Positive and negative polar questions in discourse

Roelofsen, F.; Venhuizen, N.; Sassoon, G.W.

Publication date
2013
Document Version
Final published version
Published in
Sinn und Bedeutung 17

Positive and negative questions in discourse

Floris Roelofsen — *University of Amsterdam*
Noortje Venhuizen — *University of Groningen*
Galit Weidman Sassoon — *Hebrew University Jerusalem*

Abstract. This paper presents a number of experiments assessing the felicity of positive and negative polar questions in various types of discourse contexts.

Keywords: polar questions, negation, felicity conditions.

1. Introduction

This paper is concerned with positive and negative polar questions, illustrated in (1)–(3).

(1) Did Lucy go to Greece? [positive polar question, PPQ]

(2) Did Lucy not go to Greece? [low negation polar question, LNPQ]

(3) Didn't Lucy go to Greece? [high negation polar question, HNPQ]

A high negation polar question (HNPQ) is a polar question that contains a negative clitic, attached to the inverted auxiliary (*didn't*). A low negation polar question (LNPQ) is a polar question that contains sentential negation, not attached as a clitic to the inverted auxiliary. Finally, a positive polar questions (PPQ) is a polar question that does not contain negation. We refer to the clause 'Lucy went to Greece' as the *prejacent* of the questions in (1)–(3). When considering a certain question, we often simply use p to refer to its prejacent, and $\neg p$ to refer to its negation.

It has been claimed in much previous work that positive and negative polar questions have different felicity conditions (Ladd, 1981; Büring and Gunlogson, 2000; van Rooij and Šafářová, 2003; Romero and Han, 2004; Reese, 2005, 2007; Reese and Asher, 2007; AnderBois, 2011, among others). However, the precise felicity conditions for each question type have remained controversial. This paper presents a series of experiments testing the felicity of positive and negative polar questions in various discourse contexts.

The paper is organized as follows: section 2 presents an experiment concerning PPQs, section 3 presents an experiment concerning HNPQs, and section 4 presents an experiment concerning LNPQs; finally, section 5 provides some general discussion and sketches a theoretical account of the main results of the experiments. Statistical details are provided in the appendices.

2. Experiment 1: positive polar questions

2.1. Hypotheses

Büring and Gunlogson (2000) claim that PPQs are subject to the following felicity condition:

(4) **Felicity condition for PPQs in terms of contextual evidence**
A PPQ is only felicitous in a context if there is no compelling contextual evidence against p in that context.

In formulating this generalization, Büring and Gunlogson assume the following notion of *compelling contextual evidence for or against* p:

(5) **Compelling contextual evidence for or against** p
   a. Contextual evidence is evidence that has just become mutually available to the participants in the current discourse context.
   b. Compelling contextual evidence *for* p is contextual evidence that, if considered in isolation, would allow the participants to assume that p is the case.
   c. Compelling contextual evidence *against* p is contextual evidence that, if considered in isolation, would allow the participants to assume that p is not the case.

The suggested felicity condition for PPQs is motivated by the following examples.

(6) Scenario: A and B are talking long-distance on the phone
$\Rightarrow$ no contextual evidence concerning the weather at B's location
   a. B: What's the weather like out there? Is it raining?
   b. B: What's the weather like out there? Is it sunny?

(7) Scenario: A enters B's windowless office wearing a dripping wet raincoat
$\Rightarrow$ compelling contextual evidence for rain and against sunshine
   a. B: What's the weather like out there? Is it raining?
   b. #B: What's the weather like out there? Is it sunny?

We conducted an experiment testing the hypothesis that PPQs are more felicitous in contexts with positive or neutral contextual evidence (CE) than in negative CE contexts.

Another factor that seems to affect the felicity of PPQs, especially in the absence of contextual evidence, are the *beliefs* that the speaker has concerning the prejacent of the question. This is illustrated in (8).

(8) Scenario: A and B are talking long-distance on the phone; A is traveling in Egypt, and B knows that it hardly ever rains in Egypt.
⇒ no contextual evidence as to whether it rains, but negative speaker belief
a. B: What's the weather like out there? Is it sunny?
b. ??B: What's the weather like out there? Is it raining?

To the best of our knowledge there are no explicit hypotheses in the literature as to how speaker belief (SB) affects the felicity of PPQs. We conducted an experiment testing the hypothesis that PPQs are more felicitous in positive and neutral SB contexts than in negative SB contexts.

2.2. Method

Participants. Participants were recruited using Amazon Mechanical Turk (AMT), an online labor market place where people are paid to complete small online tasks. AMT offers an efficient way to acquire high-quality experimental results that do not differ significantly in performance from standard experimental settings (Buhrmester et al., 2011). The tasks in our experiment were only visible to AMT participants with an American IP address. The reward per question was 0.03 cents. All in all 110 participants answered an average of 75 questions per participant (SD=102), with average hourly rate of approximately 5.7 American dollars. Every question was answered by 25 different participants.

