Plurality effects in an exhaustification-based theory of embedded questions

Cremers, A.

Link to publication

CrossMark

Plurality effects in an exhaustification-based theory of embedded questions

Alexandre Cremers[1] ⬦

Abstract

Questions embedded under responsive predicates and definite descriptions both give rise to a variety of phenomena which can be grouped under the term *plurality* effects: quantificational variability, cumulativity, and homogeneity effects. This similarity has not gone unnoticed, and many proposals have taken inspiration in theories of definite plurals to account for these effects with embedded questions (Dayal in Locality in WH quantification: questions and relative clauses in Hindi, Kluwer, Dordrecht, 1996; Lahiri in Questions and answers in embedded contexts, Oxford studies in theoretical linguistics 2, Oxford University Press, New York, 2002; a.o.). Recently these phenomena have received less attention, as the field has focused on the so-called *intermediate* exhaustive reading of embedded questions instead, after Spector (Exhaustive interpretations: what to say and what not to say, Presentation at LSA Workshop: 'Context and Content', 2005) called into question the traditional dichotomy between weak and strong exhaustive readings. As a result, the intermediate exhaustive reading has been accounted for at the expense of empirical coverage in other areas. In this paper, I propose a modular theory which derives the currently much discussed exhaustive readings without giving up the rich semantics necessary to account for plurality effects. My account of quantificational variability, cumulativity, and homogeneity effects builds on recent work on these phenomena in the nominal domain by adopting a categorial approach to embedded questions, while the strong and intermediate exhaustive readings are implemented using an independent strengthening mechanism suggested in Klinedinst and Rothschild (Semant Pragmat 4(2):1–23, 2011). The resulting theory not only recovers important results on plurality effects; it offers new, simple solutions for some puzzles presented in George (Question embedding and the semantics of answers, Ph.D. dissertation, UCLA, 2011; Thought J Philos 2(2):166–177, 2013) and Paillé and Schwarz (in: Stockwell (ed) Proceedings of WCCFL 36, vol 36, Cascadilla Proceedings Project, Somerville, 2018), naturally derives readings that had been postulated in previous literature (Preuss in Issues in the semantics of questions with

✉ Alexandre Cremers
  alexandre.cremers@gmail.com

[1] ILLC – Universiteit van Amsterdam, Amsterdam, The Netherlands

quantifiers, Ph.D. dissertation, Rutgers University, 2001), makes correct predictions in many unexplored cases, and is compatible with recent results in psycholinguistics. In the last sections I justify my assumptions and show how possible limitations I inherit from the theories I build on can be accommodated under standard assumptions.

1 Plurality effects with definite descriptions and questions

1.1 The basic data

Embedded questions share many properties with definite descriptions. Dayal (1996) was the first to fully acknowledge this parallel and was particularly interested in the uniqueness presupposition of singular *which*-questions. In this paper, I will mostly concern myself with the parallel between plural *which*-questions (and questions headed by a neutral *wh*-word) and definite plural DPs. Gajewski (2005) pointed out that questions give rise to all the phenomena associated with definite plurality. In particular, each of the phenomena affecting definite plural DPs in (1a)–(3a) has an equivalent found in embedded questions (1b)–(3b). For each phenomenon, the first mention of the effect in the literature on questions is indicated.

(1) Quantificational Variability Effect (QVE)[1] / Modification by an adverb of quantity (Berman 1991):

   a. The children mostly sang.
     ≈ Most children sang.

   b. John mostly knows which children sang.
     ≈ For most children who sang, John knows it.

(2) Semi-distributive / cumulative readings (Dayal 1996: 149):

   a. The girls chased the boys.
     ≈ Each girl chased some boy and each boy was chased by some girl.

   b. The teachers know which children sang.
     ≈ Each teacher knows at least a partial answer and each child who sang is known to have done so by at least one teacher.

(3) Homogeneity (Krifka 1996: 147):

   a. The children didn't sing.
     ≈ No child sang.

   b. Mary doesn't know which children sang.
     ≈ Mary has no idea which children sang.

While the phenomenon illustrated by (1) has been studied extensively from its discovery by Berman (1991) to Lahiri's (2002) major contribution, much less has been said about the two others. To my knowledge, only Lahiri (2002) proposed a detailed

---

[1] Berman (1991) makes an analogy with modification of an indefinite by adverbs of frequency, as in *A professor usually likes his students*. This led to a theory of questions as free variables lacking a quantificational force. Although I won't pursue such an analysis, I will stick to the widely accepted term 'QVE'.

analysis of cumulativity and, with the exception of recent work by Križ (2015b), no detailed analysis has been proposed for homogeneity effects with questions. Given the limited attention paid to these effects in the literature on questions, readers unfamiliar with the literature on definite plurals may want to see more data before they are convinced that anything interesting is going on in (3).

