The pragmatics of embedded questions. An experimental comparison of four verbs of embedding.

August 20, 2020

Abstract

We present an experimental study investigating the exhaustivity properties of questions embedded under German *wissen* 'know', *korrekt vorhersagen* 'predict', *sich einig sein* 'agree' and *überrascht sein* 'be surprised'. In contrast to previous empirical studies on such verbs in English and French, we did not test the general availability of the different exhaustive interpretations (strong, intermediate and weak exhaustivity) but targeted optimal interpretations from a communication-oriented perspective. To that end, we employed a novel method: We conducted an interactive lab experiment which provided a financial incentive through performance-based compensation. Participants did not judge the truth/falsity of a statement in vacuo, but in the context of a 'role play' in which they had to consider whether another person would accept their interpretation as well. The target sentence was presented as the object of a bet, the outcome of which participants had to evaluate. We found that intermediate exhaustive readings of questions embedded under *wissen* received considerably less acceptance than in previous studies by Cremers and Chemla (2016) and Cremers et al. (2017) and conclude for this matrix verb that the strongly exhaustive reading constitutes the pragmatically optimal interpretation. With regard to the three other verbs, we overall replicated for German the findings from previous experiments. In particular, same as Cremers et al. (2017) we found evidence for the communicative relevance of the strongly exhaustive interpretation of questions embedded under *überrascht sein*, the existence of which is often disputed in the theoretical literature. These insights add to the empirical basis that a semantic and pragmatic theory of embedded questions should build on.

Keywords: Experimental pragmatics, embedded questions, exhaustivity, communicative relevance, German

1   Introduction

The exhaustivity of embedded questions has been a major topic in theoretical linguistics in the past decades. Several truth-conditionally distinct readings are discussed in the literature. For example, for questions embedded under the verb *know* as in (1), they are the ones presented in (2).

(1)   Ali knows who danced at the party.

(2)   a.   **Strongly exhaustive reading (SE)**
              Ali knows for all people who danced a the party that they danced and he knows that they are the only ones who danced.
         b.   **Intermediate exhaustive reading (IE)**
              Ali knows for all people who danced at the party that they danced and he does not have false beliefs about any non-dancers.
         c.   **Weakly exhaustive reading (WE)**
              Ali knows for all people who danced at the party that they danced.

A number of experimental studies on English and French investigated the question which exhaustive readings exist under different verbs of embedding, such as *know, surprise, agree* etc. (Cremers and Chemla, 2016, 2017; Chemla and George, 2016; Cremers et al., 2017). The results of these studies were partially at odds with long-standing judgments in the theoretical literature. In particular, IE readings appeared to be more prevalent than previously assumed. This raises an interesting puzzle: why would trained linguists, who, for instance, often assume that wh-questions embedded under *to know* only have a SE reading, have different intuitions when compared to lay experimental subjects, who appear to be more liberal in their interpretations. One possibility is that the acceptance of a given reading in a certain experimental setting, on the one hand, and actually having a reading actively available in the sense that interlocutors rely on that reading as resolvable in communication, on the other, may not be the same thing. When confronted with the question of whether or not a given interpretation is available for a sentence in experimental settings in which nothing is at stake for them, lay participants may liberally accept interpretations that they themselves do not consider optimal and would therefore not actively employ in communication. This possibility is the starting point of the current paper. Instead of investigating the general availability of different exhaustivity readings this paper approaches the issue from a pragmatic angle by posing the question: Which readings are people actually ready to commit to?

To gain more insight into the pragmatics of embedded questions, we employed an experimental design that aimed to remove the assumed participants' tendency for allowing a wider range of interpretations than they would actually use themselves. The experiment confronted participants with the task of deciding whether a given interpretation is such that other language users would entertain it as well. This line of participant reasoning was induced by making target sentences the objects of bets, and by actively engaging participants through a performance-based financial incentive. Thereby, we investigated which readings are of communicative relevance, and hence particularly salient. In contrast, we will eschew from making any claims about the semantic status of the different exhaustive readings. The central goal of this study was to collect novel data that will contribute to a robust empirical basis for semantic and pragmatic theory building. Besides the application of a novel experimental method for the investigation of exhaustivity of embedded questions, this study is - to the best of our knowledge - also the first one to test

four different embedding predicates in a single comprehensive set-up. The four predicates are the German embedding verbs *wissen*, 'know', *korrekt vorhersagen*, 'correctly predict', *sich einig sein*, 'agree', and *überrascht sein*, 'be surprised'.

The paper is structured as follows. Section 2 gives a short introduction to the topic, it presents the four predicates that are investigated in the experiment, and it surveys previous research. Section 3 presents the goals and expectations of our experiment, and Section 4 describes the experimental study. Finally, Section 5 discusses the experimental results and concludes with an outlook.

2   Background

It is impossible to define the truth conditions of matrix questions. A question like (3) cannot be said to be true or false under any circumstances. Things change, though, when the question is embedded inside a matrix declarative sentence, as in (1).

(3)   Who danced at the party?

It is possible to decide whether (1) is true or false by assessing whether Ali knows the correct answer to the question *Who danced at the party?*. Thus, for a semantic theory of questions, it is essential to investigate their answerhood conditions. The central issue here is the question of how exhaustive Ali's knowledge has to be for (1) to be true. As presented above in (2), the theoretical literature on question semantics distinguishes three levels of exhaustivity.[1]

Early approaches to the semantics of embedded questions analyzed such questions as either WE (Karttunen, 1977) or SE (Groenendijk and Stokhof, 1984). However, Heim (1994) argued that the exhaustive interpretation of a question varies with the embedding verb in the matrix clause: While questions embedded under *know* receive SE readings, questions embedded under the emotive factive verb *surprise* or under speech act verbs also give rise to WE readings (see also Beck and Rullmann (1999)).[2] According to this view, given that Ali is aware of all the people in the domain, (4-b) follows from (4-a) whereas (5-b) does not follow from (5-a).

(4)   a.   Ali knows who danced at the party.
        b.   Ali knows who didn't dance at the party.

(5)   a.   Ali was surprised about who danced at the party.
        b.   Ali was surprised about who didn't dance at the party.

A crucial difference between *know* and *surprise* is that the former is obligatorily distributive whereas the latter has been argued to be only optionally distributive (Lahiri, 2002). Thus, (4-a) will only be true if Ali's knowledge applies to the entire positive answer space (WE-interpretation) or to the complete positive and negative answer space

---

[1]In addition to the three exhaustive readings, there are also non-exhaustive (NE) interpretations, or so-called mention-some readings. Such readings are typical for sentences containing existential modals, or for sentences occuring in particular (non-exhaustive) contexts. For instance, the sentence *Kim knows who can lend me a drill.* is generally considered true in a situation in which Kim knows one person who would be willing or able to lend out their drill. In this paper, we will not be further concerned with mention-some readings with the exception of NE-readings with *sich einig sein*, 'agree'. See, for example, Xiang (2016) and Moyer and Syrett (2019) on this topic.

