The contribution of context and contour to perceived belief in polar questions

Meghan E. Armstrong a,*, Pilar Prieto b,c

a Department of Languages, Literatures and Cultures, University of Massachusetts Amherst, United States
b Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Spain
c Institució Catalana de Recerca i Estudis Avançats, Spain

Received 9 December 2014; received in revised form 21 February 2015; accepted 12 March 2015

Abstract

In the past decades, we have advanced significantly in our knowledge of intonational meaning, but few studies have tested experimentally the way in which discourse contexts affect intonational meaning. In this work we were specifically interested in how listeners use both intonation and discourse context to infer information about speaker belief states. We examined the effect of five bias types on two intonation contours used for polar questions (PQs) in Puerto Rican Spanish (PRS). The bias types consisted of unbiased, mild positive bias, strong positive bias and mismatch bias contexts. The intonation contours had been previously claimed to differ in the belief state information they convey – iH*L% is known to mark utterances as PQs without encoding specific belief state on the part of the speaker, while L*HL% is known to convey a state of disbelief on the part of the speaker (Armstrong, 2010). We hypothesized that the lack of belief of the presence of disbelief for these contours, respectively, would be perceived by listeners when these contours were heard in an unbiased context. We also predicted that listeners would rely on contextual bias more for the iH*L% contour than the L*HL% contour, and that the disbelief meaning would persist regardless of discourse context. Perceived belief scores were analyzed, and results show that different bias types affected perceived belief scores in different ways. Mild positive bias did not seem to affect perceived belief for the two contours, while strong positive bias and mismatch bias did. Since L*HL% conventionally conveys disbelief, a reversal effect was shown when it was heard in the strong belief context. Participants’ comments indicate that in such cases, an ironic interpretation of the contour is available. These results, in addition to the comments provided by participants, show that perceived belief will depend both on the type of contextual bias, as well as the type of information conveyed intonationally. This work provides more evidence for the dynamic relationship between specific context types and intonation contours that differ in terms of the amount and type of meaning they convey. Published by Elsevier B.V.

Keywords: Intonation; Polar questions; Epistemic meaning; Belief states; Incredulity; Bias

1. Introduction

Over the last forty years, studies in intonational meaning have made clear that speaker attitude and belief states can be encoded intonationally. A classic claim comes from the seminal work of Pierrehumbert and Hirschberg (1990), who propose that intonation contours are chosen by speakers (1) to create a relationship between propositional content and previous/upcoming utterances and (2) to create a relationship between the propositional content of an utterance and the beliefs of the speaker and the hearer. This paper is concerned with the latter function of intonation. Following

* Corresponding author at: 419 Herter Hall, University of Massachusetts Amherst, Amherst, MA 01003, United States. Tel.: +1 413 595 4910.
E-mail address: armstrong@spanport.umass.edu (M.E. Armstrong).
Pierrehumbert and Hirschberg’s observations, we expect that from the hearer’s perspective, intonation is used to glean information about speaker belief states. The role of intonation in encoding belief is especially relevant when considering polar questions (PQs), an area of the grammar where we can find varying levels of epistemic asymmetry in terms of speaker commitment to propositional content, what the speaker knows, what the hearer knows, the attitudes the speaker may have about propositional content, and so forth (Enfield et al., 2013). For instance, when the speaker produces a true out-of-the-blue PQ, s/he may have no assumptions about whether the answer is yes or no. Enfield et al. claim that for PQs, speakers generally have a low commitment to the truth of a proposition, while the hearer is assumed to have a higher commitment (since the speaker directs his/her question to the hearer). However, we know that often times, speakers have evidence that leads them to draw conclusions about the truth value of propositions, and therefore PQs are frequently not so out-of-the-blue. Speakers are able to encode their own beliefs and attitudes about propositions, and can use the grammar to make more fine-grained distinctions within a ‘gradient possibility space’ (Enfield et al., 2013:194). Enfield et al. discuss the semantics of sentence-final particles and their use for ‘tilting the epistemic gradient’. For instance in Lao, a tonal language spoken in Thailand, the sentence-final particles böô3, vaa3, tii4 and nọq1 are used for PQs. The first means ‘I want to know if p is the case’, the second means ‘I want to know if p is the case; I’d say it is, based on current evidence’, the third means ‘Maybe ~p is the case, I don’t know; I’d say p is the case, based on independent evidence’, and the last means ‘I’m saying p is the case; I’d think you say this as well’ (p. 204). The semantics of these particles are then modulated by pragmatic context. The underpinnings of sentence-final particles mirror quite closely the function of intonation in PQs. For instance, Vanrell et al. (2013) show that through intonation, speakers can encode their greater degree of commitment to propositional content in PQs. Vanrell et al. (2014) showed that when Mallorcan Catalan speakers have physically perceived evidence for a proposition, they are likely to mark PQs with the particle que and the L+H* L% nuclear configuration. We thus expect that the meaning of intonation contours will be modulated by pragmatic context much like sentence-final particles are. One example of the pragmatics of PQ intonation can be found considering the high-rise pitch contour (H* H–H%) in American English, which may be used for PQs, but has other, more context-specific meanings. Pierrehumbert and Hirschberg (1990) give the example of the utterance My name is Mark Liberman produced with the H* H–H% contour. As mentioned, this contour is used for PQs in American English. However, in the context given by Pierrehumbert and Hirschberg, the speaker is not questioning whether his name is Mark Liberman, which would be the case if H* H–H% were being used as a PQ. Rather, in a given context, where the speaker has approached the receptionist window at a doctor’s office, the tune can be used to convey something like ‘My name is Mark Liberman and are you expecting me, or am I in the right place?’. The speaker is able to convey information (he asserts his identity), but the contour allows him convey additional information, giving the utterance another dimension of meaning in context. This same contour H* H–H% is analyzed by Hirschberg and Ward (1995) as being used by the speaker to convey that propositional content should be added to the mutual belief space but at the same time allows the speaker to question whether the hearer can relate this propositional content to their own private beliefs. Based on this analysis of H* H–H%, they propose that “the tonal features associated with phrases should better be seen as relating the interpretation of phrases in some way to speaker or hearer’s private beliefs” (411). Later, Escandell-Vidal’s (1998) analysis of interrogatives in Peninsular Spanish assumes that speakers make intonational choices based on their assessment of their beliefs, as well as the beliefs of others. Her proposal for “attributed interrogatives” is based on the speaker’s attribution of propositional content as the belief of a speaker vs. the belief of someone else (often times the hearer). She finds that in Peninsular Spanish, the low-rise conveys only one level of meaning (PQ-marking), while the high rise and the rise-fall, in addition to marking a sentence as a PQ, mark the proposition as the belief of some individual. These “attributed interrogatives”, according to Escandell-Vidal’s account, convey two levels of meaning.