1.2 A few words on homogeneity

Let us first rule out a simplistic explanation of the data in (3). The examples in (4) below show that the "negation of universal" reading of (3a,b) cannot be explained by giving the definite description/question wide scope over negation. Indeed, the binding of the pronoun forces it to stay in the scope of the negative DP, yet the homogeneity effect remains. Example (24) in the next section will show that questions embedded under *know* cannot take scope over the subject of the sentence anyway.

(4) a. No professor$_i$ talked with the students she$_i$ likes.
     b. No professor$_i$ knows which of her$_i$ students cheated.

The intuition behind homogeneity effects is that (5a) is true if all children sang, false if none did, but neither true nor false otherwise. Similarly, (5b) would be true if Mary knew a complete answer to the question *Which children sang?*, false if she didn't know any partial answer, and neither true nor false if she had partial knowledge about which children sang. The fact that the partial cases receive an indeterminate truth value (a truth value gap) rather than being plain false explains why (3a,b) receive a reading stronger than the negation of a universal statement.

(5) a. The children sang.
     b. Mary knows which children sang.

As noted by Löbner (1985, 2000), homogeneity effects disappear when an overt universal adverb is added to the sentence: if some but not all children sang, (6a) is plain false and (7a) is clearly true. Again, we observe a similar effect with questions: if Mary knows for some but not all children who sang that they did, (6b) is plain false and (7b) is clearly true.

(6) a. The children all sang.
     b. Mary completely knows which children sang.
(7) a. The children didn't all sing.
     b. Mary doesn't completely know which children sang.

The first intuition that questions are subject to a homogeneity constraint can be traced back to Krifka (1996). Gajewski (2005: 118) proposes to fully transpose the theory of definite plurals to embedded questions, conditionals, and other domains by generalizing the notion of plurality in Link (1983). George (2011: 109) independently observed the effect and, on a suggestion of Benjamin Spector, proposes to include a presupposition of homogeneity in the embedding rule. Finally, Križ (2015a) proposes a very different analysis, whereby the homogeneity effect of questions is simply the

result of the homogeneity of answers about plural individuals projecting through the embedding verb.

1.3 Plan

In the next section, I will quickly review the literature on plurality effects and questions. In Sect. 3 I will propose an implementation of plurality effects with questions. Sect. 4 will introduce the different exhaustive readings which have been proposed for questions and present Klinedinst and Rothschild's (2011) theory (henceforth K&R). In Sect. 5, I will show how K&R's ideas can be implemented in the theory of Sect. 3. In Sect. 6 I will apply the theory to examples discussed in the introduction. In Sect. 7 I discuss and justify the use of lexical restrictors, and in Sect. 8 I will review potentially problematic cases and show that they can in fact be accommodated with minor revisions of the theory.

2 Embedded questions as definite plurals in the literature

As briefly mentioned above, several proposals already link questions with plural nouns, and embedded questions with definite plurals. In this section, I will introduce the standard semantics for plural nouns and discuss the theories of questions which draw on it.

2.1 Plurals and definite descriptions

Link (1983) proposed a very influential account of plurality in natural languages, which I summarize here (leaving aside material fusion and the m-part relation, which are not relevant for *which*-questions in my approach). The essential assumption is that the domain of individuals, D_e, contains not only *atomic* individuals $(a, b, \dots)$, but also all the *plural* individuals which can be obtained by (individual) sum formation, such as $a \oplus b$. D_e is assumed to be closed under the sum operation $\oplus$. An order relation called i-part can then be defined on elements of D_e: $x \preceq y$ iff $x \oplus y = y$ (for instance, $[\![Ann]\!]$ is an individual part of $[\![Ann\ and\ Bill]\!]$). An individual is atomic if it has no proper i-part. Finally, we can define a $\star$-operator which represents the pluralization of predicates of atomic individuals (e.g., *girl*) into plural predicates (*girls*). For simplicity, I will often use the symbols '$\oplus$', '$\preceq$' and '$\star$' to represent both the elements of the object language and their translations.

(8) $[\![^\star P]\!](X) = 1$ iff $\forall y \preceq X, Atomic(y) \rightarrow [\![P]\!](y) = 1$
  $[\![girls]\!] = [\![^\star girl]\!]$: 'girls' is true of any plurality whose atoms are all girls.