[2]Cf. Sharvit (2002), Spector and Egré (2007), and Guerzoni and Sharvit (2007), who assume that questions embedded under *know* allow for both WE and SE readings.

(SE-interpretation). In contrast, there are two possibilities for (5-a) to be true under the weak exhaustive reading. He could be distributively surprised at each and every dancer in the positive answer space, but it would also suffice for him to be non-distributively surprised by some elements in the positive answer space, i.e. Ali could be surprised by some of the dancers only. Moreover, if *surprise* also allowed for SE-interpretations in ranging over the complete positive and negative answer space, (5-a) would also be true in a situation in which Ali expected some people to dance who in fact did not dance. In this case, Ali's surprise would be exclusively directed at the negative answer space. Theiler (2014) argues that such SE-readings do in fact exist under a deductive interpretation of *surprise*. On this account, the emotional state of surprise is directed at the overall composition of the answer, i.e. the positive answer space may be larger or smaller than previously assumed. The approach in Ginzburg and Sag (2000) can also account for the existence of SE-readings with *surprise*. These authors contend that emotive-factive verbs take facts, and not questions as arguments (cf. Abenina-Adar (2019) and Zimmermann et al. (2020)).

Based on the semantic behavior of speech act verbs such as *predict* Klinedinst and Rothschild (2011) postulated the existence of IE-readings (cf. Spector (2005), Spector (2006) who also argues for the existence of IE-readings under *know* and Nicolae (2013) who uses the label WE for this reading but assumes false answer sensitivity). Example (6) (their example [12], p. 7) shows that the WE-interpretation does not have the appropriate truth conditions as it contains two false positive beliefs. Such false positive beliefs are ruled out on the IE-interpretation.

(6)   Context: Frank and Emilio sang.
       A phone survey is taken to assess audience interest in last night's episode of a televised talent show, by checking their recall of the contestants. Arthur says, "I'm sure Frank sang and Emilio sang, and I'm not sure about anyone else." Bert says, "Frank, Emilio, Bill and Ted sang." It is decided to send a thank-you prize to Arthur but not Bert on the grounds that it's true that Arthur told us who sang, but Bert didn't.

In the case of *agree*,[3] truth does not play a role. What matters is the mutual alignment of the attitude holders' beliefs. There have been different predictions concerning the extent to which the beliefs have to be aligned (see Chemla and George (2016) for more extensive discussion). Kratzer (2006) and Spector and Egré (2007) take the view that a sentence like (7) is only true (i.) if Kim and Ali have opinions about every dancer, and (ii.) if their opinions are completely aligned both with regard to the dancers they consider to lack talent and with regard to the dancers they do not consider to lack talent. This corresponds to a SE interpretation. Under this SE-interpretation, (7) is considered false if Ali or Kim are uncertain about some of the dancers even if their positive beliefs as well as their uncertainties are aligned.

(7)   Kim and Ali agree on who of the dancers lack talent.

In contrast, Beck and Rullmann (1999) and Sharvit (2002) consider (7) to be true in this scenario as well, which corresponds to an IE-interpretation. On an even weaker interpretation, which we call the WE-reading, *agree* would allow for disagreement on

---

[3]In this paper, we will only be concerned with reciprocal *agree on*. The variant *agree with* is argued to behave differently in some respects (see for example Lahiri (2002))

the negative answer space (but complete agreement on the believed-to-be positive answer space). To illustrate, Ali could be unsure about a dancer who Kim believes to have talent. This WE-interpretation is predicted to be possible on the account in Lahiri (2002).

Chemla and George (2016) present results of an experimental study on the truth conditions of questions embedded under *agree* and found that participants were ready to accept belief alignments corresponding even to the weakest WE-construal from above, whereas incomplete, or non-exhaustive (NE) positive belief alignments were rejected. With regard to the verb *predict*, Cremers and Chemla (2017) present experimental evidence for both SE- and IE-readings.

Other experimental results are more at odds with the judgments found in the theoretical literature: For instance, the acceptance rate of IE readings for questions embedded under *know* was found to be around 90 percent in Cremers and Chemla (2016) and Cremers et al. (2017), which is similar to the acceptance rate of SE-readings. Moreover, contrasting with this positive evidence for the availability of IE readings, IE readings were not attested in another experimental study by Cremers and Chemla (2017). Finally, Cremers and Chemla (2017) found that next to non-distributive WE-readings, SE readings were also accepted with embedded question under *surprise*. This finding stands in contrast to a common conception in the theoretical literature (Berman, 1991; Heim, 1994; George, 2011), but is in line with claims in Theiler (2014) and Zimmermann et al. (2020).

Tables 1–4 provide for each verb of embedding an overview over the literature. The various experimental and theoretical papers come with different predictions concerning the availability of SE/IE/WE readings for different embedding verbs. Summarizing predictions on this issue comes with difficulties, partly because the papers under discussion have different (historical) backgrounds. In order to deal with this difficulty, we show in the tables whether the theories predict an embedded question construction to be true (symbol ✓) in a particular scenario (shown in each column).[4] To illustrate, Karttunen (1977) assumes a weak exhaustive question semantics. According to this view, the minimal requirement for a scenario that makes (1) true, is that Ali knows for all people who danced at the party that they danced. However, if Ali additionally does not have false beliefs about non-dancers (IE-scenario) or is aware that no other person danced (SE-scenario), (1) is also true. In contrast, in the question semantics in Groenendijk and Stokhof (1984), (1) is only considered true in a context that corresponds to the SE-reading. Note that in the case of *agree* the literature does not use consistent terminology. Therefore, we adjusted the labels according to the classification given in the previous paragraph.

Table 1: Literature overview: Questions embedded under ***know*** (italics indicate experimental study).

Theory	SE	IE	WE
Karttunen (1977), Berman (1991) Sharvit (2002), Spector and Egré (2007), Guerzoni and Sharvit (2007), *Cremers and Chemla (2017)*	✓	✓	✓

Continued on next page

---

[4]We only report explicit judgments in the literature. The question semantics by Groenendijk and Stokhof (1984), for example, is generally applicable to all verbs but as each verb could be modelled to have a different semantic contribution, which manipulates the exhaustivity of the embedded question construction, we only report judgements on verbs that the respective authors actually discuss or on verbs that are very similar to the ones discussed.

Table 1 – continued from previous page

Theory	SE	IE	WE
Groenendijk and Stokhof (1984), Heim (1994), Beck and Rullmann (1999), Lahiri (2002), George (2011), Klinedinst and Rothschild (2011), Theiler (2014)	✓	*	*
Spector (2006), Nicolae (2013), Spector and Egré (2015), *Cremers and Chemla (2016)*, *Cremers et al. (2017)*, Uegaki (2015), Theiler et al. (2018), Zimmermann et al. (2020)	✓	✓	*

Table 2: Literature overview: Questions embedded under **predict** (italics indicate experimental study).