If speakers convey information about speaker and hearer beliefs intonationally, then we also expect that this information is accessible to the hearer. After all, why would the speaker go through the trouble of marking such information intonationally? To this end, Gravano and his colleagues (2008) used experimental methods to test the contribution of intonation to listeners’ assessment of speaker certainty. In this study, they showed that just as the epistemic modal would conveyed more certainty than the verb is, the downstepped contour in American English conveyed more certainty than the declarative or yes-no question contours. Their findings confirm that like modal verbs, listeners depend on intonation in assessing the degree of speaker certainty. Within the domain of PQs, few studies have explored the effect of context and intonation contour on perceived belief experimentally, especially within the realm of PQs. Gunlogson’s (2003) pioneering work on PQ intonation in American English made clear the crucial role of discourse context for the interpretation of PQ intonation. However, experimental research examining how intonation and context conspire to lead a listener to a given meaning is rather limited, and we find even less work relating these factors to the perception of speaker belief states. Nevertheless, we review some of the relevant literature here. The L*+H L–H% contour was investigated experimentally by Ward and Hirschberg (1988, 1992) to understand which phonetic details were necessary to distinguish between uncertainty vs. incredulity meanings, both of which are available for L*+H L–H% in American English. These authors found that while listeners were influenced by other spectral characteristics, they were most influenced by pitch range. However, the authors examined ambiguous contexts in which both the uncertainty and incredulity interpretations were
available, and not how context might influence the listeners’ choice of uncertainty vs. incredulity, but rather a range of phonetic details.\footnote{See Avesani et al. (1995) and Hirschberg and Avesani (2008) for production work on intonational disambiguation outside the realm of PQs.} Later Nilsenová (2006) investigated perceived bias based on intonation contour in American English questions. Listeners were presented with positive PQs produced with any of eight intonation contours (H* H−H%, L* H−H%, L* L−H%, L* L−L%, H* L−H%, H* L−L%, H* H−L%). Using a 5-pt Likert scale, listeners had to make a decision about the answer the speaker expected: the speaker definitely expects ‘no’; probably expects ‘no’; [has] no expectations; probably expects ‘yes’; definitely expects ‘yes’. Expectations of a negative answer were found to be related to L* L−H%, while expectations of a positive answer tended to be associated with high phrase and high boundary tones. Listeners tended to perceive the speaker as expecting a positive answer with low-rises and high-rises. One limitation in this design that the author points out is that these questions were presented to listeners decontextualized, and thus it is not clear how the tunes would have been perceived in context. Speech corpora have also been used in experimental work on perceived certainty. Lai (2010) investigated the effect of context type, intonation and cue words on perceived certainty. Using the Switchboard I Release 2 corpus of telephone conversational speech, Lai examined the effect of different types of contexts (factual, evaluative, attributed, inferred), prosodic contours (rises vs. falls) and cue words (e.g. really, well, okay, sure, yeah). Participants were presented with written conversations and audio response (from the corpus, but later synthesized for the experiment) which they then rated using a 7-pt. Likert scale. Participants gave ratings based on expectedness, credibility and evidence. Lai found that rising intonation did not affect perceived credibility of the speaker, but did lead participants to rate speakers as sounding more uncertain. She also found that ratings varied depending on the specific cue word used. For instance, the cue word yeah was found to be highly susceptible to sounding uncertain with rising intonation, but not the cue word well. Hara et al. (2013) used naturalness ratings to explore the role of prosody in negative polar questions in Mandarin Chinese as well as Japanese. They found that in Mandarin, sentence-final stress in PQs conveys an epistemic conflict between speaker and addressee. In Japanese, on the other hand, deaccenting in negative polar questions can be used when the speaker seeks agreement. They characterize both non-canonical uses as instances of prosody being used to convey a meta-question about the discourse. Interestingly, both of the prosodic strategies investigated by the authors are more prototypically known in the literature as prosodic correlates of focus or information structure, but the authors show that listeners associate them with belief and discourse evidence. Most recently, Arvaniti et al. (2014) explored the pragmatics of two types of wh-question intonation in Greek: LH* L−L% and L*H L−iH*% tunes (labels are from the GRTtoBI system). With the goal of showing evidence for a phonological difference between the two contours, they used two experimental designs: one where participants listened to one contour or the other with different contexts (default, information-seeking wh-contexts vs. rhetorical wh-contexts), and another in which participants listened to a question produced with either contour and picked an appropriate response. Response patterns for both experiments clearly differed depending on the intonation pattern, allowing the authors to argue for a phonological distinction between the two tunes.

Puerto Rican Spanish (henceforth PRS) is a Caribbean variety of Spanish known to exhibit variation in the contours possible to convey polar questions. In the Spanish Tones and Breaks Indices System (Sp_ToBI) (Armstrong, 2010; Hualde and Prieto, 2015) this variety of Spanish has an intonational inventory comprised of (L*, H*, iH*\footnote{A nuclear pitch configuration with an extra-high tone iH* L% was proposed by Armstrong (2010) since narrow focus declaratives in PRS may be produced with H*L%. In general, PQs in PRS are produced with global expanded pitch range, but to date no studies have examined the specific thresholds in terms of pitch height differences for discriminating iH* versus H*. Such a study would require examining the spectral characteristics of the same vowel (for example /a/) in contexts of both PQs and narrow focus, in addition to perception experiments designed to identify the perceptual boundary for iH* and H*. The iH* category has also been proposed by Cabrera Abreu and Vizcaíno Ortega for Canary Islands Spanish, a variety that behaves quite similarly to PRS both in the segmental and suprasegmental domains.} L+<H*, L+H, L+H) and five boundary tones (L%, H%, H%, LH% and HL%). The i and l diacritics indicate phonological upstep and downstep, respectively. Like other languages (D’Imperio, 2002 for Neopolitan Italian; Grabe et al., 2004 for varieties of British English; Monte, 2012 for Brazilian Portuguese; Vanrell et al., 2013 for Majorcan Catalan; Henriksen et al., 2015 for Manchego Spanish), an utterance-final fall is used to mark polar questions in PRS. Speakers produce an extra-high tone in the nuclear stressed syllable, followed by a fall to a low boundary. The schematic in Fig. 1 illustrates the iH* L% nuclear configuration used for polar questions in PRS.