The plural definite article *the* takes a plural predicate and returns its maximal element (the biggest plural individuals which satisfies the predicate):

(9) $\begin{aligned}[\![the\ ^\star P]\!] &= \sigma x.[\![^\star P]\!](x) \\ &= \iota x.\big[[\![^\star P]\!](x) \wedge \forall y([\![^\star P]\!](y) \rightarrow y \preceq x)\big]\end{aligned}$

While this is not frequent in the literature on plural definites nor very useful at this point, we can see *the* as an overt type-shifter from predicates ($\langle e, t \rangle$) to individuals (e).

2.2 Plurality in questions

Conjunction and material implication for propositions allow an analogy with sum formation and the i-part relation for individuals. The first proposal to apply Link's (1983) model of plurality to questions is Dayal (1996). In order to account for multiple *wh*-questions and the unicity presupposition of singular *which*-questions, she argued that we must distinguish singular *which*-questions from plural-marked *which*-questions; she proposed to do so by translating the plural marking on phrases like 'which child**en**'. However, she did not consider the implications for the plurality effects presented in Sect. 1, except for a short discussion of sentences similar to (2b).

In (10), we see more concretely how the plural morphology on a *which*-phrase can be translated within a Hamblin (1973) style denotation for questions. The correspondences in (11) show that the question abstract—the function which takes an element x in D_e and returns the proposition $\lambda w.\llbracket {}^\star sang \rrbracket^w(x)$—is a semi-lattice homomorphism from $\langle D_e, \succeq, \oplus \rangle$ to $\langle Q, \rightarrow, \wedge \rangle$.[2]

(10) $Q = \llbracket \text{Which children sang} \rrbracket$ $=$ $\lambda p.\exists X \big[\llbracket {}^\star child \rrbracket(X) \wedge p = \lambda w.$
$\llbracket {}^\star sang \rrbracket(w)(X) \big]$
$=$ $\{$Mary sang, Peter sang,
Peter and Mary sang…$\}$

(11)   a. $\llbracket \text{Peter} \oplus \text{Mary sang} \rrbracket \equiv \llbracket \text{Peter sang} \rrbracket \wedge \llbracket \text{Mary sang} \rrbracket$
     b. $p \oplus m$ $\succeq$ p
$\llbracket \text{Peter} \oplus \text{Mary sang} \rrbracket$ $\rightarrow$ $\llbracket \text{Peter sang} \rrbracket$

Lahiri (2002) introduced a sophisticated theory of questions designed to explain plurality effects (although he does not discuss homogeneity). In his theory, an answer operator maps questions onto *Proposition Conjunction Algebras*, which have a structure similar to that of a starred predicate in Link's (1983) theory that allows the derivation of QVE and semi-distributive readings.

Finally, Gajewski (2005) proposed to generalize Link's (1983) theory to propositions by having a domain of plural propositions $D_{\{\langle s,t \rangle\}}$, but this proposal is not made fully explicit. In particular, we will discuss an important difference between propositions and individuals which is overlooked in Gajewski (2005), namely that the notion of atomic proposition does not make sense outside proposition conjunction algebras.

2.3 Answer operators

Most theories of embedded questions make use of an answer operator to combine questions with declarative-embedding predicates such as *know*.[3] Examples from a few influential theories are presented in (12).

---

[2] These correspondences are possible only because ${}^\star sang$ is a distributive predicate. If the *which*-phrase was associated with a collective predicate, the correspondences would not necessarily hold.

[3] A famous exception is Groenendijk and Stokhof (1982, 1984), who treat the extension of a question as a proposition, and recent work in inquisitive semantics (Ciardelli et al. 2013) which builds on Groenendijk and Stokhof's intuition that questions and propositions should share the same logical type.

(12)   a. Heim (1994): $ans_1(Q, w) = \cap(Q(w))$
       b. Dayal (1996): $Ans(Q, w) = \iota p.[p \in Q \land p(w) \land \forall p' \in Q, [p'(w) \rightarrow p \subset p']]$
       c. Beck and Rullmann (1999): $answer_1(w)(Q) = \cap\{p : Q(w)(p) \land p(w)\}$

Beyond their differences, these operators have a few common properties:

– They restrict the set of answers to true answers (if this is not done already in the denotation of the question, as proposed by Karttunen 1977).
– They type-shift the question from a set of propositions (or a higher type for multiple-*wh*-questions) to a proposition.
– They usually do so either by selecting the most informative proposition among the true answers or by constructing it (taking the conjunction of all true answers).

Some answer operators are clearly transpositions of the definite article *the* defined in (9) and extend the analogy between questions and plurals. This is particularly visible in the case of Dayal (1996), who first used the 'ι' operator to define *Ans*, and it is explicit in Gajewski (2005) and Fox (2012); the latter defines *Ans* as "*The True*", building on von Fintel et al. (2014).

