RESEARCH

Bias in polar questions: Evidence from English and German production experiments

Filippo Domaneschi[1], Maribel Romero[2] and Bettina Braun[3]

[1] University of Genoa, DISFOR – Department of Educational Sciences, Psychology Unit, IT

[2] Universität Konstanz, Fachbereich Sprachwissenschaft Raum G212, DE

[3] Universität Konstanz, Fachbereich Sprachwissenschaft Raum G221, DE

Corresponding author: Filippo Domaneschi (filippo.domaneschi@unige.it)

Different polar question forms (e.g., *Do you / Do you not / Don't you / Really? Do you… have a car?*) are not equally appropriate in all situations. The present experiments investigate which combinations of original speaker belief and contextual evidence influence the choice of question type in English and German. Our results show that both kinds of bias interact: in both languages, positive polar questions are typically selected when there is no original speaker belief and positive or non-informative contextual evidence; low negation questions (*Do you not…?*) are most frequently chosen when no original belief meets negative contextual evidence; high negation questions (*Don't you…?*) are prompted when positive original speaker belief is followed by negative or non-informative contextual evidence; positive questions with *really* are produced most frequently when a negative original bias is combined with positive contextual evidence. In string-identical forms, there are prosodic differences across crucial conditions.

1 Introduction

A polar question (PQ) is a question that expects only two possible answers: an affirmative answer or a negative one (Karttunen 1977; Groenendijk & Stokhof 1984; see also Krifka 2013; Roelofsen & Farkas 2015). However, even if we just want an affirmative or negative answer, there are different ways to phrase the question. Four possibilities are shown in examples (1)–(4), together with their linguistic classification. For the fourth possibility, two alternative realizations are given, with the adverb *really* and with focus on the finite verb, which have been treated in a parallel way (Romero & Han 2004).

(1)   Is there a good restaurant nearby?   Positive polar question (PosQ)

(2)   Is there no good restaurant nearby?   Low negation question (LowNQ)

(3)   Isn't there a good restaurant nearby?   High negation question (HiNQ)

(4)   a.   Is there really a good restaurant nearby?
        b.   [Is]$_{Focus}$ there a good restaurant nearby?   Positive polar question with *really* (*really*-PosQ)

The question forms in (1)–(4) raise the same issue – the issue of choosing between a given proposition p and its negation $\neg p$ – and have thus the same resolution conditions.[1] Following this intuition, Karttunen's (1977) classical analysis of questions assigns the same semantic representation to a PosQ like (1) and to the corresponding LowNQ like (2): both denote the set of possible answers (or resolutions) $\{p, \neg p\}$.[2] Furthermore, the standard analysis of negation derives the same semantic representation for LowNQs like (2) and HiNQs like (3): regardless of whether the full form *not* or the cliticised form *n't* is used and regardless of whether negation occupies a low (*not*) or high (*n't*) position in the syntactic tree, negation simply combines with a proposition p to yield its negation $\neg p$.[3] As for the forms in (4) (*really*-PosQ), they are not covered by traditional analyses. However, in order to capture their resolution conditions in a parallel way, they can be assigned the same semantic denotation $\{p, \neg p\}$ in addition to some requirement for newness or contrast arising from focus (Rooth 1992; Schwarzschild 1999), as in (4b).

Yet, despite the uniform resolution conditions of the polar question forms in (1)–(4), we will see that these polar question forms are not equally appropriate in any given situation. Each form conveys different additional information on what the speaker's belief or expectation in the situational context is with respect to the basic proposition p that they all have in common (i.e., the proposition p among the resolution conditions $\{p, \neg p\}$), which is the proposition 'that there is a good restaurant nearby' in our examples (1)–(4) (Ladd 1981; Büring & Gunlogson 2000; Romero & Han 2004; a.o.).

The question arises, how – beyond the truth-conditional core that these PQs share – the use-conditions of these PQ forms should be characterized. Specifically, what contextual factors are these PQ forms sensitive to? What PQ forms are preferred in which pragmatic conditions? What surface PQ forms constitute separate PQ types based on their use-conditions and what PQ forms should be grouped together as a single PQ type?

The answers to these questions offered in the current literature are fragmented in several ways and, at times, contradictory. Contextual factors that have been argued to alter the form of polar questions are: (i) the original bias of the speaker with respect to the truth of the proposition and (ii) the newly-acquired bias arising from contextual evidence.[4] There is, to date, no study that investigates the role of both these factors on the use of all the PQ types above. Consequently, there is currently no consensus on the mapping between pragmatic factors and particular surface forms, with authors diverging widely on what form-function combination constitutes a separate PQ type.[5]

The goal of the present paper is to resolve some of the disagreement in the empirical characterization of polar interrogatives by using experimental methodology. We present two psycholinguistic experiments that test which of the surface forms is used most

---

[1] More technically (Ciardelli et al. 2013; Farkas & Roelofsen subm.), and leaving probabilistic epistemic states aside (Lassiter 2001; Yalcin 2012), for any set of worlds S constituting an information state, if S resolves one of the four question forms, S also resolves the others.

[2] See Groenendijk & Stokhof (1984: 331ff.) for discussion leading to the same intuition.

[3] Note that, when low negation merges with the indefinite determiner, it is pronounced as the negative determiner *no* (see, e.g., Penka 2011). PQ forms with *not* and *no* are equally considered LowNQs.

[4] The biases discussed in the PQ literature are typically epistemic, but they may also be bouletic or deontic in nature (Huddleston & Pullum 2002; van Rooij & Šafárová 2003; Reese 2006). In this paper, we concentrate on the effects of epistemic biases, leaving non-epistemic biases for future research.

[5] We are aware of the large psychological literature on the notion of "bias" (e.g., Baron 1994; Haselton et al. 2005). In this paper, for the purpose of our experiments, we will adopt a linguistic definition based on information that is already part, or that has become part, of the epistemic state of the speaker. We will concentrate on speaker bias about events in the world and leave aside higher order bias, like the bias found in rising declaratives (e.g., *There is a good restaurant nearby?*), by which a given answer is attributed to the addressee in the speaker's mind (Gunlogson 2003), or the bias found in rhetorical questions (e.g., *Did he lift a finger to help you?*), by which a given answer is attributed to both the speaker and the addressee in the speaker's mind (Caponigro & Sprouse 2007).

frequently when the two kinds of biases are manipulated to generate different combinations of original speaker bias and contextual evidence bias. We investigate this issue both in English, and in German, the two languages that the debate on biased polar questions has mostly concentrated on. In particularly interesting (or unexpected) cases, we will also examine the prosodic realization of the most frequently chosen form(s). We will not attempt any theoretical modelling at this stage, leaving open how and to what extent the resulting descriptive characterization of the data in terms of biases should be reflected in a theoretical account. The studies and results to be presented in this paper make a contribution in the vast enterprise of characterizing empirically the role of biases in questions in general.

1.1 Empirical characterization of bias in polar questions

Ladd (1981) and Romero & Han (2002, 2004) focus on *original speaker bias*, characterized in (5). In English HiNQs like (3) mandatorily express original speaker bias (Ladd 1981), whereas LowNQs can be used without this bias (Romero & Han 2004). This is shown in (6): in a context where the speaker has no previous belief for or against *p*, the LowNQ (6S) can be uttered felicitously but the HiNQ (6S') cannot.[6]

(5)   *Original speaker bias* (for a proposition *p*) (cf. Ladd 1981: 166)
        Belief or expectation of the speaker that *p* is true, based on his epistemic state prior to the current situational context and conversational exchange.

(6)   Scenario: The speaker is organizing a party and she is in charge of supplying the non-alcoholic beverages for teetotallers. The speaker is going through the list of invited people. The speaker has no previous belief or expectation about their drinking habits.
        A: Jane and Mary do not drink.
        S: OK. What about John? Does he not drink (either)?   LowNQ
        S': #OK. What about John? Doesn't he drink (either)?   HiNQ

Furthermore, while expressing original speaker bias, HiNQs are argued to be ambiguous between two readings (Ladd 1981): a so-called outer negation reading in which the speaker wants to double-check *p* (= *that there is a good restaurant nearby*), illustrated in (7), and a so-called inner negation reading in which the speaker wants to double-check ¬*p* (= *that there is no good restaurant nearby*), exemplified in (8).[7]

(7)   A: You guys must be starving. You want to get something to eat?
        S: Yeah, isn't there a vegetarian restaurant around here?   *Outer-HiNQ*

(8)   S: I'd like to take you guys out to dinner while I'm here – we'd have time to go somewhere around here before the evening session tonight, don't you think?
        A: I guess, but there's not really any place to go in Hyde Park.
        S: Oh, really, isn't there a vegetarian restaurant around here?   *Inner-HiNQ*

---

[6] See Trinh (2014) on what discourse ingredients facilitate the use of LowNQs as unbiased.

[7] In the inner reading, negation is part of the proposition being checked whereas in the outer reading it is not. The two readings can be disambiguated by the presence of positive polarity items (PPIs, e.g., *some, already, too*) vs. negative polarity items (NPIs, e.g., *any, yet, either*). PPIs force the outer negation reading (double-checking *p*) and NPIs force the inner negation reading (double-checking ¬*p*) (Ladd 1981): (i). For a potential formalization of the function of double-checking in a discourse model, see Krifka (to appear).

(i)   a.   Isn't there some good restaurant around here?   Outer-HiNQ (double-check *p*)
      b.   Isn't there any good restaurant around here?   Inner-HiNQ (double-check ¬*p*)

Building on Ladd (1981), Romero & Han's (2002; 2004) resulting generalizations for English (and partially for a small survey of other languages) are: (i) PosQs and LowNQs are compatible with there being no original bias, (ii) *really*-PosQs require original bias for ¬p and are used to double-check p; (iii) HiNQs require original bias for p and are ambiguous between an inner negation reading double-checking ¬p and an outer negation reading double-checking p. These generalizations appear in Table 1.

Büring & Gunlogson (2000) investigate PosQs, LowNQs[8] and Outer-HiNQs in English and German and classify them with respect to their (in)compatibility with *bias* induced by immediate compelling *contextual evidence*. This bias is defined in (9) and its effects for selecting a particular PosQ are illustrated in (10).

(9)   *Contextual evidence bias* (for a proposition p) (Büring & Gunlogson 2000: 7)
        Expectation that p is true (possibly contradicting prior belief of the speaker) in-
        duced by evidence that has just become mutually available to the participants in
        the current discourse situation.

(10)   Scenario: A enters S' windowless computer room wearing a dripping wet rain-
        coat (contextual evidence for p = *it is raining*).
        a.   S: What's the weather like out there? Is it raining?
        b.   S' #What's the weather like out there? Is it sunny?

Büring & Gunlogson's generalizations are: (i) a PosQ *p?* (e.g., *Is it raining?*) is incompatible with evidence bias against *p*, (ii) a LowNQ (e.g., *Is it not raining?*) is compatible only with evidence against *p*, and (iii) an Outer-HiNQ (e.g., *Isn't it raining?*) is incompatible with evidence for *p*. Table 2 summarizes these generalizations.

