Intonation and pragmatic enrichment: how intonation constrains ad-hoc scalar inferences.

John M. Tomlinson, Jr. (tomlinson@zas.gwz-berlin.de)
Nicole Gotzner
Zentrum für Allgemeine Sprachwissenschaft (ZAS)
Berlin 10117 Germany

Lewis Bott
School of Psychology, Cardiff University, Cardiff, UK

Abstract

Pragmatic inferences require listeners to both derive alternatives and negate relevant alternatives to arrive at the speaker's intended meaning. The vast majority of research in Experimental Pragmatics has examined how pragmatic inferences arise from lexical triggers, however less has been done on how intonation can influence pragmatic inferences. Our investigation tests how pitch accents affect the processing of ad-hoc scalar implicatures in English. In the first experiment, L+H* accents led to more direct response towards pragmatically strengthened interpretations (compared to H* accents). The second experiment replicated this finding, in addition to showing a garden path effect towards pragmatically strengthened meanings. Together, we interpret these findings as evidence for the incremental integration of intonation into the processing of pragmatic inferences.

Introduction

The relationship between intonation and meaning is complex. To better understand the systematic contribution of intonation, a central aim in linguistics has been to distinguish the properly linguistic and non-linguistic functions of intonation (Lehiste, 1970; Ladd, 1996; Gussenhoven, 2002). This approach has lead to considerable advancements in our understanding of how intonation contributes structurally to referential meaning and discourse comprehension. Our investigation, and that of this special issue, examines cases in which intonational meaning goes beyond reference and also blurs this distinction between linguistic and non-linguistic functions of intonation. That is, we examine how intonation works at the level of inferential meaning. We start by discussing several reasons for shifting away from referential meaning towards pragmatic inferences. We then later integrate this into our broader goal, and that of experimental pragmatics in general, which is to develop both linguistic and psychological mechanisms that underlie pragmatic inferences.

Intonation can affect referential meaning and information flow in discourse in very predictable ways. For example, *pitch accenting*, the placement of pitch contours onto stressed syllables, correlates strongly with availability of referents in common ground, e.g. *new* vs. *given* information (Clark & Haviland, 1977; Prince, 1980, Pierrehumbert and Hirschberg, 1990; Baumann and Grice, 2006). However, intonation clearly affects other aspects of pragmatic communication and these have received much less attention. Consider the following example:

(1a) A: "I saw you stumble out of that new hip bar around the corner.
The one with the overpriced cocktails and 30 or so craft beers on
tap. Looked like you had a little of everything, right?"

(1b) B: "I drank some of the BEERs"

(1c) B: "I drank SOME of the beers"

Here, Speaker A teases Speaker B about how drunk she got at a local bar. While the suggestion that she drank a bit of everything is meant to be sarcastic, B understands A's comment as a request to list what and how much she drank. The example shows two replies that differ in their prosody. In (1b), B uses a pitch accent to increase the prominence of "beers". The communicated meaning is that she only drank beers, and did not have any cocktails. In (1c), however, the accent is placed on the quantifier *some*. In this case the communicated meaning is different. B now implies something about the quantity of available beers that she drank e.g. not all of the 30 beers, but it is harder in this example to derive that she did not drink any cocktails. In other words, different implications arise depending on the intonation.

The inferences in these examples arise because the listener reasons about *alternatives* to what the speaker said, that is, material that would have been relevant, but that the speaker did not say (Grice, 1975). For example in (1b), B's emphasis on *beers* means that the listener generates alternatives to *beer*, such as *cocktails*, and in (1c), the listener generates alternatives to *some*, such as *all*. Since the speaker did not say the alternatives, the listener potentially derives an enriched meaning of the utterance, concluding that the alternatives are false.[1] In our study we investigate how intonation alters this process. We present two experiments that test the role of intonation in deriving a specific type of pragmatic inference, namely *ad-hoc scalar implicatures* such as the inference that the speaker did not drink cocktails, arising from Example (1b). We start out by presenting an overview on the role of intonation in establishing reference and evoking alternatives. Then, we discuss work on the derivation of scalar implicatures and how intonation affects this process.

Evoking alternatives through intonation

---

[1] Note that evoking alternatives is the primary function of intonational focus marking (Rooth, 1992). Additionally, certain contexts license a pragmatic inference, negating relevant alternatives (see the section on scalar implicatures for further discussion.)

Intonation can perform a variety of linguistic functions across different contexts. We restrict our focus here to how intonation helps listeners evoke linguistic alternatives. The derivation of alternatives is essential for implicatures, in that the computation of implicatures requires listeners to reason about things that the speaker could have said, but did not. We review several pioneering and recent psycholinguistic studies on how different pitch accents can both evoke and integrate alternatives in identifying referents during comprehension. First, we note that one challenging aspect in every empirical investigation of intonational meanings is defining the exact nature of the intonational contour in question; namely these minimal pairs can be hard to uniquely identify in the continuous acoustic stream filled with both other meaningful prosodic parameters as well as noise. Like most psycholinguists, we are ultimately interested in meaning differences introduced by (perceived) categorically different intonational phenomena. Because of this, we limit our investigation to two pitch contours, namely the H* (L-L%) and the L+H* (L-L%) patterns[2], that have not only received much attention in the experimental phonetic literature, but also whose functions have been investigated under the rigorous scrutiny of psycholinguistic research using various online measures such as eye-tracking (see Watson, Gunlogson, & Tanenhaus, 2006 and Wagner & Watson, 2010 for a review).

The so-called H* accent is generally characterized as an increase in the relative prosodic prominence of a stressed syllable (Pierrehumbert, 1980). Dahan, Tanenhaus, & Chambers (2002) have shown that referents marked with such accents facilitate the processing of discourse new referents (as originally predicted by Pierrehumbert & Hirschberg, 1990). A second complex pitch accent (L+H*) has received both much at-

[2] We note quickly that we are not explicitly endorsing the Tone and Break Indices (ToBI - Beckman & Ayers, 1993) over other transcriptional systems, RaP (Dilley & Brown, 2005) for example, rather are simply using this notation as shorthand for structural differences between pitch accents that have been extensively researched in psycholinguistics and also appear to have reliable phonetic differences.

tention in the experimental phonetics literature as well as in psycholinguistics[3]. While the H* accent is characterized by a simple high target on the stressed syllable, the L+H* accent starts with a low target followed by a steep rise in pitch contour (late peak). Phonetic studies have shown that the L+H* accent has a higher pitch excursion, longer duration and greater intensity compared to the H* accent (e.g., Bartels & Kingston, 1994 and Krahmer & Swerts, 2001). An example of the two accent types is shown in Figure 2, depicting the average F0 values for our experimental items. Moreover, psycholinguistic work has shown that listeners distinguish these two accents during online processing. Yet the meanings of these accents seem to overlap in certain cases. For example, Watson, Tanenhaus, & Gunlogson (2008) investigated whether H* and L+H* accents help listeners predict referents among a set of alternatives in a visual world paradigm. In their experiment, listeners heard discourses such as "Click on the camel and the dog (establishes alternative set). Move the dog to the right of the square. Now, move the *camel/candle* (L+H* vs. H*) below the triangle". The last sentence tests whether the pitch accent pushes listeners to contrastive referents (look back towards camel) or new referents (candle) as camel was previously mentioned and established in a contrastive set, whereas candle was not. Critically, listeners' eye-movements anticipated contrastive targets and cohorts upon hearing L+H* on the first (overlapping) syllable, however were significantly less likely to look at unmentioned (new) targets and cohorts, e.g. candle. When hearing the H* accent on the other hand, listeners were equally likely to to look at both new and contrastive targets. This finding supports the idea that H* might be a more general prosodic marker for increased salience of referent in the common ground (a replication of Dahan et al 2002's findings), whereas the L+H* seems to create a bias to-

[3] The controversy has surrounded whether this accent represents a phonological contrast to the H* accent or whether these accents are indeed phonologically the same, albeit with phonetic differences. For example, Kügler & Gollrad (2015) show that listeners are more susceptible to the alignment and scaling of the f0 peak than to the preceding low tone (L+), suggesting that the difference between these two accents is more phonetic than phonological. We are agnostic regarding the exact nature of this difference (phonetic or phonological), as several psycholinguistic studies show that listeners can distinguish the meanings of these two contours in terms of the informational contribution.

wards an alternative referent that has been previously mentioned in the discourse. In addition, an experiment by Ito & Speer (2008) showed that the L+H* led listeners on a garden path such that they anticipated an alternative referent even when the display did not present an item contrasting in the relevant dimension. For example, when the noun phrase *blue angel* was followed by *GREEN ball*, listeners erroneously fixated on angels for 200 ms and only afterwards fixations turned to the target.

As we have seen, the H* and L+H* accents help listeners establish reference, albeit in different ways, by allowing listeners to rapidly integrate alternatives in common ground during online processing. However, how do these accents evoke alternatives when they are not explicitly provided in the context? Braun & Tagliapietra (2010) found that the L+H* primed alternatives to the accented item when no set of elements was contextually-introduced (e.g., the word *slipper* activated the word *flip flop*). The H* accent, on the contrary, did not lead to priming effects. These results indicate that the L+H* accent automatically activates alternatives. In addition, a study by Husband & Ferreira (2015) indicates that the L+H* can help restricting the set of alternatives. More specifically, the study showed that initially listeners activated a broad set of elements and only at a certain amount of time decay the L+H* made relevant alternatives salient. If we compare these results to those reviewed above (in which alternatives are already active in the context), we see that listeners need more time, e.g. processing effort, to evoke alternatives and converge on the relevant ones when they are not already provided in the context. One way to reconcile these findings is to assume that the activation of alternatives underlies two mechanisms: One mechanism, which activates a cohort of alternatives and a second mechanism which selects the contextually-appropriate alternatives (see especially Husband & Ferreira, 2015 and Gotzner, Wartenburger & Spalek, 2016). Intonation seems to play a crucial role in both of these processes. If, however, alternatives are already active in the context, the L+H* primarily makes these contextual alternatives more salient, as found in a probe recognition experiment by Gotzner et al. (2013) (see also Fraundorf et al., 2010 demonstrating such effects in long-term memory). To summarize, previous studies indicate that the L+H* can introduce additional alternatives as well as make specific alternatives more salient, depending on whether alternatives are contextually available or not. We now address how intonation affects how listeners integrate alternatives when deriving pragmatic inferences. Specifically, we argue that listeners have to not only integrate these alternatives as in the case of establishing referential meaning, but also need to recognize which alternatives are pragmatically stronger than the original statement and negate these alternatives (and combine it with the basic meaning) to derive the inference.