Theory	SE	IE	WE
Karttunen (1977), Berman (1991), Heim (1994), Beck and Rullmann (1999), Sharvit (2002), *Klinedinst and Rothschild (2011)* (small pilot study)	✓	✓	✓
Groenendijk and Stokhof (1984)	✓	*	*
Spector (2006), Spector and Egré (2015), Theiler (2014), Uegaki (2015), *Cremers and Chemla (2016)*, Theiler et al. (2018)	✓	✓	*

Table 3: Literature overview: Questions embedded under **agree** (italics indicate experimental study).

Theory	SE	IE	WE
Lahiri (2002), Guerzoni and Sharvit (2007), Uegaki (2015), *Chemla and George (2016)*, Theiler et al. (2018)	✓	✓	✓
Kratzer (2006), Spector and Egré (2007), Theiler (2014)	✓	*	*
Beck and Rullmann (1999), Sharvit (2002), Spector and Egré (2015)	✓	✓	*

Table 4: Literature overview: Questions embedded under **surprise** (italics indicate experimental study).

Theory	SE	WE
Berman (1991), Heim (1994), Beck and Rullmann (1999), Sharvit (2002), Guerzoni and Sharvit (2007), Spector and Egré (2007), George (2011) (assumes mention-some answer), Nicolae (2013), Uegaki (2015), Spector and Egré (2015), Theiler et al. (2018)	*	✓

Theory	SE	WE
Klinedinst and Rothschild (2011), Theiler (2014), *Cremers and Chemla (2017)*, Zimmermann et al. (2020)	✓	✓

All previous experimental studies have in common that they tested for the general availability of different EXH-readings through different variants of truth-value judgment tasks. The studies had a low empirical threshold in the sense that they aimed at testing for all readings that are acceptable one way or other. Still, it is not entirely clear what exactly it means for a reading to be accepted in a given experimental setting. Would the participants themselves actually use the target sentence with the relevant interpretation in the respective context? Or do they merely consider it possible that somebody else may understand the sentence in this way, possibly on a more liberal use of the language? Since it is possible to conceive of several (in part non-linguistic) reasons for why participants accept an interpretation in a given experimental setting, it is not trivial to decide what can be safely concluded from such truth-value judgment experiments. To be clear, when a reading is accepted in an experiment, this does not necessarily mean that the availability of this reading is relevant from a semantic perspective. Let alone, that interlocutors would reliably use such expressions for conveying that particular interpretation.

3   Goals and Expectations

Given the inconclusive empirical evidence so far, the central aim of this study was to gather independent evidence for the availability and relevance of different exhaustivity readings in natural and goal-oriented communication. We did so by testing four different embedding predicates in the same experimental setting, thereby contributing to the construction of a robust data base required for semantic theory building. To be concrete, in our experimental task, participants had to take the perspective of an addressee into account when evaluating the availability of a given reading in a given situation. This setup was thus specifically designed for targeting readings that are optimal from a communication-oriented perspective. This should enable us to differentiate between dominant, and marginal, and impossible interpretations of embedded questions.

We tested each of the the four embedding predicates *wissen* 'know', *korrekt vorhersagen* 'correctly predict'[5], *sich einig sein* 'agree', and *überrascht sein* 'be surprised' for the availability of three different readings. For *wissen* and *korrekt vorhersagen*, these were SE, IE and WE. With *agree*, we tested for the availability of SE- and IE-readings, and in order to set a negative baseline, we also tested for the non-exhaustive (NE) reading, on which positive belief alignment is incomplete. As for *überrascht sein*, we tested for both distributive and non-distributive WE readings, as well as for the non-distributive SE reading. Recall that on this reading the surprise is directed at parts of the negative answer space.

As we apply a novel approach to experimentally investigating exhaustivity properties of embedded questions, and not a particular theory, we will eschew from making predictions

---

[5]We tested *korrekt vorhersagen* instead of bare *vorhersagen* as the latter is ambiguous between a literal interpretation of uttering a predictive speech act, and a non-literal, truth-based interpretation that this prediction was actually correct, cf. Theiler (2014)

in the classical sense. Still, certain readings were of particular interest to us against the backdrop of previous research on the availability of different exhaustive readings. For these, we would like to spell out some expectations. To begin with, concerning *wissen*, it seems intuitively clear that if one says (1), and if Ali knows the domain, it must follow that he also knows who did not dance. In contrast, we find it harder to accept statement (1) if Ali is unsure whether some people in the domain danced, even if they, in fact, did not dance. Thus, we expect SE to be the pragmatically optimal interpretation, in line with large parts of the theoretical literature, and we expect a diminished acceptance for IE-readings of questions embedded under *wissen*. Finally, we do not expect participants to tolerate false beliefs with factive and veridical *wissen*, and therefore we expect rejection of WE-readings. Similarly, questions embedded under *korrekt vorhersagen* should likewise not allow for the acceptance of false beliefs on the WE-reading, whereas the other two readings (IE, SE) should be likely interpretive options in natural linguistic settings. With regard to *sich einig sein*, it seems unlikely that the incomplete belief alignments found with NE-interpretations will be accepted. In the case of *überrascht sein*, we do not have any fixed expectations. Here, we are particularly interested in finding out whether there is a difference in acceptance between distributive and non-distributive WE-readings, on the one hand, and to what extent the debated (non-distributive) SE reading will be accepted in our communication-oriented experimental setup.

4 Experiment

4.1 Methodological considerations

Semantic experiments present participants with a cognitively demanding task. Participants have to concentrate for a considerable length of time, and they must generate interpretation-based judgments on numerous stimuli that are often quite similar to one another. Lack of motivation and/or increasing fatigue may lead to superficial reading and shallow processing of the experimental stimuli, and, at worst, result in the selection of a random answer. This type of behavior is described by Krosnick (1991, 1999), who subsumes it under the label *satisficing*, a notion originally coined by Simon (1957). According to Vanette and Krosnick (2014) "when faced with [...] demanding information-processing tasks[...][,] people often expend only the amount of effort necessary to make an acceptable or satisfactory decision" (see also Kool et al. (2010) on avoidance of cognitive demand). In particular, such experimental settings in which the participant has to click through a repetitive experiment while seated alone in front of a computer seem prone to the induction of satisficing strategies. It seems that often a compensation for mere participation is an insufficient incentive for participants to attempt to come up with the best response in each and every experimental trial.

In view of this methodological problem, we designed an experiment that maximizes rational behavior in participants, and which makes them want to find the optimal response in each and every experimental trial. In this experiment, every individual response of the participant had financial consequences for them.[6] This way, it was in participants' own interest to think carefully about each individual stimulus. Moreover, we aimed at creating a diverse and entertaining item set, and we chose an interactive lab setting with

---

[6]In economics, it is common practice to use such performance-based compensations. See in particular Camerer and Hogarth (1999) on the effort increasing potential of money.

a dialogue partner in order to prevent fatigue in participants, and in order to increase their commitment.