3 A possible implementation

3.1 First attempt: Hamblin-style approach

Building on Hamblin (1973), Dayal (1996), Lahiri (2002), and Fox (2012), we can use the question denotation in (13). I refer to Kotek (2015) for a compositional derivation. I will assume that this question denotation combines with the answer operator in (14), which takes as a second argument a set of propositions C.

(13) $[\![\text{Which children sang}]\!] = \lambda p.\exists X[^\star child(X) \land p = \lambda w.^\star sang(w)(X)]$

(14) $[\![Ans]\!](Q)(C) = \iota p.[Q(p) \land C(p) \land [\forall p', Q(p') \land C(p') \rightarrow p \subseteq p']]$

Concretely, C will be a restrictor provided by the embedding verb and $Ans(Q)(C)$ can be understood as 'the C p in Q'. In this, I am following Lahiri (2002). In Sect. 7, I will provide more arguments to support the assumption that C is determined by the embedding verb. For veridical responsive verbs such as *know*, C is simply the set of all true propositions in the world of evaluation, and we recover Fox's answer operator: $Ans(Q)(C)$ = 'the true Q'.

For pair-list readings of multiple-*wh* questions, which denote *families of questions*, we would need to recursively define a generalized answer operator compatible with questions of arbitrarily higher types $\langle \sigma, t \rangle$ (Dayal 1996; Fox 2012; Kotek 2014, 2015):

(15) $[\![Ans]\!](Q_{\langle \sigma, t \rangle})(C) = \lambda w.\left[\forall P \in Q, [\![Ans]\!](P)(C)(w)\right]$

The last thing we need is a definition of 'atomic answers', similar to the definitions of 'atomic individuals' in (16), for a set or a plural individual.

(16) $At(P_{et}) = \lambda x.[P(x) \land \forall y.[(P(y) \land y \preceq x) \rightarrow y = x]]$
       $At(X_e) = At(\lambda x_e.x \preceq X)$

Although we showed that a strong analogy can be drawn between definite descriptions and questions, it is not perfect. Indeed, if we follow Link (1983) in assuming that D_e is an atomic Boolean algebra, we can always retrieve the atoms from a plural individual. However, we cannot do so from an arbitrary proposition, because there is no such thing as an 'atomic proposition' in general (unlike D_e, D_{st} is not an atomic Boolean algebra). Denotations of typical questions do form atomic algebras on D_{st}, so atomic answers can be defined but only relative to a Hamblin-set. For a concrete example, note that (17) below would be an atomic answer to the question (18a) but not to the question (18b).

(17)   Ann and Bill talked to John.

(18)   a.   Who did Ann and Bill talk to?
       b.   Who talked to John?

This means that we can define atoms for a Hamblin-set Q, but not for $Ans(Q)(C)$.

3.2 A solution based on categorial approaches

Xiang (2016, 2018) proposes a hybrid approach between the propositional and categorial approaches to questions. An important motivation for this move is to propose a new account of QVE, addressing some challenges to Lahiri (2002) raised in Williams (2000) which are not unrelated to the atomicity issue.[4] I will adopt an account close to hers, designed to derive cumulative and homogeneity effects as well.

For the purpose of the present proposal, Xiang's theory can be summarized as follows (I am ignoring complexities introduced to account for mention-some questions):

– The root denotation of a question is a topical property, as in (19) below.
– Two type-flexible answerhood operators are defined on these topical properties: Ans^S returns a short answer (in the previous case, a plurality of children who sang), and Ans returns a propositional answer (the proposition that the children denoted by the short answer sang).
– Question embedding under responsive predicates uses the propositional Ans, but the domain of quantification for QVE is provided by Ans^S.

Let me define answerhood operators inspired by Xiang (2016) for a topical property such as (19). Note that there is no need to give a recursive definition for multiple-*wh* questions, since everything is already built into the fact that Q can be of any type $\langle \tau, st \rangle$. I will use the underline notation to mark presuppositions.

(19) $[\![\text{Which children sang}]\!] = \lambda x.\underline{[\![{}^\star child]\!](x)} : [\![{}^\star sang]\!](x)$

---

[4] Williams points out that examples with collective predicates where only one true answer (about a plural individual) is true, such as (i) below, still allow QVE readings where the quantification is over atomic individuals which are part of the answer. (ii) is a similar example discussed in Lahiri (2002: 216).

(i)   For the most part, Al knows which soldiers surrounded the fort.

(ii)   For the most part, Sarah knows which students form the bassoon quintet.

His solution is to adopt a denotation for *which* such that 'which students P' really means 'which students are part of a plurality y such that $P(y)$', thereby reconstructing a distributive predicate.