Scalar implicatures and the role of intonation

The paradigmatic case of how listeners use alternatives to derive pragmatic inferences are *scalar implicatures* (Grice, 1969, Horn, 1972, Gazdar, 1979). Scalar implicatures arise when a speaker uses a weak expression when a stronger expression would have been relevant. There are a variety of different approaches to how scalar implicatures are derived but most researchers agree on the following: (i) the listener computes the basic meaning of the phrase, (ii) recognizes that an alternative phrase could have been used, but that it was not, (iii) negates the alternative and combines it with the basic meaning (modern and developed theories can be found in van Rooij and Schulz (2004), Sauerland (2004), and Chierchia, Fox & Spector (2012), amongst others). In (1c), for example, the speaker said *some*, and the listener recognizes that they could have used the stronger expression, *all*. The listener then negates the alternative (*not all*) and combines it with the basic meaning of the sentence to form *some but not all*.

Scalar implicatures are *optional* components of meaning. The listener is not obliged to derive an implicature in the same way that they are obliged to derive the basic meaning of a sentence. Thus the listener must make a decision about whether the speaker intends an implicature. This decision depends on a range of pragmatic and semantic factors. From a pragmatic perspective, the goals and knowledge of the speaker are relevant (Grice, 1975). For example, if the speaker is not knowledgeable enough to know whether the stronger expression is true, the listener should not draw conclusions on the basis that they did not use the stronger expression. Structural factors also seem to play a role in that certain grammatical environments block scalar implicatures (see e.g., Chierchia, 2004). Although implicatures are context-dependent, an underlying question about scalar implicatures has been the extent to which the derivation of scalar implicares is derived by default (Levinson, 2000).

Indeed, researchers in experimental pragmatics have spilled much ink over this question. To this end, most previous psycholinguistic studies have focused on implicatures in which the relevant alternatives could be stored in the lexicon, such as quantifiers (*some*) or disjunction (*or*) (e.g., Bott & Noveck, 2004; Bott, Bailey & Grodner, 2012; Tomlinson, Bailey & Bott, 2013; Chevallier et al., 2008; Huang & Snedecker, 2009; Degen & Tanenhaus, 2014; among many others). The general consensus is that scalar implicatures are not derived by default as Levinson (2000) originally claimed. However, there is still much disagreement about the exact mechanism and constraints that underlay the derivation of scalar implicatures. While lexical triggers have received much attention, very little work has examined how intonation factors into the derivation of scalar implicatures and the mechanisms underlying them.

Intuitively, intonation should contribute to the decision about whether implicatures are derived. This can be seen from (1c), in which stress on *some* makes it clearer that the implicature is relevant. Two previous studies have investigated how intonation can affect the accessibility to alternatives during the derivation of scalar implicatures. Chevallier et al. (2008) found that participants were more likely to derive the *not both* implicature associated with *or* (e.g., “You can have the fish or the meat course” implies *not both* courses). They presented participants with letter strings, such as, “TABLE”, for very brief display times and asked questions about the letters in the word. For the crucial questions, the answer was “false” if the participant derived an implicature, but “true” if they did not. For example, for “TABLE”, the question was, “Is there are a T or a B?” The participant answered “false” if they understood *or* to mean *T or B but not both* (the implicature; since both T and B were present), or “true” if they understood *or* to mean *T or B and possibly both* (the semantic meaning). They found that when *or* was stressed there were more false responses (implicature responses) than when there was no stress on *or*. As such, their findings show that intonation increased the rate of implicatures (see Schwarz et al., for a similar finding). One critical drawback of this study is that no phonetic or phonological characterization of the stimuli is provided by the authors, hence it is difficult to see which intonational aspects were driving the effect and how this work relates to previous studies concerning the role of pitch accents in alternative activation (e.g., Braun & Tagliapietra, 2010).

A recent study by Gotzner and Spalek (2014), which does provide a phonetic and phonological breakdown of their stimuli, tested *ad hoc* scalar implicatures (Hirschberg, 1991; and see Katsos & Bishop, 2011, for an example). For these, the scale is constructed according to the context (on an *ad hoc* basis, similar to example 1b). Gotzner and Spalek presented short discourses to participants in which the critical sentence contained an L+H* accent or an H* accent on a noun in subject position (e.g., *The JUDGE followed the argument*). They asked true/false interpretation questions that were false if the participant derived the implicature and found that the rate of implicature was higher in the L+H* condition than the H* condition (just as in Chevallier et al.).

While both of these studies demonstrate a link between scalar implicatures and pitch accenting, they suffer two limitations. First, the effects of intonation on the derivation of pragmatic inferences might be conflated with lexical retrieval. For example, Chevallier et al. (2008) only investigated lexical scalar implicatures triggered by *or*. Since *or* and its alternative *and* are conceivably linked in the lexicon (e.g., Horn, 1972; Levinson, 2000), their results might be restricted to lexical scalar implicatures. To remedy this, we test ad-hoc scalar implicatures as did Gotzner and Spalek (2014). Second, the methodology limits the range of conclusions that can be made about the underlying processing mechanisms. For example, both studies focused on final interpretation judgments and do not report online processing measures, such as response times, eye fixations or mouse paths. Although these are likely to be related, a higher rate of implicature does not necessarily equate to faster processing (see e.g., Bott et al. 2012). Our study takes the novel step by addressing how intonation affects processing of scalar implicatures, and not solely interpretation

Critically, processing data can shed light on the exact mechanisms involved in the inferential process. In the case of intonation, there are at least three ways pitch accents could affect these mechanisms. First, intonation might not play a role in processing, rather it might affect only the rate of implicature derivation. For example, L+H* could simply be an additional signal to the listener that they derive the implicature, in which case L+H* would lead to a greater rate of implicature (as in the results of Chevallier et al. and Gotzner & Spalek) but not necessarily a processing advantage. Second, L+H* might reduce the number or complexity of the mechanisms used to arrive at the decision, which would lead to a processing advantage. For example, L+H* intonation could speed up how quickly alternatives are negated. Finally, L+H* could direct the listener to derive the implicature at a different time during sentence processing, but otherwise leave the mechanisms involved in derivation unchanged. For example, L+H* might license the listener to derive implicature early in sentence processing rather than waiting until the speaker has finished the utterance.

Our studies investigate these possibilities by addressing two questions: 1) whether a particular type of pitch accent (L+H*) affects only the rate of implicatures or both the rate and processing of scalar implicatures (Experiments 1 and 2) and 2) whether this pitch accent might license the early derivation of and creates a preference for stronger alternatives (Experiment 2). To obtain more fine-grained information on the mechanisms behind implicature computation, we use a novel experimental paradigm with mouse tracking. While the speed of processing is an important part of our investigation, the hypotheses mentioned above cannot be distinguished solely on how quickly a listener derives an enriched meaning. By simultaneously presenting alternatives to the listeners, we can better isolate any effects that intonation might have on mechanism involved in deriving the implicature, e.g. rejecting alternatives. To achieve this, we created a sentence picture matching task to see how a competitor image depicting an alternative meaning interferes with how listeners select the correct target image depicting the implicature meaning. As such, any interference of a competitor image on participants' mouse paths towards the target would allow a better window into how listeners consider and exclude a meaningful alternative before arriving at the enriched meaning.

In our study participants heard a sentence that could be interpreted with an implicature, a *strong* interpretation, or without an implicature, a *weak* interpretation. Two pictures were paired with the sentence, one of which was consistent with the weak interpretation and one with the strong interpretation. Participants' task was to click on the picture which matched the sentence and we measured their mouse trajectories towards the picture representing the strong interpretation. Our critical manipulations were the extent to which participants would be distracted by the weak interpretation when choosing the strong interpretation and whether the sentence included the L+H* intonation or the H* intonation. If L+H* affects the mechanisms underlying scalar implicatures by eliminating processing steps or licensing the derivation of the inference earlier on in processing, participants should decide for strong interpretations more readily and be distracted less by a weaker interpretation with the L+H* accent than the H* accent, as we describe below.

Experiment 1

Participants heard a sentence about a boy called Mark and some objects in his possession. They had to say which of two pictures best matched the meaning. In *inferential* conditions, both pictures were consistent with the basic meaning of the sentence, and participants had to make an inference to form an unambiguous interpretation. One picture was the *weak* interpretation and one was the *strong* interpretation. The weak picture included the reference object and an object that was not mentioned in the sentence. For example, the weak picture for "Mark has a candle," was a candle and a camel. The interpretation was weak because if participants did not enrich "Mark has a candle", then it was consistent with Mark having a candle and a camel, and indeed any other object. The strong interpretation included only the reference object. For example, the strong picture of "Mark has a candle," was, simply, a candle (and nothing else). This interpretation required the participant to reason along the lines of, "since the speaker did not say that Mark had two objects, and they are being truthful and informative, they must believe that Mark has only one object", that is, the interpretation required a scalar implicature. For comparison, we added *control* conditions, in which one picture was consistent with the sentence and one was not, and so the response was unambiguous. For example, "Mark has a candle," was accompanied by a picture of a candle on the left and a camel on the right, and so the correct interpretation (the *target*) was the candle picture.

We also manipulated the pitch accent on the referent (e.g., "candle"), which received either an H* or L+H* accent. Thus if the L+H* accent interacts with the implicature, there should be a greater effect of the L+H* in the experimental trials compared to the control trials.

