ZEBRA as long as they can access an alternative form. In their experiment, four-year-olds were at chance when they only heard the contrast of "It looks like a ZEBRA" (noun-focus) and "It LOOKS like a zebra …" (verb-focus). However, children from the same age group could reliably derive the ¬ZEBRA meaning when the same speaker produced an alternative form "It is a zebra!" to convey the alternative meaning i.e., ZEBRA. The authors consider that the underlying mechanism guiding the children's inferences can be explained in the same manner as the spoon-whisk example. When the more familiar form "it is a zebra" is used to encode one of the alternative meanings (i.e., ZEBRA), the unfamiliar form "It LOOKS like a zebra …" becomes more likely to encode an alternative meaning available in context. This reasoning may enable young children to construct the form-meaning mapping even before they become capable of drawing the contrastive interpretation through the intonation pattern (i.e., L + H★ on the verb and a rising tone at the end) alone. (For a similar finding in understanding of scalar adjectives, see Barner, Brooks, and Bale (2011).) Comparing alternatives thus allows learners to leverage inferences to make otherwise novel mappings between forms and meanings.

Alternatives provide a powerful framework that makes it possible for listeners and learners to infer what the speaker must have meant by choosing a linguistic form in context.

ACKNOWLEDGMENTS

We thank Andrés Buxó-Lugo for data visualization, and Bethany Gardner and Crystal Lee for feedback during editing. We also thank Eve Clark, Marisa Casillas, Madeleine Long, Kara Federmeier, Alice Turk, and Duane Watson for their feedback on an earlier version of the manuscript. H. Rohde is supported in part by a Leverhulme Trust Prize in Languages and Literatures.

REFERENCES

Ackerman, B. P. (1981). When is a question not answered? the understanding of young children of utterances violating or conforming to the rules of conversational sequencing. *Journal of Experimental Child Psychology, 31*(3), 487–507.

Ackerman, B. P., Szymanski, J., & Silver, D. (1990). Children's use of the common ground in interpreting ambiguous referential utterances. *Developmental Psychology, 26*(2), 234–245.

Akhtar, N., & Tomasello, M. (1996). Two–year–olds learn words for absent objects and actions. *British Journal of Developmental Psychology, 14*(1), 79–93.

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

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*, 502–518.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Ariel, M. (1990). *Accessing noun phrase antecedents*. Routledge.

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

Arnold, J. E. (2008). The bacon not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition, 108*(1), 69–99.

Arnold, J. E., Hudson Kam, C., & Tanenhaus, M. (2007). If you say thee uh- you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 914–930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science*, 578–582.

Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. *Papers in Laboratory Phonology, 9*, 547–576.

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.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition, 118*(1), 84–93.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In S. Santi, I. Guaïtella, C. Cave, & G. Konopczynski (Eds.), *Oralité et gestualité, communication multimodale, intéraction* (pp. 597–600). Paris: L'Harmattan.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language & Cognitive Processes, 25*(4), 441–455.

Beckman, M. E., & Pierrehumbert, J. B. (1986). *Intonation structure in Japanese and English*.

Bergen, L., & Grodner, D. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology-learning Memory and Cognition, 38*(5), 1450–1460.

Blunsom, P., Cohn, T., Dyer, C., & Osborne, M. (2009). A gibbs sampler for phrasal synchronous grammar induction. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP* (pp. 782–790).

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26*, 119–137.

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). London: Academic Press.

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no… How the brain interprets the pregnant pause in conversation. *PLoS One, 1–15*.

Bortfield, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 32*, 229–259.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language, 75*, 104–116.

Braine, M. D., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology, 31*(1), 46–70.

Breheny, R. R., Ferguson, H. H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language & Cognitive Processes, 28*(4), 443–467.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274–296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383–398.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science, 32*, 643–684.

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition, 14*, 237–273.