(20) Answerhood operators:

    a. $[\![Ans]\!](Q)(C) = \sigma\, p\, [p \in Range(Q) \cap C]$

    b. $[\![Ans^S]\!](Q)(C) = \iota\alpha\big[Q(\alpha) \in C \wedge \forall\beta[Q(\beta) \in C \to (Q(\alpha) \subseteq Q(\beta))]\big]$

Given the question denotation (and this is crucial), the two answer operators are inter-definable, as shown in (21). In what follows, I will exclusively use Ans^S, which means that Ans will need to be derived from Ans^S and the question.

(21) a. $[\![Ans]\!](Q)(C) = Q([\![Ans^S]\!](Q)(C))$

    b. $[\![Ans^S]\!](Q)(C) = \iota\alpha[Q(\alpha) = [\![Ans]\!](Q)(C)]$

Xiang proposes that the QV sentence (1b) receive an interpretation along the lines of (22).

(1b) John mostly knows which children sang.

(22) $[\![(1b)]\!]^w = most\ y[y \in At(Ans^S(Q)(C^w_{know}))]$
$\big[K\big(\lambda w'.[y \preceq Ans^S(Q)(C^{w'}_{know})], j, w\big)\big]$

She does not offer a derivation of this interpretation, so our first task will be to fill in the missing details. The main difficulty is that two copies of the short answer to Q are needed: one to provide the restrictor of *most*, the other to define a mapping from y_e to the proposition that y is part of the answer. Note that the function $\lambda y.\lambda w.[y \preceq Ans^S(Q)(C^w_{know})]$ is usually identical to the topical property denoted by Q, except in Williams's (2000) examples, where an individual can be part of the true complete short answer without being a true answer.

The simplest solution may be to adapt Cooper's (1979) analysis of e-type pronouns, whereby *it* is understood as THE $R_j(x_i)$ with x_i a (usually bound) variable and R_j a contextually salient relation (of type $\langle e, et \rangle$). Here we would assume that a silent variable f is responsible for the most embedded occurrence of Q. It would be of the same type as the question (usually $\langle e, st \rangle$) and would naturally pick the function described above as its value (or for simplicity, the topical property denoted by the question).[5] We can then assume that the basic structure is as in (23).

---

[5] That this antecedent for f is contextually available seems to be required for independent reasons. For instance, Aloni et al. (2007) argue that this is the mechanism that allows short answers to be interpreted after matrix questions. Note that embedded questions seem to have the same effect:

  (i)   Mary knows which children sang: John and Sue.

  (ii) Mary knows which children sang. They wore blue hats.

  (iii) Which children form the bassoon trio? Ann and Bill for sure, and maybe Chris.

However, as shown by (ii), the question also makes the short answer a potential antecedent for subsequent pronouns, which means that an analysis where the short answer in the restrictor of *most* would be given by some silent pronoun may work as well. (iii) shows that the correct function for the interpretation of short answers seems to be more complex than the topical property denoted by the question, in line with Williams' examples.

(23) Mary knows which children sang.
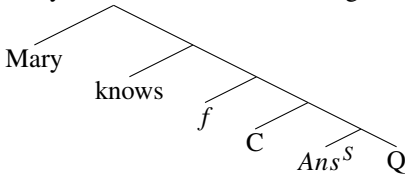
However, two issues arise with this structure. First, assuming that Principle C holds for this kind of variable, f cannot be coreferential with Q because it c-commands it. Nevertheless, I will soon assume that f receives a value which is more complex than Q, avoiding this issue. More problematically, if the restrictor C stayed in the scope of f, it would be bound by the world variable introduced by f, and (at least for veridical verbs) this would make the output of f a tautology. As a consequence, as we will see in a moment, the constituent $[C[Ans^S Q]]$ moves above the embedding verb.

3.3 Application

We now have the basic structure to offer an account of plurality effects with questions. All we need is a proper interpretation for the adverbs of quantity, and the mechanisms at the source of cumulativity and homogeneity. The derivation of QVE will take inspiration from Lahiri (2002), but will result in the interpretation suggested by Xiang (2016). Cumulativity and homogeneity will derive naturally from the structure I am about to propose, since it will introduce a relation between arguments of type e. More specifically, I will adopt the approach of Beck and Sauerland (2000), because they allow the **-operator to apply to complex expressions (see Champollion 2017 for a review of the different approaches). Križ (2015b) discusses how homogeneity is predicted for embedded questions under the assumption that the propositional answers they denote are homogeneous. In the present account this will be even more straightforward, since everything can be reduced to a relation between plural individuals. For homogeneity, I will adopt the approach of Križ (2015a), which is the most detailed account currently available (in particular, it offers a discussion of cumulative relations and collective predicates).