Our main-dependent measure was the mouse-trajectory of the response. The cursor appeared at the bottom-centre of the screen at the beginning of each trial and the participant moved the mouse from the start to the response option (top left or top right corner of the screen - see Figure 1). Ease of processing was measured by the directness of the mouse path from initial position to response (see Spivey, 2008). Mouse-trajectories that are indirect, e.g., involving a wide arc from start to finish, indicate extensive processing, while those that are direct, e.g., a straight line from start to finish, indicate minimal processing (Freeman & Ambady, 2010). We used mouse trajectories as a measure of processing rather than button-press reaction time for the following reasons. First, mouse dynamics provide information about cognitive processes intervening between stimulus and completed action (e.g., Tomlinson et al., 2013), whereas button-presses only provide information about the completed action. Second, and relatedly, mouse trajectories provide a measure of incremental sentence interpretation. Participants change their mouse paths as the interpretation of the sentence changes. Finally, mouse tracking allows multiple response options whereas button-press paradigms become difficult for the participant with more than two. In Experiment 2 we made use of all three of these properties.

To summarize, we had two types of interpretation conditions, control and inferential, and two pitch accents, L+H* and H*. These were crossed in a 2 x 2 design, with mouse trajectories as the dependent measure. If processing of scalar implicatures is facilitated by the L +H* accent, we should observe more direct mouse paths in the inferential condition than the control condition.

Method

Participants. Twenty-eight Cardiff University students participated for course credit or 3 pound Sterling. The experiment took roughly 15 minutes to complete.

Sentences and pictures. Inferential trials and control trials used the same sentence frame, “Mark has a [X]” where X was an object (see Table 1). The response options differed however. For inferential conditions (see Figure 1), the strong response option was a picture of one object, X, and the weak response option was a picture of two objects, X and Y, where X was an object named in the sentence and Y was not. The strong response option was consistent with the strong meaning of the sentence, *X and nothing else*, and the weak response option was consistent with the weak meaning, *X regardless of whether something else was present*. While both meanings were logically consistent with the sentence we anticipated that participants would generally choose the strong meaning, and so we designated the strong response option as the target. For control conditions (see Figure 1), both response options were pictures of only one object. However, only one response option was a picture of X, the object named in the sentence, and so this was designated the target.

Roughly half of the experimental items (22 items) were adapted from Dahan, Tanenhaus, & Chambers (2002) and the other half (20 items) were created to increase the number of items. Of these, half of the sentence and picture combinations were phonological competitors, e.g. candle vs. camel and the other half were semantic competitors, e.g. pencil vs. eraser. This was done to help disguise the purpose of the experiment. Similarly, 30 filler items were created and varied among several dimensions to prevent listeners developing strategies during the experiment.

The pictures consisted of black and white clip art pictures found from an internet search engine. The objects were the same size in one object pictures as in two object pictures. This was done to control the salience of a one-object and two-object pictures. Response options were placed in the top left and the top right corners of screen.

Prosody. A male speaker of British English with no noticeable regional variety recorded the sentences. Sentences were recorded in a sound attenuated booth using a unidirectional microphone and digitized with a USB sound capture device. All utterances were first recorded in carrier phrase ("Mark has a candle") in both H* and L+H* forms. A trained phonetician inspected these recordings and made sure that utterances had either H*L-L% patterns or L+H*L-L% patterns. These patterns were further tested by an analysis measuring the stylized F0 difference between the patterns across all items at 5 normalized time points across the stressed syllable as used in Watson et al. (2008) and Spalek, Gotzner, & Wartenburger, (2014) (see Figure 2). The F0 values were measured at 5 equal intervals in the syllable receiving the pitch accent. Figure 2 shows that the H* tone of the L+H* pitch accent is realized as a plateau between the 3rd and the 4th F0 interval. The mean F0 was not statically different between H* and L+H* items at the first and last intervals (111.5 vs. 118.9 Hz, t= 0.43, p=.66; 108.8. vs 122.1 Hz, t= 0.78, p=.44), however significantly differed across the 2nd, 3rd ,and 4th intervals (t's > 2.21 and p's <.03). L+H* and H* items differed also in where in the syllable the F0 peak was reached: F0 maxima were found on average 62 ms later in the syllable for L+H* items than H* items, t= 2.71, p<.01. The accented syllables for L+H* items had slightly longer durations than accented syllables in H* items (332ms vs. 319ms), however this difference was not significant, t= 0.64, p=.54. Further acoustic measurements are summarized in Table 2.

Next, a single sentence frame ("Mark has a") was spliced into each of the recordings, so that each item had the same sentence frame across conditions[4]. This was done to assure that the acoustic stimuli only differed in pitch accent and not in acoustic infor-

[4] Upon inspection during the review process, we noticed that 4 of the 42 items had the frame "Mark has" with a pitch accent on Mark, e.g. MARK has a CANDLE/candle. Because this was present both versions of the 4 items (L+H* and H*), it would equally affect both conditions.

mation preceding the referent. All items were scaled for intensity at 68dBs using Praat (Boersema & Weenick, 2015).

Counter-balancing. The assignment of item to condition was counter-balanced so that (i) each item occurred equally often in each condition across participants and (ii) participants never saw the same items twice. A Latin square design was used so that participants only saw one version of an item across the 4 cells (H* control, L+H* control, H* inferential, L+H* inferential). This meant there were four counterbalancing lists. Each list had 20 or 21 experimental items and a given participant saw 5 or 6 items in each condition as well as 30 filler items. The position of the target response was also counterbalanced, so that it appeared equally at the left top and right top positions throughout the experiment.

Filler trials. In addition to experimental trials there were filler trials. Filler trials made up roughly 60% of all trials. All filler sentences had two referents, e.g., "Mark has a [X] and a [Y]" where X and Y were objects. They were included to prevent participants from adopting strategies in which they only considered one of the objects in the sentence. There were 3 versions of filler trials (10 of each). Participants either saw a one object vs. two object picture display as in the inferential conditions, however unlike the inferential conditions heard utterances mentioning two referents. This type of filler made it equally possible that a participant would hear a one referent or two referent utterance when seeing a picture display that had one vs. two object response options. Participants also saw picture combinations in which both responses were two object responses. Half of these trials had cases, in which the response options shared the first initial object, and therefore participants had to use the second referent to distinguish between the two responses. The different pitch accents (L+H* and H*) were intermixed amongst them, so that listeners would be equally likely to hear both pitch accents across all picture combinations[5].

Procedure. The experiment was conducted using the Runner program available in the Mousetracker suite (Freeman & Ambady, 2010). The instructions described to participants that they were overhearing someone describe which objects a fictitious child (Mark) had on his desk. To start each trial, participants had to click on a START button at the bottom of the screen. Two pictures subsequently appeared at the upper corners and 2s later the audio stimulus was presented. The participants could start moving their mouse at the onset of the audio stimulus and their movements were recorded at the onset of the word “has”.

Results

Participants’ responses were analyzed for both accuracy and the directness of the mouse-path towards the target response. Accuracy rates were at ceiling (over 97%) for both inference and control conditions (See Table 3). The data files were pre-processed using the Analyser program, which normalized participants’ raw mouse trajectories into 101 time steps in order to compare the geometrical spatial attraction towards the targets across responses with different response times. Responses with reaction times 3 standard deviations outside of the grand mean of all responses were excluded. This amounted to roughly 1.3% of the entire data.

The main dependent measure was Area under the Curve (AUC). This measure calculates the total geometrical area between a straight line from the starting point of a response (the START button) and the correct response and the participant’s actual response. Participants’ mouse trajectories (average y- over x-coordinates as well as the raw tra-

[5] A reviewer pointed out that participants could predict based on some of the picture combinations (A vs. B and AB vs. AC or AB vs. CD) whether the upcoming utterance contained a single referent or two referents. While this could have affected the control conditions in Experiment 1, the same type of predictions could not be made for the inferential conditions. Ideally, we could have created another filler condition, in which participants saw two object response options and heard an utterance with a single referent. Such a condition (non-inferential condition) was added in Experiment 2 because a four response option display was better suited to test such a condition.

jectories) are shown in Figures 3 and 4, for the control and the inferential conditions respectively.

Two linear mixed-effect regression models were fitted to the data. The first model made use of the 2 x 2 design and used the two experimental factors as predictors, pitch accent (H* vs. L+H*) and sentence-picture combination (inferential vs. control). Similarly, a simple effects model, in which the 4 levels were collapsed into one fixed factor (condition), was also constructed to examine pair-wise comparisons and for consistency across the two experiments[6]. All of the models included random slopes per item and per participant (as suggested by Barr et al 2013). The formula for the simple effects model as well as the coefficients, t-values, and p-values for the AUCs values are summarized in Table 3.

First, the omnibus mixed-model used two fixed factors (pitch accent and picture condition) to predict the area under the curve values shown in Table 3. In this model, the factors were sum coded. Critically, there was an interaction of pitch-accent and sentence-picture combination, $\beta = 0.15$, $t = 2.05$, $p < .05$. In the inferential condition, listeners' mouse paths were significantly more direct when hearing L+H* accents than when hearing H* accents. In the control condition, however, participants' mouse paths towards the target image did not differ significantly.

This interpretation was further supported by two simple effects models, which provided pair-wise comparisons in each level of the picture combination factor. In the first model (Model 1 in Table 5), the H* Control condition was used as the reference variable (intercept) in the model to test the effect of pitch accent in the control condition (treatment coding). There was no significant difference for AUC values between H* and

[6] We only provide the details of the simple effects models in Tables 3 and 5. This was done for two reasons. First, we do not have an a priori theoretical reason to compare the main effects between the levels of the picture conditions (control vs. critical), rather only care about the differences or lack thereof within the levels of picture conditions. Second, the design of Experiment 2 renders the interpretation of the interaction term more difficult.

L+H* accents in the control condition, β = -0.06,t = 0.68, p=.49, suggesting that participants' mouse paths towards the target image did not differ significantly in the control conditions. In the second model (Model 2 in Table 5), the reference variable (intercept) was changed to the H* Inferential condition to test the effect of pitch accent on AUC values in the inferential picture condition. In the inferential conditions, there was a difference between pitch accents, β = -0.28, *t =3.15*, *p*<.002, showing that listeners had more direct mouse paths towards the target picture when hearing the L+H* than the H* accent.