Van Rooij & Šafářová (2003) merge the two types of biases into one and argue that there is no grammatical distinction between LowNQs, Inner-HiNQs and Outer-HiNQs in English. Whether an original speaker bias or a newly-acquired contextual evidence bias is

	PosQ p?	*really*-PosQ *Really p?*	LowNQ *Not p?*	Inner-HiNQ *n't p + NPI?*	Outer-HiNQ *n't p + PPI?*
Mandatory bias	no	yes	no	yes	yes
Original bias for …		¬p		p	p
Q double-checks…		p		¬p	p

Table 1: Ladd's (1981) and Romero & Han's (2002; 2004) typology and characterization.

Contextual evidence	PosQ p?	LowNQ *Not p?*	Outer-HiNQ *n't p + PPI?*
for p	✓	*	*
neutral	✓	*	✓
against p	*	✓	✓

Table 2: Büring & Gunlogson's (2000) typology and characterization. The hack mark indicates that the question form can be used with the contextual evidence at issue and the asterisk indicates that it cannot.

---

[8] Büring & Gunlogson (2000) label their second PQ group as "inner negation polar questions", inspired by the inner/outer distinction observed by Ladd (1981). But their examples in this group are all with *low* negation (not with Ladd's Inner-High negation), as exemplified in (i). Hence, for the sake of coherence in the text, we re-labeled their second group as "LowNQs".

(i)   Is there no vegetarian restaurant around here?

	PosQ	LowNQs, Inner-HiNQ, Outer-HiNQ
Original or evidence bias for...	p	$\neg p$

Table 3: van Rooij & Šafářová's (2003) typology and characterization.

perceived depends on the pragmatic context and the pragmatic contribution of negative polarity items (NPIs). Their main generalizations are sketched in Table 3.

After these pioneering works, authors have assumed a partial mixture of the characterizations above. The current state of the art is as follows.

In terms of kinds of biases, most authors investigate one kind of bias and remain silent (or say very little) about the other (Büring & Gunlogson 2000; Romero & Han 2004; Asher & Reese 2005; Reese 2006; Asher & Reese 2007; Trinh 2014; Krifka to appear). Only a few works actively address both kinds of bias, some arguing for a conflation of the two (van Rooij & Šafářová 2003) and some arguing for a complete separation between the two biases (AnderBois 2011; Sudo 2013).

In terms of which forms of polar questions constitute a genuine PQ type with its own pragmatic function and what question forms are simply used in free variation, the current picture is dramatically divergent: (i) LowNQs, Inner-HiNQs and Outer-HiNQs are sometimes merged as one single PQ type (van Rooij & Šafářová's 2003); (ii) other authors do not separate LowNQs and Inner-HiNQs but distinguish them from Outer-HiNQs (Asher & Reese 2007; Krifka to appear); (iii) other researchers vindicate the separation between LowNQs and HiNQs but classify Inner-HiNQs and Outer-HiNQs as one single grammatical type (AnderBois 2011); and (iv) yet others treat LowNQs, Inner-HiNQs and Outer-HiNQs as a separate type each (Romero & Han 2004; Reese 2006; Walkow 2009).

In sum, there is substantial disconnection and at times blatant disagreement about what kinds of bias PQ surface forms are sensitive to, what use-conditions each PQ form has and what PQ types should be distinguished.

1.2 Previous experimental studies

Some valuable initial steps have been taken to test the conflicting empirical generalizations and the predictions about the empirical behaviour of the different polar question types experimentally.

Hartung (2006), for instance, investigated the effect of original speaker belief (positive or no belief) on the acceptability of negative polar questions in English and German (English: LowNQ and HiNQ, both crossed with *too* and *either,* German: HiNQ with German equivalent of *too, LowNQ* with *too/either*/no polarity item). Contextual evidence bias was not considered in this study. Participants received a written presentation of a context setting up the original bias and a written version of a target question. In both languages, only one of the conditions showed an effect of speaker belief: HiNQ with *too* (higher acceptability with positive than with neutral speaker belief). The English HiNQs with *either* were degraded. This experiment represents a first attempt to experimentally evaluate how a particular kind of bias (here, original speaker bias) affects the acceptability of a set of negative polar question types. Unfortunately, Hartung's work takes into account only a small subset of negative polar questions (in relation to *either* and *too*), and only considers one kind of bias.

To our knowledge, Roelofsen, Venhuizen and Weidman Sassoon (2012) were the first who crossed original speaker bias (positive, neutral, negative) and contextual evidence bias (positive, neutral and negative) in a web-based rating study. They tested the acceptability of three English polar question forms: PosQ, LowNQ and HiNQ. As materials, they used a sequence of three cartoons, indicating a discourse between two participants with speaker balloons. The first utterance in the first cartoon was intended to manipulate the

original speaker bias, the second the contextual evidence and the last cartoon contained the context question (e.g., Original Bias: *Kate to Rose*: "I am going to get a cat/pet/dog."; Contextual Evidence: *Rose to Jennifer*: "Did you hear? Kate got a cat/pet/dog."; Question: *Jennifer to Rose*: "Did she get a cat?" / "Didn't she get a cat?" / "Did she not get a cat?"). The results of this comprehension study suggest an interaction between the two kinds of biases on the appropriateness ratings of all three question forms tested. However, as Roelofsen et al. themselves note in their paper (2012: 460), their design faced two problems: (i) the positive contextual evidence they provided was apparently too strong, so that every question form was considered inappropriate in those conditions (the question felt redundant); and (ii) the intended negative contextual evidence was mediated by a conversational implicature (e.g., *She got a dog* implicates *She didn't get a cat* in the context used in the experiment), which led to unexpected results in some conditions. Note that there is a further combination of biases in which none of the PQ forms was judged acceptable. This concerns the neutral/¬*p* condition. We suspect that the linguistic provision of the contextual evidence in Roelofsen et al.'s study is problematic here as well. In this condition, the absence of a speaker bias (*Kate to Rose*: "I'm going to get a pet.") followed by negative contextual evidence (*Rose to Jennifer*: "Kate got a dog.") renders all PQs (*Jennifer to Rose*: "Did she get a cat?" / "Didn't she get a cat?" / "Did she not get a cat?") rather unnatural (after all, the "cat" in the question comes out of the blue).

Both of these experimental studies provide first insights into the interplay between pragmatic factors and polar question form. One drawback of the use of written materials in these studies is that one cannot control the prosody with which participants silently read the items (Fodor 2002). That the prosodic realization also plays a role in signalling speaker bias (independent of morphosyntactic marking) has been shown in a number of studies on different languages (Escandell Vidal 1998; Kügler 2003; Savino & Grice 2011; Savino 2012; Vanrell et al. 2013; Vanrell, Armstrong & Prieto 2014; Borràs-Comes & Prieto 2015). For instance, Vanrell et al. (2013) showed that the pitch scaling of the leading tone in an H + L* accent in Catalan was a major cue for the perceptual distinction between information- and confirmation-seeking polar questions (i.e., neutral vs. positive speaker bias). Specifically, a higher scaled leading tone ¡H + L* in the Catalan system was perceived as information-seeking, while a lower-scaled leading tone (H + L*) as confirmation-seeking. Vanrell et al. (2014) elicited positive and negative polar questions with and without bias towards the proposition of the sentence radical in Sardinian. Apart from lexico-syntactic differences, they report two different accentual patterns: an ¡H + L* L% pattern was frequently produced in conditions without speaker bias, while positive bias questions were primarily produced with an ¡H* + L L- pattern. For German, Kügler (2003) analyzed a set of 66 polar questions in Saxon German and argued that polar questions with a rising boundary tone are information-seeking while those with a falling boundary tone are confirmation-seeking (i.e., the answer is available in the prior discourse, see also Savino & Grice 2011 for the prosodic effects of this pragmatic difference in Bari Italian polar questions). Perception data have also shown that increases in pitch range change the perception of a question from an information-seeking to an incredulity question in English (Ward & Hirschberg 1985).

1.3 Rationale of the current studies

In the present paper, we report two experiments (Experiment 1 in English and Experiment 2 in German), in which we cross original speaker bias and contextual evidence bias. Participants were asked to select and produce the most appropriate polar question form out of five possibilities (PosQ, *really*-PosQ, LowNQ, HiNQ and Other).

The rationale for having a selection task rather than an acceptability task is the following. The two kinds of biases are not on a par. Contextual evidence is, by definition (9), mutually available to both the speaker and the addressee, but the information that led to the original speaker bias is not. Hence, while the speaker's contextual evidence bias is in the common ground, the speaker's original bias is not and the speaker can choose to express it or to leave it unexpressed. Given this, in an acceptability study, the more unmarked form (which does not express the original bias) may always be as appropriate as a marked form (which expresses the original bias) and thus we may miss the difference between the two. For example, in (11), both the PosQ continuation (11.S1) and the HiNQ continuation (11.S2) are acceptable.[9] Nevertheless, (11.S2) communicates a bias (the addressee learns that the speaker thinks Frege has reviewed for us) that (11.S1) does not (from the point of view of the addressee, the speaker may be suggesting to check Frege simply because she thinks highly of him). This difference will not be detected by a simple acceptability task.

(11)   *Context*: Unbeknownst to A, S thinks that Frege has previously reviewed for us.
         A: I'd like to send this paper to a senior reviewer.
         S1: Has Frege reviewed for us? He'd be a good one.
         S2: Hasn't Frege reviewed for us? He'd be a good one.

In our selection studies, in order to tap into what question forms are possible in a given condition when all biases are communicated, we depict informal conversations among peers where a cooperative speaker would naturally use a question form that is both (i) acceptable in the context and (ii) maximally expresses the existing biases. Subjects are asked to select the question form they would use in such a situation. We are interested in their *preferred* choice(s): i.e., the form(s) that is/are acceptable and match(es) the biases best. Note that there is a choice "Other", so that, if none of the explicit forms was a good match, subjects were free to propose a different question form. Regarding forms that are chosen less frequently, the task is not suited to make conclusions about these forms. These forms may be interpreted as inacceptable forms or as acceptable forms that, however, do not express the two biases.

   Given the problems that Roelofsen et al. (2012) encountered, we paid particular attention to make the positive contextual evidence not too strong and to create negative contextual evidence without conversational implicatures, so that a polar question seemed natural in all conditions.

   Finally, we had participants utter the selected question form to be able to document the prosodic realization in the different polar question types (e.g., for later perception experiments with auditory materials) and to map possible prosodic differences of otherwise string-identical question forms in different conditions. Note, however, that prosody was not the main aim of the experiment. Thus, in the design of the experiment, pragmatic considerations (appropriate control of biases, coherence of the stories, etc.) were ranked higher than prosodic considerations (matching syntax, word-prosodic structure, number of syllables, using sonorous sounds, etc.). Consequently, not all items are equally well suited for prosodic analysis.

---

[9] Indeed, in Roelofsen et al.'s (2010) acceptabiltiy study, PosQs and HiNQs are both rated as highly natural in scenarios like (11) exemplifying the *p*/neutral condition (1.41 and 1.78 respectively on a scale from 1 (highly natural) to 7 (highly unnatural)).

1.4 Research questions, hypotheses and predictions

The present paper addresses two main questions currently under debate in the literature:

(RQ1) Which pragmatic bias(es) is the surface form of a polar question sensitive to? More concretely, do the use-conditions of polar question forms depend on:
  (**Hyp 1**)   only evidence bias
  (**Hyp 2**)   only original bias, or
  (**Hyp 3**)   a combination of original bias and evidence bias?