3.3.1 Quantificational variability

An important idea from Lahiri (2002) is that questions embedded under responsive predicates can undergo movement to the restrictor position of an adverb (overt or covert), and must do so when the adverb is overt. This movement, modeled after Quantifier Raising (QR), is called *Interrogative Raising* (IR). I will stick to the idea that QVE involves IR, albeit to a position lower than what Lahiri (2002) assumed. My main motivation for postulating a shorter move comes from the example in (24). The default reading predicted by Lahiri's theory, which is paraphrased in (24), would have the question take scope over the subject of the sentence. This reading does not seem

to be available, and so, following Preuss (2001: 149), I will assume that adverbs of quantity adjoin to VP rather than IP, and therefore IR targets Spec,VP.[6,7]

(24) For the most part, a professor knows which student cheated.
 $\neq$ For most students who cheated, some professor or other knows that they did.

Diverging a bit more from Lahiri's account, I will assume that it's the short answer $Ans^S(Q)(C)$ that is moving, leaving behind a trace of type e rather than st. As discussed earlier, a variable f denoting a function of type $\langle e, st \rangle$ must be present to convert short answers into propositions. The resulting structure is presented in (25b) for sentence (1b), repeated below.

(1)   a. The children mostly sang.
      b. John mostly knows which children sang.

(25) a. Structure for (1a):
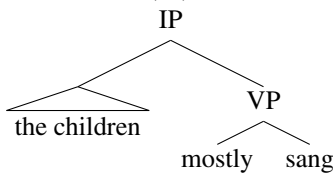
      b. Structure for (1b):
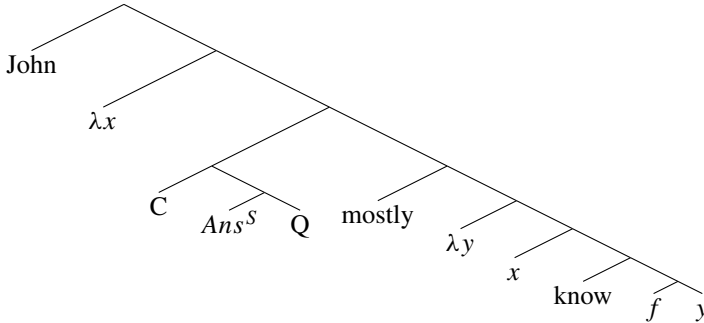
In order for λy to end up *below* the adverb in (25b), I need to assume something along the line of Lahiri's rule AB (see Lahiri 2002: 85). A possible way to ground Lahiri's rule would be to adopt Sauerland's (1998a) hypothesis that movement generates two new nodes: starting with the configuration $[_{YP} \ldots XP \ldots]$, moving XP to YP yields $[_{YP''} XP [_{YP'} \lambda x [_{YP} \ldots x \ldots]]]$. As we will see in the next section, Sauerland uses this hypothesis to generate binary relations by moving a second element from YP between XP and λx, which requires targeting YP′ instead of YP″ (I will adopt this

---

[6] Lahiri (2002: 76) assumes that IR can adjoin the interrogative complement to IP or VP, but the interpretation derived from VP-adjunction is very different (this is what he calls the focus-affected reading).

[7] Preuss's motivation for this move was to make adverbial binding sensitive to the embedding predicate, because her theory assumed that the embedding verb determines the strength of exhaustivity (as first suggested by Heim 1994). As will become clear in Sect. 5, I will treat exhaustivity as largely independent from the embedding verb, but my theory still assumes that the embedding verb plays an important role in the derivation of QVE by fixing the value of the restrictor C.

assumption when dealing with cumulative readings in the next section). A very similar idea was independently proposed in Barker (2007) to account for the "parasitic scope" of adjectives like *same*. We could imagine that floating adverbs of quantity likewise can attach to YP′ (though in this case it is unlikely to happen through movement from a lower position).

Together with rule AB comes the assumption that the adverb combines with its scope before its restrictor. In (26), I present my semantics for *mostly*. Importantly, the denotation contains a free parameter α. At this point, we can simply assume that $\alpha = \frac{1}{2}$ (Hackl 2009), but in the next section I will treat it as a vague threshold. I also assume that *mostly* presupposes that its restrictor isn't atomic (i.e., is a plurality of at least two individuals).[8]

(26) $[\![\text{mostly}]\!] = \lambda P_{\langle e,t \rangle}.\lambda X_e : |At(X)| \geq 2.\big[|At(X) \cap P| > \alpha \times |At(X)|\big]$
      $|\cdot|$ is the cardinality function, which simply takes a set and returns the number of elements in it.