This differs from the type of fall found for declaratives, for example, which are also characterized by final falls. However, declarative falls are characterized by a fall through the nuclear stressed syllable, labelled H+L* L% in the Sp_ToBI system. As can be seen in Fig. 2, the pretonic syllable is high, followed by an F0 fall through the nuclear accented syllable to a low boundary tone.
Armstrong (in press) has also described a contour used for PQs that express ‘disbelief’ in PRS. This contour is described as a low, flat tone in the nuclear tonic syllable (L*), followed by a bitonal boundary tone HL% in the post-tonic syllables. Fig. 3 illustrates this contour.

In production work using the Discourse Completion Task (Prieto, 2001; Nurani, 2009; Prieto and Roseano, 2010) paradigm, it has been shown that in PRS, the jH*L% contour in the Sp_ToBI system (Armstrong, 2010; Hualde and Prieto, 2015) is used as a general PQ marker, (sentence-level marking, but no belief marking) while the L*HL% contour is used to grammatically encode the speaker’s state of incredulity or disbelief with respect to a proposition (Armstrong, in press).

While many varieties of Spanish convey incredulity using the pitch shape of a general PQ-marking intonation contour in addition to expanded pitch range or non-modal voice quality, PRS is unique in that a specific intonation contour, L*HL%, has been found to be used rather reliably to express incredulity or disbelief. As far as we know, the relationship between the L*HL% tune and incredulity/disbelief meaning has not been shown for other Spanish dialects (though it is true that not all Spanish dialects have intonational descriptions). Intonational encoding of speaker disbelief, or incredulity, has been documented for other Romance languages such as Valencian and Central Catalan (Crespo-Sendra, 2011; Crespo-Sendra et al., 2013), Buenos Aires Spanish (Ar Lee et al., 2010), Brazilian Portuguese (Truckenbrodt et al., 2009; Frota et al., 2015), Friulian (Roseano et al., 2015), and French (Michelas et al., 2013). The PQs in Figs. 1 and 3 would differ then, in the amount and type of information they convey, as shown in Table 1.

We can refer to L*HL% questions as multidimensional, i.e. conveying two “layers” of meaning in that they seem to convey both information about sentence type (i.e. PQ-marking) and belief state (i.e. conveying the speaker’s state of
disbelief). For the case of ıH*L%, it is possible that the speaker’s state of “unknowing” would be inferred, by nature of the utterance being marked as a question.

We do not restrict the idea of layered intonational meaning to PQs, since we assume that such strategies would also available for other sentence types. If it is the case that the default PQ-marking contour ıH*L% in PRS does not encode any specific belief information/speaker bias apart from the sense of “wondering” or “unknowing” associated with questions in general, then it follows that this contour would be felicitous in unbiased contexts, as in (1):

(1) José and Felicia are chatting at the airport as they wait for their flight.
José: I can’t believe I’m missing my daughter’s dance recital.
Felicia: That’s too bad.
José: Do you have children? ıH*L%

Here José has never met Felicia and therefore has no way of knowing whether or not she has children. The default contour, but not the disbelief contour, would be felicitous here. A context where the disbelief contour might arise is shown in (2):

(2) Chris and Rebecca have been coworkers for 10 years. Rebecca has never once mentioned that she has children.
Rebecca: I can’t stay late tonight, I need to pick up my kids from soccer practice.
Chris: You have children?!? L*HL%

In (2) Chris reacts to the presupposition in Rebecca’s prior utterance, that she has a child. Since Chris has known Rebecca for a decade and she has never mentioned having children, the idea that Rebecca has children does not match with Chris’ expectations about the world, resulting in his state of incredulity. The PRS L*HL% intonation would allow a speaker in this context to both question the propositional content and express this state of disbelief at the same time. Thus, the contour carries two layers of meaning: sentence-level (PQ marking, or perhaps more minimally, response-eliciting) and belief state (disbelief marking). There are no restrictions in terms of how the information is activated for L*HL% questions – p may be activated linguistically (as in (2)) or extra-linguistically (as in (3)):

(3) Jessica, who lives and works in New York City, comes home from work and opens the door to discover her husband holding a goat and a sign that says “Happy Birthday!”.
Jessica: You got me a goat!?

Since a goat is not a typical gift for city-dwellers nor is it a pet typically kept by New York City apartment-dwellers, Jessica conveys her state of disbelief that her husband got her a goat. However, the proposition is activated extra-linguistically – it is something Jessica has perceived through sight.
Given the ways in which the PRS PQ contours have been claimed to be pragmatically restricted (Armstrong, in press), we would expect that when listeners hear PQs produced with iH*L% they should not perceive any specific belief on the part of the speaker (other than a state of unknowing), but should perceive disbelief when they hear PQ questions produced with L*HL%. But to what extent does discourse context play a role? The meanings proposed above were based on production experiments and in context, and therefore the first goal of this work was to confirm whether the belief states that listeners perceived in a perception experiment could confirm earlier claims about the meaning of the two contours described above. Below we outline a rating experiment designed to examine the meanings of iH*L% and L*HL%, in addition to how pragmatic context might modulate their meanings. We reasoned that if the default contour iH*L% does not convey belief state information, then participants should perceive neutral belief states when the contour is heard in an unbiased context. Disbelief, on the other hand, should be perceived even in an unbiased context for utterances produced with L*HL%, if disbelief meaning is encoded through the contour. We also hypothesized that perceived belief states would be more affected by context when the default contour is heard than when the disbelief contour is heard, since there would be no encoded belief state information competing with contextual bias, i.e. since no belief is encoded through the default contour, it would be easier to base perceived belief purely on context. However, since disbelief intonation (L*HL%) in PRS encodes two levels of meaning, we expect context to affect its meaning differently from the default contour. In order to understand how context might affect belief states that are encoded intonationally, we used an experimental methodology that used contexts differing in degree of bias. We qualified the degree of bias based on the type of evidence for a given proposition available to the speaker in the discourse context. The contexts we chose were by no means exhaustive. We pitted one unbiased context against four different biased contexts. Our predictions and hypotheses are tested in the experiment outlined in section 2.