(RQ2) What is the exact mapping between pragmatic conditions and the surface forms PosQ, LowNQs, HiNQs and *really*-PosQ? In particular:
  (i)   What surface PQ forms present the same selection pattern – i.e., they are selected across the same pragmatic conditions – and thus constitute together a single PQ type? And what forms differ in selection pattern – i.e, they are selected in different pragmatic conditions – and thus constitute separate PQ types?
  (ii)   For each of the resulting PQ types, what are the pragmatic conditions (combination of original bias and evidence bias) under which they are selected?

Let us elaborate on our two research questions and hypotheses. We start by plotting the two kinds of bias on two axes, as in Table 4. The horizontal axis corresponds to original bias and the vertical axis to evidence bias, with three possible values for each kind of bias: bias for p, neutral (i.e., no bias for or against p) and bias for $\neg p$ (i.e., bias against p). We are interested in the distribution of preferred PQ forms over the white cells. The grey cells are left out for the following reason. A speaker with an original bias for p that receives contextual evidence for p (cell p/p) will most naturally assume that p is true and will not feel prompted to inquire about the truth of p. A parallel reasoning applies to the cell $\neg p/\neg p$. Indeed, all question PQ forms received very low naturalness ratings in these conditions in Roelofsen et al. (2012). As for the cell $\neg p$/neutral (original bias for $\neg p$ / no evidence bias), the PQ form predicted by (part of) the literature is a complex form that combines two of the surface factors (Romero & Han 2004; AnderBois 2011): high negation and low negation. Complex PQs will not be tackled in the present study.[10] This leaves us with six experimental conditions to be tested in a first step.

Let us turn to the hypotheses in RQ1 and the corresponding predictions they make for RQ2. First, recall Büring & Gunlogson's (2000) data on evidence bias from Table 2. Assuming that they manipulated contextual evidence while leaving any potential original speaker bias aside, their data should be plotted as in Table 5. Now, under the hypothesis that

		ORIGINAL BIAS		
		p	Neutral	¬p
CONTEXTUAL EVIDENCE	p			
	Neutral			
	¬p			

Table 4: Crossing original bias and contextual bias.

---

[10] One example illustrating the pragmatic cell $\neg p$/neutral comes from Romero & Han (2004: 28):

  (i)   A: I'd like to send this paper to a senior reviewer, but I'd prefer somebody new.
         B: Has*n't* Frege *not* reviewed for us yet?

This and other complex PQ forms are left out of the present paper for two reasons: (i) to keep the experimental design manageable and (ii) because we think it is important to first pin down the pragmatic contribution of each surface cue separately (presence/absence of high negation, presence/absence of low negation and presence/absence of *really*) and then, in a next step, to investigate the joint combination of two surface cues and see whether this joint contribution can be derived compositionally.

evidence bias is the only kind of bias that matters for the selection of a PQ form (**Hyp 1**), the p-column and the $\neg p$-column should look identical to the middle neutral column (leaving out the grey cells). This is shown in Table 6. This means that, if only evidence bias matters, the predicted answers to our RQ2 are: (i) each of the three plotted surface forms – PosQ, LowQ and HiNQ – constitutes a separate PQ type, one of them possibly jointly with the left-out form *really*-PosQ, and (ii) leaving aside *really*-PosQs, the distribution of preferred PQ forms across pragmatic conditions is as in Table 6.

Second, recall Ladd's (1981) and Romero and Han's (2004) data on original speaker bias. If we plot the data from Table 1 and add their considerations about contradiction and non-contradiction scenarios (namely, that *really*-PosQ and Inner-HiNQs are felicitous only in contradiction scenarios, whereas Outer-HiNQs are acceptable with or without contradiction), we arrive at Table 7. Now, under the hypothesis that original bias is the only one that matters (**Hyp 2**), the blank cells in the p-row and in the $\neg p$-row should be identical to the corresponding cell on the neutral row. This results in Table 8. This hypothesis leads to the following answers to RQ2: (i) PosQs and LowNQs together constitute a joint PQ type, separate from the HiNQ type and from the *really*-PosQ type, and (ii) the distribution of PQ forms across pragmatic cells will be as depicted in Table 8.

Finally, under the hypothesis that both evidence bias and original bias are crucial for the choice of PQ form (**Hyp 3**), we need to merge Tables 6 and 8. The merge is done as follows. In a first step, for each cell, we take the intersection of the question forms considered by both theories, namely, the intersection among the forms PosQ, LowNQ and Outer-HiNQ. In the $\neg p/p$ cell, this results in the empty set. In a second step, we add the question forms studied by only one of the approaches. These are the form *really*-PosQ, which is added to the $\neg p/p$ cell, and the form Inner-HiNQ, which is added to the $p/\neg p$ cell.

CONTEXTUAL EVIDENCE	ORIGINAL BIAS		
	p	Neutral	**¬p**
p		PosQ	
Neutral		PosQ / HiNQ (outer)	
¬p		LowNQ / HiNQ (outer)	

Table 5: Plotting Büring and Gunlogson's (2000) data on a two-bias matrix.

CONTEXTUAL EVIDENCE	ORIGINAL BIAS		
	p	Neutral	**¬p**
p		PosQ	PosQ
Neutral	PosQ / HiNQ (outer)	PosQ / HiNQ (outer)	
¬p	LowQ / HiNQ (outer)	LowNQ / HiNQ (outer)	

Table 6: Predicted distribution of PQ forms under **Hyp 1** (only evidence bias matters).

CONTEXTUAL EVIDENCE	ORIGINAL BIAS		
	p	Neutral	**¬p**
p			*really*-PosQ
Neutral	HiNQ (outer)	PosQ / LowNQ	
¬p	HiNQ (outer / inner)		

Table 7: Plotting Ladd (1981) and Romero and Han's (2004) data on a two-bias matrix.

The result is Table 9. This predicts the following answers for our RQ2: (i) each of the four surface forms – PosQ, LowQ, HiNQ and *really*-PosQ – constitutes a separate PQ type, and (ii) the distribution of PQ forms will be as indicated in Table 9.[11]

Note that, by concentrating on what kind(s) of bias play(s) a role, the distinction between Ladd's (1981) inner and outer readings on English HiNQs exemplified in (7–8) is only incompletely captured in the predicted tables. While some hypotheses predict a particular pragmatic condition – the *p*/neutral cell – to host solely HiNQs in the outer reading, under no hypothesis is there a pragmatic cell exclusively dedicated to HiNQs in their inner reading. In particular, the *p*/¬*p* cell in Tables 7 and 8 ambiguously allows for the outer and inner reading of HiNQs. We will call the *p*/¬*p* cell, henceforth, the ambiguity cell.

Here the comparison between English and German will be valuable. To see this, compare the four PQ surface forms to be tested in the two languages, given in Tables 10 and 11:

Interestingly, though the syntax of the English and German PQ forms is largely parallel (see Schwarz & Bhatt 2006 on the scopal position of high negation in German), their pragmatic use is partly misaligned. Of the two readings that Ladd (1981) describes for English HiNQs, the outer reading is translated into German using high negation (*nicht…ein Restaurant* 'n't…a restaurant') and the inner reading with low negation (*kein Restaurant* 'no restaurant') (Büring & Gunlogson 2000). This cross-linguistic difference opens a critical

CONTEXTUAL EVIDENCE	ORIGINAL BIAS			
		p	**Neutral**	**¬p**
	p		PosQ / LowNQ	*really*-PosQ
	Neutral	HiNQ (outer)	PosQ / LowNQ	
	¬p	HiNQ (outer/inner)	PosQ / LowNQ	

Table 8: Predicted distribution of PQ forms under **Hyp 2** (only original bias matters).

CONTEXTUAL EVIDENCE	ORIGINAL BIAS			
		p	**Neutral**	**¬p**
	p		PosQ	*really*-PosQ
	Neutral	HiNQ (outer)	PosQ	
	¬p	HiNQ (outer/inner)	LowNQ	

Table 9: Predicted distribution of PQ forms under **Hyp 3** (both evidence bias and original bias matter).

PosQ		Is there a good restaurant?
really-PosQ		Really?! Is there a good restaurant?
LowNQ		Is there no good restaurant?
HiNQ	inner reading	Isn't there a good restaurant?
	outer reading	

Table 10: English PQ forms discussed in the literature.

---

[11] In the scarse theoretical works that distinguish and investigate both kinds of bias, the intuitive characterization of the PQ forms that Sudo's (2013) examines – namely, PosQ, Inner-HiNQ and Outer-HiNQ – is as in Table 9. As for AnderBois (2011), his distribution of HiNQs and *really*-PosQs matches the *p*-column and ¬*p*-column of Tables 7 and 8; but it is not clear what his analysis of what he calls "weak" bias (a combination of evidence bias and bouletic-like bias) predicts for the middle column when evidence bias and bouletic bias play against each other.

PosQ		Gibt es ein gutes Restaurant?
really-PosQ		Wirklich?! Es gibt ein gutes Restaurant?
LowNQ		Gibt es kein gutes Restaurant?
HiNQ	inner reading	
	outer reading	Gibt es nicht ein gutes Restaurant?

Table 11: German PQ forms discussed in the literature.

window to examine what participants are doing in the ambiguous $p/\neg p$ condition. If subjects are producing solely Outer-HiNQ readings, English and German subjects will select *n't…a* and *nicht…ein* respectively at comparable rates. But, if a percentage of Inner-HiNQ readings is produced, this percentage will surface as a differential between the English speakers – which will select *n't…a* for those readings – and the German speakers – who will express them with *kein*.

We hope to be able to use prosody to disentangle the different readings. More concretely, if there are two possible readings for a given form, we expect to find two different prosodic realizations. On the other hand, if there is just one reading, we expect less prosodic variation and, ideally, just one single prosodic form.

One further point needs to be made concerning the forms to be used in our experiments. Both in the English and German materials, the adverb *really* precedes the full polar question rather than being inserted in it. In scenarios described as prototypical for *really*-PQs by Romero and Han (2004), we found that the function of double-checking p can be carried out by a prefixed *really* as well. We chose the prefixed version (i.e., *Really? Is there a good restaurant nearby?*) in order to allow participants to select the *really*-PosQ form no matter whether their preferred realization was via *really* or via focus on the verb.

2 Experiment 1
2.1 Methods
2.1.1 Participants

Forty-two English native speakers (between 20 and 36 years, $M = 25.0$; SD $= 2.9$; 17 males, 25 females) participated for a small fee or course credit. Two further participants were tested but they did not complete the experiment. All participants were from the participant pool of the *Division of Psychology and Language Sciences – University College London* (UK). None of them was aware of the goal of the experiment. Informed consent was obtained from every participant.

2.1.2 Stimuli

We created 46 written scenarios (one used as practice trial, 30 as target trials and 15 as filler trials). The scenarios presented ordinary fictional conversations (e.g., two friends preparing the dinner, two students looking for the library). Each story was composed of two caption/picture pairs, followed by the selection of the most appropriate polar question.