Putting everything together, we get the denotations in (27a,b) for (1a,b), which correspond to the usual meaning for these sentences (of course, in the case of embedded questions, this is the WE reading).

(27) a. $[\![1a]\!]^{w_0} = \big|\{x | [\![\text{child}]\!]^{w_0}(x) \wedge [\![\text{sang}]\!]^{w_0}(x)\}\big| > \alpha \times |[\![\text{child}]\!]^{w_0}|$
          $\approx$ The proportion of children who sang is more than α.
      b. $[\![1b]\!]^{w_0} = \big|\{y | K(Q(y))(j) \wedge y \in At(Ans^S(Q)(C))\}\big|$
          $\approx > \alpha \times |At(Ans^S(Q)(C))|$
          $= \big|\{y | K(\lambda w.[\![\text{sang}]\!]^w(y))(j) \wedge [\![\text{sang}]\!]^{w_0}(y) \wedge [\![\text{child}]\!]^{w_0}(y)\}\big|$
          $> \alpha \times |\{y | [\![\text{sang}]\!]^{w_0}(y) \wedge [\![\text{child}]\!]^{w_0}(y)\}|$
          $\approx$ Among children who sang, the proportion of those for whom John knows that they sang is more than α.

3.3.2 Cumulative readings

Cumulative readings arise when a transitive verb takes two pluralities as its arguments (Scha 1981). In the usual case, the two pluralities are simply plural definite descriptions (see (2a)). The same phenomenon occurs when some responsive verbs take a plural individual and a plural-marked question (see (2b)). Many different theories have been proposed for (2a) (Winter 2000; Beck and Sauerland 2000; Kratzer 2008; Champollion 2010, 2014; Winter and Scha 2015; a.o.). In what follows I will use a cumulation operator '**' (first introduced in Krifka 1986) and assume that this operator can apply to whole expressions, as argued by Beck and Sauerland (2000). Let me first give a definition of this cumulation operator (from Sternefeld 1998):

(28) Given a two-place predicate R defined on $E \times E$ where E is a join-semilattice, $^{**}R$ is the smallest two-place predicate such that:

---

[8] In practice, since *most* entails more than half, restrictors of cardinality 2 will be deviant as well, because they would make *mostly* vacuous (the only way to know more than half of a set of two is to know the whole set).

a. $\forall x, y \in E, Rxy \rightarrow {}^{\star\star}Rxy$

b. $\forall x_1, x_2, y_1, y_2 \in E, ({}^{\star\star}Rx_1y_1 \wedge {}^{\star\star}Rx_2y_2) \rightarrow {}^{\star\star}R(x_1 \oplus x_2)(y_1 \oplus y_2)$

The goal will be to apply this operation to the relation '$\lambda y.\lambda x.K(f(y))(x)$', which in our example simply translates as 'x knows that y sang'. This will require certain assumptions about the LF of the sentence, which will end up somewhat similar to my structure for QVE. Sauerland (1998a) and Beck and Sauerland (2000) argue that DP can QR to let a complex expression be closed by the cumulation operator, as in (29).

(29) a. The two women want to marry the two men.

     b. (the two women) (the two men) ${}^{\star\star}\lambda y\lambda x[x$ want to marry $y]$

Note that this is not straightforward: just like in the case of adverbs of quantity, (29b) requires 'the two men' to land between 'the two women' and the corresponding λx, so that the operator ** can apply to a relation. As discussed in the previous section, Sauerland (1998a) achieves this by assuming that QR generates two projections, and the derivation allows an insertion of ** at the intermediate site, between the moved DP and the λ. An important difference between the structure for QVE in (25b) and the structure for cumulative predication in (30) is that in the former IR targets the traditional position below the λx introduced by movement of the subject, and the adverb of quantity is then attached in the non-standard intermediate position between the moved question and its λy, while in (30) IR targets Sauerland's intermediate position between the moved subject and its λx to create the binary relation. Like the adverb of quantity, the cumulation operator is then inserted between the moved question and its λy.

(2b) The teachers know which children sang.

(30) Structure for (2b):

Let us see how this applies to example (2b), assuming that there are two teachers and that three children sang in the actual world:

(31) $[\![(2b)]\!] = {}^{\star\star}[\lambda y.\lambda x.K(f(y))(x)](Ans^S(Q)(C))(\sigma X^{\star}[\![\text{teacher}]\!](X))$

       $= {}^{\star\star}[\![\cdot \text{ know that } \cdot \text{ sang}]\!](c_1 \oplus c_2 \oplus c_3)(t_1 \oplus t_2)$ (schematically)

Given the definition for '**', we correctly derive that each teacher must know that at least one of the three children sang, and for each child who sang, at least one teacher must know that they did.