Buxó-Lugo, A., Toscano, J. C., & Watson, D. G. (2018). Effects of participant engagement on prosodic prominence. *Discourse Processes, 55*(3), 305–323.

Calhoun, S. (2004). Phonetic dimensions of intonational categories: The case of L+H* and H*. *Proceedings of Speech Prosody, 2004*, 103–106.

Calhoun, S. (2006). *Information structure and the prosodic structure of English: A probabilistic relationship* (Ph.D. thesis). University of Edinburgh.

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

Casillas, M. (2013). *Learning to take turns on time: Perception and production processes involved in keeping inter-turn gaps short* (Ph.D. thesis). Stanford University.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language, 43*(6), 1310–1337.

Chafe, W. (1974). Language and consciousness. *Language, 50*, 111–133.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (July 2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47*(1), 30–49.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 687–696.
Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2006). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do (Ed.), *BUCLD 25 proceedings* (pp. 157–168). Cascadilla Press.
Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Erlbaum.
Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*(2), 417–431.
Clark, H. H. (1994). Managing problems in speaking. *Speech Communication, 15*, 243–250.
Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
Clark, E. V. (1999). Acquisition in the course of conversation. *Studies in the Linguistic Sciences, 29*(2), 1–18.
Clark, E. V. (2009). *First language acquisition* (vol. 2). Cambridge University Press.
Clark, E. V. (2015). Common ground. In B. MacWhinney, & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 328–353). London: Wiley-Blackwell.
Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111.
Clark, E. V., & Grossman, J. B. (1998). Pragmatic directions and children's word learning. *Journal of Child Language, 25*(1), 1–18.
Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201–242.
Clark, E. V., & Wong, A. (2002). Pragmatic directions about language use: Words and word meanings. *Language in Society, 31*, 181–212.
Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*(3), 804–809.
Cohen, S. B. (2016). *Bayesian analysis in natural language processing. Synthesis lectures on human language technologies*. Morgan and Claypool.
Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience, 30*(1–2), 1–31.
Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology, 1*(2).
Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition, 105*, 658–668.
Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronoun. *Journal of Psycholinguistic Research, 4*, 245–264.
Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language, 12*, 643–661 (ST – Intonation comprehension in ten–year).
Cruttenden, A. (1997). *Intonation*. Cambridge University Press.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.
Cutler, A. (2015). *Native listening: Language experience and the recognition of spoken words*.
Cutler, A., Dahan, D., & Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141–201.
Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language, 14*(01), 145–167.
Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(5), 441–452.
de Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515–535.
de Ruiter, J. P., & Cummins, C. (2012). A model of intentional communication: AIRBUS (Asymmetric Intention Recognition with Bayesian Updating of Signals). In S. Brown-Schmidt, J. Ginzburg, & S. Larsson (Eds.), *Proceedings of SemDial (SeineDial), 16th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 149–150).