Participants' bias towards optimal solutions was further fostered by the necessity to engage in recursive thinking when responding to the trial stimuli. Participants had to consider the question of whether another person would share their judgment, or not. Thus, the notion of 'optimal response' should not only be understood in terms of the best solution arrived at after intensive pondering, but it is also that interpretation that constitutes a communicative equilibrium (cf. Benz and Stevens (2018) on Nash equilibria in game theory).

Finally, we were interested in gaining further insights into participants' reasoning procedures. To this end, we conducted short interviews after the actual experiment. The goal of these interviews was to find out whether decisions on critical stimuli were difficult, and to gain insight into which reasoning procedures resulted in the judgments that were actually made.

4.2   Method

Participants did not judge the test items in vacuo, but in the context of a role play. They were presented with the following context: In a reality show called *The Glass House*, the five contestants, Alessa, Carlo, Freddy, Mara and Sophie, are filmed during their activities in the house and while doing certain things for a dare. In a special episode at the end of the season, the presenters, Tim and Tiffany, are looking back at the season. In doing so, they have to answer questions about the contestants and their activities in the house. Beforehand, viewers could place bets on what Tim and Tiffany know, what they correctly predicted, what they agree on, and what they are surprised at. The participants' concrete task in the experiment was to evaluate the outcome of the bets.

4.2.1   Participants

We tested 24 participants, 17 female, 7 male, between 20 and 31 years old (M = 24.37 years). Participants were recruited via postings on university-related facebook groups and via printed posters on campus. 20 of them were University students. All of them were native speakers of German (mostly Austrian German). The financial compensation was performance-based and contingent on the individual responses. It varied between 9.40 and 10.40 Euro.

4.2.2   Materials

As the target sentences were the objects of bets, the experimental stimuli were presented on betting slips, as shown in the template in (8). Concrete sample items are presented in (10)–(13) in Section 4.2.3. The front side of the betting slip showed (i.) the bet in the form of a question embedded under one of the four relevant matrix verbs, and (ii.) a dialogue or monologue in which the attitude holder in question expresses her beliefs. The backside of the slip showed a table displaying what had actually happened. The different modes of presentation were chosen to prevent participants from using a low-level answering strategy[7], i.e. they could not simply look for differences between two sentences printed on the same page.

---

[7]Chemla and George (2016) also discuss this issue.

(8)   *Frontside*
        Lina bets: X <+/- negation> <verb> <Q>

        Dialogue/Monologue: Contextual information that X <attitude> that <answers to Q>

        Backside
        <facts in the world>

We manipulated the variable *verb* in the template by inserting one of the four verbs *wissen* 'know', *korrekt vorhersagen* 'correctly predict', *sich einig sein* 'agree' and *überrascht sein* 'be surprised'. For each verb, we created a set of 6 lexicalizations, yielding a total of 24 test items. In addition to the 24 test items, we created 26 fillers that also served as controls. The fillers also involved questions embedded under the four verbs. Eight of the fillers featured *wissen*, the other verbs occurred six times each. The filler questions were not only constituent questions like the target items but were of various types (adjunct questions, polar questions...). In addition, half of the fillers consisted of bets that were uncontroversially won, and the other half were bets that were uncontroversially lost. Same as the test items, fillers varied concerning the factor NEGATION (see Section 4.2.3). The experimental items were distributed over six experimental lists with four participants per list. The combination of the six conditions (3 readings x [+/-] negation, see below) and six lexicalizations varied systematically within each verb between lists. That is, one lexicalization occurred in a different condition on each list.

4.2.3   Factorial design

For each verb, the design was 3 (READING) x 2 (NEGATION) x 2 (ROLE). The first factor READING had three levels that depended on the verb as described in Section 3. The factor was tested within subjects and within items (per verb). The contents of the dialogue/monologue and the reported actual facts in the world were manipulated in the template to yield the three different readings by changing the values of the variables *attitude, answers to Q* and *facts in the world*. Table 5 shows the manifestations of the three readings that were tested for each verb. Note that for the verb *sich einig sein* 'agree', there was no display of facts in the world on the backside since in this case there is no objective factual base against which to measure the attitude holders' subjective agreement.

Table 5: Readings tested for each verb.

Verb	Reading	Statement of attitude holder	Facts in the world
wissen 'know'	SE	A, B, C have the property. D, E do not have the property.	A, B, C have the property. D, E do not have the property.
	IE	A, B, C have the property. Not sure whether D, E have the property.	A, B, C have the property. D, E do not have the property.

Continued on next page

10

Table 5 – continued from previous page

Verb	Reading	Statement of attitude holder	Facts in the world
	WE	A, B, C, D have the property. E does not have the property.	A, B, C have the property. D, E do not have the property.
korrekt vorhersagen 'correctly predict'	SE	A, B, C will do the activity. D, E won't do the activity.	A, B, C did the activity. D, E didn't do the activity .
	IE	A, B, C will do the activity. D, E will maybe do the activity.	A, B, C did the activity. D, E did not do the activity.
	WE	A, B, C, D will do the activity. E won't.	A, B did the activity. C, D, E did not do the activity.
sich einig sein 'agree'	SE	Both believe that A, B, C have the property and that D, E don't have the property.	-
	IE	Both believe that A, B, C have the property and don't have an opinion about D, E.	-
	NE	Both believe that A, B, C have the property and that E does not have the property. Different opinions about D	-
überrascht sein 'surprise'	SE	Expected that C, D would do the activity, too.	A, B did do the activity. C, D, E didn't
	WEdis	Didn't expect A, B, C to do the activity.	A, B, C did do the activity. D, E didn't.
	WEnondis	Did not expect A, B to do the activity.	A, B, C, D did do the activity. E didn't.

The second factor, NEGATION, was tested within subjects and within items. As indicated in (8), the embedding matrix predicate was either negated or not. We included this factor to see whether participants' judgments were consistent. To illustrate, if participants judge consistently, they should accept the target sentence in the condition SE, -neg, shown in (9-a), and reject the target sentence in the condition SE, +neg, shown in (9-b), and vice versa.

(9)   a.   **-neg** Tiffany knows who of the participants wears contact lenses.
       b.   **+neg** Tiffany doesn't know who of the participants wears contact lenses.

(10)–(13) present sample test items for each verb.

(10)   *Know*, IE, -neg
         Frontside
         Lina wettet: Tiffany weiß, wer von den Teilnehmerinnen und Teilnehmern Sushi mag.

'Lina bets: Tiffany knows who of the participants likes sushi.'