Discussion

When participants heard sentences where a strong interpretation was relevant, that is, the inferential conditions, they overwhelmingly selected the strong interpretation as the most plausible meaning (97%). Participants thus saw nothing unusual about the task nor had difficulties understanding the intended meaning. More importantly, mouse paths were more direct when referents had L+H* pitch accents than when they had H* pitch accents. When the strong interpretation was not relevant, that is, the control conditions, there was no advantage for the L+H* accent. Our results therefore show that L+H* facilitates the derivation of scalar implicatures.

While our findings are consistent with those of Chevallier et al. (2008), Schwarz (2008), and Gotzner (2013), they also add to them. Because we observed very high rates of implicature under both L+H* and H*, our results cannot be explained by an account that predicts implicatures are only more likely under L+H*. Our effects relate to the processing of implicatures and not the overall rate of derivation. Furthermore, we also demonstrate that facilitation is not restricted to lexical implicatures such as *or* or *some*. Instead, L+H* must affect mechanisms that are not related only to lexical retrieval of alternatives. In Experiment 2 we explore the mechanisms underlying the facilitatory effect of L+H*.

Experiment 2

Implicatures are optional components of meaning, and as such the listener needs to decide when to incorporate an implicature and when not. The decision is complex and requires considering many contextual factors such as the speaker's knowledge (e.g., Grice, 1969; Sauerland, 2004). One explanation for the results of Experiment 1 is that the L+H* reduces the optionality of the implicature, in effect making it obligatory. This would simplify the formation of the strong interpretation because the listener would need to assess fewer contextual variables before incorporating the additional meaning into the sentence (although the content of the implicature would still need to be derived). L+H* would act rather like an explicit focus operator, such as *only*, as in "Mark has only a candle." A simpler process would suggest the direct mouse paths we observed in Experiment 1. Support for this hypothesis can be found in the offline data by Gotzner & Spalek (2014), who show that implicature rates are equally high with the L+H* and explicit *only*. Furthermore, it is very difficult to cancel an implicature with an L+H* accent relative to an unmarked accent. For example, (4) is less felicitous than (5).

(4) I drank SOME of the beers. In fact I drank them all. (!)

(5) I drank some of the beers. In fact I drank them all.

Since it is only possible to defease a meaning if it is optional, the contrast between (4) and (5) suggests that L+H* influences optionality.[7] One prediction of such an account is that the implicature should arise immediately after the pitch accent is recognized. There would be no need to wait for further evidence of the speaker's knowledge etc. because

[7] In a similar vein, Fretheim (1992) argues that with focus intonation the enriched meaning of numerals is a semantic entailment.

the prosody signals that this is guaranteed. Further support for this prediction comes from Kim et al. (2015) and Filik, Patterson & Liversedge (2009), who demonstrate that (explicit) *only* is processed incrementally during sentence comprehension in that listeners generate and disregard alternatives prior to having processed the entire sentence. In Experiment 2 we test this prediction by investigating the incremental nature of the L+H* effect we observed in Experiment 1.

We made two changes to the design of Experiment 2. First, we embedded the sentences from Experiment 1 in a prepositional phrase. Objects were either on a table or a shelf, such as, “Mark has candle on the table/shelf.” This meant that the reference object and accompanying prosody were no longer sentence final. Furthermore, since the location of the objects varied from trial to trial, participants needed to listen to the complete sentence before they were able to select the correct interpretation (note that increasing the variability in the sentence also meant that we had to increase the number of response options from two to four). Second, in addition to the inferential and control conditions from Experiment 1, we included *non-inferential* conditions. These were trials in which deriving the implicature mid-sentence would “garden-path” participants to the incorrect sentence interpretation. Sentences were the same as in the other conditions but the pictures indicated either a strong interpretation at an incorrect location, or a weak interpretation at a correct location. For example, “Mark has a candle on the table,” would be accompanied by pictures of a candle on the shelf, and a candle and a camel on the table (and two irrelevant response options). Deriving the implicature on “candle” mid-sentence would direct participants to the picture of only a candle (on the shelf), whereas if participants waited until the end of the sentence, they would go directly to the correct interpretation, that is, the picture of a candle and a camel (on the table). L+H* or H* accents were placed on the reference object, just as in the other conditions. Thus, if L+H* obliges participants to derive the implicature mid-sentence, mouse trajectories should first be directed at the strong response option, but subsequently double back towards the weak (and correct) interpretation. Of course, *ad-hoc* implicatures might be derived incrementally even with an unmarked accent (e.g., Breheny, Ferguson, Katsos, 2013), in which case garden-path effects will be observed for both accents. However, there should be a greater garden-path effect with L+H* accent than the H* accent.

Method

Participants. Sixty Cardiff University students participated for course credit or a 3 pound Sterling reimbursement. The experiment took roughly 25 minutes to complete.

Design. The pitch factor involved two levels, L+H* and H*, as in Experiment 1. However, there were now three levels of the sentence interpretation: inferential, control, and non-inferential. Thus the design was a 2 x 3 factorial. Table 3 illustrates the design in more detail.

Sentences and pictures. The sentences for the inferential, control and non-inferential conditions all followed the same form, "Mark has a [X] on the table", where X was an object from the list used in Experiment 1. As in Experiment 1, the pictures in the response options distinguished between the conditions, however instead of two pictures at the top corners of the screen, there were four response options in total, e.g. on each corner of the screen. There were also filler trials so that participants had to pay attention to all of the objects in the sentence and the prepositional phrase, as we describe below.

The addition of the non-inferential condition and the prepositional phrase meant that we needed to increase the response options from two to four. Thus there were pictures in all four corners of the screen. In the inferential conditions (see Figure 5), the options were (i) X and Y on the shelf, (ii) X on the table, (iii) A on the shelf, (iv) A and B on the shelf, where X was the object mentioned in the sentence and A, B and Y were objects not mentioned in the sentence. In the non-inferential condition, the options were (i) X on the shelf, (ii) X and Y on the table, with (iii) and (iv) as above. In each case option (ii) was the target. Finally, in the control conditions (see Figure 5), the options were (i) X on the shelf, (ii) X on the table, with (iii) and (iv) as above. Table 3 shows examples. The same 30 filler items from Experiment 1 were used and these trials also consisted of different type of picture combinations: roughly one fourth using the same combination as experimental trials (i) X and Y on the shelf, (ii) X on the table, (iii) A on the shelf (iv), A and B on the shelf and the other three fourths consisting of different patterns picture arrangements.

Counterbalancing. The assignment of items to conditions was counterbalanced in a similar way to Experiment 1. This meant that there were six counterbalancing lists and each list showed 42 experimental items (8 in each condition) along with 30 filler items. The position of the correct response (top-left, top-right, bottom-right, and bottom-left) was rotated randomly across lists so that participants were equally likely to see the correct response at each of the four corners across the experiment.

Prosody. The same experimental items and fillers were used from Experiment 1. They were modified to fit the four-picture response paradigm by splicing a prepositional phrase indicating the object location ("on the shelf" and "on the table") to the carrier phrase. The prepositional phrases were recorded by the same speaker and with the carrier phrase. All items were again scaled for intensity using Praat (Boersema & Weenick, 2015) after the prepositional phrases were spliced onto the items and fillers.

Procedure. The experimental procedure was the same as in Experiment 1. However, the number of counterbalanced lists was increased to six. Participants were presented with all four picture targets 2000ms before the onset of the auditory stimulus. The only difference was that the START button found located in the middle of screen as opposed to the bottom of the screen as in Experiment 1.

Results

The accuracy rates for the inferential and control conditions used in Experiment 1 were at ceiling (over 97%) but the non-inferential condition had slightly lower accuracy rates, around 90% (see Table 6). The dependent measure and data preprocessing for the mouse-tracking data were the same as in Experiment 1.

The plots of average response trajectories over x- and y- coordinates for the experimental conditions are shown in Figures 6, 7, and 8. An omnibus mixed-effects model with two factors (pitch accent, 2 levels, x picture condition, 3 levels), revealed a significant interaction between pitch accent and picture, $\beta = -0.35$, $t = 3.62$, $p < .001$, as would be predicted if L+H* facilitated processing of the inference sentences and impaired processing of the non-inference sentences. However, a variety of other patterns could also explain the effect. We therefore conducted three, one factor simple effects models to identify our effects in greater detail. The statistics for these models are shown in Table 7.

The simple effects models replicated the findings from Experiment 1 for inferential and control conditions: Listeners had more direct mouse trajectories with L+H* patterns than H* patterns in the inferential condition, whereas there was no difference between these two accents in the control conditions (Figures 7 & 6). As in Experiment 1, the L+H* accent in the control condition did not differ from the H* accent in the control condition (see Model 1 in Table 7), $\beta = 0.0003$, $t = 0.014$, $p = .98$. To test whether the L+H* pattern in the inferential condition differed significantly from the H* pattern in the inferential condition, the model was rerun (see Model 2 in Table 7) by setting the H* accent in the inferential condition as the reference variable (intercept). This was indeed the case and replicated the findings from Experiment 1: the AUC values for the L+H* in the inferential condition were significantly smaller than those for the H* condition in inferential condition, $\beta = 0.062$, $t = 2.72$, $p < .01$.

In the non-inferential condition, both pitch accents showed a garden-path effect as can be observed in Figure 8. To test, whether the L+H* accent induced a stronger garden path effect than the H*, a third simple effects model (see Model 3 in Table 7) was constructed by setting the H* accent in the non-inferential condition as the reference variable (intercept). The L+H* accent had significantly larger AUC values than the H* accent in the non-inferential condition, $\beta = 0.075$, $t = 3.13$, $p < .002$. This suggests that the L+H* accent serves as a stronger trigger to derive the implicature than the H* accent.

Discussion

Experiment 2 showed similar results to Experiment 1. In the inference condition, mouse paths to the strong interpretation were more direct with L+H* than with H* but in the control condition they were not. Experiment 2 therefore provides further evidence that the L+H* accent facilitates the derivation of scalar implicatures. In both experiments, we found that L+H* accents facilitates the processing of ad-hoc scalar implicatures.