Degen, J. (2013). *Alternatives in pragmatic reasoning* (Ph.D. thesis). University of Rochester.
Degen, J., & Tanenhaus, M. K. (January 2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science, 40*(1), 172–201.
Dilley, L., Breen, M., Gibson, E., Bolivar, M., & Kraemer, J. (2006). A comparison of inter-transcriber reliability for two systems of prosodic annotation: Rap (rhythm and pitch) and tobi (tones and break indices). In *Proceedings of the international conference on spoken language processing. Pittsburgh, PA*.
Dilley, L., Breen, M., Gibson, E., Bolivar, M., & Kraemer, J. (2012). A comparison of inter-transcriber reliability for two systems of prosodic annotation: RaP (rhythm and pitch) and ToBI (tones and break indices). In *Corpus linguistics and linguistic theory* (vol. 8, pp. 277–312).
Dilley, L. C., Ladd, D. R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics, 33*(1), 115–119.
Drager, K. (2010a). Sensitivity to grammatical and sociophonetic variability in perception. *Laboratory Phonology, 1*(1), 473–480.
Drager, K. (July 2010b). Sociophonetic variation in speech perception. *Language and Linguistics Compass, 4*(7), 473–480. https://doi.org/10.1111/j.1749-818X.2010.00210.x.
Endress, A. D. (2013). Bayesian learning and the psychology of rule induction. *Cognition, 127*(2), 159–176.
Eskritt, M., Whalen, J., & Lee, K. (2008). Preschoolers can recognize violations of the gricean maxims. *British Journal of Developmental Psychology, 26*(3), 435–443.
Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491–505.
Feldman, N. H., & Griffiths, T. L. (2009). A rational account of the perceptual magnet effect. *Proceedings of the 29th Annual Conference of the Cognitive Science Society, 1*, 6.
Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709–738.
Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62*, 151–167.
Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science, 336*, 998.
Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology, 75*, 80–96. https://doi.org/10.1016/j.cogpsych.2014.08.002.
Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*(5), 578–585.
Frederiksen, J. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes, 4*, 323–347.
Fukumura, K., & van Gompel, P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language, 62*, 52–66.
Ganong, W. F. (1980). Phonetic categorization in auditory word 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.
Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27*, 699–717.
Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of the first mention versus the advantage of clause recency. *Journal of Memory and Language, 28*, 735–755.
Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative, cross-language study* (pp. 1–42). Amsterdam: John Benjamins.

Goldwater, S., Griffiths, T., & Johnson, M. (2009). A bayesian framework for word segmentation: Exploring the effects of context. *Cognition, 112*, 21–54.
Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology, 28*(1), 99–108.
Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*.
Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17*, 311–347.
Grice, P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics* (vol. 3, pp. 41–58). New York: Academic Press.
Grice, H. P. (1989). *Studies in the way of words* (vol. 65). Harvard University Press.
Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). Some, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition, 116*(1), 42–43.
Grodner, D. J., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter, & E. Gibson (Eds.), *The processing and acquisition of reference* (pp. 239–272). MIT Press.
Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21*, 203–225.
Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. In *Proceedings of SALT XI. Ithaca, NY* (pp. 231–247).
Gualmini, A., Maciukaite, S., & Crain, S. (2003). Children's insensitivity with contrastive stress with only. In *Proceedings of the 25th Penn linguistics colloquium* (vol. 9, pp. 87–110). University of Pennsylvania.
Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*(2), 274–307.
Hagoort, P., & van Berkum, J. (May 2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 362*(1481), 801–811.
Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*(4), 609–642.
Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878–887.
Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*(4), 865–892.
Hay, J., & Drager, K. (September 2007). Sociophonetics. *Annual Review of Anthropology, 36*(1), 89–103. https://doi.org/10.1146/annurev.anthro.34.081804.120633.
Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458–484.
Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K. (April 2015). Inferring difficulty: Flexibility in the real-time processing of disfluency. *Language and Speech, 58*(2), 190–203.
Hemforth, B., Konieczny, L., Scheepers, C., Colonna, S., Schimke, S., Baumann, P., & Pynte, J. (2010). Language specific preferences in anaphor resolution: Exposure or gricean maxims? In S. Ohlsson, & R. Catarambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 2218–2223).
Hirschberg, J. (1985). *A theory of scalar implicature* (Ph.D. thesis). Philadelphia: University of Pennsylvania.
Horn, L. R. (1972). *On the semantic properties of logical operators in English* (Ph.D. thesis). UCLA.
Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology, 58*, 376–415.
Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language & Cognitive Processes, 26*, 1161–1172 (933044199).

Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology, 102*, 105–126.
Ito, K., Bibyk, S. A., Wagner, L., & Speer, S. R. (2013). Interpretation of contrastive pitch accent in six- to eleven-year-Old English-speaking children (and adults). *Journal of Child Language, 1*, 27.
Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*(2), 541–573.
Jäger, G., & Franke, M. (2014). Pragmatic back-and-forth reasoning. In S. P. Reda (Ed.), *Pragmatics, semantics and the case of scalar implicature* (pp. 170–200). Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
Järvikivi, J., van Gompel, R., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first mention and subject preference accounts. *Psychological Science, 16*, 260–264.
Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20*(2), 137–194.
Jurafsky, D. (2008). *Pragmatics and computational linguistics*. Wiley-Blackwell.
Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics*. Prentice-Hall.
Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition, 120*, 67–81.
Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*(1), 1–44.
Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics, 39*, 1–37.
Kehler, A., & Rohde, H. (2018). Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics: Special Issue on Prominence in Pragmatics*.
Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes, 52*(4), 255–289.
King, J. P. J., Loy, J. E., & Corley, M. (2018). Contextual effects on online pragmatic inferences of deception. *Discourse Processes, 55*(2), 123–135.
Kleinschmidt, D. F., & Jaeger, T. F. (April 2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203.
Köhne, J., & Demberg, V. (2013). The time-course of processing discourse connectives. In *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 2760–2765).
Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107*(1), 54–81.
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59.
Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition, 133*(2), 335–342.
Kurumada, C., Brown, M., Bibyk, S., & Tanenhaus, M. K. (2018). *Probabilistic inferences and adaptation in pragmatic interpretation of contrastive prosody* (under review).
Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Prosody and pragmatic inference: It looks like speech adaptation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 647–652).
Kurumada, C., Brown, M., & Tanenhaus, M. K. (2017). Effects of distributional information on categorization of prosodic contours. *Psychological Bulletin and Review*.
Kurumada, C., & Clark, E. (2016). Pragmatic inferences in context: Learning to interpret contrastive prosody. *Journal of Child Language*.
Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics, 34*(4), 500–515.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press.
Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics, 25*(3), 313–342.
Ladd, D. R., & Schepman, A. (2003). Sagging transitions between high pitch accents in English: Experimental evidence. *Journal of Phonetics, 31*(1), 81–112.
Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review, 18*, 135–140.
Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Levinson, S. C. (1983). *Pragmatics (Cambridge textbooks in linguistics)*. Cambridge, UK: Cambridge University Press.
Levinson, S. (2000). *Presumptive meanings*. The MIT Press.
Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 731.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177.
Liberman, M. (1960). Some acoustic correlates of word stress in {A}merican {E}nglish. *Journal of the Acoustical Society of America, 32*, 451–454.
Lidz, J., & Musolino, J. (2002). Children's command of quantification. *Cognition, 84*(2), 113–154.
Loy, J., Rohde, H., & Corley, M. (2017). Effects of disfluency in online interpretation of deception. *Cognitive Science, 41*, 1434–1456.
MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language & Cognitive Processes, 9*(2), 157–201.
Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology, 47*(3), 241–275.
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., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118*(2), 219–246.
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*(3), 283–312.
Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*(5), 838.
Neely, J. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.
Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Assessing the contribu- tions of syntax, semantics and plausibility in resolving ambiguities. *Language & Cognitive Processes, 11*(3), 283–334.
Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62–85.
Norris, D., Cutler, A., & McQueen, J. M. (2000). The optimal architecture for simulating spoken-word recognition. In *Cognitive science in Australia, 2000: Proceedings of the fifth biennial conference of the Australasian Cognitive Science Society*. Adelaide: Causal Productions.
Norris, D., & McQueen, J. (2008). Shortlist b: A bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238.