In the experimental trials, the first caption/picture combination manipulated the original speaker bias (for instance, *He usually takes a train in the early morning, before 7:00* to generate a bias for the proposition p, see left display in Figure 1), the second manipulated contextual evidence (e.g., the question mark in Figure 2, together with the sentence *Do you have any preference?* to provide no contextual evidence). The captions were the same across experimental conditions, while pictures varied in such a way as to induce the six experimental conditions introduced above (i.e., $p/\neg p$, p/neutral, neutral/neutral, $\neg p$/neutral, neutral/p, $\neg p/p$). The pictures providing the contextual evidence were subtle enough in order not to prevent participants from using a question at all. In the conditions $p/\neg p$, $\neg p/p$,

neutral/¬*p* and neutral/*p,* the use of an information-seeking question was triggered by the epistemic mismatch between the original bias and the provided contextual evidence. More precisely, we designed the pictures in a way that the degree of evidence provided by the pictures that generated the original biases was stronger than the degree of evidence provided by the pictures that induced the contextual evidence. The contextual evidence provided was therefore strong enough (but not too strong) to generate an epistemic mismatch that prompted the use of a question. All pictures were pretested with a paper-and-pencil questionnaire with 18 participants. In the questionnaire, participants were presented with one caption and one picture at a time (cf. Figure 1 and 2), together with a question asking about the likelihood of a certain event (e.g., "How likely is it that there is a train from Nottingham to Sheffield in the early morning?"). Participants rated the probability of the event described by the critical proposition *p* (from 1 – very unlikely to 5 – very likely). Pictures that obtained scores close to 1 were selected as original bias/contextual evidence against, close to 3 as neutral, close to 5 as in favour.

Each caption/picture pair was followed by a fixed set of four questions, one positive polar question, a positive polar question preceded by *really*, a high-negation question and a low negation question, see Table 12. The full list of PosQs is shown in Table A1 in the Appendix. The four question types were presented together with an "other" option after each scenario. The order of choices (PQ forms) was constant across trials.

The filler trials also had two caption/picture pairs and a set of questions. In filler trials, the questions were always four constituent questions (how, where, when, who), plus an "other" option.
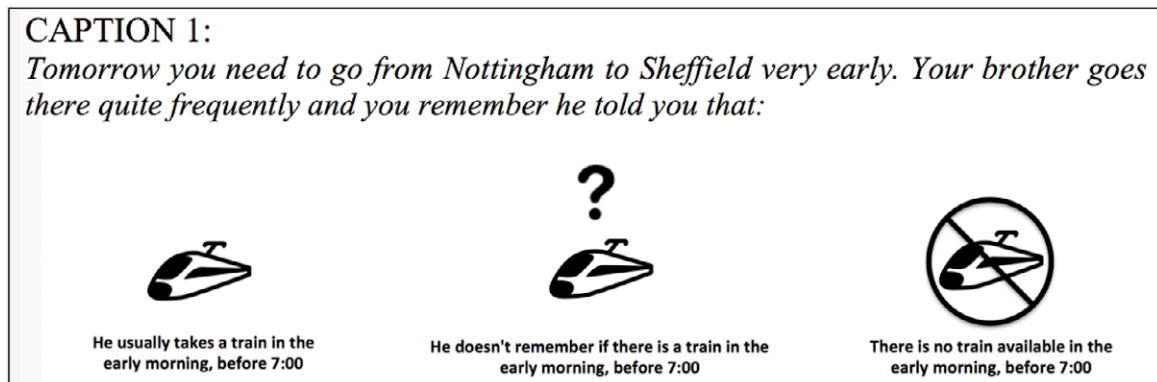
Figure 1: Example of captions and the pictures used to generate three different original biases (i.e., *p*, neutral, ¬*p*) towards the proposition *p* = *there is a train in the early morning* in one of the scenarios of the English version of the experiment.
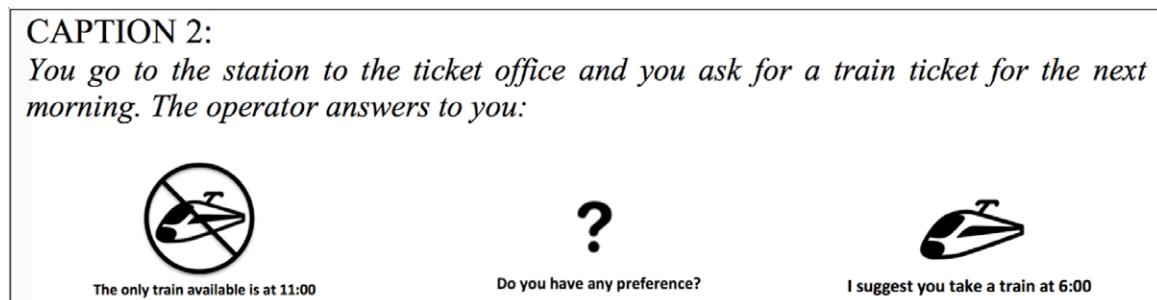

Figure 2: Example of captions and the pictures used to generate three evidence biases (i.e., ¬*p*, neutral, *p*) towards the proposition *p* = *there is a train in the early morning* in one of the scenarios of the English version of the experiment.

QUESTION TYPE	EXAMPLE QUESTION
PosQ	Is there a train in the early morning?
Really-PosQ	Really?! Is there a train in the early morning?
LowNQ	Is there no train in the early morning?
HiNQ	Isn't there a train in the early morning?
Other option	Other ways of asking if there is a train in the early morning

Table 12: Example list of questions for the English version.

The linguistic materials, the coherence of the stories, and the naturalness of the questions used in the experiments were evaluated by seven participants beforehand and altered, if necessary.

2.1.3 Procedure

There were six experimental lists, rotating the relevant levels of original bias and contextual evidence across trials following a Latin Square Design. Consequently, each participant got each of the 30 experimental items, but each item in only one of the six conditions (resulting in 5 items per condition). The trials were pseudo-randomized, repeating a certain condition at most once. Each list further included all the filler items, approximately evenly distributed in the list. The practice trial was put at the beginning of the list. Participants were randomly assigned to one of the experimental lists (7 participants for each list).

Participants were tested individually in a quiet laboratory at the *Division of Psychology and Language Sciences – University College London* (UK). The experiment started with written instructions informing them that they were expected to read the two captions attached to the pictures, and to select the question that sounded most natural by producing it aloud. Then, an example was given for each phase of a trial, followed by a practice trial. If none of the questions was considered appropriate, participants were instructed to formulate a different question or to say "Other ways of asking".

Instructions, stimuli, response recording and data collection were controlled by a laptop computer with an 13 inch display. Participants sat approximately 60 cm from the display. The room had normal lighting. Only a keyboard (no mouse) was available for performing the experiment. The experiment was self-paced. In all trials, the first caption was shown on the screen together with the first picture whose purpose was to generate a positive/negative/neutral original bias towards the proposition p. To proceed, participants had to click the space bar on the keyboard. The second caption was shown on the screen together with the second picture whose purpose was to generate a positive/negative/contextual evidence bias towards the proposition p. After clicking the space bar, the list of questions appeared on screen (see Figure 3). After producing a question, participants could proceed by pressing the space bar again. Their response was recorded directly by a *Olympus LS-5* microphone (sampling rate 44.1kHz, 16 Bit). The average duration of the whole experiment was 16 minutes.

2.2 Results

Participants' choices were coded online, during the experiment, for response option (PosQ, *really*-PosQ, HiNQ, LowNQ, other). Responses were then manually cut, and the online coding was checked by as a second coder and corrected, if necessary. Changes in
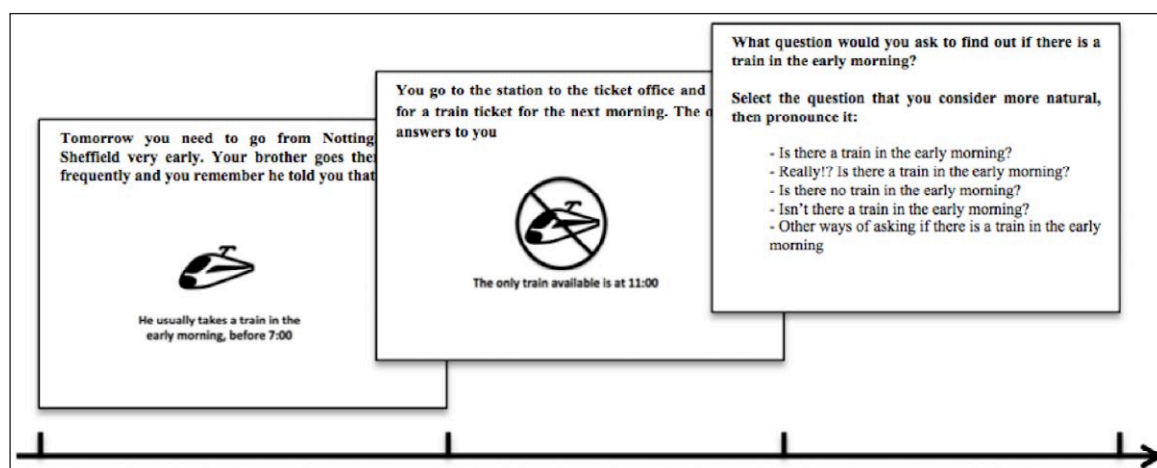
Figure 3: Example of a trial in condition *p/¬p*.

word order (no verb inversion), tense, and number as well as addition of particles were ignored if they did not affect the two biases at issue.[12]

There were 41 cases in which participants chose the "other" category. Of these, in 12 cases they produced declaratives (29% of the "other" cases), in 10 cases they produced double codings such as *Isn't central street not open?, Really, don't you have a big tent?* (24%), in 7 cases wh-questions (17%), in 10 cases productions with very different wordings (24%) and in two cases tag questions (5%).

For each condition, we calculated the average proportion of responses for any of the five response options and the 95% confidence interval CI (Sison & Glaz 1995), as can be seen in Figure 4.

As shown in Figure 4, there are clear preferred and non-preferred question forms for each of the six conditions. Usually, there is just one or at most two question forms that were chosen in more than 25% of the cases (which would represent chance level if we only consider the four proper question response options and exclude "other"). For the statistical analysis, we are not interested in whether the preferred choice differs from chance, but instead, whether the most frequently chosen question type was chosen in the *majority* of cases (more than 50% in the respective conditions).[13] To this end we averaged the percentage of the preferred choice by participants and items and subjected them to two separate one-sample t-tests (t1 and t2). The results are summarized in Table 13.

These analyses show that the proportion of the preferred choice constituted the majority in five out of the six conditions. Only in the neutral/*p* condition was there a strong competition between PosQs and *really*-PosQs, so that the expected PosQ condition was

---

[12] There were 17 cases with verb-second (V2) word orders (8x in condition ¬*p/p*, 4x in condition neutral/*p*, 4x in condition neutral/¬*p* and once in *p/¬p*). V2 word orders occurred with *really*-PosQs 7 times, with PosQs 6 times, with LowNQ 3 times and once with a HiNQ. Note that eleven of the V2 cases were produced by two participants, suggesting a participant-specific preference. According to the literature, V2 signals that the speaker attributes a particular piece of information (*p* in positive rising declaratives and ¬*p* in negative rising declaratives) to the addressee (Gunlogson 2003). This does not seem to affect the speaker biases we tested. Furthermore, if we classify these V2 cases as "other", the general pattern of findings does not change. There were 18 cases with additional interjections, particles or adverbials (3 each *oh/so*, 2 each *any/now/then*, 1 each *sometimes/also/definitely/too/yet*) and one case in which *really* occurred in sentence-medial position.