2. Methods

2.1. Participants

Twenty-six speakers of Puerto Rican Spanish between the ages of 18 and 65 participated in the study. In order to be included in the analysis, the participants had to have grown up in Puerto Rico and could not have lived in any other country for a period of more than 3 consecutive months. All of the speakers, with the exception of one, were from the San Juan Metropolitan Area (within a 25 mile radius of San Juan). The other speaker was from Orocovis, which is located in the Central Mountain Range area of Puerto Rico.

2.2. Materials

2.2.1. Pragmatic contexts

Five distinct pragmatic contexts were included in the materials. We created one neutral context and 4 biased contexts that differed from each other in terms of contextual evidence for the p(reposition). Table 1 shows descriptions of the different types of contexts that were presented to participants.4

The first context (Context 1), was considered unbiased, since there is no contextual evidence for p. Including an unbiased context was critical for allowing us to assess whether either of the two target contours (iH*L% and L*HL%) grammatically encode some type of (dis)belief. We characterize Contexts 2 and 3 as having a mild positive bias. In Context 2 evidence for p is activated linguistically in the present discourse context; in Context 3 the same holds but this activation can be considered discourse-old (i.e. was activated at a point in time prior to the discourse context). We considered Context 4 to have a strong positive bias, since evidence for p is generally known, and additional evidence is present (through visual information) in the present discourse. Finally, we consider Context 5 as a case of mismatch bias. The speaker has a prior belief that ¬p and is provided with new information, which clashes with the speaker’s prior belief: in contrast with Context 4, Context 5 is expected to interact with the appearance of the neutral contour.

For the rating experiment, the discourse contexts shown in Table 1 were followed by responses produced with the L*HL% or the iH*L% intonation contours. These paired combinations provide the ground to empirically test the relative semantic weight of intonation and context on the hearer’s interpretation of the speaker’s belief.

An example of one of the pragmatic contexts (Context 4) is given in (4)

---

4 In order to explore various types of positive contexts, we limited the experiment to positively biased contexts. In general, the way in which bias arises in negative polar questions is still disputed (Hara et al., 2013), and we therefore restricted the experiment to positive polar questions. As an anonymous reviewer has pointed out, it will also be important to observe how the contours interact with contexts where evidence for ¬p is available.
(4) **Contexto:** Carlos y Jorge son compañeros de casa. Carlos sabe que Jorge normalmente duerme hasta la 1 o las dos todos los domingos. Un domingo a la 1:30, Jorge sale de su cuarto y empieza a hablar con Carlos.

Jorge: Hola Carlos, ¿qué hay?
Carlos: Na, aquí, cogiéndolo suave...
Jorge: ¿Qué has hecho hoy?
Carlos: Fui al gimnasio, hice compra, y tú... (mirando la cara de sueño de Jorge): ¿Acabas de despertarte?

**Context:** Carlos and Jorge are roommates. Carlos knows that Jorge normally sleeps until one or two on Sundays. One Sunday, at 1:30, Jorge comes out of his room and starts talking to Carlos.

Jorge: Hi Carlos, what's up?
Carlos: Nothing, just taking it easy...
Jorge: What have you done today?
Carlos: I went to the gym, went to the store and you... (looking at Jorge’s sleepy face): **Did you just wake up?**

In this case, p (Jorge habitually sleeps into the afternoon) is discourse old, since it is something that both Carlos and Jorge have in their mutual belief space. But the contextual evidence of Jorge emerging from his room and his sleepy face (in addition to the background knowledge that Carlos has about his sleeping habits) also provide evidence that Jorge had just woken up. We determined that this context had a strong positive bias, since the prior belief matches the evidence in the context. All discourse contexts and target utterances are detailed in the Appendix.

2.2.2. Audio materials

The speech stimuli were produced by two male speakers of PRS. There were five unique segmental strings that corresponded to each of the pragmatic contexts. Each speaker produced the five segmental strings with each of the different contours: jH*L% and L*HL% (5 × 3 = 15 utterances per speaker). After the stimuli were recorded, the phonetic implementations were checked by the first author. Representative implementations of the contours are shown in Figs. 4 and 5. The nuclear configuration in Fig. 4 shows a question produced extra-high tone, jH* in nuclear position, followed by a fall to a low boundary tone.

The L*HL% nuclear configuration is phonetically realized as a flat low tone in the nuclear stressed syllable with a subsequent rise and fall in the post-tonic segmental material, and is shown in Fig. 5. The stimuli were checked by the first author to ensure that for each nuclear configuration, the stimuli were consistent in terms of the prenuclear pitch accents that appeared before the nuclear configuration. Prior to the nuclear configuration, iH*L% questions always had an initial L*+H prenuclear pitch accent followed by one or more H* pitch accents (see Fig. 4). L*HL% questions always had an initial L+<H* prenuclear pitch accent followed by one or more H* prenuclear pitch accents (see Fig. 5).

2.2.3. Contexts

The contexts were designed with a native speaker of PRS to ensure that they (1) conveyed the desired meanings and (2) were culturally acceptable. Three additional speakers of PRS read each context and confirmed to the first author that the contexts were acceptable based on these two criteria.

![Waveform, spectrogram and F0 trace for the polar question ¿Tú tienes cita médica el viernes? "Do you have a doctor's appointment on Friday?" produced with an L*+H prenuclear accent on **citas** and a H* on **médica.** A jH* nuclear pitch accent is found on **viernes** followed by a fall to a low boundary tone.](image-url)
2.3. 

Participants were sent a hyperlink to access the survey on the website surveygizmo.com. After consent, participants were brought to a training page where each of the elements of the experiment was explained. They were told that this was a study about the intonation of Puerto Rican Spanish, and that the word “intonation” refers to the melody of speech. The participants were given an example of how depending on the intonation contour, the speech act type for Es alto. ‘S/he is tall.’ vs. ¿Es alto? ‘Is s/he tall?’ is determined. Participants were then told that in the experiment, they would hear utterances with different types of question intonation, and that it was very important to have a way (speakers, headphones) to listen to the sound files in the experiment well, and that it was additionally important to read each of the contexts well to ensure comprehension. After the participants completed the training page, they moved on to the test trials. There were a total of 20 test trials (2 contours × 5 contexts × 2 speakers). The order of presentation was randomized.