Stimuli: target items. The stimuli were 3-picture cartoons, exemplified in Figure 1. Participants were asked to judge the naturalness of the PPQ in the third picture. In the example below, the prejacent of the PPQ is 'Kate got a cat' and the speaker is Jennifer. The first picture was used to manipulate *speaker belief*:

- Kate to Jennifer: "I'm going to get *a cat*." ⇒ positive SB
- Kate to Jennifer: "I'm going to get *a pet*." ⇒ neutral SB
- Kate to Jennifer: "I'm going to get *a dog*." ⇒ negative SB

The second picture was used to manipulate the *contextual evidence*:

- Rose to Jennifer: "Kate got *a cat*." ⇒ positive CE
- Rose to Jennifer: "Kate got *a pet*." ⇒ neutral CE
- Rose to Jennifer: "Kate got *a dog*." ⇒ negative CE

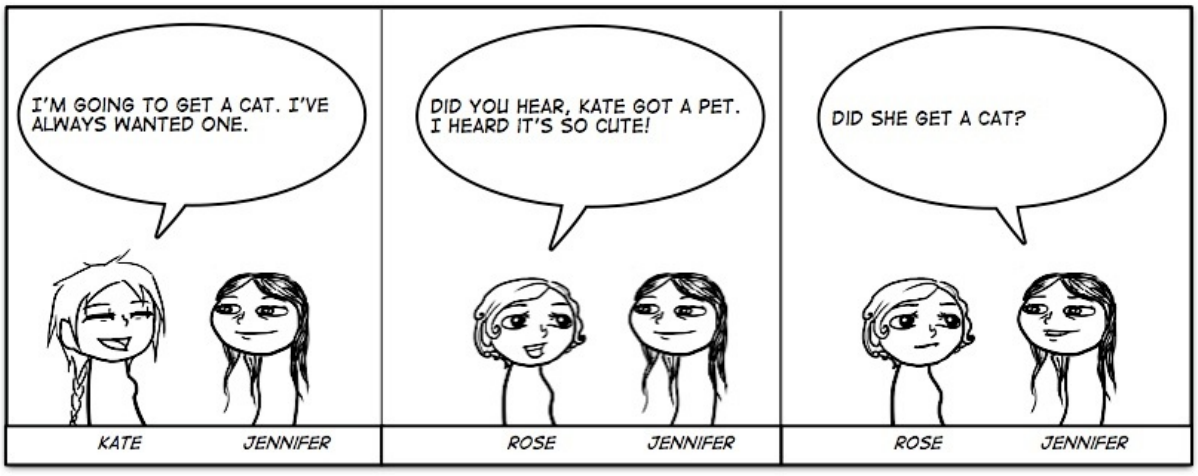

Figure 1: Example target item for testing PPQ felicity.

Thus, for each PPQ that we tested, there were $3 \times 3 = 9$ different cartoons, corresponding to the 9 types of SB x CE discourse contexts. We tested 6 PPQs, so in total there were $6 \times 9 = 54$ target items. The first and the second picture always contained an extra sentence (e.g., *I've always wanted one* or *I heard it's so cute!*) in order to reduce the effect that might be caused by the repetition of specific nouns on the felicity of the PPQ.

Stimuli: fillers. As fillers we used: (a) The target items for experiment 2-3, with LNPQs and HNPQs in the last picture, and (b) 164 additional fillers with *wh*-questions in the last picture, e.g. *Which pet did she get?*. The additional fillers were based on 41 different questions. They were balanced w.r.t. the polarity of the question involved (positive vs negative) and the naturalness of the question (according to our own judgement). There were 326 stimuli in total (54 PPQ targets, 108 HNPQ/LNPQ targets, and 164 *wh*-fillers).

Procedure. The stimuli were presented in a random order except that a *wh*-filler separated each two target items. After every test item, participants had the option to continue with the next test item or to stop. Thus, different participants filled in different sets of test items, with randomly different orders. Participants were asked to rate the naturalness of the PPQ in the third picture on a scale from 1 (completely natural) to 7 (completely unnatural). Each test item was presented together with instructions and a 1 to 7 scale to rate the naturalness of the question in the cartoon.

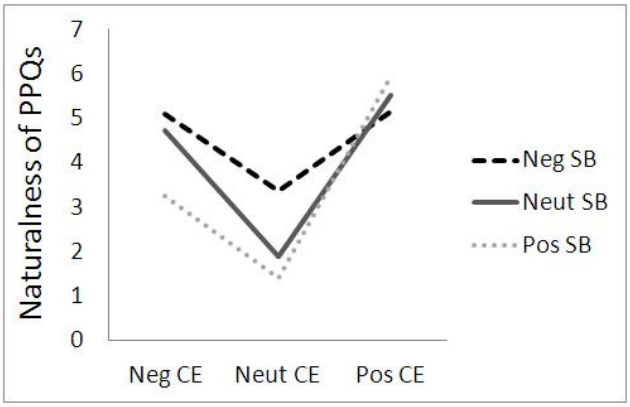

Figure 2: Averaged naturalness ratings for experiment 1.

2.3. Results

The results of experiment 1 are presented numerically in Table 1 and are visualized in Figure 2. The ratings for the *wh*-fillers are listed in Table 2. The results show that PPQs are most natural in neutral CE contexts with neutral or positive SB. They are less natural in negative CE contexts, and even worse in positive CE contexts. Positive SB increases the naturalness of PPQs in neutral and negative CE contexts, while it decreases it in positive CE contexts. Negative SB does exactly the opposite: it decreases the naturalness of PPQs in neutral and negative CE contexts, while it increases it in positive CE contexts. Finally, if CE and SB are not neutral, felicity of the PPQ increases when CE and SB are contrasting. There is a significant interaction between SB and CE. A detailed statistical analysis of the results can be found in Appendix A.