Tiffany: Asiatisches Essen finde ich persönlich klasse, besonders Sushi! Ich habe bemerkt, dass Carlo, Mara und Freddy auch Sushi mögen. Die drei hatten ja mal ein ausführliches Gespräch über Sushi. Was Alessa und Sophie betrifft, bin ich mir unsicher. Ich habe nicht mitbekommen, ob die beiden Sushi mögen oder nicht.
'Tiffany: I love Asian food, especially sushi! I noticed that Carlo, Mara and Freddy like sushi as well. The three of them once had quite a lengthy discussion about sushi. Concerning Alessa and Sophie, I am not sure. I didn't figure out whether they like sushi or not.'

Backside (Actual Facts)

Personen 'Persons'	**Sushi mögen** 'Likes sushi'
Alessa	nein 'no'
Carlo	ja 'yes'
Freddy	ja 'yes'
Mara	ja 'yes'
Sophie	nein 'no'

(11)   *correctly predict*, IE, +neg
         Frontside
         Lina wettet: Tim hat nicht korrekt vorhergesagt, wer von den Teilnehmerinnen und Teilnehmern in der Sendung einen Wutanfall bekommen würde.
         'Lina bets: Tim didn't correctly predict who of the participants would throw a tantrum on the show.'

         Dialog aus der ersten Sendung, in der die fünf Teilnehmerinnen und Teilnehmer neu in das gläserne Haus eingezogen sind.
         Tiffany: Tim, wie ist deine Prognose? Wer von den Teilnehmerinnen und Teilnehmern wird in der Sendung einen Wutanfall bekommen?
         Tim: Also, Carlo, Mara und Alessa scheinen ein wenig cholerisch. Die drei werden in der Sendung einen Wutanfall bekommen und vielleicht werden auch noch andere einen Wutanfall bekommen.
         'Dialogue from the first episode, in which the five participants had newly moved into the glass house.
         Tiffany: Tim, what's your prediction? Who of the participants will throw a tantrum on the show?
         Tim: Well, Carlo, Mara and Alessa seem to be a bit choleric. The three of them will throw a tantrum on the show, and maybe others might throw a tantrum as well.'

Backside (Actual Facts)

Personen 'Persons'	**Wutanfall bekommen** 'Lost temper'
Alessa	ja 'yes'
Carlo	ja 'yes'
Freddy	nein 'no'
Mara	ja 'yes'
Sophie	nein 'no'

(12)   *agree*, SE, -neg
Frontside
Tiffany und Tim sind sich einig, wer von den Teilnehmerinnen und Teilnehmern sich stilvoll kleidet.
'Lina bets: Tiffany and Tim agree on who of the participants dresses stylishly.'

Tiffany: Reden wir mal über den Kleidungsstil der Teilnehmerinnen und Teilnehmer. Ich finde ja, dass sich Freddy sehr stilvoll kleidet.
Tim: Das sehe ich genauso. Freddy hat einen tollen Geschmack. Ich finde aber auch, dass sich Mara und Alessa sehr schön kleiden.
Tiffany: Ja, da stimme ich dir zu. Sophie und Carlo allerdings laufen total unmöglich rum, wie ich finde.
Tim: Oh ja, die beiden kleiden sich überhaupt nicht stilvoll.
'Tiffany: Let's talk about the way the participants dress. I think Freddy dresses very stylishly.
Tim: I agree. Freddy has a great taste. However, I also think that Mara and Alessa dress nicely.
Tiffany: Yes, I am with you on that. Sophie and Carlo, by contrast, have an impossible style of dressing.
Tim: Oh yeah, the two of them do not dress stylishly at all.'

(13)   *surprise*, SE, -neg
Frontside
Lina wettet: Tiffany war überrascht, wer von den Teilnehmerinnen und Teilnehmern in der Sendung eine Heuschrecke gegessen hat.
Lina bets: Tiffany was surprised about who of the participants ate a grasshopper on the show.

Tiffany: Ich muss noch oft an die Sendung denken, in der Freddy und Alessa mutigerweise eine Heuschrecke gegessen haben und die anderen drei sich geweigert haben. Ich hatte erwartet, dass auch Carlo und Sophie in der Sendung eine Heuschrecke essen würden. Schließlich sind die beiden im Allgemeinen essenstechnisch sehr flexibel.
'Tiffany: I often think back to the episode in which Freddy and Alicia bravely ate a grasshopper and the other three refused to do it. I had expected Carlo and Sophia to also eat a grasshopper on the show since neither of them are picky eaters.'

Backside

Personen 'Persons'	**Heuschrecke gegessen** 'Ate grasshopper'
Alessa	ja 'yes'
Carlo	nein 'no'
Freddy	ja 'yes'
Mara	nein 'no'
Sophie	nein 'no'

Next to the item manipulation, we included a third factor on the participant level, ROLE, which was tested between subjects and within items. This factor served to control for an answering bias. Participants in role 1 had the task to decide whether or not to cash in the bet at a betting office. They were told that their friend, Lina, had placed some bets but did not have time to go to the betting office to cash in her bets. Instead, the participants were asked to go on her behalf, for which they would earn a share in the profits. They received 5 Euros starter cash. Submission of bets cost a fee of 10 cents each. For each submitted bet that was actually won the participants received 30 cents in return. Thus, in effect, the participants gained 20 cents for a submitted bet which was won, and they lost 10 cents for a bet that was lost. In this role, the participants profit financially from bets that are won, for which reason they may be biased towards judging difficult borderline cases as won bets. However, this bias is harnessed in by the fee for submission, as the participants will lose money by randomly submitting bets. Participants in role 2 acted as the clerk in the betting office and had to decide for each submitted bet whether it was won or not. Participants in role 2 received 15 Euro starter cash. For a won bet, they had to cash out 20 cents. If they decided incorrectly that a bet was lost which was actually won, there was a deduction of 30 cents. Participants in role 2 profit from lost bets, for which reason they could be biased towards rejecting bets. However, in this case, too, the financial deduction for incorrect decisions served to harness in the bias.

4.2.4 Procedure

The experiment was conducted in the lab. Before the start of the experiment, participants were offered a glass of water, a cup of coffee, and a sweet (Gailliot et al. (2007) offer experimental evidence for the dependence of self-control, as required for cognitively demanding tasks, on glucose). Depending on the role they were assigned, they received 5 or 15 Euro starter cash in stacks of 10 and 20 cent coins. After completing three practice items, which served to get the participants accustomed to the task, the actual experiment began[8]. The experimenter handed the betting slips one by one and in randomized order to the participant, who then had to decide whether to submit the betting slip or to cash out the gains. If they wanted to submit or cash out, depending on the role, they had to return the betting slip together with the money to the experimenter. The experimenter entered the participant's decision into an excel sheet that automatically calculated the sum the participant received as financial compensation after the experiment and the interview. The participant did not receive any feedback as to his/her gains and losses from individual bets. After processing of the first half of the betting slips there was a short

---

[8]Participants did not receive any feedback for their decisions on the practice items.

break in which the participant was offered a second sweet. The interview was conducted on completion of the experiment. It proceeded as follows: The experimenter gave the betting slips in question back to the participants one by one, s/he reminded them of their decision, and s/he asked them whether the decision had been easy, what had caused their decision, and whether they could imagine a different line of reasoning. The experiment and interview took 25 to 40 minutes in total. We collected the participants' personal data only after the experiment in order to prevent stereotype threat.