In Experiment 2 we added non-inferential conditions. If implicatures were derived incrementally, mouse trajectories should have initially been directed at the strong interpretation, but subsequently redirected towards the weak interpretation. In other words they should have shown a garden path effect. This was indeed the pattern of results. More importantly, the garden path effect in the L+H* condition was greater than in the H* condition. Consequently we argue that the L+H* pitch accent facilitates implicature processing by facilitating an earlier derivation of the implicature. The more direct mouse paths observed in the inference condition and the less direct paths in the non-inferential condition were due to the less complex processing necessary to consider whether to include the inference in the completed sentence interpretation.

Experiment 2 also allows us to exclude an alternative explanation for the facilitative effects of L+H*. In Experiment 1, the steeper fall in F0 under L+H* could have served as a signal to the listener that the speaker was finishing her turn (De Ruiter, Mitter, Enfeld. 2006). This meant that the more direct mouse paths for L+H* accents could have been due to turn-taking cues instead of mechanisms involved with pragmatic inference. Embedding the referential expression in a prepositional phrase, as we did in Experiment 2, meant that L+H* was no longer at the end of the sentence. Thus, even when participants could no longer use L+H* to predict the end of a speaker's turn, we still observed more direct mouse paths to the strong response option.

General Discussion

Our goal was to investigate how pitch accents affect the processing of ad-hoc scalar implicatures. In our experiments the alternatives were readily available to participants, which meant that we could test how pitch accents affect the inferential process beyond the retrieval of alternatives. In Experiment 1, we found that L+H* facilitates the processing of ad-hoc scalar implicatures. In Experiment 2, we replicated the findings from Experiment 1 as well as showing that listeners integrate information from pitch accents incrementally. On this basis, we suggest that L+H* accent makes it more likely that listeners will derive the implicature early during sentence processing. Below we explain how this might work and the implications of our findings.

Intonation facilitates incremental implicature processing

In Experiment 2, we argued that the L+H* accent signaled that the listener need not consider contextual factors associated with speaker knowledge. Just as with (explicit) *only*, L+H* signals that the listener should derive the implicature as soon as possible during sentence interpretation. L+H* would cause the listener to bypass the consideration of speaker's beliefs, thereby speeding up the implicature process (but at the cost of potentially over-committing to the implicature, as in the case of our experiment). We note that the garden path effect was also found for utterances with H*, but to a lesser extent. Thus, participants derived the implicature early on some occasions even without

L+H*. L+H* increases the probability of an early implicature but this might happen anyway to some extent.

While this account of L+H* is consistent with the intuitive difficulty of defeasing L+H* and the offline results of Gotzner & Spalek (2014), there are no other studies linking L+H* with speaker knowledge effects in implicatures (moreover, we have no direct evidence that knowledge effects play a role in our study). This suggests that other accounts of our findings might be more parsimonious overall. We can see two possibilities. The first is that L+H* greatly elevates the salience of the relevant alternatives relative to H*. More salient alternatives might make the implicature more likely overall and difficult to cancel. This account would be consistent with the conclusions of Husband and Ferreira (2015), Fraundorf et al. (2010), and Gotzner et al. (2013), all of whom claim that L+H* makes relevant alternatives salient. The second is that L+H* activates the mechanism that negates the alternatives (or combines the negated alternatives with the basic meaning). The alternatives might already be present in the listener's representation (L+H* does not make them more salient) but there might still be doubt about how to use them. L+H* might remove this doubt to a larger extent compared to H*.

In summary there are three possibilities. L+H* could (i) elevate the salience of the alternatives (ii) promote their use, or (iii) eliminate the need to consider speaker knowledge. We suggest that (i) is unlikely to account for the findings of the current study. While it is parsimonious, in that it draws together previous findings with ours, there are a number of factors that remain to be explained. First, Gotzner & Spalek (2014) found very similar inference rates with *only* as with L+H*. Even though L+H* could raise the saliency of the alternatives by any degree, the level of saliency might still differ for L+H* and *only* (as was the case in Gotzner et al., 2013). In contrast, an account that assumes *only* and L+H* both oblige the listener to derive the implicature makes exactly the prediction that L+H* should facilitate implicatures and lead participants to overcommit to it. Another reason to doubt that saliency alone can explain the facilitation of implicatures is that the alternatives in our task were presumably very salient to participants, even without L+H*. The alternatives were easily derived from the response options and we included filler trials that explicitly introduced the alternatives. Furthermore, rates of implicature were extremely high throughout the experiment, which would not have been the case if the alternatives were difficult to retrieve. Under these circumstances it is difficult to see how L+H* could make the alternatives any more salient than they already were. Finally, there is no experimental evidence that salient alternatives make the implicature difficult to defease. It is an interesting possibility, but one that so far remains to be tested.

Deciding whether L+H* affects the use of the alternatives (ii), or the knowledge mechanisms (iii), is more difficult. Our opinion is that (iii) is more likely. The reason is that consideration of the speaker's knowledge is generally assumed to be a separate and necessary mechanism in the derivation of implicatures (e.g., Grice, 1969; Sauerland, 2004; Geurts, 2010; Bergen & Grodner, 2012). In contrast, there is no evidence that there is autonomous mechanism that determines whether to use alternatives. While clearly some processes integrates the enrichment material with the basic meaning, the mechanism responsible for this may not be independent from the knowledge mechanisms or even the saliency of the alternatives (use of the alternatives might be automatic under knowledge conditions). Thus, there may not a mechanism for the L+H* accent to affect.

We hope that future experiments can clarify how L+H* interacts with knowledge effects (or not, as the case may be). Interesting possibilities include testing whether speaker knowledge manipulations affect L+H* marked implicatures differently to H* implicatures, as they do with *only* and unmodified implicatures (Bergen & Grodner, 2012), and collecting offline interpretation judgments about the defeasibility of L+H* marked implicatures.

Gradient vs. categorical effects of the pitch accents L+H and H*?*

Why did the L+H* accent have more direct mouse paths to the strong response option? While the design of Experiment 2 rules out the possibility that listeners interpret L+H* accents to signal turn finality, it could be the case the phonetic differences between the two accents (higher overall F0 and larger pitch range for L+H*) could have drawn listeners' attention towards more salient alternatives. Because our control conditions rule out this a general low-level effect for salience at the level of word recognition, this effect must be post-referential, hence supporting our general claim that pitch accents can work at the level of inferential processing. Nonetheless, it is still possible that these effects could have been driven by phonetic and not phonological differences. This criticism, however, would be valid for not only our study, but also for other phonetic and psycholinguistic studies researching contrast effects for the L+H* (Ladd & Morton, 1997; Watson, Gunlogson, & Tanenhaus, 2008; Braun & Tagliapietra, 2010). Regardless of whether this effect is more phonetic or phonological in nature, both ideas are consistent with our account.

While we do not take a stand on the phonological status of these two accent, we do endorse the idea that differences in meaning between L+H* and H* are probably gradient and not categorical. The garden path effects from Experiment 2 support Watson, Gunlogson, & Tannenhaus (2008)'s conclusion that the meaning differences between L+H* and H* can overlap. We build on their study in that, our findings show that these gradient differences in meaning are also found in more complex meanings such as the derivation of pragmatic inferences, not just in establishing reference.

Where does intonation fit into theories of pragmatic inferences?

Our studies show that pitch accents are integrated incrementally into the derivation of ad-hoc scalar implicatures. We now explain how our findings could be accounted for in different theories of implicature. Although these accounts are not processing theories per se, how listeners integrate intonation in pragmatic inferences speaks to both the availability of linguistic alternatives and, especially in our studies, the mechanisms that operate over these alternatives. In other words, we see processing data as acting as constraints on these theories as opposed to explicit tests of the core assumptions of these theories.

In grammatical accounts (Chierchia, 2004; 2006; 2013; Fox, 2007) implicatures arise by insertion of a grammatical operator, a silent counterpart of *only*. Further, it is assumed that focus activates alternatives and feeds into the implicature generation mechanism (see especially Chierchia, 2013). At first glance, our findings are quite consistent with this account: L+H* could encourage the insertion of the grammatical operator and thereby facilitate the derivation of the implicature. But this would be predicted with any pitch accent that can lead to focus marking. In our experiments, referents were in focus position and received an H* accent, which should be sufficient for focus marking. While the garden path effects from Experiment 2 show that with both accents participants derive the implicature to a high degree, it does not clearly follow from grammatical accounts why different pitch accents should have differing gradient effects on the insertion of covert operators. One way to reconcile this is to assume that the L+H* makes the alternatives more salient and therefore a covert operator can be inserted earlier. For example, in the version of the grammatical account by Chierchia (2013), it is assumed that once alternatives are active, they have to be consumed/taken up by an overt or covert operator (e.g., *only* or its silent counterpart). We find plausible that the L+H* facilitates implicatures because it has increased the salience of one alternative over another, however it is largely an open question about the nature of the information that makes one alternative more salient than another in the first place.

Neo-Gricean accounts (e.g., Sauerland, 2004; Geurts, 2010) hold that listeners have to complete the so-called epistemic step to arrive at the strengthened interpretation (derive primary and secondary implicatures). The "epistemic step" account holds that listeners consider alternatives that are relevant to a speaker's conversational goals. Whether listeners choose to derive an inference or not will depend on listeners recognizing these goals and whether a speaker has the necessary knowledge to intend one alternative over another. This accounts is usually portrayed as a set of additional considerations above and beyond structural aspects of an utterance such as linguistic scales and information structure (see Breheny, Ferguson, & Katsos, 2013) As we discussed previously, the L+H* accent might remove the necessity to do derive primary implicatures first or reduce the need to consider contextual factors relevant for deriving the implicature (Sauerland, 2004; Breheny,et al, 2013). Both the facilitation by the L+H* accents in Experiment 1 and the garden path effects in Experiment 2 could equally be explained by such an account. Breheny et al (2013) show that the epistemic step can be done incrementally: listeners do not need to wait until the end of an utterance to derive pragmatic inferences. The Neo-Grician account seems to differ from grammatical accounts in that grammatical accounts assume a role of speaker knowledge (and relevance) in the derivation of alternatives, however the mechanisms by which alternatives are disregarded should be impervious to non-linguistic factors. Therefore, future research should examine the role of speaker knowledge in both the derivation of alternatives with L+H* and well as the process of negating alternatives.