[13] Note that next to the preferred choice there often is a form that is in-between the preferred choice and seemingly irrelevant choices. The secondary choices in the conditions neutral/¬*p*, *p*/neutral, *p/¬p* and ¬*p/p* did not differ from chance (p > 0.6, p > 0.2, p > 0.25 and p > 0.15 respectively), the secondary choice in condition neutral/neutral was significantly lower than chance (p = 0.01), and the secondary choice in condition neutral/*p* was significantly above chance (p < 0.01). So only the *really*-PosQ in condition neutral/*p* seems a serious competitor to the preferred choice (here PosQ).
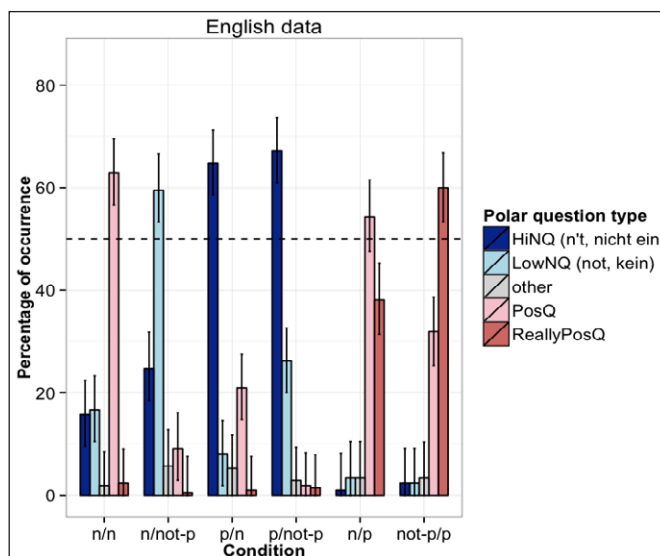
Figure 4: Percentage of selected questions across condition in English. Whiskers represent the 95% confidence interval.

	preferred choice	average value [95% CI]	analysis by subjects	analysis by items
neutral/neutral	PosQ	63% [56;70]	t1(41)= 3.7, p < 0.001	t2(25) = 2.0, p <= 0.052
neutral/¬p	LowNQ	59% [50;68]	t1(41) = 2.1, p < 0.05	t2(29) = 2.4, p < 0.03
neutral/p	*PosQ*	*54% [45;63]*	*t1(41) = 1,0, p > 0.3*	*t2(25) = 1.3, p > 0.19*
¬p/p	*really*-PosQ	60% [51;69]	t1(41) = 2.3, p = 0.02	t2(29) = 3.2, p < 0.005
p/neutral	HighNQ	65% [55;74]	t1(41) = 3.1, p < 0.005	t2(29) = 2.8, p < 0.01
p/¬p	HighNQ	67% [59;75]	t1(41) = 4.6, p < 0.0001	t2(29) = 4.2, p =< 0.0005

Table 13: Average percentage of the most frequently chosen responses for each condition in English, together with the 95% CI (averaged by subjects) and results of one-sample t-test of the by-subjects and by-items analysis.

not the unique most preferred choice. One possible explanation is that the preposed *really* was interpreted as a discourse particle, which was produced in order to signal interest in the proposition. If that was the case, we would expect different prosodic realizations of the *really*-PosQs in this condition, in which it was unexpected, and in the *¬p/p* condition, in which it was expected.

To test this post-hoc explanation, we analyzed a subset of the data prosodically: we selected eight cases that had a similar syntactic and word-prosodic structure (*Do you like beer?, Do you have a bike?, Do you have a tent?, Do you have dogs?, Do we have a printer?, Do you have a backpack?, Do you like flying?* and *Have you bought the ticket?*) and classified the intonational realization of *really*, the distribution of pitch accents and the nuclear tune (last pitch accent and boundary tone) in the polar question (see Table 14).

Table 14 suggests some differences in the realization of *really*-PosQs when produced in the expected *¬p/p* condition as compared to the neutral/*p* condition. For the word *really*, there are three typical realizations: a plain rise, a fall and a fall-rise. There do not

	¬p/p (N = 30)	neutral/p (N =19)
fall-rise *really*	9 (30%)	4 (21%)
falling *really*	6 (20%)	5 (26%)
rising *really*	14 (47%)	10 (53%)
accent on non-finite verb	13 (43%)	3 (10%)
falling nuclear contour	19 (63%)	9 (47%)
rising nuclear contour	6 (20%)	8 (42%)
fallrise nuclear contour	5 (17%)	2 (10%)

Table 14: Absolute numbers of particular prosodic realizations (numbers in brackets refer to percentages relative to the analyzed items in each condition).

seem to be distributional differences across conditions. However, in the ¬*p/p* condition, participants produced numerically more accents on the non-finite verb (e.g., *Do you LIKE beer,* where capitals signal the nuclear pitch accent) than in the neutral/*p* condition. Furthermore, the question ended in a falling nuclear contour in the majority of the ¬*p/p* cases (63%), while rising and falling contours were equally frequent in the neutral/*p* condition (42% and 47%, respectively).

2.3 Discussion

Plotting the most frequently selected PQ choices from our experimental results in a two-bias table, we obtain Table 15.

RQ1 tackled the issue which pragmatic bias(es) the surface form of polar questions is sensitive to: only evidence bias, only original speaker bias or both. Table 15 shows that both kinds of bias matter. If only evidence bias was relevant, we would expect to find homogenous rows: the cells neutral/*p* and ¬*p/p* would show the same selection pattern, and so would the cells *p*/neutral and neutral/neutral and the cells *p*/¬*p* and neutral/¬*p*. If only the original speaker bias was relevant, we would expect homogenous columns, with the cells neutral/*p*, neutral/neutral and neutral/¬*p* in the middle column crucially showing the same choice pattern. Instead, what we find is heterogeneity, showing that the selection of PQ form depends on a combination of both kinds of bias.

RQ2 was concerned with the preferred choice of PQ form in each pragmatic cell or condition. The pattern of results is almost identical to the predictions displayed in Table 9 in Section 1.4, where we had taken the fragmentary approaches in the literature and merged their predictions. We briefly discuss each pragmatic condition in turn.

In the neutral/neutral condition, the favoured PQ choice is a PosQ (63%), whereas HiNQs (16%) and LowQs (17%) are dispreferred. Unless some other ingredient (e.g., non-epistemic biases mentioned in Footnote 4) is factored in, the low selection rate of HiNQs and of LowQs in this condition is unexpected for an only-evidence bias approach and for an only-original bias approach, respectively.

We turn to the pragmatic conditions where a negative question form was expected, namely the cells neutral/¬*p*, *p*/neutral and *p*/¬*p*. Comparing the neutral/¬*p* and *p*/neutral cells, our results show that HiNQs in their outer reading and LowNQs are selected differently. HiNQs are the preferred choice in the *p*/neutral condition where the speaker is double-checking *p* (65%) but dispreferred in the neutral/¬*p* condition (24%). Conversely, LowNQs are the preferred form in the latter condition (59%) and dispreferred in the former condition (8%). In terms of the typology of PQs, this means that Outer-HiNQs and LowNQs cannot be merged into one single type, contra van Rooij and Šafářová (2003). Rather, they constitute two truly distinct form-function types.

CONTEXTUAL EVIDENCE	ORIGINAL BIAS		
	p	**Neutral**	**¬p**
p		PosQ / *really*-PosQ	*really*-PosQ
Neutral	HiNQ (outer)	PosQ	
¬p	HiNQ (outer/inner)	LowNQ	

Table 15: Results for preferred PQ form per pragmatic cell in English.

As for the *p/¬p* condition, HiNQs are the preferred choice (67%) and LowNQs are dispreferred (26%), as predicted by **Hyp 3**. However, this is the ambiguity cell allowing both for an Outer-HiNQ reading double-checking *p* and for an Inner-HiNQ double-checking *¬p*. Since we have no evidence on whether participants produced any Inner-HiNQs in this condition at all just by looking at the English data, we cannot reach any conclusion on their grouping or separation from other PQ forms. As far as these data go, Inner-HiNQs might constitute a joint grammatical type with Outer-HiNQs (AnderBois 2011), they may merge with LowNQs into a single type (Asher & Reese 2007; Krifka to appear) or they may differ from both (Romero & Han 2004; Reese 2006; Walkow 2009). We will return to this ambiguity cell in Section 4, where English and German will be compared.

Finally, we examine the conditions *¬p/p* and neutral/*p*. In the former condition, *really*-PosQ was the preferred choice, as expected. In the condition neutral/*p*, PosQ was the preferred choice, again as expected, but there was also competition from the form *really*-PosQ. One potential explanation for the frequent selection of *really*-PosQs is that, when *really* precedes its clause, it can be understood as an epistemic adverb asking for confirmation about *p* (which was intended) or as a discourse particle that signals interest or engagement in the information conveyed by *p*. In the condition that is expected to trigger *really*-PosQ unambiguously (*¬p/p* cell) (Romero & Han 2004), we found considerably more *really*-PosQ productions, suggesting that prefixed *really* was understood in its intended epistemic use. In the neutral/*p* condition, we find fewer *really*-PosQ productions and these compete with the PosQ productions, suggesting that prefixed *really* was interpreted here as the discourse particle. Our preliminary prosodic analysis suggests that the prosodic realization may be used to distinguish string-identical question types as shown for *really*-PosQs. For instance, in the *¬p/p* condition, *really*-PosQs were more frequently produced with an accent on the non-finite verb and with a falling pitch contour, as compared to the neutral/*p* condition. An accent on the non-finite verb (*Do you LIKE beer?*) has been argued to signal negative original bias (Asher & Reese 2007; Turco, Dimroth & Braun 2013) comparable to that triggered by an accent on the finite verb (*DO you like beer?*) (Romero & Han 2004, building on Höhle 1992). Note that the different proportion of accents on the verb cannot be attributed to differences in information status of the object across conditions. The caption/picture pairs that manipulated contextual evidence bias were matched for information status across conditions. Further research is necessary to arrive at a more complete understanding of the pragmatic effects of the final pitch rise vs. fall in these contexts.

3 Experiment 2
3.1 Methods
3.1.1 Participants
Forty-two German native speakers (between 18 and 37 years, M = 24.7; SD = 2.9; 15 males, 27 females) participated for a small fee or course credit. They were all recruited from the participant pool of the *Department of Linguistics – University of Konstanz* (*Germany*). None of the participants was aware of the goal of the experiment. Informed consent was obtained from every participant.

3.1.2 Materials

The materials were similar to the English materials. Three student assistants translated the English materials into German. However, given the structure of the two languages it was not always possible to have the same caption/picture pairs across languages (in German the distinction between low and high negation can be realized overtly only in certain conditions, such as quantified sentences, cf. Schwarz & Bhatt 2006). The full list of questions for the two languages is shown in Table A1 in the Appendix. In total, 17 items were conceptually and structurally similar across languages (see boldface in Table A1). Thirteen new items had to be constructed.