Participants had to read the context/dialogue, and clicked on the soundfile in order to hear the target question, as shown in Fig. 6:

Next, the speakers had to judge the belief state of the speaker they heard using a 7-pt Likert scale. Rietveld and Chen (2006:294) note the advantages of a 7-pt Likert scale for obtaining perceptual judgments for intonational meaning:

- Gradations in judgments are possible.
- The resulting scale is claimed to be at the interval level.
- A yardstick is given, which enables the researcher to assess pairwise whether objects differ significantly.

---

5 We acknowledge that perhaps a better way of instructing participants would have been to ask subjects to judge the belief state without telling them which information to use. As an anonymous reviewer has pointed out to us, subjects often try to identify the meaning of an intonation contour and apply that decision to subsequent contours of the same type.
The speakers were presented with seven statements, and were told to choose one as their answer. Instructions were written so that participants would choose the sentence that best described the attitude of the speaker with respect to the propositional content of the question. The participants were always given the instructions in the format shown in (5). They did not see the explanations we provide for the reader in parentheses.

(5) Choose the option that best describes the attitude of (i.e. speaker name from dialogue) with respect to this phrase:

Participants then picked a number corresponding to the following scale:

1. Sabe que no es verdad. ‘He knows it is not true.’
2. Está convencido de que no es verdad. ‘He is convinced it is not true.’
3. Duda que sea verdad. ‘He doubts that it is true.’
4. No sabe si es verdad o no. ‘He doesn’t know whether it is true or not.’
5. Cree que es verdad. ‘He believes it is true.’
6. Está convencido de que es verdad. ‘He is convinced it is true.’
7. Sabe que es verdad. ‘He knows it is true.’

Participants also had the option to provide comments for each test item. In this way the participants could optionally explain their decisions, allowing for both a qualitative and quantitative analysis of the data. A total of 520 trials were submitted to analysis (26 participants × 2 contours × 5 contexts × 2 speakers).

3. Results

3.1. Effect of contour and context on perceived belief state

Fig. 7 shows means and standard error of perceived belief ratings for the two contours (default and disbelief) for each of the five contexts.

A linear mixed-effects model was fit to examine the effect of Context and Contour on perceived belief ratings, which were done on a 7-point Likert scale (lowest belief = 1, highest belief = 7). For the model presented here, the dependent
The variable was **perceived belief score** (Likert score) and the independent variables were **contour** (two levels: iH*L% and L*HL%) and **context** (five levels, one for each context). **Contour** and **context** were coded as discrete variables, and **participant** was included as a random slope by contour and context, using the lmer function in the lme4 package of the R statistical package (Bates et al., 2008; R Development Core Team, 2011). Given we used a within-subjects design, the Likert scores were not normalized. The best-fit model included **contour** and **context**, as well as the interaction between the two, is shown in Table 2. Since contour 1 was designed to be neutral, it was taken as the model’s baseline. This allowed us to compare the contours heard with the biased contexts as compared to a neutral context. The model’s fixed effects are shown in Table 2, alongside the factors’ estimated coefficients and their predictive significance. We show results for both contours as baseline – first iH*L% and then L*HL%.

What is perhaps most striking about the data is that context has about the same influence on contour interpretation for most contexts, with the exception of the strong positive bias, context 4. This context completely changes the direction of the ratings for the disbelief contour, L*HL%, and also increases the positive belief ratings for iH*L%. An effect was also found for the mismatch context (context 5) driving down the ratings for both contours. In this case the mean ratings for the disbelief contour L*HL% reflected even lower perceived belief than we find in the other contexts, showing that this congruent context seems to enhance the meaning already contributed by L*HL%.

4. Discussion

4.1. Unbiased context

The results summarized in Table 2 show that when we compare the two contours in the unbiased context (Context 1), belief scores for L*HL% are significantly lower than those for iH*L%, showing that the belief state information conveyed through L*HL% is not context-specific. Perceived belief scores were slightly above 4 (No sabe si es verdad o no ‘He doesn’t know if it’s true or not’) for iH*L%, showing that even in unbiased contexts there is a slight bias for p. This is not surprising since, as Büning and Gunlogson (2000) point out, positive polar questions are not always neutral. The fact that someone would even ask about the truth of a proposition at all allows us to assume that the speaker may have some belief about p in positive polar questions. Enfield et al. (2013) in fact claim that there is always (our emphasis) some kind of expectation as to the truth of a proposition in PQs, and we believe that the slight positive bias shown here reflects this. Questions heard in a neutral context with L*HL%, on the other hand, had a mean Likert score of about 3 (Duda que sea verdad), showing the contour’s independent contribution of belief meaning.

4.2. Mild positive bias contexts

We now turn to the biased contexts (i.e. Contexts 2–5). With respect to iH*L%, we hypothesized that this contour would not contribute any meaning in addition to PQ-marking. That is, iH*L% would only convey one level of meaning (i.e. a PQ-marker). In general, Fig. 5 shows that the mean perceived belief score does not vary much for iH*L% across Contexts 1, 2 and 3. No statistical differences are observed for perceived belief scores for iH*L% heard with Context 2 (mild bias) vs. Context 1 (unbiased), nor iH*L% heard with Context 3 (mild bias), vs. Context 1. The mean perceived belief score for iH*L% heard with Context 2 was about the same as that of iH*L% heard in Context 1, i.e. just above 4 (no specific belief). The mean perceived belief score for iH*L% heard with Context 3 was slightly higher than the other two contexts, but still
between 4 and 5 (somewhere between no belief, and a slight positive belief). This difference was not significant. These results suggest that the effect of mild positive bias contexts was not sufficient to trigger significant changes in perceived belief for PQs produced with jH*L%.

As noted before, when L*HL% is heard in an unbiased context (Context 1), a significantly lower degree of perceived belief is observed compared to jH*L%. This means that listeners tended to choose the point on the Likert scale corresponding to Duda que sea verdad ‘Doubts that it is true’. The results for the mild positive bias contexts are quite similar. Regardless of the Context type, mean perceived belief ratings stay consistent with very little variation, with mean perceived belief scores at or around 3 for both Contexts 2 and 3, respectively. Thus regardless of the mild positive bias in the context, listeners tend to rate utterances produced with L*HL% as conveying doubt. No significant differences are observed for L*HL% heard with Context 2 vs. Context 1, nor Context 3 vs. Context 1. Therefore neither of the contours examined showed significant effects of context when heard with the mild positive bias contexts.