5   Results

Fillers
Table 6 shows the submission rate of filler bets by type of filler in count and percentage. Due to the overall low number of mistakes, no participant had to be excluded from further data analysis.

Table 6: Results fillers

Type of filler	Count	Percentage
Won bet	308	99
Lost bet	3	1

Decriptive statistics
We collected 24 data points per condition. Figure 1 shows the mean submission rates per condition for each verb. As expected, SE-readings for questions embedded under *wissen* 'know' had at-ceiling acceptance rates, whereas the acceptance rate of the corresponding IE-readings was considerably lower. In contrast, WE-readings were rejected, as expected for veridical *wissen*. Likewise, WE-readings of questions embedded under *korrekt vorhersagen* were also rejected. In the case of this verb, SE- and IE-readings received almost identical acceptance rates. Questions embedded under *sich einig sein* show the same pattern as those under *korrekt vorhersagen*. Again, there is no difference between SE- and IE-readings, whereas the NE-readings were completely rejected. With regard to *überrascht sein* 'be surprised', we observe that the acceptance rate of distributive and non-distributive WE-readings was high, but the SE readings also had a fairly high acceptance rate of almost 60 percent. Moreover, with the exception of the SE-reading for *überrascht sein*, the results of the non-negated and negated versions roughly match up for all other conditions.

   Turning to the factor ROLE, there was no essential difference between the two roles across all test items and fillers. Participants in role 1, who had to decide whether to submit bets, submitted 50 percents of all bets as won, and participants in role 2, who had to decide whether to cash out gains for submitted bets considered 49 percent of all bets as won. As there was hardly any variance in most conditions, we will only look in more detail at the behavior of the two roles in the IE-condition of *wissen* and the SE-condition of *surprise*. In condition IEnoneg of *wissen*, 42 percent of the participants in role 1 (submit bets) and 50 percent of the participants in role 2 (cash out gains) decided the bet was won. In condition SEnoneg of *überrascht sein*, participants in both roles decided that bets are won in 58 percent of cases.
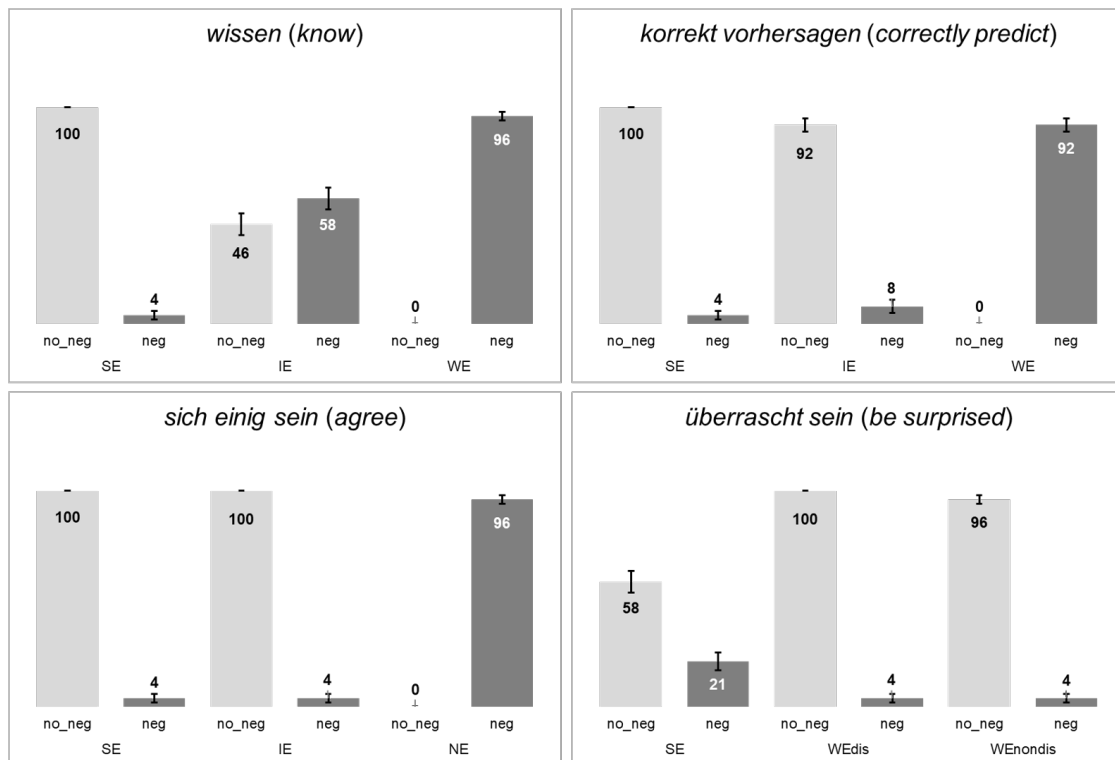
Figure 1: Mean submission rates per condition for each verb. Error bars represent standard error of the mean.

Inferential statistics

For the inferential statistical analysis, we used the software *R* (R Core Team, 2017) and the package *multcomp* (Hothorn et al., 2008). For each embedding predicate, we created nested GLMs.[9] The full model contained READING and NEGATION with an interaction term as factors (*glm(response ∼ reading * negation, family = binomial)*. In order to test for significance of the interaction and for main effects of the two factors, we conducted comparison of nested GLMs via a chi-square test. The model with the effect in question was compared against the model without the effect in question. We found that for each verb, there was a significant interaction between the factors READING and NEGATION (*wissen*: p < 0.001, *korrekt vorhersagen*: p < 0.001, *sich einig sein*: p < 0.001, *überrascht sein*: p < 0.01). Moreover, there were significant main effects of READING in the case of *korrekt vorhersagen* (p < 0.01), *sich einig sein* (p < 0.001), and *überrascht sein* (p < 0.001).

In addition, we used a Tukey post-hoc to test for significant differences between conditions in cases with differences of interest.[10] To this end, we created the following GLM for each verb: *glm(response ∼ condition, family = binomial)*. With regard to *wissen*, the difference between the SE- and the IE-reading was significant (p < 0.05) as well as the difference between the IE- and the WE-reading (p < 0.05). Concerning *korrekt vorhersagen*, the small difference between SE and IE was not significant (p = 0.92). In the case of *überrascht sein*, the difference between the distributive and the non-distributive WE-readings was not significant (p = 0.99) nor were the differences between the SE-reading

---

[9]We had to add data of a dummy participant to generate minimal noise due to cases in which variance was zero. In each condition, this dummmy participant did exactly the opposite of what the majority did. Thus, there is no conceivable way in which a type I error could result from this approach.

[10]The results reported below always refer to the conditions without negation.

and the distributive (p = 0.16) and the non-distributive WE-reading (p = 0.23).