A final consideration is whether pitch accents affect different implicatures in the same way. This question relies heavily on whether on the same processing mechanisms are used for different types of implicatures (Bott & Chemla, 2016). If intonation affected the processing different types on implicatures in similar ways, this would suggest a shared mechanism needed to integrate information from pitch accents. As mentioned, intonation has been shown to increase the likelihood of standard cases of scalar implicatures, e.g., with the quantifier *some* (Schwarz, 2008) or the connective *or* (Chevallier et al, 2008) as well as *ad-hoc* scalar implicatures (Gotzner & Spalek, 2014). Unlike our studies, these studies do not seek to test the exact processing mechanisms, which are ultimately responsible for increase in implicature rates. Future research should also examine how pitch accents work across different types of implicatures to better understand how intonation factors into the inferential procedures, which ultimately influence whether listeners opt for one alternative over another.

References

Barr, D.J., Levy, R., Scheepers, C., and Tily, H.J. (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, 68(3). 255-278.

Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. The Journal of the Acoustical Society of America, 95(5), 2973–2973

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*(10), 1636-1675.

Beckman, M. E., & Ayers, G. M. (1993). Guidlines for ToBI Labeling. OSU.

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1450.

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.21, retrieved 29 September 2015 from http://www.praat.org/

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423-440.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437-457.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. Journal of Memory and Language, 66(1), 123-142.

Bott, L. & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. Journal of Memory and Language..

Braun, B. & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25 (7,8,9), 1024-1043

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse comprehension and production* (pp. 1-40). Norwood: Ablex.

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *The Quarterly Journal of Experimental Psychology*, *61*(11), 1741-1760.

Chierchia, G., Fox, D., & Spector, B. (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. *Unpublished manuscript*.

Chierchia, G., Fox, D., & Spector, B. (2012) Scalar implicature as a grammatical phenomenon. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning, 3*, 2297–2331. Berlin: De Gruyter Mouton.

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

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*, 515-535.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*, 667-710.

Filik, R., Paterson, K. B., & Liversedge, S. P. (2009). The influence of only and even on on-line semantic interpretation. *Psychonomic Bulletin & Review, 16,* 678-683

Fox, D. (2006). Free Choice Disjunction and the Theory of Scalar Implicatures. In *Presupposition and Implicature in Compositional Semantics*, ed. by Uli Sauerland and Penka Stateva, 71-120. New York: Palgrave Macmillan.

Freeman, J.B. & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*, 226-241.

Fretheim, T. (1992). The effect of intonation on a type of scalar implicature. *Journal of Pragmatics*, *18*(1), 1-30.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.

Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press.

Gotzner, N., Spalek, K., Wartenburger, I. (2013). How pitch accents and focus particles affect the recognition of contextual alternatives. In M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Eds.), Proceedings of the 35th Annual Meeting of the Cognitive Science Society, 2434-2440. Austin, TX: Cognitive Science Society, 2013.

Gotzner, N. & Spalek, K. (2014). Exhaustive inferences and additive presuppositions: Interplay of focus operators and contrastive intonation. In J. Degen, M. Franke, and N. Goodman (Eds.), *Proceedings of the Formal & Experimental Pragmatics Workshop*, 7-13, Tübingen

Gotzner, N., Wartenburger, I. & Spalek, K. (2016). The impact of focus particles on the recognition and rejection of contrastive alternatives. *Language and Cognition, 8*, 59-95.

Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review*, 147-177.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hirschberg, J. (1991). *A theory of scalar implicature*. Garland Press. New York.

Horn, L. R. (1972). On the semantic properties of logical operators in English. Ph.D. thesis, University of California, Los Angeles.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376-415.

Husband, E.M. and Ferreira, F. (2015) The role of selection in the comprehension of focus alternatives. Language, Cognition and Neuroscience

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.

Katsos, N.& Bishop, D. V. M. (2011) Pragmatic Tolerance: Implications for the Acquisition of Informativeness and Implicature. *Cognition, 20*, 67-81.

Kim, C. S. and Gunlogson, C., Tanenhaus, M. K. and Runner, J. T. (2015). Context-driven expectations about focus alternatives. *Cognition, 139,* 28-49.

Krahmer, E.,&Swerts, M. (2001). On the alleged existence of contrastive accents. Speech Communication, 34, 391–405.

Ladd, R. D. (1996). *Intonational Phonology*. Cambridge: Cambridge Univeristy Press.

Ladd, R. D., & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, *25*(3), 313 - 342. doi:DOI: 10.1006/jpho.1997.0046

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Levinson, S. C. (2000) *Presumptive meanings*. Cambridge, Mass.: MIT Press.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation.* Ph. D. MIT.

Pierrehumbert, J., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen, J. Morgen, & M. E. Pollack (Eds.), *Intentions in Communications* (pp. 271-311). Cambridge, Mass.: The MIT Press.

van Rooij, R. & Schulz, K. (2004). Exhaustive Interpretation of Complex Sentences. *Journal of Logic, Language and Information 13* (4), 491-519.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and philosophy*, *27*(3), 367-391.

Spalek, K., Gotzner, N. & Wartenburger, I. (2014). Not only the apples. Focus-sensitive particles improve memory for information-structural alternatives. *Journal of Memory and Language, 70*, 68-84

Tomlinson, J. M, Jr., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, *69*(1), 18-35.

Wagner, M. & Watson, D.G. (2010) Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*(7), 905–945

Watson, D.G., Gunlogson, C. A., & Tanenhaus, M. K. (2006). Online methods for the investigation of prosody. In S. Sudhoff, D. Lenertov, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, J. Schlieer (eds.), Methods in Empirical Prosody Research. Walter de Gruyter: New York. 259-282

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H*. *Cognitive Science*, *32*(7), 1232-1244.

Acknowledgments. John Michal Tomlinson, Jr. and Lewis Bott were supported by ESRC Grant RES-062-23-2410. We are grateful for the help of Elena Chepucova for the data collection and stimuli construction.

Figures

Figure 1. Trial set up for Experiment 1

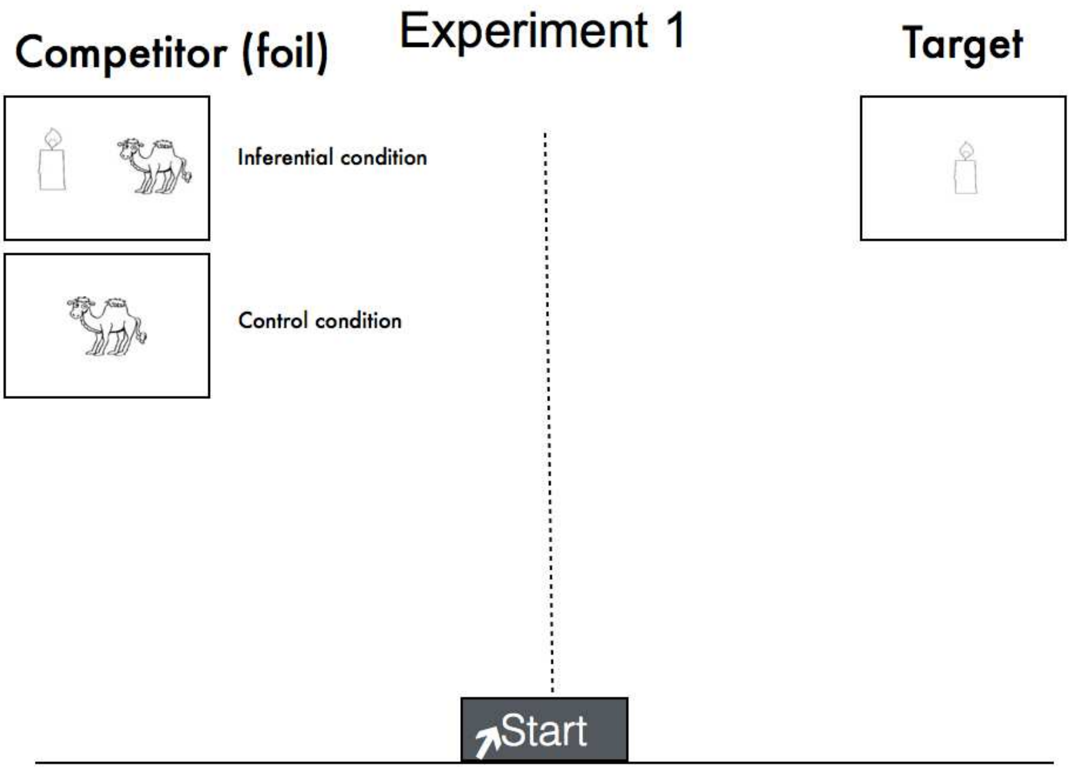

Figure 2. Mean F0 values for stressed syllables.