4.3. Strong positive bias and mismatch bias contexts

The results for Contexts 4 (strong positive bias) and 5 (mismatch bias) differ from the mild positive bias contexts. For Context 4, when p was in the speaker’s private belief space, and evidence for p was available in the discourse context (i.e. strong positive bias), a significantly higher mean belief score is found for jH*L% when compared to jH*L% heard with Context 1 (\(p < 0.001\)). The mean perceived belief score for jH*L% heard with the strong positive bias context was between 5 (Cree que es verdad ‘He believes that it is true’) and 6 (Está convencido de que es verdad ‘He is convinced that it is true.’) Therefore there was a significant effect of context on perceived belief for jH*L% questions heard with this context. This finding confirms that all biased contexts are not equal. That is to say, different types of bias will affect perceived beliefs in different ways. For Context 5, the mismatch bias context, ~p was in both the speaker and the hearer’s mutual belief spaces, but p was activated in the discourse context, creating a mismatch between a prior belief of the speaker, and linguistically-activated evidence in the discourse context. Results show that a significantly lower perceived belief score is found when jH*L% is heard in the mismatch context compared to the same contour heard in the unbiased context. This difference resides only in the fact that the mean perceived belief score was slightly less than 4 for Context 5, while it was slightly over 4 for Context 1. A look at participant comments, however, reveals that there may have been disagreement among speakers with respect to how appropriate jH*L% was for this context.

(5)  
... no suena sorprendido.  
‘... he doesn’t sound surprised.’

(6)  
Hace falta complementar esto con el gesto para estar seguro de lo que quiere decir.  
‘This needs to be complemented with a gesture to be sure of what he means.’

On the other hand, the comment in (7) justifies its felicity:

(7)  
Está bien también, suena sorprendido pero tranquilo.  
‘It sounds fine too, it sounds surprised but calm.’

(8)  
Debe de sonar más sorprendido y la intonación no cuadra tampoco.  
‘He should sound more surprised and the intonation doesn’t fit either.’

These comments show that participants had different opinions regarding how appropriate the jH*L% contour was in a scenario like Context 5, where we would actually expect the speaker to be surprised and show disbelief. It seems that some participants felt that this intonation pattern was not reflecting the degree of surprise that a speaker would be expected to have in such a situation, and that additional cues to surprise were necessary (gesture, expanded pitch range), though one participant seemed to accept the speaker sounding “surprised but calm”. The ‘calmness’ may be related to the speaker not conveying enough information about his belief state, since jH*L% does not convey belief state information.

Similar to the case of jH*L%, a significant effect of context is observed for L*HL% heard with the strongly biased context, Context 4, as expected. The perceived belief ratings flipped in direction – where we found low perceived belief for the unbiased context and the mild positive bias contexts, (as well as the mismatch bias context), the mean perceived belief rating now crosses well into the threshold of positive perceived belief. The mean rating falls somewhere between the Likert ratings 5 and 6, corresponding to Cree que es verdad ‘He thinks that it’s true’ and Está convencido de que es verdad ‘He’s convinced that it’s true’. This resulted in a significant effect of context on perceived belief rating. While the contour seems to convey doubt or low belief in unbiased contexts, in a context where the speaker should be in the opposite belief state (i.e. the speaker would have to believe that \(p\), based on prior knowledge and evidence), we find that the mean perceived
belief rating reflects the contextual bias rather than the meaning conveyed by the contour. Participants’ comments suggest that an ironic meaning is available under the conditions we created with Context 4:

(9) *Tiene sentido si estaba siendo sarcástico.*
It makes sense if he was being sarcastic.

(10) *Tiene sentido si estaba diciendo lo contrario, o sea, si está siendo irónico está bien. Suena como ‘¡aja?’, irónico.*
It makes sense if he’s saying the opposite, that is, if he’s being ironic it’s fine. It sounds like “mmmm..”, ironic.

Therefore, our results show that the disbelief meaning of L*H*L% can be exploited as a strategy of irony or sarcasm, as a subject points out in (10)—even though the contour conventionally conveys doubt regarding \( p \), an ironic interpretation is available where the exact opposite meaning is implied. Finally, as expected, there was a significant difference between perceived belief scores for L*H*L% heard with Context 5 (mismatch bias) vs. the unbiased context. The congruent context-contour pairing shows an intensifying effect on the interpretation of the contour, forcing the mean perceived belief score down closer to 2, *Está convencido de que no es verdad* ‘He is convinced that it is not true’.