Interviews
In the following, we will provide anecdotal information on participants' comments during the interviews following the completion of the actual experiment. We focus on the conditions *wissen* 'know', IE, and *überrascht sein* 'surprise', SE, as these were the conditions in which participants' judgments diverged most.

Regarding *wissen*, IE, half of the participants expressed that they found the decision hard. For comparison, in the case of *wissen*, WE, the majority of participants indicated that they found the decision easy. Three of the participants who accepted the IE-reading under *wissen* commented that Tim/Tiffany had merely guessed correctly with the effect that the bet is won, but that they do not possess the actual knowledge. With regard to this condition, most participants could imagine the opposite standpoint to be valid. One participant insisted that completeness of knowledge is crucial for saying that someone knows something, corresponding to the SE-interpretation.

In the case of *überrascht sein*, SE, fewer participants found the decision hard than in the case of *wissen*, IE. Out of the participants who accepted this interpretation, there were three that stated that there was a general feeling of surprise and non-fulfilled expectations, for which reason they considered the bet as won. Moreover, three participants that accepted this reading said afterwards that they would revise their decision if they still could. The reason was that they only realized during the interview that the surprise was directed at the negative answer space, i.e. at those persons that did not engage in the relevant activity. Contrasting with this, there were also two participants from those who rejected the interpretation first time round who would in hindsight change their decision to acceptance. Overall, more than half of the participants expressed that they could imagine the opposite standpoint to be valid.

6   Discussion

An important general finding is that our novel experimental methodology is applicable and empirically sound in allowing for obtaining clean data: The participants in the experiment made only very few mistakes on the filler items, and their responses were overall consistent, which indicates that participants were able to work on the experimental task in a focused manner without suffering from cognitive overload. Moreover, we could not detect a bias stemming from the experimental factor role (submitting the bet vs. cashing out). This finding suggests a semantic basis for many of the interpretations obtained.

Turning to the experimental results, importantly the IE-reading of *wissen* 'know' received considerably less acceptance in our participant-engaging design than in the judgment tasks of Cremers and Chemla (2016) and Cremers et al. (2017). In addition, the subsequent interviews showed that for many participants it was not easy to make a decision on the test items in the IE-condition. Still, the significant differences between the IE-responses, on the one hand, and the SE- and the WE-responses, on the other, shows that the IE-interpretation is a reading of its own. In our experiment, participants had to take into account whether another person that was an authority on whether the bets considered were actually won, would share their semantic judgment. According to our results, then, the IE-reading appears to be of less communicative relevance that the SE-reading in the case of *wissen*. The SE-reading is the preferred interpretation constituting a communicative Nash-equilibrium. A full semantic and pragmatic theory of questions

will have to account for this finding.[11]

As for the speech act verb *korrekt vorhersagen* 'correctly predict', there was no significant difference between the SE- and the IE-readings. In line with judgments in the more recent literature, the IE-reading seems to constitute a perfectly acceptable reading for speech act verbs. Our experiment showed that this also holds true under an communication-oriented perspective. As in the case of *know*, it became clear that false beliefs, which the WE-reading allows, are unacceptable reconfirming the need for a false belief constraint in the semantics of embedded questions.

In line with Chemla and George (2016), we found that sentences with questions embedded under the verb *sich einig sein* 'agree' are accepted in contexts in which belief alignment is complete with respect to the positive answer space, i.e. in SE and IE-scenarios. The IE-interpretation allows for aligned uncertainties. As expected, participants consistently rejected NE-interpretations in which there was no agreement on the positive answer space. We did not test for WE-readings on which the interlocutors disagree about the negative answer space, such that one interlocutor states that she does not believe person x to exhibit a certain property while the second interlocutor is unsure whether x has the property in question; see above. This reading was attested in the experiment by Chemla and George (2016). It would be interesting to see whether WE-readings are also accepted in our experimental setup targeting only readings that are optimal from a communication-oriented perspective.

Concerning questions embedded under *überrascht sein* 'be surprised', we found evidence for the communicative relevance of three different interpretations: Next to distributive and non-distributive WE-readings, SE-readings involving surprise at the negative answer space could also be accessed to a significant degree, in line with the finding in Cremers and Chemla (2017). However, participants' responses in SE-scenarios were not as consistent as in the other two conditions. This may be an artefact of the experimental design. It may be argued that the test items in this condition were formulated less clearly than in the other conditions. In the wording "I would have expected Carlos and Sophia to **also** eat a grasshopper on the show" it may have been possible to miss the *also*, which expresses the fact that Tim/Tiffany expected the other participants to engage in the activity as well. Still, the subsequent interviews provide some evidence that several participants consciously accepted the SE reading, and that they were content with their decision even in hindsight. The results indicate that questions embedded under *überrascht sein* have more liberal truth conditions than often assumed in the classical theoretical literature. As mentioned in Section 2, some theoretical approaches account for SE readings with questions under *surprise* by assuming that the cognitive state of surprise can be directed at the general composition of the answer proposition on a deductive re-interpretation of *surprise* (Theiler, 2014). Others argue, in contrast, that surprise takes facts or situations as arguments (Ginzburg and Sag, 2000), and that SE-surprise is conditioned by missing but expected ingredients of the situation characterized by the wh-complement (Zimmermann et al., 2020). The comments made by some participants that there was a general feeling of surprise can be taken as support for such a view.

For comparison with the predictions in the literature, Tables 1–4 from Section 2 are shown again in 7–10. The rows corresponding with our results are highlighted. As we

---

[11]Zimmermann et al. (2020) make a first step in this direction. They formulate a pragmatic *Principle of attitude verification (PARV)*, which is based on theory-of-mind-related considerations on how to evaluate the validity of knowledge attributions, and which does account for the prevalence of SE-readings of questions embedded under *know*.

did not test the WE-reading of questions embedded under *agree*, two rows in Table 9 are potentially compatible with our results. Note that the predictions in Spector and Egré (2015), Uegaki (2015) and Theiler et al. (2018) match our experimental results except for questions embedded under *surprise*.