Noveck, I. A. (2000). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*(2), 165–188.
Ochs, E., & Schieffelin, B. (1979). *Developmental pragmatics*. Academic Press.
Orena, A. J., & White, K. S. (2015). I forget what That's called! Children's online processing of disfluencies depends on speaker knowledge. *Child Development, 86*(6), 1701–1709.
Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech & Language, 9*, 19–35.
Papadopoulou, D., & Clahsen, H. (2006). Ambiguity resolution in sentence processing: The role of lexical and contextual information. *Journal of Linguistics, 42*, 109–138.
Papafragou, A. (2006). From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language, 33*(4), 721–757.
Papafragou, A., & Tantalou, N. (2004). *Children's computation of implicatures*.
Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120*(3), 302–321.
Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language, 37*(3), 607–642.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Ph.D. thesis). MIT.
Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press.
Pisoni, D. B., & Levi, S. V. (2007). Representations and representational specificity in speech perception and spoken word recognition. *The Oxford Handbook of Psycholinguistics*, 3–18.
Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press UK.
Rabagliati, H., & Robertson, A. (2017). How do children learn to avoid referential ambiguity? insights from eyetracking. *Journal of Memory and Language, 94*, 15–27.
Recanati, F. (1989). The pragmatics of what is said. *Mind & Language, 4*(4), 295–329.
Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience, Special Issue on Production of Referring Expressions: Models and Empirical Data*, 912–927.
Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language, 94*, 43–60.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735.
Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348–351.
Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass, 6*(1), 36–49.
Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language, 46*, 341–370.
Selkirk, E. (1996). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Blackwell Publishing.
Shriberg, E. (1996). Disfluencies in switchboard. In *Proceedings international conference on spoken language processing (ICSLP)* (pp. 11–14). Addendum.
Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., … Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *International conf. On spoken language processing* (vol. 2, pp. 867–870). Banff: International Conf. on Spoken Language Processing.
Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In *Papers in laboratory phonology I* (pp. 72–106). Cambridge University Press.
Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology, 30*(2), 191–205. http://www.sciencedirect.com/science/article/pii/0022096580900570.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*, 25–38.
Sobel, D., Tenenbaum, J., & Gopnik, A. (2010). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*(3), 303–333.
Solan, L. (1980). Contrastive stress and children's interpretation of pronouns. *Journal of Speech & Hearing Research, 23*(3), 688–698.
Sonderegger, M., & Yu, A. C. L. (2010). A rational account of perceptual compensation for coarticu- lation. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 375–380).
Speer, S. R., & Ito, K. (2009). Prosody in first language acquisition – acquiring intonation as a tool to organize information in conversation. *Language and Linguistics Compass, 3*(1), 90–110.
Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29), 10393–10398.
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* (pp. 799–804).
Stephens, N. (2010). *Given-before-new: The effects of discourse on argument structure in early child language* (Ph.D. thesis). Stanford University.
Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development, 11*(2), 176–190.
Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (June 2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*(26), 10587–10592.
Stivers, T., Sidnell, J., & Bergen, C. (2018). Children's responses to questions in peer interaction: A window into the ontogenesis of interactional competence. *Journal of Pragmatics, 124*, 14–30.
Surian, L. (1991). Do children exploit the maxim of antecedent in order to interpret ambiguous descriptions? *Journal of Child Language, 18*(2), 451–457.
Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *The Behavioral and brain sciences, 24*, 629–640. discussion 652–791.
Turnbull, R., Royer, A. J., Ito, K., Speer, S. R., & Turnbull, R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience, 0*(0), 1–17.
Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.
Vieira, T., & Eisner, J. (August 2017). Learning to prune: Exploring the frontier of fast and accurate parsing. *Transactions of the Association for Computational Linguistics (TACL), 5*, 263–278.
Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences, 109*(32), 12933–12937.
Warren, P. (2017). The interpretation of prosodic variability in the context of accompanying sociophonetic cues. *Laboratory Phonology, 8*(1), 1–21.

Watson, D., Tanenhaus, M. K., & Gunlogson, C. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science, 32*(7), 1232–1244.
Wells, B., Peppe, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31*(04), 749–778.
Wightman, C. W. (2002). Tobi or not tobi?. In *Proceedings of Speech Prosody. Aix-en-Provence, France* (pp. 25–29).
Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience, 30*(6), 648–672.
Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 220–251.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245–272.
Yoon, S. O., & Brown-Schmidt, S. (2016). *Influence of the historical discourse record on language processing in dialogue*. Discourse Processes.