PPQs	Neg CE	Pos CE	Neut CE	Total
Neg SB	5.09 (0.76)	5.16 (0.2)	3.37 (0.29)	4.54 (.97)
Pos SB	3.24 (1.08)	5.95 (0.29)	1.41 (0.2)	3.54 (2.02)
Neut SB	4.72 (1.1)	5.51 (0.36)	1.87 (0.46)	4.04 (1.74)
Total	4.35 (1.24)	5.54 (.43)	2.22 (.91)	4.04 (1.66)

Table 1: Averaged naturalness ratings plus standard deviations for experiment 1.

wh-fillers	Good	Bad
Positive	1.91 (1.59)	6.49 (1.24)
Negative	3.06 (2.19)	6.75 (0.91)
Total	2.48 (2.00)	6.62 (1.09)

Table 2: Averaged naturalness ratings plus standard deviations for the *wh*-fillers.

2.4. Discussion

Contextual evidence. The results are partly in line with the hypothesis of Büring and Gunlogson (2000). In particular, if we restrict our attention to positive and neutral SB contexts we find (a) that PPQs are completely natural in neutral CE contexts, and (b) that PPQs are not completely natural in negative CE contexts, as expected. However, we also find (c) that PPQs are completely **un**natural in positive CE contexts, and (d) that PPQs are not completely unnatural in negative CE contexts with positive SB, which was not expected.

The first unexpected result, (c), may be due to the fact that in our experimental setup, contextual evidence was manipulated linguistically; as a result, PPQs may have seemed *redundant* in positive CE contexts, and may have been judged unnatural because of this. This is illustrated in (9):

(9) [picture 2] Rose to Jennifer: Kate got a cat.
[picture 3] Jennifer to Rose: Did she get a cat?

In order to verify this possible explanation, a follow-up experiment would have to be conducted with contexts where positive CE is given non-linguistically.

The second unexpected observation, (d), may be due to the fact that, strictly speaking, most of the negative CE contexts that we created were compatible with the prejacent of the PPQ in question. The incompatibility only arose through a *conversational implicature*. This is illustrated in (10).

(10) Rose: Kate got a dog. ⇒ implicates but does not entail that she didn't get a cat
Jennifer: Did she get a cat?

Given a positive speaker belief, in this case the belief that Kate got a cat, it would be reasonable for the speaker to check whether the 'no cat' implicature is indeed intended. This may explain why PPQs were not judged completely unnatural in CE contexts with positive SB. To verify this possible explanation, a follow-up experiment would have to be conducted where the incompatibility between the prejacent of the question and the contextual evidence in negative CE contexts is really entailed.

Speaker belief. We hypothesized that PPQs would be more felicitous in positive and neutral SB contexts than in negative SB contexts. The picture that we found is a bit more complicated, because SB interacts with CE. Positive SB improves the felicity of PPQs in neutral and negative CE contexts, but decreases felicity in positive CE contexts. By contrast, negative SB decreases the felicity of PPQs in neutral and negative CE contexts, but improves it in positive CE contexts. Hence, to predict the felicity of a PPQ in a given context we need to know both the polarity (negative,

positive, or neutral) of the available contextual evidence and of the polarity of the speaker's belief with respect to the prejacent of the question.

3. Experiment 2: high negation polar questions

3.1. Hypotheses

It is often assumed that HNPQs generally allow for two different 'readings,' which are called the *outer* negation question reading and the *inner* negation question reading, respectively (Ladd, 1981; Büring and Gunlogson, 2000; Romero and Han, 2004; AnderBois, 2011, a.o.). Whether this distinction is real and what it amounts to exactly has remained controversial (cf. van Rooij and Šafářová, 2003; Hartung, 2007). Büring and Gunlogson (2000) suggest the following generalizations, distinguishing between outer negation and inner negation readings:

(11) **Inner negation HNPQs**
Under the inner negation reading, HNPQs are only felicitous if there is compelling contextual evidence against p.

(12) **Outer negation HNPQs**
Under the outer negation reading, HNPQs are only felicitous if there is no compelling contextual evidence for p.

For the purposes of this paper, we wish to remain agnostic as to whether the distinction between inner and outer negation readings is real. Therefore, we derive from the above generalizations the following general hypothesis about the felicity of HNPQs (on either reading).

(13) **Felicity condition for HNPQs in terms of contextual evidence**
HNPQs (on either reading) are only felicitous if there is no compelling contextual evidence for p.

It is widely assumed that the felicity of HNPQs does not only depend on *contextual evidence*, but also on *speaker belief*. Based on Ladd (1981), Büring and Gunlogson (2000), Romero and Han (2004), and AnderBois (2011), we formulate the following hypothesis.

(14) **Felicity condition for HNPQs in terms of speaker belief**
HNPQs are only felicitous if the speaker believes that p is the case.