In the translation and construction process, the three student assistants were instructed to compile naturally sounding German questions. One of the student assistant opted for *really*-PosQs with a V2 syntax (i.e., *Wirklich? Es gibt…?* 'Really? There is…?'), one for *really*-PosQ with V1 syntax (i.e., *Wirklich? Gibt es…?* 'Really? Is there..?') and the third produced a majority of V2 patterns (28 V2-items out of 30), which is the option we finally opted for. As in the English version of the experiment, the linguistic materials were evaluated by six additional participants before the start of the experiment and modified if necessary.

3.1.3 Procedure

The procedure was identical to Experiment 1. Participants were tested in the *PhonLab at the University of Konstanz*. The experiment lasted 19 minutes on average.

3.2 Results

The data were coded as in Experiment 1. In four cases, participants did not produce anything. In total, participants chose the "other" option in 34 of the 1260 trials. What participants produced in the "other" cases were double codings in 14 cases (e.g., *Wirklich, gibt es keinen Bus nach 21 Uhr*?, 'Really? Is there no bus after 11p.m.?', 41%), wh-questions in 5 cases (15% of the cases), declarative questions in 3 cases (9%), 10 cases with very different wording (29%) and two tag questions (6%).

Like the English participants, the German participants also added extra particles to the polar question types. This occurred in 34 cases overall. Most frequent were *also* (discourse particle) (9 x), *auch* 'too' (5 x), *dann* 'then' (4x), *denn* (discourse particle) (3x), and *mehr* 'more' (3x).

The data were analysed as in English. The percentage of selected question types across conditions is shown in Figure 5. The results of the t-tests comparing the most frequently selected PQ form against 50% are summarized in Table 16. Similar to English, the proportion of the preferred choice was above 50% in five out of the six conditions. Again, the only condition in which the preferred choice was not chosen significantly more often than 50% was the PosQ in the neutral/*p* condition.[14]

In English, the prosodic analysis indicated differences in the realization of *really*-PosQ across conditions (neutral/*p* and ¬*p/p*), which suggests that the *really*-PosQs produced in the neutral/*p* condition did not have the same function as in the ¬*p/p* condition.

To corroborate the prosodic differences found for English also for German, we prosodically analyzed a subset of the German data (*Gibt es einen Supermarkt?* 'Is there a supermarket?', *Gibt es einen Tutor?* 'Is there a tutor?', *Gibt es einen Direktflug?* 'Is there a direct

---

[14] Analyses for the secondary choices showed that for conditions neutral/neutral and ¬*p-p*, the secondary choice did not differ from chance (p = 0.1 and 0.6, respectively). In the neutral/¬*p* and *p*/neutral conditions, the secondary choices were significantly below chance level (both p-values < 0.01). In the neutral/*p* and *p*/¬*p* conditions, the secondary choices were significantly above chance level (both p-values < 0.005). As in English, the *really*-PosQ seems to be a strong competitor to PosQ in the neutral/*p* condition. Unlike English, the LowNQ seems to present some competition to the HiNQ in the *p*/¬*p* condition. We will look closely at the difference between English and German in the latter condition in Section 4.
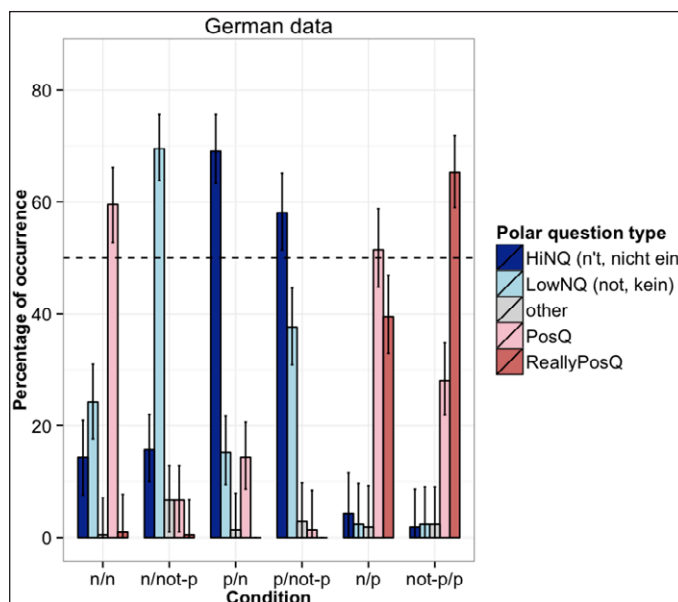
Figure 5: Percentage of selected questions across condition. Whiskers represent the 95% confidence interval.

	preferred choice	average value	analysis by subjects	analysis by items
neutral/neutral	PosQ	59% [52;67]	t1(41) = 2.7, p < 0.01	t2(29) = 2.1, p = 0.04
neutral/¬p	LowNQ	69% [63;76]	t1(41) = 6.4 p < 0.0001	t2(29) = 5.2, p < 0.0001
neutral/p	PosQ	51% [43;60]	t1(41) = 0.3, p > 0.7	t2(29) = 0.5, p > 0.6
¬p/p	really-PosQ	65% [57;73]	t1(41) = 3.8, p < 0.0005	t2(29) = 2.7, p = 0.01
p/neutral	HighNQ	69% [61;77]	t1(41) = 5.1, p < 0.0001	t2(29) = 4.5, p < 0.0001
p/¬p	HighNQ	58% [50;66]	t1(41) = 2.0, p = 0.05	t2(29) = 2.2, p = 0.03

Table 16: Average percentage of the most frequently chosen responses for each condition in German, together with the 95% CI (averaged by subjects) and results of one-sample t-test of the by-subjects and by-items analysis.

	¬p/p (N = 25)	neutral/p (N=13)
accent on non-finite verb	10 (40%)	1 (8%)
rise ending in a high plateau (H-%)	11 (44%)	7 (54%)
high rise (H-^H%)	11 (44%)	6 (46%)
low rise (L-H%)	1 (4%)	0
nuclear fall (H* L-%)	2 (8%)	0

Table 17: Absolute and relative number of accents on the finite verb and the nuclear tunes of the question in conditions ¬p/p and neutral/p.

flight?'*, Gibt es eine Mensa?* 'Is there a Mensa?'*, Hast du ein Fahrrad?* 'Do you have a bike?'*, Hast du ein Auto?* 'Do you have a car?'*, Hast du einen Ausweis?* 'Do you have a passport?'*, Hast du eine Garage?* 'Do you have a garage?').

The results of the prosodic analysis (see Table 17) showed that *wirklich* 'really' was produced with a high-rising tune in all of the cases (unlike in English, where there was

more variation). Regarding the distribution of accents in the second part of the question, there was, importantly, a higher proportion of accents on the finite verb (e.g., *Wirklich? Es GIBT einen Supermarkt* 'Really? There IS a supermarket?') in the ¬*p/p* condition than in the neutral/*p* condition. Note that not all accents on the finite verb were the sole accents in the phrase, but some occurred together with an additional accent on the object. Other than that, there are no strong differences across conditions in the realization of nuclear tune, besides a slightly higher proportion of questions ending in a high plateau (H-%) in the neutral/*p* condition. Conversely, we have a few cases of falling and low-rising questions in the condition ¬*p/p,* two prosodic forms that are not attested in the neutral/*p* condition. In the neutral/*p* condition, the most frequent pattern is a high-rise.

3.3 Discussion

The German experimental results are mapped to the two-bias matrix in Table 18.

The distribution of primary PQ choices across pragmatic conditions is identical to the one we found for English. The overall results are, thus, the same.

With respect to RQ1, both original bias and evidence bias affect the selection of PQ forms in German.

With respect to RQ2, the biases influence the selection in the two languages in the same way, since the primary choices for each cell are the same. The predictions concerning the neutral/neutral cell are confirmed for German too: the favoured PQ choice is a PosQ (59%), while HiNQs (14%) and LowQs (24%) are dispreferred (contra an only-evidence bias approach and an only-original bias approach, unless non-epistemic bias is factored in). Among the German negative question forms, Outer-HiNQs and LowNQs constitute separate form-function types, since they are selected in different conditions: HiNQs are the preferred choice in the *p*/neutral condition (69%) but dispreferred in the neutral/¬*p* condition (16%), while LowNQs are the preferred choice in the latter condition (70%) and dispreferred in the former condition (15%). Finally, next to the intended epistemic interpretation, prefixed *wirklich* 'really' arguably has a use as discourse particle that allows it to be selected in the neutral/*p* condition. The two interpretations of prefixed *really* are echoed in the prosodic structure to some extent. When *really*-PosQs are selected in the ¬*p/p* condition, we find more accents on the finite verb than in the neutral/*p* condition. This is what we would expect, since this prosodic realization has been argued to signal a negative original bias (Asher & Reese 2007). Overall, more research is needed on the prosodic realization of *really*-PosQ in the two languages.

4 The ambiguity cell in English and German

As discussed in Section 1.4, our experimental design does not allow us to trace whether Inner-HiNQs are being produced if we look exclusively at English. But, given the partial misalignment of PQ forms and functions between English and German that we saw in Tables 10 and 11, a comparison between the two languages delivers some insights. The relevant condition is the *p/*¬*p* condition, which hosts an ambiguity, as it is predicted

		ORIGINAL BIAS		
		p	**Neutral**	**¬p**
CONTEXTUAL EVIDENCE	**p**		PosQ / *really*-PosQ	*really*-PosQ
	Neutral	HiNQ (outer)	PosQ	
	¬p	HiNQ (outer/inner)	LowNQ	

Table 18: Results for preferred PQ form per pragmatic cell in German.

to allow both for Inner-HiNQs and Outer-HiNQs in English. The results for English and German for this cell are plotted together in Figure 6. To corroborate the seeming difference in the proportion of HiNQs in English and German, the individual responses for HiNQ were analysed using a binomial mixed-effects regression model with language as fixed factor and participants and items as crossed random factors, allowing for random intercepts and slopes (Baayen 2008; Jaeger 2008; Cunnings 2012). The results showed a significant effect of language group (ß = 0.46, SE = 0.24, t = 1.9, p = 0.05). In addition, in the English data set, there was a significantly higher number of speakers that predominantly (in more than 80% of the $p/\neg p$-cases) chose the HiNQ than in German (22 English speakers vs. 12 German speakers, $\chi^2(1)$ = 4.9, p = 0.03).

How can we explain that the difference between the selection rate of HiNQs vs. LowNQs is considerably larger in English (67% vs. 26%) than in German (58% vs. 38%)? A potential explanation of this differential is the following. Both in English and in German, while holding an original bias for p and being faced with evidence against p, in most trials the original belief p is double-checked (outer reading), but in some of the trials the opposite proposition ($\neg p$) is double-checked (inner reading). The percentage of double-checking $\neg p$ readings gives us the differential: the reading is expressed using a HiNQ in English and a LowNQ in German.