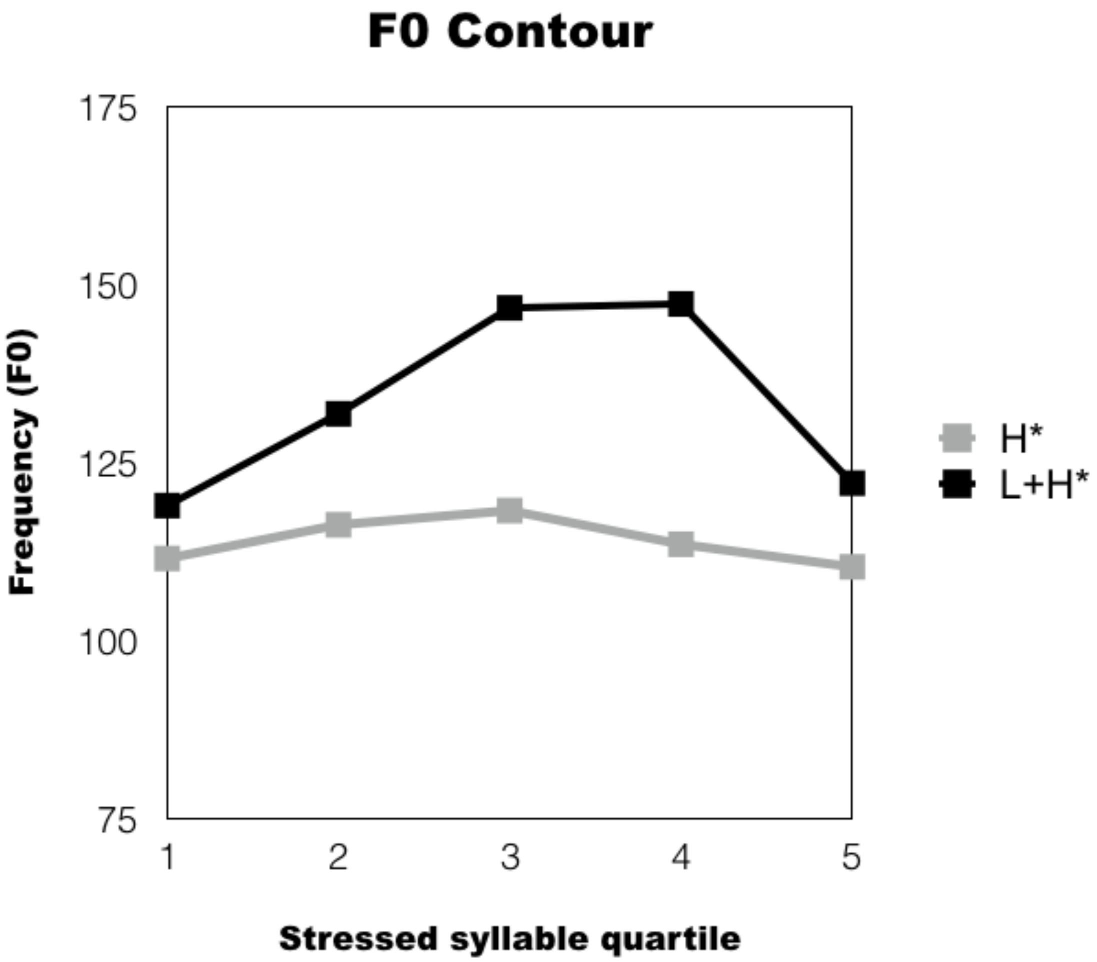

Figure 3 Raw and average mouse trajectories for the control conditions in Experiment 1. The x-coordinates were transformed by multiplying by a factor of -1 for trials with reversed potions of the target and competitor pictures.

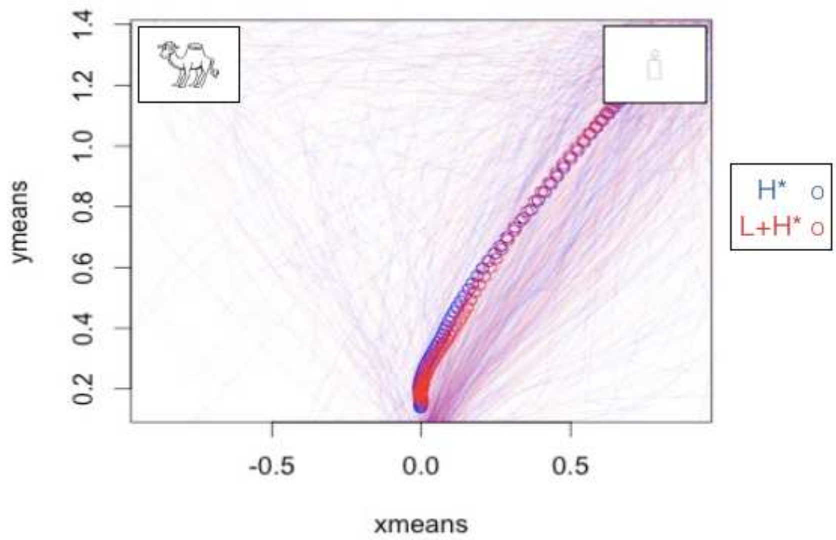

Figure 4 Raw and average mouse trajectories for the inferential conditions in Experiment 1. The x-coordinates were transformed by multiplying by a factor of -1 for trials with reversed potions of the target and competitor pictures.

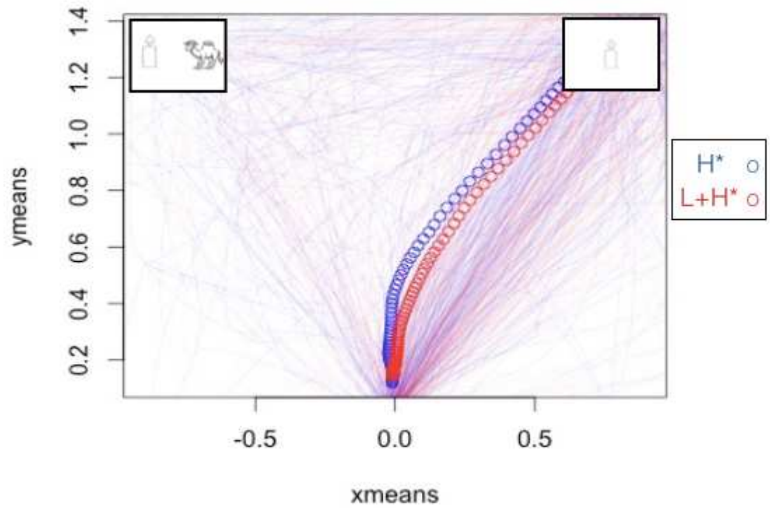

Figure 5. Trial set up for Experiment 2

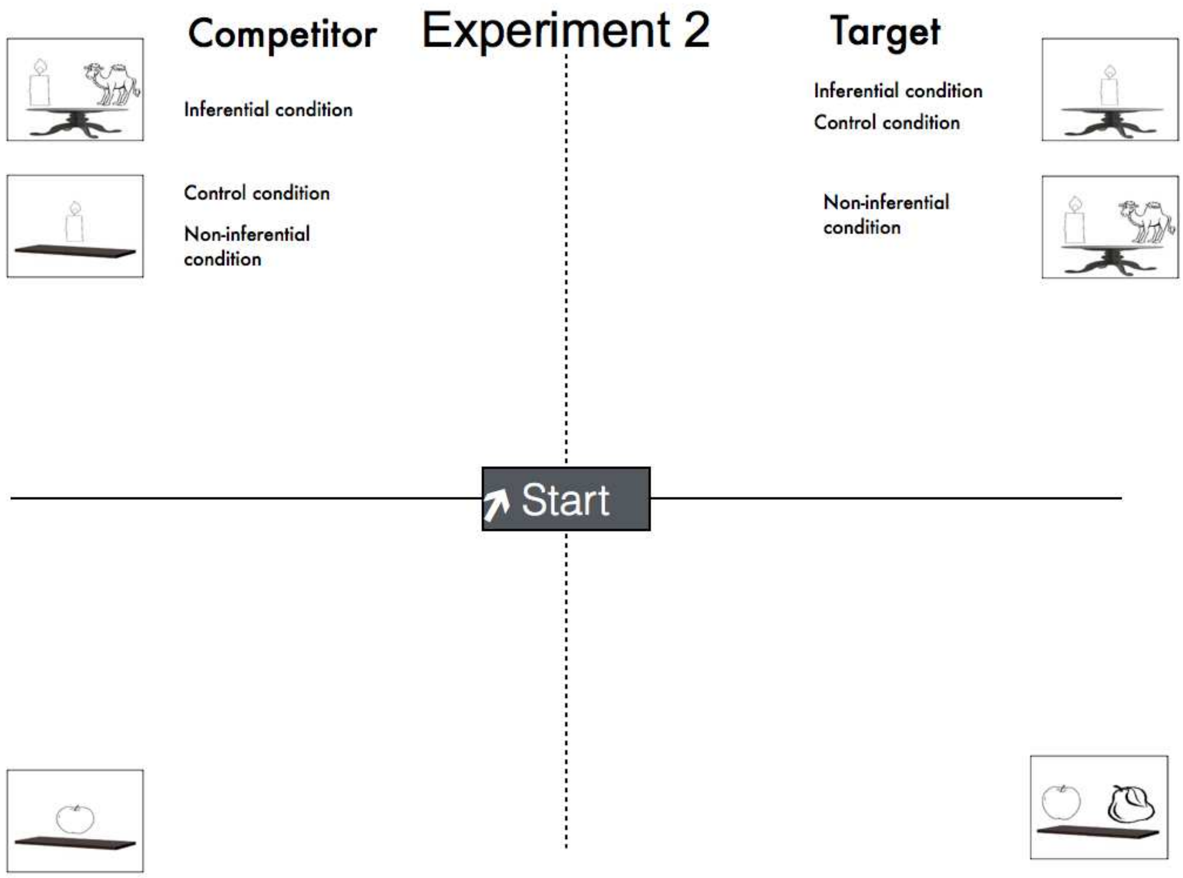

Figure 6 Raw and average mouse trajectories for control conditions in Experiment 2. The x- and y-coordinates were transformed by multiplying by a factor of -1 for trials with reversed potions of the target and competitor pictures.

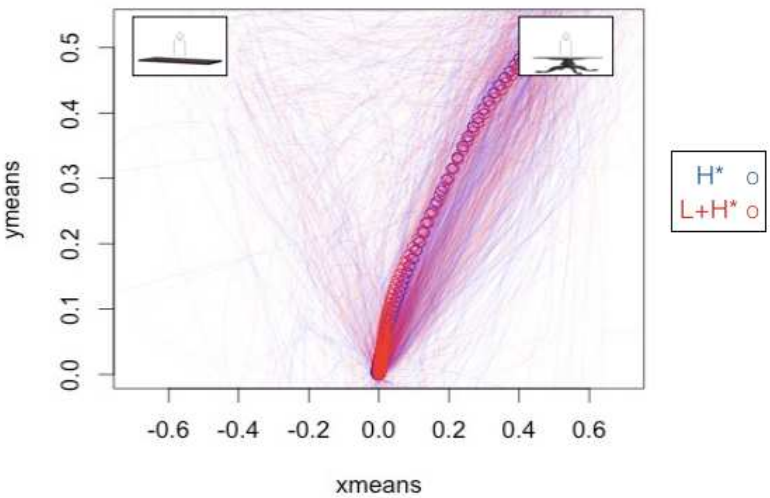

Figure 7 Raw and average mouse-trajectories for inferential conditions in Experiment 2. The x- and y-coordinates were transformed by multiplying by a factor of -1 for trials with reversed potions of the target and competitor pictures.