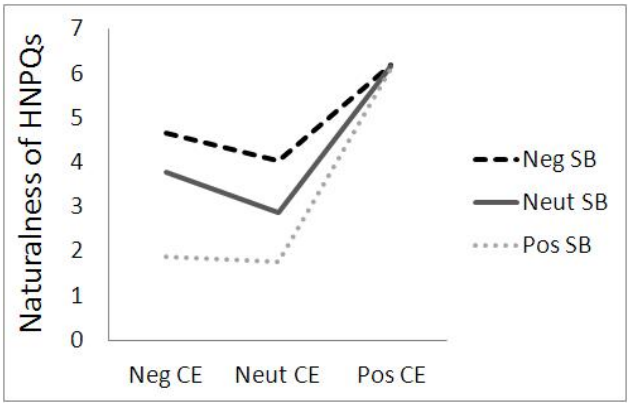

Figure 3: Averaged naturalness ratings for experiment 2.

We conducted an experiment testing the hypotheses in (13) and (14).

3.2. Method

Participants were recruited using AMT, as explained in section 2.2. The target items were cartoons with dialogues identical to those of experiment 1, but this time with an HNPQ in the third picture of the cartoon (e.g., *Didn't she get a cat?*). The fillers consisted of the target items of experiment 1 and 3, as well as 164 additional fillers with *wh*-questions, which are described in section 2.2 together with additional details about the procedure.

3.3. Results

The results of experiment 2 are presented numerically in Table 3 and are visualized in Figure 3. We see that HNPQs are completely natural in negative and neutral CE contexts with positive SB. They are less natural with neutral SB, and even less so with negative SB. Finally, they are completely unnatural in positive CE contexts, no matter what the SB level is. We found a significant interaction between SB and CE. A detailed statistical analysis of the results is provided in Appendix B.

HNPQs	Neg CE	Pos CE	Neut CE	Total
Neg SB	4.66 (0.28)	6.19 (0.16)	4.03 (0.58)	4.96 (1)
Pos SB	1.87 (0.29)	6.1 (0.19)	1.76 (0.29)	3.24 (2.09)
Neut SB	3.79 (0.39)	6.17 (0.17)	2.86 (0.16)	4.27 (1.54)
Total	3.44 (1.24)	6.15 (.17)	2.88 (1.02)	4.16 (1.7)

Table 3: Averaged naturalness ratings plus standard deviations for experiment 2.

3.4. Discussion

Contextual evidence. The results clearly support hypothesis (13): HNPQs are only natural if there is no compelling contextual evidence for the prejacent.

Speaker belief. The results also support hypothesis (14): HNPQs are only completely natural in positive SB contexts. Notice that in negative and neutral CE contexts we have the following scale:

$$\text{positive SB} >> \text{neutral SB} >> \text{negative SB}$$

HNPQs are only completely natural with positive SB, less natural with neutral SB, and even less natural with negative SB. This scale may be explained by the fact that, strictly speaking, neutral and negative SB contexts are compatible with a positive speaker belief. For instance, if Kate says that she is going to get a pet, then Jennifer may think that she is in fact going to get a cat. And even if Kate says that she is going to get a dog, Jennifer may still think that she is (also) going to get a cat. Note that a positive speaker belief is easier to accommodate in a neutral SB context than in a negative SB context. This may explain the marginal acceptability of HNPQs in neutral SB contexts and their even more marginal acceptability in negative SB contexts.

4. Experiment 3: low negation polar questions

4.1. Hypotheses

LNPQs have received less attention in the literature than PPQs and HNPQs. However, one concrete generalization has been formulated by AnderBois (2011, p.133):

> "LNPQs are inconsistent with scenarios where the speaker has a contextually clear neutral or positive stance."

This generalization is motivated by the following variants of Büring and Gunlogson's examples:

(15) Scenario: A and B are talking long-distance on the phone
⇒ no contextual evidence concerning the weather at B's location
a. #B: What's the weather like out there? Is it not raining?
b. #B: What's the weather like out there? Is it not sunny?

(16) Scenario: A enters B's windowless office wearing a dripping wet raincoat
⇒ compelling contextual evidence for rain and against sunshine
a. #B: What's the weather like out there? Is it not raining?

b. ?B: What's the weather like out there? Is it not sunny?

It is not entirely clear whether AnderBois' generalization is intended to make reference to the speaker's beliefs, or to the contextual evidence. We formulate two hypotheses loosely based on AnderBois' generalization and examples, one in terms of contextual evidence and one in terms of speaker beliefs.

(17) **Felicity condition for LNPQs in terms of contextual evidence**
LNPQs are only felicitous in contexts with compelling contextual evidence against p.

(18) **Felicity condition for LNPQs in terms of speaker belief**
LNPQs are only felicitous in contexts with a negative speaker belief w.r.t. p.

We carried out an experiment to test these hypotheses.

4.2. Method

Participants were recruited using AMT, as explained in section 2.2. The target items were cartoons with dialogues identical to those of experiment 1 and 2, but this time with a LNPQ in the third picture of the cartoon (e.g., *Did she not buy a cat?*). The fillers consisted of the target items of experiment 1 and 2, as well as 164 additional fillers with *wh*-questions, which are described in section 2.2 together with additional details about the procedure.

4.3. Results

The results of experiment 3 are presented numerically in Table 4 and are visualized in Figure 4. We see that LNPQs are most natural in negative CE contexts with positive SB, followed by neutral CE contexts with positive SB. LNPQs are completely unnatural in positive CE contexts, no matter the level of SB.