In other words, while the German subjects selecting a HiNQ in this condition uniformly produced the outer reading, some of the English subjects choosing a HiNQ produced an outer reading and some an inner reading. This leads us to expect a single intonational pattern for German HiNQs versus two intonational strategies for the English HiNQs in this condition. This prediction can be tested directly. We therefore analysed the HiNQ in the two subsets of the two languages prosodically, paying particular attention to the nuclear tune, see Table 19 and English example realizations in Figure 7. As hypothesized, there is no variation in the German data (in which there is no ambiguity), but there are two distinct nuclear tunes in the English data; about half of the items were produced with a rising and half with a falling nuclear tune.[15]
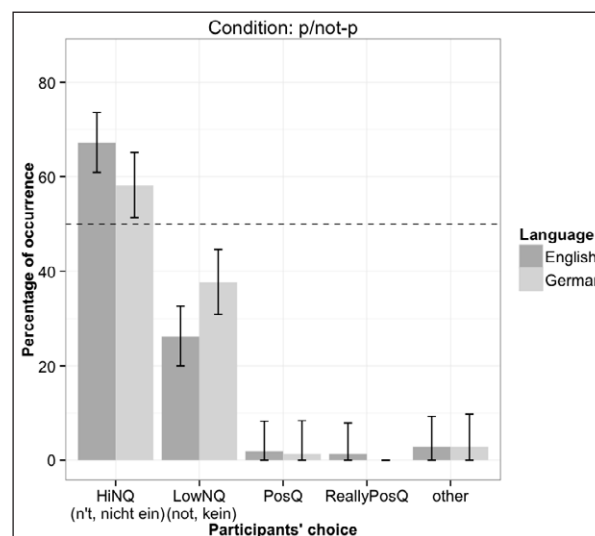

Figure 6: Comparison of choices in the ambiguity cell ($p/\neg p$) across languages.

---

[15] An alternative explanation of this phonological finding would be that the German participants are just reading off the sentences with one intonation pattern while the English participants are more lively and play-act the questions better. We do not think that such an objection holds, however. Indeed, in German LowNQ (in which there are two potential readings as shown in Table 11 in the Introduction), we do find prosodic differences: In half of the cases, participants produced the negative element *kein/keine/keinen* 'no' with a prominent rising accent (L*+H); in half of the cases, the negative element was unaccented or produced with a very weak accent.

	Language	
	English	**German**
nuclear rise	14	30
nuclear fall	16	0

Table 19: Proportion of rising and falling nuclear tunes in the HiNQs in the two languages.
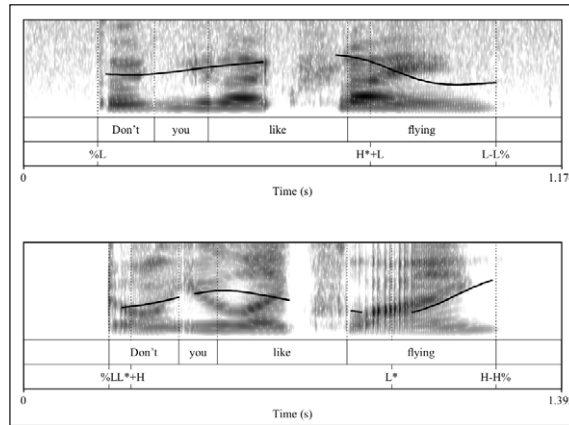

Figure 7: Two sample realizations of the English HiNQ, with a final pitch fall (upper panel) and a final pitch rise (lower panel).

Other than this explanation of the selection differential between English and German, we see two alternative, but less preferred, interpretations for the difference across languages. First, one may try to explain it away by saying that Germans, as opposed to English speakers, have a preference for the low negation form *kein* 'no' over the high negation form *nicht ein* 'n't... a' for some reason unrelated to bias or to our experimental goals, e.g., because *kein* occurs much more often than *nicht ein* in declaratives in their language. But then, the Germans' dispreference towards *nicht ein* relative to the English' acceptance of *n't* should be apparent throughout all conditions. This is contrary to fact, however. In the condition *p*/neutral, for instance, HiNQs are produced in 69% of the cases in German and in only 65% of the cases in English, see Tables 13 and 16. Furthermore, this line of argumentation would leave the prosodic findings unexplained.

Second, one might argue that Germans are potentially more circumspect than English speakers and therefore have a tendency to leave their original epistemic biases unexpressed. Hence, they default more often to LowNQs with *kein* than English to LowNQs with *not*. But, then, Germans' inhibition to express their original epistemic bias relative to the English' expressivism should be found throughout all conditions. This is, again, contrary to fact. In condition ¬*p*/*p*, for example, German speakers used the overtly biased form *really*-PosQ more often than the English speakers did (65% in German vs. 60% in English).

Importantly, if our first explanation is on the right track, then grouping LowNQ and Inner-HiNQ as one single type is supported for German (as generally assumed) but not for English (contra Asher & Reese 2007; Krifka to appear). Further experiments are needed to test this possibility in more depth, comparing LowNQs and NPI-hosting HiNQs, which have only the intended inner reading (see Section 5). Note, furthermore, that the presence of different prosodic realizations in the *p*/¬*p* cell in English does not allow us to determine whether a specific realization signals the inner or outer reading. Further experiments are needed in order to investigate (i) which is the preferred reading in the condition *p*/¬*p* and (ii) what specific prosodic realizations accompany Inner- and Outer-HiNQs.

5 General discussion

Two main results emerge clearly from the discussion presented above.

First, the surface form of a polar question is sensitive both to original speaker bias and to contextual evidence bias. This finding contradicts van Rooij & Šafářová's (2003) proposal, which tries to derive the data from Büring & Gunlogson (2000) as well as the data from Romero & Han (2002; 2004) from one single pragmatic factor (namely, the utility value of p vs. $\neg p$ with respect to the speaker's epistemic state). But, more importantly, this result goes contra what is by and large modelled in the literature. While not denying that other kinds of bias may be at play, most authors so far have formalized the use-conditions of PQ types based solely on one kind of bias or the other (Romero & Han 2004; Asher & Reese 2005; Reese 2006; Asher & Reese 2007; Reese 2007; Krifka to appear). Our experimental results prove these analyses partial, showing that a comprehensive pragmatic modelling needs, in the end, to take both kinds of bias into account. How exactly the two types of bias should be factored into the theoretical analysis is left open as far as our experimental results go. One possibility would be to interpret the preferred choice in each condition as the only fully acceptable form; in other words, the inherent use-conditions of the non-preferred forms runs against one or both biases in the input context. A second possibility, already hinted at in Section 1.3, would be to consider (at least some of) the non-preferred choices as still acceptable but not optimal because they fail to maximize the expression of bias; since another PQ form exists that expresses a higher number of the biases present in the context, the forms would compete and the latter would be preferred over the former. Yet a third possibility would be that (at least some of) the non-preferred choices are acceptable but not optimal due to other factors. For example, a given form may be additionally associated with an emotional attitude (e.g. impatience) that makes it more plausible in a given context than another form. Crucially, the relevant factor would have to interact with biases somehow, since otherwise the preference for that form should be across the board.

Second, results reveal that each possible combination (under examination) of speaker original bias and contextual evidence bias prompts the use of specific polar question types. As shown, the preferred primary choices are the same in both languages, as summarized in Table 20.

While in almost all the conditions both in English and in German we have identified only one primary choice, in condition neutral/p it emerged that in both languages there is one favoured choice (PosQ) with a strong competitor (*really*-PosQ), arguably due to an additional reading of prefixed *really*.

As for PQ types, we extract the following conclusions from Table 20. PosQs and *really*-PosQs differ from each other and from the negative question forms in terms of PQ types, as expected from the literature. More interestingly, with respect to the four conflicting partitions (i)–(iv) of negative questions in the literature (see end of Section 1.1), we have seen that: (a) since LowNQs and Outer-HiNQs are chosen in different pragmatic conditions in English and German, they constitute two separate PQ types (contra van Rooij & Šafářová 2003); and (b) based on preliminary evidence arising from comparison between

	ORIGINAL BIAS			
CONTEXTUAL EVIDENCE		**p**	**Neutral**	**¬p**
	p		PosQ/ *Really*-PosQ	*Really*-PosQ
	Neutral	HiNQ (outer)	PosQ	
	¬p	HiNQ (outer/inner)	LowNQ	

Table 20: Overview of the primary choices in English and German.

English and German, LowNQs and Inner-HiNQs seem to constitute different PQ types in English (contra Asher & Reese 2007; Krifka to appear).

How do the results of our selection studies compare to those obtained in Roelofsen at al.'s (2012) acceptability studies? What we would expect is that the forms that are chosen most frequently in our experiment would be judged as very natural in an acceptability study. Recall that in Roelofsen et al. (2012), the linguistic encoding of positive contextual evidence made the evidence too strong so that all questions were considered as "redundant" (all conditions with a positive contextual evidence were indeed judged as very bad). For this reason, we will leave the cells neutral/*p* and ¬*p*/*p* out of the comparison. Furthermore, the neutral/¬*p* cell was problematic because the mention of the *p*-proposition came out of the blue ("Did she get a cat?" following negative contextual evidence arising from the utterance "She got a dog"). This leaves three cells for comparison, as shown in Table 21: *p*/neutral, *p*/¬*p* and neutral/neutral. In these cells, the preferred choices in our studies are rated as very acceptable in Roelofsen et al.'s (2012) study.

The present experimental work makes a first but necessary step towards a more accurate empirical characterization of the role of biases in the polar question forms considered. In the future, we plan to run acceptability studies with our materials to be able to compare the results from the two types of tasks more directly.

Two immediate open issues that we expect to tackle in future research are the following. First, the typology of negative questions needs to be further investigated. In a first future experiment, we will address the distinction between Inner-HiNQs and LowNQs in English more directly. We will test, using a comprehension study, whether English LowNQs differ from Inner-HiNQs with NPIs in terms of mandatory original speaker bias. In a second future experiment, we plan to examine Ladd's (1981) intuitive ambiguity more closely by comparing the behavior of HiNQs hosting PPIs, HiNQs hosting NPIs and ambiguous HiNQs without PPIs/NPIs.

Second, from a prosodic perspective, it is not clear which prosodic forms relate to which kinds of bias and whether the prosodic realization alone can disambiguate different PQ types in string-identical forms. In future studies, we plan to use recordings of polar questions to a) analyze the prosodic realizations in more detail and b) to train a classifier that allows us to predict the underlying bias given certain morphosyntactic forms and prosodic realizations.

Once these and other empirical issues are settled, the question arises, how the bias-related use-conditions of the studied question forms should be modelled theoretically. In the last years, the linguistic literature has seen important refinements of Stalnaker's (2002) Common Ground (CG) and its dynamics, distinguishing between the discourse commitments of each conversational participant (Gunlogson 2003), the "Table" (Farkas & Bruce 2010) and strong *vs.* tentative commitments (Malamud & Stephenson 2015). With this richer system in place, the issue is how to model the complex meaning of each

	ORIGINAL BIAS			
CONTEXTUAL EVIDENCE		**p**	**Neutral**	**¬p**
	p		NA	NA
	Neutral	HiNQ (outer) rating in Roelofsen et al: 1.76	PosQ rating in Roelofsen et al: 1.41	
	¬p	HiNQ (outer/inner) rating in Roelofsen et al: 1.87	NA	

Table 21: Comparison of our results with the acceptability ratings (1 – very natural, 7 – very unnatural) in Roelofsen et al. (2012). The problematic cells are marked as NA.

question type so that the right bits of the CG are updated. One possibility is to maintain very lean semantics for these question forms and derive their biases from their pragmatics, using e.g., Gricean implicatures (Romero & Han 2004; Lauer 2014) or a decision-theoretical approach (van Rooij & Šafárová 2003). Another possibility is to enrich their semantic representation. This, in turn, can be done by making the propositional content more complex (using e.g., Inquisitive Semantics as in Farkas & Roelofsen 2014; to appear) or by adding a layer of semantic content to it: e.g., a speech act layer (Krifka to appear) or a context-management layer (Repp 2013; Romero 2015).