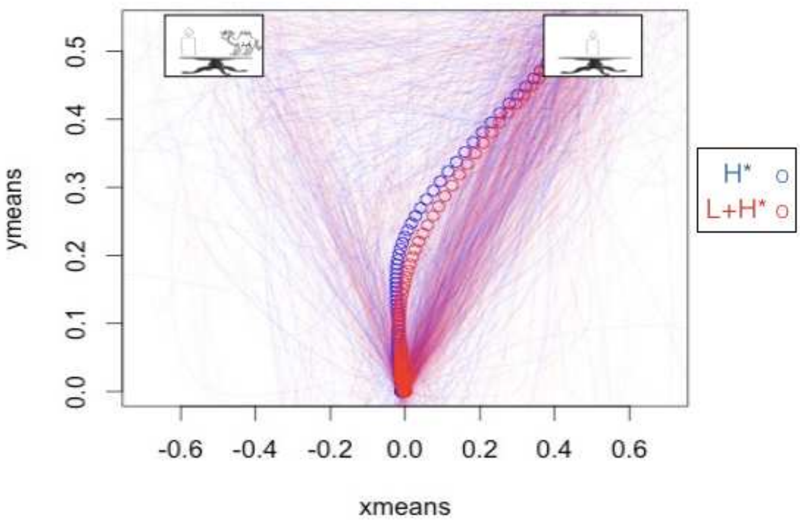

Figure 8. Raw and average mouse trajectories for non-inferential conditions in Experiment 2. The x- and y-coordinates were transformed by multiplying by a factor of -1 for trials with reversed po- tions of the target and competitor pictures.

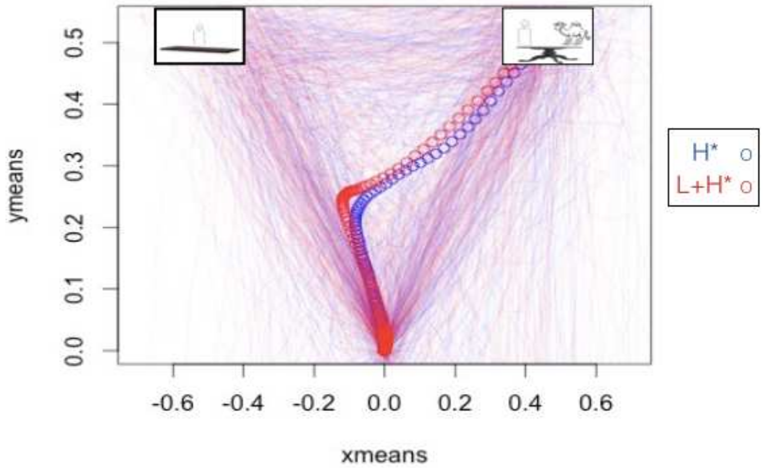

Table 1. Experiment 1 Conditions

Targets	Condition	Pitch accent	Target picture	Competitor
Mark has a can-dle (A)	Inferential	H*	A	A+ B
		L+H*	A	A + B
	Control	H*	A	B
		L+H*	A	B
Fillers				
Mark has cat (A) and a carrot (B)	Fillers	H*	A + B	A (or B)
		L+H*	A + B	A (or B)
		H*	A + B	A + C
		L+H*	A + B	A + C
		H*	A + B	C
		L+H*	A + B	C
		H*	A + B	C + D
		L+H*	A + B	C + D

** all items had fallling (L-L% boundary tones)*

Table 2. F0 and duration parameters for experimental items

Pitch accent	Durati-on (ms)	F0max (Hz)	Average F0 1st Interval (Hz)	Average F0 3rd Interval (Hz)	Average F0 5th Interval (Hz)	Average point of F0max (ms)	Relative-point of F0max (F0max ms/duration)
H*	318.9	123.45	114.46	118.2	105.72	121.3	0.36
L+H*	332.5	162.25	118.89	146.67	108.86	183.2	0.65

Table 3. Experiment 2 Conditions

Targets	Condition	Pitch accent	Target picture	Competitor 1	Competitor 2	Competitor 3
Mark has a candle (A) on the table	Inferential	H*	A on table	A + B on shelf	A (or C) on shelf	C + D (or B) on table
		L+H*	A on table	A + B on shelf	A (or C) on shelf	C + D (or B) on table
	Control	H*	A on table	A on shelf	A +B (or C) on shelf	C + D (or B) on table
		L+H*	A on table	A on shelf	A +B (or C) on shelf	C + D (or B) on table
	Non-inferential	H*	A + B on table	A on shelf	A (or C) on table	C + D (or B) on shelf
		L+H*	A + B on table	A on shelf	A (or C) on table	C + D (or B) on shelf
Fillers						
Mark has cat (A) and a carrot (B)	Fillers	H*	A + B on shelf	A (or B) on table	A (or C) on shelf	C + D (or B) on table
		L+H*	A + B on shelf	A (or B) on table	A (or C) on shelf	C + D (or B) on table
		H*	A + B on table	C +D on shelf	A on shelf	C on table
		L+H*	A + B on table	C +D on shelf	A on shelf	C on table

		H*	A + B on table	A on shelf	A on table	C + D (or B) on shelf
		L+H*	A + B on table	A on shelf	A on table	C + D (or B) on shelf

* *Target and competitor images were randomly rotated for screen positions. Only target images and competitor one images remained next to each other, although these positions were randomized (e.g. either below/above or left of right).*

Table 4

Accuracy and Mean Area Under the Curve Values for Experiment 1

Conditions	Accuracy	AUCs M(SD)	Initiation times (ms)	Click times(ms)
H* Control	99%	0.26 (0.68)	354.72	1091.42
L+H* Control	98%	0.21 (0.64)	363.5	1055.71
H* Inferntial	97%	0.48 (0.81)	350.61	1214.37
L+H* Inferenti-al	98%	0.28 (0.65)	345.47	1181.63

Table 5.Fixed effect and variance estimates for AUCs values for simple effects models in Experiment 1

Model 1 - *(AUC ~ conditions + (1 + pitchaccent|subject) + 1 + pitchaccent|item), item n = 42, subject = 28, reference variable = H* Control*

fixed effect	Estimate	SE	t	p (MCMC)
Intercept	0.26	0.1	2.76	
H* Inferential	0.32	0.08	3.9	<.002*
L+H* Control	-0.06	0.09	0.68	0.52
L+H* Inferential	0.03	0.09	0.36	0.71
random slopes	s^2			
Subject	0.03			
Item	0.02			

Model 2 *(AUC ~ conditions + (1 + pitchaccent|subject) + 1 + pitchaccent|item), item n = 42, subject = 28,reference variable = H* Inferential*

fixed effect	Estimate	SE	t	p (MCMC)
Intercept	0.59	0.1	6.01	
L+H* Inferential	-0.28	0.09	3.15	<.003*
H* Control	-0.32	0.08	3.91	<.001*
L+H* Control	-0.38	0.09	4.21	<.001*
random slopes				
Subject	0.03			
Item	0.02			

**not all of the models converged using condition in the random slope term. Because of this pitch accent was used because this was the comparisons of interest acroos the experiments.*

Table 6

Accuracy and Mean Area Under the Curve Values for Experiment 2

	Accuracy	AUCs M(SD)	Initiation times (ms)	Click times(ms)
H* Control	99%	0.07 (0.23)	497.23	1397.04
L+H* Control	99%	0.07 (0.28)	456.22	1386.94
H* Inferential	93%	0.21 (0.35)	458.91	1425.9
L+H* Inferenti-al	95%	0.13 (0.31)	443.57	1383.01
H* Non-inferential	87%	0.28 (0.43)	441.35	1737.01
L+H* Non-inferential	89%	0.32 (0.42)	479.11	1770.03

Table 7. Fixed effect and variance estimates for AUCs values for simple effects models in Experiment 2

Model 1 *(AUC ~ conditions + (1 + pitchaccent|subject) + (1 + pitchaccent|item), item n = 42, subject = 66, reference variable = H* Control*

fixed effect	Estimate	SE	t	p (MCMC)
Intercept	0.07	0.02	3.02	
L+H* Control	0.003	0.02	0.01	0.98
H* Inferential	0.14	0.03	6.26	<.001*
L+H* Inferential	0.08	0.03	3.48	<.001*
H* Non-inferential	0.21	0.02	9.15	<.001*
L+H* Non-inferential	0.29	0.02	12.35	<.001*
random slopes	s^2			
Subject	0.02			
Item	0.01			

Model 2 *(AUC ~ conditions + (1 + pitchaccent|subject) + 1 + pitchaccent|item), item n = 42, subject = 66, reference variable = H* Inferential*

fixed effect	Estimate	SE	t	p (MCMC)
Intercept	0.21	0.02	10.22	
L+H* Inferential	-0.06	0.02	-2.82	<.006*
H* Control	-0.14	0.02	-6.28	<.001*
L+H* Control	-0.14	0.02	-6.25	<.001*
H* Non-inferential	0.07	0.02	3.14	<.001*

Model 1 *(AUC ~ conditions + (1 + pitchaccent|subject) + (1 + pitchaccent|item), item n = 42, subject = 66, reference variable = H* Control*

L+H* Non-inferential	0.15	0.02	6.38	<.001*
random slopes	s^2			
Subject	0.02			
Item	0.01			

Model 3 *(AUC ~ conditions + (1 + pitchaccent|subject) + 1 + pitchaccent|item), item n = 42, subject = 66, reference variable = H* Non-inferential*

fixed effect	Estimate	SE	t	p (MCMC)
Intercept	0.28	0.02	13.16	
L+H* Non-inferential	0.08	0.02	3.12	<.002*
H* Inferential	-0.07	0.02	3.14	<.002*
L+H* Inferenti-al	-0.13	0.02	5.8	<.001*
H* Control	-0.21	0.02	9.15	<.001*
L+H* Control	-0.21	0.02	9.11	<.001*
random slopes	s^2			
Subject	0.02			
Item	0.01			