Negative and neutral CE contexts exhibit the SB scale familiar from experiment 2:

$$\text{positive SB} >> \text{neutral SB} >> \text{negative SB}$$

The rating of LNPQs in negative CE contexts does not differ significantly from their rating in neutral CE contexts. However, if we consider only contexts with positive SB, then the rating of LNPQs in negative CE contexts does differ significantly from their rating in neutral CE contexts.

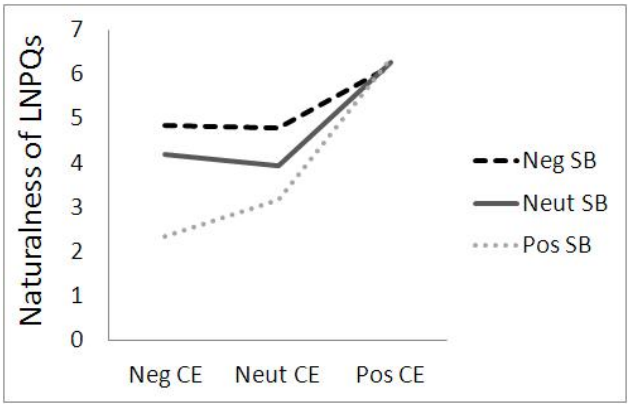

Figure 4: Averaged naturalness ratings for experiment 3.

LNPQs	Neg CE	Pos CE	Neut CE	Total
Neg SB	4.85 (0.63)	6.19 (0.16)	4.78 (0.2)	5.27 (.76)
Pos SB	2.33 (0.23)	6.36 (0.18)	3.18 (0.3)	3.96 (1.8)
Neut SB	4.19 (0.47)	6.26 (0.21)	3.93 (0.3)	4.79 (1.12)
Total	3.79 (1.18)	6.27 (.19)	3.96 (.72)	4.67 (1.39)

Table 4: Averaged naturalness ratings plus standard deviations for experiment 3.

We found a significant interaction between SB and CE. A detailed statistical analysis of the results is provided in Appendix C.

4.4. Discussion

Contextual evidence. The results do not support hypothesis (17): LNPQs are more natural in negative CE contexts than in positive CE contexts; however, they are not significantly more natural in negative CE contexts than in neutral CE contexts. Only if we restrict ourselves to contexts with positive SB, hypothesis (17) is supported.

The pattern that we found for LNPQs is strikingly similar to the pattern we found for HNPQs. There is a difference in neutral CE contexts, where HNPQs are more natural than LNPQs, but the overall pictures do not differ as much as expected.

Speaker belief. The results clearly reject hypothesis (18), which says that LNPQs are only natural in contexts with negative SB. In fact, we see that LNPQs are least natural in negative SB contexts. Again, the pattern that we find is similar to that of HNPQs:

$$\text{positive SB} >> \text{neutral SB} >> \text{negative SB}$$

where only positive SB results in complete acceptability. Under the assumption that LNPQs require

a positive speaker belief, the observed scale may be explained just like we did in the case of HNPQs: the neutral and negative SB contexts in the experiment suggest the absence of positive speaker belief, but strictly speaking, they do not exclude it.

As in the previous experiments, the effects of CE and SB exhibit a significant interaction. Thus, to predict the felicity of LNPQs in a given context we need to know both the polarity (negative, positive, or neutral) of the contextual evidence and the polarity of the speaker's beliefs with respect to the prejacent of the question.

5. General discussion

We saw that the felicity of all question types is affected by both CE and SB, with significant interactions between these factors. This suggest that future theorizing should consider both CE and SB, as well as the relation between the two, as prominent factors affecting the licensing of different types of polar questions.

The hypotheses that we tested were mostly supported by the experimental results. However, some of the results were not (entirely) expected in light of the given hypotheses: (a) the complete unacceptability of PPQs in positive CE contexts, (b) the marginal acceptability of PPQs in negative CE contexts with positive SB, (c) the 'SB acceptability scale' found for HNPQs and LNPQs, and (d) the marginal acceptability of LNPQs in neutral CE contexts with positive SB. We provided possible explanations for these results, which may be tested in further experimental work.

The experimental results obtained here also provide the basis for further theoretical work. A comprehensive account is beyond the scope of this paper, but we suggest that the following factors play a role in how speakers formulate polar questions, and therefore indirectly determine the felicity of the different types of polar questions in a given context.

1. DON'T ASK ANYTHING IF NOT NEEDED

   This explains why all types of PQs are infelicitous in positive CE contexts with positive or neutral SB, and in negative CE contexts with negative or neutral SB. Marginal acceptability in negative CE contexts with neutral SB can be explained by the fact that neutral SB contexts are strictly speaking compatible with a positive speaker belief, in which case the speaker does have a reason to ask a polar question.

2. AVOID REVERSING RESPONSES

   Farkas and Roelofsen (2012) make a general distinction between *agreeing* and *reversing* responses, and argue that reversing responses are more marked than agreeing responses. Furthermore, they hypothesize that, whenever possible, speakers formulate their questions in such a way that a marked reversing response can be avoided. This explains why PPQs are infelicitous in negative CE contexts, and why HNPQs and LNPQs are infelicitous in positive CE contexts, even with negative SB. It also explains the very marginal felicity of PPQs in

neutral CE contexts with negative SB.