Our experimental data show that the different contours and contexts chosen for the experiment affect how listeners perceive speaker belief states. The results make clear the fact that \( \text{H}^*L\% \) and L*H*L% differ in terms of the type and amount of information they convey. In general, we see a slight positive bias for PQs heard with \( \text{H}^*L\% \) in an unbiased context. As stated before, we believe this to be an effect of the nature of positive polar questions, since if the speaker felt compelled to ask whether \( p \) is the case, s/he may have already formed a hypothesis regarding \( p \). We do not, however, see \( \text{H}^*L\% \) as contributing specific information about speaker belief state, but rather view the contour as an intonational morpheme used to mark utterances as PQs in PRS. In contrast, when L*H*L% is heard in an unbiased context, the contour conveys disbelief (or doubt, according to our scale options). Therefore when we hold one aspect of the contours’ meaning (i.e. PQ-marking) constant (since speakers were told that all utterances in the experiment were questions), it appears that disbelief meaning is an independent contribution of the L*H*L% contour. Next, our mild positive bias contexts (Contexts 2 and 3) did not have a significant effect on the meaning of either of the contours. For \( \text{H}^*L\% \), the mean perceived belief scores remained close to the score corresponding to the option *No sabe si es verdad o no* ‘He doesn’t know whether or not it is true’, while for L*H*L% the mean scores tended to be around *Duda que sea verdad* ‘He doubts it is true.’ for the unbiased context and the contexts showing mild positive bias. The strong positive bias context had an effect on the interpretation of both intonation contours. For \( \text{H}^*L\% \), perceived belief scores rose significantly when the contour was heard with the strong positive bias. The strong positive bias affected belief ratings for L*H*L% in a significant way as well, this time completely reversing the directionality of perceived beliefs for this contour. As indicated by subjects, an ironic meaning could be calculated when L*H*L% was heard with the strong positive bias. This shows that speakers of PRS can exploit the meaning conveyed through the contour (disbelief) in a context where they could not possibly be in a state of disbelief, given common ground and evidence at hand. We would like to point out here that in many languages, evidential markers are used to convey unawareness, surprise, admiration, amazement or irony (Tomič, 2006), and we believe that it is the interplay between beliefs and evidence in this combination that gives rise to a sarcastic interpretation. The combination of the belief state information conveyed through L*H*L% ([−belief]), the visual evidence in the context ([+belief]) and the private belief of the speaker ([+belief]) results in an ironic interpretation. Importantly, no subjects mentioned an ironic interpretation for the case of \( \text{H}^*L\% \), which is not surprising if we assume that \( \text{H}^*L\% \) is only doing the job of PQ-marking, rather than explicitly marking a belief. The comments provided by subjects also show their sensitivity to two Gricean Maxims: Quality and Relevance. In the strong bias context, a literal interpretation would involve the speaker conveying disbelief in a context where everything points to positive belief. For him to assert his disbelief would be false, thus violating the Maxim of Quality. However if we interpret this as the speaker *floating* the Maxim of Quality, this can lead us to an ironic interpretation. Since the ironic interpretation is perhaps the most relevant interpretation, this satisfies the Maxim of Relevance. We also point out that although we provided unbiased contexts, we did not present decontextualized instances of L*H*L% utterances. That is, the PQ component of L*H*L%’s meaning was already specified for subjects. It is worth noting that Armstrong (2010) documents the use of the same contour L*H*L% for exhortative commands, i.e. when the speaker is trying to encourage the hearer to do something (e.g. *Vente ‘Come along’*). Interestingly, both PQs and commands require some sort of response on the part of the hearer, and we might propose that minimally, L*H*L% has some sort of response-eliciting function. It is possible that between context and the lexical content associated with L*H*L%, its meaning is activated. Further work is needed to understand the meaning of L*H*L% when sentence type is not controlled for. In any case, \( \text{H}^*L\% \) and L*H*L% are similar in that when heard in mildly biased contexts, their meaning is not significantly affected. Additionally, their meanings were both affected by the strong bias context, but in different ways: \( \text{H}^*L\% \) was intensified, while the directionality of belief ratings for L*H*L% was reversed.
Also worth mentioning is the fact that $i^\text{H*L}\%$ was rated with lower perceived belief in the mismatch bias context when compared to the unbiased context, with some disagreement between listeners regarding its felicity. Some subjects’ comments reflected that the contour did not provide sufficient belief state information, indicating that additional cues may have been necessary for the speaker to sound as surprised or disbelieving as he should have been given the context. Armstrong (in press) suggested that delayed peaks and expanded pitch range in $i^\text{H*L}\%$ are strategies that speakers might use to convey surprise, or mirativity, in PRS. In the absence of these cues, the phonetic implementation of $i^\text{H*L}\%$ was perhaps not sufficient. We also point out that if it were that case that $i^\text{H*L}\%$ independently conveyed positive belief, participants might have observed that the speaker sounded too “believing”. This did not happen, they simply spoke to the lack of surprise or disbelief conveyed. Ideally, future work would investigate how various factors (lexical information, peak timing and pitch expansion) affect the interpretation of $i^\text{H*L}\%$ in context. Apart from this specific factor it seems that generally speaking, participants’ comments reveal that listeners find it necessary for speakers to convey the appropriate amount of information for specific contexts. The comments give insight to how speakers may go about making decisions between contour types, and are quite reflective of Grice’s Cooperative Principle, specifically the Maxim of Quantity — make your contribution as informative as is required (for the current purposes of the exchange). Whether or not crossing into the threshold of sufficient informativeness for specific contexts includes modifications in pitch scaling, alignment, gesture or the choice of a completely different contour is an empirical question that could be tested using a similar methodology to the one employed here.

5. Conclusions

We initially hypothesized that multiple layers of meaning (sentence type + belief state) may be conveyed through intonation, depending on the specific contour. We had different predictions for $i^\text{H*L}\%$ vs. $L^\text{H*L}\%$. With respect to the ‘default’ contour, $i^\text{H*L}\%$, we predicted that: (a) it would carry sentence type information without conveying information about speaker belief; and (b) it would take on the contextual bias of the context (as in Gunlogson, 2003). We believe (a) to be the case, while (b) was partially confirmed. We found differences with respect to the type of context – contexts with mild positive bias did not change perceived belief ratings significantly when compared to the unbiased context. While we did find that the mean responses for $i^\text{H*L}\%$ in the unbiased context showed a slight positive bias, we believe this to be related to the fact that positive polar questions are rarely completely unbiased. The speaker has often formulated some sort of hypothesis about the truth of the proposition that has led him/her to ask the question in the first place. Later, the context with strong positive bias (e.g., Context 4) moved the perceived belief ratings in a positive direction for $i^\text{H*L}\%$, in line with our original prediction. The mismatch bias context also affected perceived belief ratings for $i^\text{H*L}\%$ in a significant way. Thus $i^\text{H*L}\%$ may indeed take on the bias of the context, but this is not a hard and fast rule. The case of $i^\text{H*L}\%$ shows that not only do contours not take on the bias of the context all the time, the degree to which they take on the bias of the context may vary. The idea that not all contextual evidence is equal is not new. Büring and Gunlogson (2000), for example, distinguish between contextual evidence and compelling contextual evidence, which by nature suggests that there are strength distinctions in contextual evidence. Our results confirm this experimentally.

With respect to $L^\text{H*L}\%$, we hypothesized that (a) if it encodes disbelieve meaning directly, this meaning should be detected in an unbiased context, and that (b) this meaning might override the effect of context. The first part of the hypothesis was confirmed. Listeners related $L^\text{H*L}\%$ with doubt in the unbiased context. Our results for $L^\text{H*L}\%$, as for the case of $i^\text{H*L}\%$, confirmed that the effect of the context depends on the type and amount of bias conveyed contextually. While the mild positive bias contexts did not affect perceived belief for $L^\text{H*L}\%$, the strongly biased and negative bias contexts did. The strongly biased context forced listeners to search for the most relevant meaning in order for cooperativeness, and it was shown that in this case an ironic interpretation for $L^\text{H*L}\%$ was available. The bias of the negative bias context worked in tandem with the context, and significantly lower perceived belief scores were found for $L^\text{H*L}\%$ when compared to its meaning in an unbiased context.