Additional File

The additional file for this article can be found as follows:

- **Appendix.** DOI: https://doi.org/10.5334/gjgl.27.s1

Acknowledgements

Competing Interests

References

AnderBois, Scott. 2011. *Issues and alternatives*. San Diego, CA: UC San Diego dissertation.

Asher, Nicholas & Brian Reese. 2005. Negative bias in polar questions. In Emar Maier, Corien Bary & Jianneke Huitink (eds.), *Proceedings of SuB9*, 30–43. www.ru.nl/ncs/sub9.

Asher, Nicholas & Brian Reese. 2007. Intonation and discourse: Biased questions. *Interdisciplinary Studies on Information Structure* 8. 1–38.

Baayen, R. Harald 2008. *Analyzing linguistic data. A practical introduction to statistics*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511801686

Baron, Jonathan. 1994. *Thinking and deciding*. 2nd edition; Cambridge: Cambridge University Press.

Borràs-Comes, Joan & Pilar Prieto. 2015. Intonation and the pragmatics of yes-no questions in central Catalan. In *Proceedings of ICPhS 2015*. Glasgow.

Büring, Daniel & Christine Gunlogson. 2000. Aren't positive and negative polar questions the same?. Ms., University of California. http://homepage.univie.ac.at/daniel.buring/phpsite/content/allpapers.html#gunlogsonburing

Caponigro, Ivano & Jon Sprouse. 2007. Rhetorical questions as questions. In Estela Puig-Waldmüller (ed.), *Proceedings of Sinn und Bedeutung 11*, 121–133, Barcelona: Universitat Pompeu Fabra.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2013. Inquisitive semantics: a new notion of meaning. *Language and Linguistics Compass* 7(9). 459–476. DOI: https://doi.org/10.1111/lnc3.12037

Cunnings, Ian. 2012. An overview of mixed-effects statistical models for second language researchers. *Second Language Research* 28(3). 369–382. DOI: https://doi.org/10.1177/0267658312443651

Escandell Vidal, Victoria. 1998. Intonation and procedural encoding: the case of Spanish interrogatives. In Villy Rouchota & Andreas H. Jucker (eds.), *Current issues in Relevance Theory,* 169–203. Amsterdam, John Benjamins.

Farkas, Donka & Floris Roelofsen. 2014. Assertions, polar questions, and the land in between: division of labor between semantics and discourse pragmatics. Talk at the *Non-Canonical Questions Workshop* in Konstanz, 17–19 February 2014.

Farkas, Donka & Floris Roelofsen. To appear. Division of labor in the interpretation of declaratives and interrogatives.

Farkas, Donka & Kim Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27. 81–118. DOI: https://doi.org/10.1093/jos/ffp010

Fodor, Janet Dean. 2002. Prosodic disambiguation in silent reading. In Masako Hirotani (ed.), *Proceedings of the North East Linguistic Society* 32(1), 113–132. Amherst, MA: GSLA, University of Massachusetts.

Groenendijk, Jeroen & Martin Stockhof. 1984. *Studies on the semantics of questions and the pragmatics of answers.* Amsterdam: University of Amsterdam Dissertation.

Gunlogson, Christine. 2003. *True to form: rising and falling declaratives as questions in English.* New York: Routledge.

Hartung, Simone. 2006. *Forms of negation in polar questions.* Tübingen: University of Tübingen MA thesis.

Haselton, Martie G., Daniel Nettle & Paul W. Andrews. 2005. The evolution of cognitive bias. In David M. Buss (ed.), *Handbook of evolutionary psychology*, 724–746. Hoboken: Wiley.

Höhle, Tilmann. 1992. Über Verum Fockus im Deutschen. *Linguistische Berichte Sonderhefte* 4. 112–141.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press.

Jaeger, T. Florian. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. DOI: https://doi.org/10.1016/j.jml.2007.11.007

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1). 3–44. DOI: https://doi.org/10.1007/BF00351935

Krifka, Manfred. 2013. Response particles as propositional anaphors. In *Proceedings of Semantics and Linguistic Theory 23*, 1–18. DOI: https://doi.org/10.3765/salt.v23i0.2676

Krifka, Manfred. to appear. Negated polarity questions as denegations of assertions. In Ferenc Kiefer & Chungmin Lee (eds.), *Contrastiveness and scalar implicatures*. Berlin: Springer. http://amor.cms.hu-berlin.de/~h2816i3x/Publications/NegatedPolarity-Questions.pdf.

Kügler, Frank. 2003. Do we know the answer? – Variation in yes-no-question intonation. In Susann Fischer, Ruben Vogel & Ralf van de Vijver (eds.), *Experimental studies in linguistics,* 9–29. Potsdam: Universitätsverlag.

Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In *Proceedings of Chicago Linguistic Society 17*, 164–171.

Lassiter, Daniel. 2011. *Measurement and modality: The scalar basis of modal semantics*. New York University dissertation.

Lauer, Sven. 2014. Mandatory implicatures in Gricean pragmatics. In Noah Goodman, Michael Franke & Judith Degen (eds.), *Proceedings of the formal and experimental pragmatics workshop*, 21–28.

Malamud, Sophia & Tamina Stephenson. 2015. Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics* 32(2). 275–311. DOI: https://doi.org/10.1093/jos/ffu002

Penka, Doris. 2011. *Negative indefinites*. Oxford: Oxford University Press.

Reese, Brian. 2006. The meaning and use of negative polar interrogatives. In Oliver Bonami & Patricia Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics* 6, 331–354.

Repp, Sophie. 2013. Common ground management: modal particles, illocutionary negation and VERUM. In Daniel Gutzmann & Hans-Martin Gaertner (eds.), *Beyond Expressives: Explorations in use-conditional meaning* (Current research in the semantics/pragmatics interface (CRiSPI) 28), 231–274. Leiden: Brill. DOI: https://doi.org/10.1163/9789004183988_008

Roelofsen, Floris & Donka Farkas. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2). 359–414. DOI: https://doi.org/10.1353/lan.2015.0017

Roelofsen, Floris, Noortje Venhuizen & Galit Weidman Sassoon. 2012. Positive and negative questions in discourse. In Emmanuel Chemla, Vincent Homer and Grégoire Winterstein (eds.), *Proceedings of Sinn und Bedeutung 17*, 455–472.

Romero, Maribel. 2015. High negation in subjunctive conditionals and polar questions. In Eva Csipak & Hedde Zeijlstra (eds.), In *Proceedings of Sinn und Bedeutung 19*, 519–536. University of Göttingen.

Romero, Maribel & Chung-hye Han. 2002. Verum focus in negative yes/no questions and Ladd's *p/¬p* ambiguity. In Brendan Jackson (ed.), *Proceedings of Semantics and Linguistic Theory XII*, 204–224. Ithaca, NY: CLC Publications.

Romero, Maribel & Chung-hye Han. 2004. On negative yes/no questions. *Linguistics and Philosophy* 27(5). 609–658. DOI: https://doi.org/10.1023/B:LING.0000033850.15705.94

Rooth, Mats. 1992. A theory of focus interpreation. *Natural Language Semantics* 1. 75–116. DOI: https://doi.org/10.1007/BF02342617

Savino, Michelina. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42(1). 23–48. DOI: https://doi.org/10.1017/S002510031100048X

Savino, Michelina & Martine Grice. 2011. The perception of negative bias in Bari Italian questions. In Sónia Frota, Gorka Elordieta and Pilar Prieto (eds.), *Prosodic categories: Production perception and comprehension* (Studies in Natural Language and Linguistic Theory 82), 187–206. DOI: https://doi.org/10.1007/978-94-007-0137-3_8

Schwarz, Bernhardt & Rajesh Bhatt. 2006. Light negation and polarity. In Raffaella Zanuttini et al. (eds.), *Negation, tense, and clausal architecture*, 175–198. Washington: Georgetown University Press.

Schwarzschild, Roger. 1999. GIVENness, AvoidF and other constraints on placement of focus. *Natural Language Semantics* 7(2). 141–177. DOI: https://doi.org/10.1023/A:1008370902407

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25. 701–21. DOI: https://doi.org/10.1023/A:1020867916902

Sudo, Yasutada. 2013. Biased polar questions in English and Japanese. In Daniel Gutzmann & Hand-Martin Gaertner (eds.), *Beyond expressives: Explorations in use-conditional mean-*

ing (Current Research in the Semantics/Pragmatics Interface (CRiSPI) 28), 275–296. Leiden: Brill. DOI: https://doi.org/10.1163/9789004183988_009

Trinh, Tue. 2014. How to ask the obvious. A presuppositional account of evidentail bias in Eng; sich yes/no questions. In Luka Crnic & Uli Sauerland (eds.), *The art and craft of semantics: A festschrift for Irene Heim* 2, 227–249.

Turco, Giuseppina, Christine Dimroth & Bettina Braun. 2013. Intonational means to mark verum focus in German and French. *Language and Speech* 56. 461–491. DOI: https://doi.org/10.1177/0023830912460506

Vanrell, Maria del Mar, Francesc Ballone, Carlo Schirru & Pilar Prieto. 2014. Intonation and its interfaces in Sardinian polar questions. *Loquens* 1(2). e014. DOI: https://doi.org/10.3989/loquens.2014.014

Vanrell, Maria del Mar, Ignasi Mascaró, Francesc Torres-Tamarit & Pilar Prieto. 2013. Intonation as an encoder of speaker's certaity: Information and confirmation yes-no questions in Catalan. *Language and speech* 56(2). 163–190. DOI: https://doi.org/10.1177/0023830912443942

Vanrell, Maria del Mar, Meghan Armstrong & Pilar Prieto. 2014. The role of prosody in the encoding of evidentiality in Catalan. In *Proceedings of the 7th international conference on speech prosody* 7, 1022–1026.

van Rooij, Robert & Marie Šafárová. 2003. On polar questions. In Robert B. Young & Yuping Zhou (eds.), *Proceedings of Semantics and Linguistic Theory 13*, 292–309. Ithaca, NY: CLC Publications. DOI: https://doi.org/10.3765/salt.v13i0.2887

Walkow, Msrtin. 2009. When to ask an inner negation polar question. In Maria Biezma & Jesse Harris (eds.), *UMass occasional papers in linguistics 39*, 125–146. Amherst, MA: University of Massachusetts.

Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of the fall-rise intonation. *Language* 61. 747–776. DOI: https://doi.org/10.2307/414489

Yalcin, Seth. 2012. Context probabilism. In Maria Aloni, Vladim Kimmelman, Floris Roelofsen, Galit W. Sasson, Katrin Schulz & Matthijs Westera (eds.), *Logic, language and meaning. Revised selected papers from the 18th Amsterdam Colloquium*, 12–21. Dordrecht: Springer. DOI: https://doi.org/10.1007/978-3-642-31482-7_2