3. IN CASE OF CONFLICT, PUT ALL ALTERNATIVES ON THE TABLE

   If there is a conflict between CE and SB, it is in the speaker's interest to make all relevant alternatives salient. This explains why PPQs are infelicitous in positive CE contexts with negative SB, as illustrated in (19).

   (19) Kate to Jennifer: I am going to get a cat.
   Rose to Jennifer: Did you hear? Kate got a dog.
   Jennifer to Rose: #Did she get a dog? / ✓Didn't she get a cat?

4. (a) IF POSSIBLE, USE THE LEAST MARKED FORM (PPQ) (AnderBois, 2011)

   (b) USE A MARKED FORM (HNPQ OR LNPQ) TO SIGNAL POSITIVE SB

   The felicity of HNPQs and LNPQs in the presence of positive SB is explained by (b); their infelicity in the absence of positive SB is explained by (a). We have no deep explanation of the connection between marked forms, involving negation, and positive SB. For now, this connection needs to be stipulated as an unexplained linguistic convention.

Of course, even though the experiments presented in this paper provide some new insight into the felicitiy of positive and negative polar questions in various discourse contexts, much more work is needed to really obtain a clear picture of this empirical domain. Besides the various follow-up experiments mentioned along the way, it would be especially worthwhile to investigate (a) the role of *intonation*, and (b) the role of positive and negative *polarity items*, especially in HNPQs and LNPQs (see Hartung, 2007, for initial work in this direction).

Acknowledgments

We are grateful to Maria Aloni, Adrian Brasoveanu, and Simone Gieselman for helpful discussion at various stages of this project, and to Tirza van Meurs for drawing the cartoons used in the experiment. We are also grateful to the SuB referees and participants for very useful feedback. The research reported here was made possible by financial support from the Netherlands Organization for Scientific Research (NWO), which is gratefully acknowledged.

A. Statistical analysis of the results of experiment 1

- **3-way comparison of the different CE levels**
  A nonparametric Friedman Test for the significance of the difference among the distributions of 3 correlated samples of averaged ratings in negative, positive and neutral CE contexts, each with 18 matched items (the values for 6 items each within 3 different SB contexts), yields that the samples are significantly different (csqr = 28.78, df = 2, P< .0001), with mean ranks Mneut = 1, Mneg = 2.3, Mpos = 2.7.

- **Pairwise comparison of the different CE levels**
  A Wilcoxon Signed-Rank Test for the difference between two samples of 18 matched pairs yields a significant difference between neutral and negative CE contexts (W = -171, $n_{s/r}$ = 18, z = -3.71, P<.0002), a significant difference between neutral and positive CE contexts (W = -171, $n_{s/r}$ = 18, z = -3.71, P<.0002), and a significant difference between negative and positive CE contexts (W = -125, $n_{s/r}$ = 18, z = -2.71, P<.0067).

- **Pairwise comparison of the different SB levels**
  A Wilcoxon Signed-Rank Test for the difference between two samples of 18 matched pairs yields a significant difference between neutral and negative SB contexts (W = -88, $n_{s/r}$ = 17, z = -2.07, P<.039), a significant difference between neutral and positive CE contexts (W = 93, $n_{s/r}$ = 18, z = 2.01, P<.045), and a significant difference between negative and positive CE contexts (W = 125, $n_{s/r}$ = 18, z = 2.71, P<.0067).

- **Main effects of CE and SB, and interaction between the two**
  A 2-factor Anova with 9 repeated measures yields a significant effect of CE (df=2, F=49.45, P<.0001), a significant effect of SB (df = 2, F=54.83, P<.0001), and a significant interaction between CE and SB (df=4, F=45.66, P<.0001).

- **Pairwise comparison of the different SB levels, given negative CE**
  Considering only negative CE contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between positive and neutral SB (W = 21, $n_{s/r}$ = 6, P<.05) and a significant difference between positive and negative SB (W = 21, $n_{s/r}$ = 6, P<.05). A significant difference between neutral and negative SB only in a directional test (W = -13, $n_{s/r}$ = 5, Pdir <.05).

- **Pairwise comparison of the different SB levels, given positive CE**
  Considering only positive CE contexts, we obtain a significant difference between positive and negative SB (W = 21, $n_{s/r}$ = 6, P<.05). A significant difference between positive and neutral SB is obtained only in a directional test (W = 17, $n_{s/r}$ = 6, Pdir <.05), and the same goes for the difference between neutral and negative SB (W = 18, $n_{s/r}$ = 6, Pdir <.05).

- **Pairwise comparison of the different SB levels, given neutral CE**
  Considering only neutral CE contexts, we obtain a significant difference between neutral and negative SB (W = -21, $n_{s/r}$ = 6, P<.05) and a significant difference between negative and positive SB (W = 21, $n_{s/r}$ = 6, P<.05). A significant difference between neutral and positive SB is obtained only in a directional test (W = 17, $n_{s/r}$ = 6, Pdir <.05).