In sum, we feel that this study makes an important contribution to research aimed at better understanding how contour and context conspire in the expression of meaning. Our hypothesis that the PQ-marking contour, $i^\text{H*L}\%$, would be more susceptible to the bias of the context did not hold true. In fact both contours behaved quite similarly for mild positive bias, showing extremely similar belief ratings to those found with neutral contexts. The fundamental difference between contours that convey one layer (e.g. $i^\text{H*L}\%$, PQ-marking) vs. two layers (e.g. $L^\text{H*L}\%$, PQ-marking + belief state) of information, does not guarantee they will also be affected differently by discourse context. They may be affected in a similar way (our mild positive bias contexts) or in different ways (our strong bias context). As recent work has pointed out (Armstrong, in press) our observations also point to parallels between final PQ contours and sentence-final particles, which have been described as being used to tilt the epistemic gradient within the domain of PQs through the semantics of intonation. SFPs may simply mark a sentence as a PQ (e.g. *bôô3 in Lao), or they may add additional, non-truth-propositional information to the PQ. For example, Tzetzal speakers may use the SFP che’e not only to mark a PQ, but also to convey that speaker has just come to realize that p might be the case, thinks the addressee knows and wants the
addressee to confirm (Enfield et al., 2013). In their discussion of SFPs in POs, the authors state “By saying that they are more than just question markers we mean that they make subtle distinctions in relative strengths of knowledge or commitment to a proposition, of both speaker and addressee, manipulating the epistemic gradient . . . Yet they are also fully grammaticalized question markers...” (219). We believe these same observations to apply to the nuclear configurations we have examined here. Both are grammaticalized question markers, with L*HL% conveying a second layer of information. L*HL% is similar to ch’i e in its pragmatic restrictions – the speaker has just come to know that p is the case, and wants the addressee to confirm. In production we expect L*HL%, then, to be more restricted, while we expect j*H*L% to be more frequent, since it is used as a more general PQ-marker. Further exploration of the parallels between intonational morphemes and SFPs could be revealing for future work in the area of intonational meaning. Our work adds to the existing body of literature that has been developing over the last decades in an effort to understand how listeners “read minds” through the dynamic interplay between intonational meaning and discourse context.

Acknowledgments

We are thankful to Gibran Delgado and Erik Rodríguez for producing the stimuli for this experiment. Many thanks also to Gilmari Gerena and Marcos Rohena-Madrazo for help with creating the contexts. Finally, we are grateful to Scott Schwenter and Patricia Amaral for their helpful insights.

Appendix

Context 1 (Unbiased)

**Contexto:** José y Enrique se acaban de conocer en su vuelo que va a Nueva York. Están hablando de los alquileres en NYC.

Enrique: Hoy día, hasta en el Bronx pagas un montón.
José: Es verdad.
Enrique: Por todos lados las rentas están más caras que nunca.
José: Tienes razón... ¿Tú vives en la isla?

Context: José and Enrique just meet each other on a flight to New York. They are talking about rent prices.

Enrique: These days you pay a lot in the Bronx.
José: It’s true.
Enrique: Rents are high everywhere.
José: You’re right... Do you live on the island?
(It is assumed that given the frequent movement between New York and the Bronx that someone on a flight from Puerto Rico to New York has an equal chance of living in either place).

Context 2 (Mild Positive Bias 1)

**Contexto:** Ramón y Eduardo comparten un apartamento. Eduardo acaba de llegar a casa.

Ramón: ¡Hola! ¡Llegaste temprano!
Eduardo: Sí, hoy la jefa me dejó salir a las 4. Tú también llegaste temprano.
Ramón: Sí, es que me llamó el plomero para decirmelo lo esperara.
Eduardo: ¿Viene el plomero?

Context: Ramón and Eduardo share an apartment. Eduardo just got home.

Ramón: Hi! You’re home early.
Eduardo: Yes, my boss let me go at 4. You’re home early too.
Ramón: Yeah, the plumber called me to tell me he was coming.
Eduardo: The plumber is coming?

Context 3 (Mild Positive Bias 2)

**Contexto:** Raúl y Ernesto están en la playa hablando.

Raúl: Oye, mano, el día sí que está bien.
Ernesto: Sí, ¿verdad? Ojalá pudiera hacer esto todos los días.
Raúl: Sí... yo dejaría mi trabajo...
Ernesto: (Viendo que Raúl está sacando dinero y viene el vendedor de mantecado) ¿Tú vas a comprar mantecado?
Contexto: Raúl and Ernesto are on the beach talking.
Rául: Dude it’s a nice day.
Ernesto: I know, right? I wish I could do this every day.
Rául: I know, I would quit totally quit my job...
Ernesto: (Seeing that Rául is talking out money and there is an ice cream vendor approaching) Are you going to buy ice cream?

Context 4 (Strong Positive Bias)
Carlos y Jorge son compañeros de casa. Carlos sabe que Jorge suele dormir hasta la 1 o las dos todos los domingos. Un domingo a la 1:30, Carlos sale de su habitación.
Jorge: Hola Carlos, ¿qué hay?
Carlos: Na, aquí, cogíéndolo suave…
Jorge: ¿Qué has hecho hoy?
Carlos: Fui al gimnasio, hice compra, y tú… (mirando la cara de sueño de Jorge): ¿Acabas de despertarte?
Carlos and Jorge are roommates. Carlos knows that Jorge always sleeps until 1 or 2 on Sundays. One Sunday at 1:30, Carlos emerges from his room.
Jorge: Hey Carlos, what's up?
Carlos: Not much, just taking it easy…
Jorge: What have you been up to today?
Carlos: I went to the gym, went to the store, and you… (looking at Jorge’s sleepy face…) Did you just wake up?

Context 5 (Mismatch Bias)
Oscar sabe que su esposa, Judith, no puede soportar los perros
Judith: He estado pensando en lo que me vas a regalar para la Navidad…
Oscar: Hmm… y has pensado en algo específico?
Judith: Pues había pensado que a lo mejor un perrito…
Oscar: ¿Quieres un perro?
Oscar knows his wife Judith can’t stand dogs.
Judith: I’ve been thinking about what you’re going to give me for Christmas…
Oscar: Hmm… and have you thought about something specifically?
Judith Well I was thinking maybe a dog…
Oscar: You want a dog?

References