Table 7: Literature overview: Questions embedded under **_know_** (italics indicate experimental study)

Theory	SE	IE	WE
Karttunen (1977), Berman (1991) Sharvit (2002), Spector and Egré (2007), Guerzoni and Sharvit (2007), *Cremers and Chemla (2017)*	✓	✓	✓
Groenendijk and Stokhof (1984), Heim (1994), Beck and Rullmann (1999), Lahiri (2002), George (2011), Klinedinst and Rothschild (2011), Theiler (2014)	✓	*	*
Spector (2006), Nicolae (2013), Spector and Egré (2015), *Cremers and Chemla (2016)*, *Cremers et al. (2017)*, Uegaki (2015), Theiler et al. (2018), Zimmermann et al. (2020)	✓	✓	*

Table 8: Literature overview: Questions embedded under **_predict_** (italics indicate experimental study)

Theory	SE	IE	WE
Karttunen (1977), Berman (1991), Heim (1994), Beck and Rullmann (1999), Sharvit (2002), *Klinedinst and Rothschild (2011)* (small pilot study)	✓	✓	✓
Groenendijk and Stokhof (1984)	✓	*	*
Spector (2006), Spector and Egré (2015), Theiler (2014), Uegaki (2015), *Cremers and Chemla (2016)*, Theiler et al. (2018)	✓	✓	*

Table 9: Literature overview: Questions embedded under **_agree_** (italics indicate experimental study)

Theory	SE	IE	WE
Lahiri (2002), Guerzoni and Sharvit (2007), Uegaki (2015), *Chemla and George (2016)*, Theiler et al. (2018)	✓	✓	✓
Kratzer (2006), Spector and Egré (2007), Theiler (2014)	✓	*	*
Beck and Rullmann (1999), Sharvit (2002), Spector and Egré (2015)	✓	✓	*

Table 10: Literature overview: Questions embedded under **surprise** (italics indicate experimental study)

Theory	SE	WE
Berman (1991), Heim (1994), Beck and Rullmann (1999), Sharvit (2002), Guerzoni and Sharvit (2007), Spector and Egré (2007), George (2011) (assumes mention-some answer), Nicolae (2013), Uegaki (2015), Spector and Egré (2015), Theiler et al. (2018)	*	✓
Klinedinst and Rothschild (2011), Theiler (2014), *Cremers and Chemla (2017)*, Zimmermann et al. (2020)	✓	✓

To sum up, except for the case of IE-readings of questions embedded under *wissen*, we were able to generally replicate the findings of previous experimental research on the meaning of embedded wh-questions. However, whereas the previous experiments focus on the question of which readings are available in principle without considering their status as semantic or pragmatic inferences, our approach goes one step further. In particular, our participant-engaging communication-based design provides insights into the pragmatics of embedded questions. By using a method that focuses on optimal readings from a communication-oriented perspective, we could show which interpretations constitute communicative equilibria. These insights should form the basis for future semantic and pragmatic theories of the meaning of questions.

References

Abenina-Adar, M. (2019). Interesting interrogatives. Talk at Sinn und Bedeutung 24, Osnabrück.

Beck, S. and Rullmann, H. (1999). A flexible approach to exhaustivity in questions. *Natural Language Semantics*, 7:249–298.

Benz, A. and Stevens, J. (2018). Game-theoretic approaches to pragmatics. *Annual Review of Linguistics*, 4:173–191.

Berman, S. (1991). *On the semantics and logical form of WH-clauses*. PhD thesis, University of Massachusetts, Amherst, MA.

Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1):7–42.

Chemla, E. and George, B. R. (2016). Can we agree about 'agree'? *Review of Philosophy and Psychology*, 7(1):243–264.

Cremers, A. and Chemla, E. (2016). A psycholinguistic study of the exhaustive readings of embedded questions. *Journal of Semantics*, 33(1):49–85.

Cremers, A. and Chemla, E. (2017). Experiments on the acceptability and possible readings of questions embedded under emotive-factives. *Natural Language Semantics*, 25:223–236.

Cremers, A., Tieu, L., and Chemla, E. (2017). Children's exhaustive readings of questions. *Language Acquisition*, 24(4):343–360.

Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, A., Tice, D. M., and Brewer, L. E. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92(2):325–336.

George, B. R. (2011). *Question embedding and the semantic of answers*. PhD thesis, University of California, Los Angeles, CA.

Ginzburg, J. and Sag, I. (2000). *Interrogative investigations: The form, meaning and use of English interrogatives*. CSLI Publications, Standford, CA.

Groenendijk, J. and Stokhof, M. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam, Amsterdam.

Guerzoni, E. and Sharvit, Y. (2007). A questions of strength: On npis in interrogative clauses. *Linguistics and Philosophy*, 30:361–391.

Heim, I. (1994). Interrogative semantics and karttunen's semantics for *know*. In Buchalla, R. and Mittwoch, A., editors, *Proceedings of the Ninth Annual Conference and the Workshop on Discourse of the Israel Association for Theoretical Linguistics*, pages 128–144, Jerusalem. Academon.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1(1):3–44.

Klinedinst, N. and Rothschild, D. (2011). Exhaustivity in questions with non-factives. *Semantics and Pragmatics*, 4(2):1–23.

Kool, W., McGuire, J. T., Rosen, Z. B., and Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, 139(4):665–682.

Kratzer, A. (2006). Exhaustive questions. Linguistics Colloquium, Massachusetts Institute of Technology.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3):213–236.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1):537–567.

Lahiri, U. (2002). *Questions and answers in embedded contexts*. Oxford University Press, Oxford.

Moyer, M. and Syrett, K. (2019). Contextual, lexical and structural factors licensing mention-some embedded questions. In *Proceedings of the 54th Annual Meeting of the Chicago Linguistics Society*.

Nicolae, A. C. (2013). *Any questions? Polarity as a window into the structure of questions*. PhD thesis, Harvard University, Cambridge, MA.

R Core Team (2017). R: A language environment for statistical computing. Technical report, R Foundation for Statistical Computing, Vienna.

Sharvit, Y. (2002). Embedded questions and 'de dicto' readings. *Natural Language Semantics*, 10:97–123.

Simon, H. A. (1957). *Models of man: Social and Rational*. John Wiley, New York, NY.

Spector, B. (2005). Exhaustive interpretations: What to say and what not to say. Unpublished paper presented at the LSA workshop on *Context and Content*.

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques*. PhD thesis, Université Paris Diderot.

Spector, B. and Egré, P. (2007). Embedded questions revisited: *an* answer, not necessarily *the* answer. Presentation at MIT Ling-Lunch Seminar and Journées Sémantique et Modélisation.

Spector, B. and Egré, P. (2015). A uniform semantics for embedded interrogatives: *an* answer, not necessarily *the* answer. *Synthese*, 192:1729–1784.

Theiler, N. (2014). A multitude of answers: Embedded questions in typed inquisitive semantics. Master's thesis, Universiteit van Amsterdam, Amsterdam.

Theiler, N., Roelofsen, F., and Aloni, M. (2018). A uniform semantics for declarative and interrogative complements. *Journal of Semantics*, 35:409–466.

Uegaki, W. (2015). *Interpreting questions under attitudes*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Vanette, D. L. and Krosnick, J. A. (2014). Answering questions: A comparison of survey satisficing and mindlessness. In Ie, A., Ngnoumen, C. T., and Langer, E. J., editors, *The Wiley-Blackwell Handbook of Mindfulness*, pages 312–327. John Wiley & Sons.

Xiang, Y. (2016). *Interpreting questions with non-exhaustive answers*. PhD thesis, Harvard University, Cambridge, MA.

Zimmermann, M., Fricke, L., and Onea, E. (2020). Embedded questions are exhaustive alright, but... Submitted.