- **Pairwise comparison of the different CE levels, given neutral SB**
  Considering only neutral SB contexts, we obtain a significant difference between neutral and negative CE (W = -21, $n_{s/r}$ = 6, P<.05) and a significant difference between positive and neutral CE (W = 21, $n_{s/r}$ = 6, P<.05). A significant difference between positive and negative CE is obtained only in a directional test (W = 13, $n_{s/r}$ = 6, Pdir <.05).

- **Pairwise comparison of the different CE levels, given negative SB**
  Considering only negative SB contexts, we obtain a significant difference between neutral and negative CE (W = -21, $n_{s/r}$ = 6, P<.05) and a significant difference between neutral and positive CE (W = -21, $n_{s/r}$ = 6, P <.05). We do not find a significant difference between negative and positive CE contexts (W = -3, $n_{s/r}$ = 6).

- **Pairwise comparison of the different CE levels, given positive SB**
  Considering only positive SB contexts, we obtain a significant difference between neutral and negative CE (W = -21, $n_{s/r}$ = 6, P<.05), a significant difference between neutral and positive CE (W = -21, $n_{s/r}$ = 6, P <.05), and a significant difference between negative and positive CE (W = -21, $n_{s/r}$ = 6, P <.05).

B. Statistical analysis of the results of experiment 2

- **Pairwise comparison of the different CE levels**
  A Wilcoxon Signed-Rank Test for the difference between two samples of 18 matched pairs yields a significant difference between neutral and negative CE (W = -129, $n_{s/r}$ = 17, z = -3.04, P<.0025), a significant difference between neutral and positive CE (W = -171, $n_{s/r}$ = 18, z = -3.71, P<.0002), and a significant difference between negative and positive CE (W = -171, $n_{s/r}$ = 18, z = -3.71, P<.0002).

- **Pairwise comparison of the different SB levels**
  A Wilcoxon Signed-Rank Test for the difference between two samples of 18 matched pairs yields a significant difference between neutral and negative SB (W = -153, $n_{s/r}$ = 18, z = -3.32, P=.0009), a significant difference between neutral and positive SB (W = 158, $n_{s/r}$ = 18, z = 3.43, P<.0006), and a significant difference between negative and positive SB (W = 163, $n_{s/r}$ = 18, z = 3.54, P<.0004).

- **Main effects of CE and SB, and interaction between the two**
  A 2-factor Anova with 9 repeated measures yields a significant effect of CE (df=2, F=400.6, P<.0001), a significant effect of SB (df = 2, F=136.3, P<.0001), and a significant interaction between CE and SB (df=4, F=64.8, P<.0001).

- **Pairwise comparison of the different SB levels, given negative CE**
  Considering only negative CE contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and negative SB (W = -21, $n_{s/r}$ = 6, P <.05), a significant difference between neutral and positive SB (W = 21, $n_{s/r}$ = 6, P <.05), and a significant difference between negative and positive SB contexts (W = 21, $n_{s/r}$ = 6, P<.05).

- **Pairwise comparison of the different SB levels, given neutral CE**
  Considering only neutral CE contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and

negative SB (W = -21, $n_{s/r}$ = 6, P<.05), a significant difference between negative and positive SB (W = 21, $n_{s/r}$ = 6, P<.05), and a significant difference between neutral and positive SB contexts (W = 21, $n_{s/r}$ = 6, P<.05).

- **Pairwise comparison of the different SB levels, given positive CE**
  Considering only positive CE contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields *no* significant difference between positive and negative SB (W = -13, $n_{s/r}$ = 6, P>.05), *no* significant difference between positive and neutral SB (W = -8, $n_{s/r}$ = 6), and *no* significant difference between neutral and negative SB (W = -3, $n_{s/r}$ = 6).

- **Pairwise comparison of the different CE levels, given neutral SB**
  Considering only neutral SB contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between positive and negative CE (W = 21, $n_{s/r}$ = 6, P <.05), a significant difference between positive and neutral CE (W = 21, $n_{s/r}$ = 6, P<.05), and a significant difference between neutral and negative CE contexts (W = -21, $n_{s/r}$ = 6, P <.05).

- **Pairwise comparison of the different CE levels, given negative SB**
  Considering only negative SB contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and positive CE (W = -21, $n_{s/r}$ = 6, P <.05), a significant difference between negative and positive CE (W = 21, $n_{s/r}$ = 6, P<.05), and a significant difference between neutral and negative CE (W = -21, $n_{s/r}$ = 6, P <.05).

- **Pairwise comparison of the different CE levels, given positive SB**
  Considering only positive SB contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and positive CE (W = -21, $n_{s/r}$ = 6, P <.05), a significant difference between negative and positive CE (W = -21, $n_{s/r}$ = 6, P <.05), but *no* significant difference between neutral and negative CE contexts (W = -4, $n_{s/r}$ = 5).

C. Statistical analysis of the results of experiment 3

- **Pairwise comparison of the different CE levels**
  A Wilcoxon Signed-Rank Test for the difference between two samples of 18 matched pairs yields a significant difference between neutral and positive CE contexts (W = -171, $n_{s/r}$ = 18, z = -3.71, P<.0002), a significant difference between negative and positive CE contexts (W = -171, $n_{s/r}$ = 18, z = -3.71, P<.0002), but *no* significant difference between neutral and negative CE contexts (W = 39, $n_{s/r}$ = 18, z = .84, P=.4).

- **Pairwise comparison of the different SB levels**
  A Wilcoxon Signed-Rank Test for the difference between two samples of 18 matched pairs

yields a significant difference between neutral and negative SB contexts (W = -113, $n_{s/r}$ = 17, z = -2.66, P=.0078), a significant difference between neutral and positive CE contexts (W = 143, $n_{s/r}$ = 18, z = 3.1, P<.0019), and a significant difference between negative and positive CE contexts (W = 136, $n_{s/r}$ = 18, z = 2.95, P<.0032).

- **Main effects of CE and SB, and interaction between the two**
  A 2-factor Anova with 9 repeated measures yields a significant effect of CE (df = 2, F = 404.14, P<.0001), a significant effect of SB (df = 2, F = 112.2, P<.0001), and a significant interaction between CE and SB (df = 4, F = 28.17, P<.0001).

- **Pairwise comparison of the different SB levels, given negative CE**
  Considering only negative CE contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and positive SB (W = 21, $n_{s/r}$ = 6, P <.05) and a significant difference between negative and positive SB (W = 21, $n_{s/r}$ = 6, P<.05). A significant difference between neutral and negative SB was obtained only in a directional test (W = -15, $n_{s/r}$ = 5, Pdir <.05).

- **Pairwise comparison of the different SB levels, given neutral CE**
  Considering only neutral CE contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and negative SB contexts (W = -21, $n_{s/r}$ = 6, P<.05), a significant difference between negative and positive SB contexts (W = 21, $n_{s/r}$ = 6, P<.05), and a significant difference between neutral and positive SB contexts (W = 21, $n_{s/r}$ = 6, P<.05).

- **Pairwise comparison of the different SB levels, given positive CE**
  Considering only positive CE contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields *no* significant difference between positive and negative SB (W = 14, $n_{s/r}$ = 6), *no* significant difference between positive and neutral SB (W = 7, $n_{s/r}$ = 6), and *no* significant difference between neutral and negative SB (W = 7, $n_{s/r}$ = 6).

- **Pairwise comparison of the different CE levels, given neutral SB**
  Considering only neutral SB contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between positive and negative CE (W = 21, $n_{s/r}$ = 6, P <.05) and a significant difference between positive and neutral CE (W = 21, $n_{s/r}$ = 6, P<.05). A significant difference between neutral and negative CE is obtained only in a directional test (W = -18, $n_{s/r}$ = 6, Pdir <.05).

- **Pairwise comparison of the different CE levels, given negative SB**
  Considering only negative SB contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and positive CE contexts (W = 21, $n_{s/r}$ = 6, P <.05), a significant difference between negative and positive CE contexts (W = 21, $n_{s/r}$ = 6, P<.05), but no significant difference between neutral and negative CE contexts (W = -1, $n_{s/r}$ = 6).

- **Pairwise comparison of the different CE levels, given positive SB**
  Considering only positive SB contexts, a Wilcoxon Signed-Rank Test for the difference between two samples of 6 matched pairs yields a significant difference between neutral and negative CE contexts (W = 21, $n_{s/r}$ = 6, P<.05), a significant difference between neutral and positive CE contexts (W = -21, $n_{s/r}$ = 6, P <.05), and a significant difference between negative and positive CE contexts (W = -21, $n_{s/r}$ = 6, P <.05).

References

AnderBois, S. (2011). *Issues and alternatives*. Ph. D. thesis, University of California Santa Cruz.

Buhrmester, M., T. Kwang, and S. Gosling (2011). Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science 6*(1), 3–5.

Büring, D. and C. Gunlogson (2000). Aren't positive and negative polar questions the same? Manuscript, UCSC/UCLA.

Farkas, D. and F. Roelofsen (2012). Polar initiatives and polarity particle responses in an inquisitive discourse model. Presented at the Amsterdam Colloquium, December 2011, and at the Newcastle Workshop on the Syntax of Answers to Polar Questions, June 2012, available via www.illc.uva.nl/inquisitive-semantics.

Hartung, S. (2007). Forms of negation in polar questions. In H. Zeijlstra and J.-P. Soehn (Eds.), *Proceedings of the Workshop on Negation and Polarity at the University of Tübingen*, pp. 58–63.

Ladd, R. (1981). A first look at the semantics and pragmatics of negative questions and tag questions. In *Proceedings of Chicago Linguistics Society, Volume 17*, pp. 164–171. CLS.

Reese, B. (2005). The meaning and use of negative polar interrogatives. In O. Bonami and P. C. Hofherr (Eds.), *Empirical Issues in Syntax and Semantics 6*, pp. 331–354.

Reese, B. (2007). *Bias in questions*. Ph. D. thesis, University of Texas at Austin.

Reese, B. and N. Asher (2007). Prosody and the interpretation of tag questions. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung 11*, pp. 448–462.

Romero, M. and C. Han (2004). On negative yes/no questions. *Linguistics and Philosophy 27*(5), 609–658.

van Rooij, R. and M. Šafářová (2003). On polar questions. In R. Young and Y. Zhou (Eds.), *Proceedings of Semantics and Linguistics Theory (SALT 13)*, pp. 292–309. CLC Publications.