The only puzzle left at this point is the fact that some responsive verbs do not seem to allow cumulative readings. In fact, Lahiri (2002) argues that non-veridical predicates never do, possibly because their lexical restrictor is dependent on the subject, and that this creates problems when the subject is plural.

(32) Lahiri (2002): John is certain about Q = John is certain about every answer to Q that he considers possible.
More formally, if Dox_j^w is the set of worlds compatible with John's beliefs in w:
$$C_j^w = \lambda p_{st}.(Dox_j^w \cap p \neq \emptyset)$$

(33) The teachers were certain about which children sang.

    a. *Available distributive reading:* Each teacher was certain about which children sang (but different teachers may disagree with each other).

    b. *Unavailable cumulative reading:* For each child that some/all teachers considered a potential singer, at least one teacher was certain that she sang, and each teacher was certain of at least one child that she sang. (compatible with each teacher being uncertain about some children)

I will leave it as an open empirical issue whether Lahiri's generalization really holds of all non-veridical predicates. An important confound is that non-veridical predicates tend not to be lexicalized as simple verbs (e.g., *be certain*), and even when they do (e.g., *agree*), seem to require a preposition when embedding questions. I have followed Beck and Sauerland (2000) in assuming that cumulation can apply to arbitrarily complex expressions, but deriving the expression which is cumulated requires QR and, depending on how the interaction between *be certain*, the preposition *about*, and the question is analyzed, QR might be blocked.

Several other explanations as to why cumulative readings are seemingly unavailable for cases like (32) are possible. First, as suggested by Lahiri (2002), it could simply be that there is no conventional way to derive the common restrictor (union, intersection?). Second, we could assume that it is in principle possible to build the restrictor, but that the complex dependency makes this too demanding cognitively, hence this reading is strongly dispreferred.

The constraint seems indeed to be related to the difficulty of extracting a restrictor or processing the complex relation, because the reading seems to become available again when there is a natural pairing between attitude holders and (short) answers, as illustrated by the example below (B. R. George, p.c.). In scenario (34), the sentence (35) is false under a distributive reading, yet it sounds like a correct description of the situation.

(34) Scenario:
Each expert is an expert on a different genus, and she is certain about which of the species in her genus are capable of asexual reproduction. Suppose further that every genus has an expert associated with it, and that no two experts are opinionated about the same species (each one explicitly reserves judgment and defers to the appropriate other expert).

(35) The experts are certain about which of the species are capable of asexual reproduction.

3.3.3 Homogeneity

Since my theory of question embedding amounts to applying a predicate to a short answer which is typically a plural individual, standard theories of homogeneity straightforwardly predict homogeneity effects for embedded questions. For concreteness, I will adopt Križ's (2015a) proposal.

Križ's theory of homogeneity: The proposal is rather technical, but the essential ingredients are (a) a trivalent logic where # represents homogeneity violations and (b) the assumption that D_t comes with the partial order in (36).

(36) $D_t = \{0, 1, \#\}$. For all $x \in \{0, 1, \#\}$, $x \preceq x$ and $x \preceq \#$.

D_e is ordered by the usual i-part relation introduced in Sect. 2.1, and—assuming that all types end in e or t—(37) shows how we can define an order inductively on arbitrary types. As a first approximation, homogeneity can then be expressed as a simple constraint, namely that the domains be restricted to functions which are monotonic with respect to the partial order on their domain and range. (38) sketches out how this would work.

(37)   For all $f, g \in B^A$, $f \preceq g$ iff for all $x \in A$, $f(x) \preceq g(x)$.

(38)   Let $a, b \in D_e$ be atomic individuals, $P \in D_{et}$ a predicate.
P is monotonic and $a, b \preceq a \oplus b$, so $P(a), P(b) \preceq P(a \oplus b)$.
If $P(a \oplus b) = 1$, $P(a), P(b) \preceq 1$. Since $0 \npreceq 1$ and $\# \npreceq 1$, it follows that $P(a) = P(b) = 1$. The same reasoning tells us that if $P(a \oplus b) = 0$, $P(a) = P(b) = 0$. In every other case, $P(a \oplus b) = \#$.

From (38), we can conclude that if $P(a) \neq P(b)$, $P(a \oplus b) = \#$. In other words, P can only be true (resp. false) of a plurality if it is true (resp. false) of each of the atoms in this plurality. A further requirement that the possible interpretation of constants, relation symbols, and determiners be atomic in their respective domains makes all predicates distributive.

In order to allow collective predicates, the monotonicity and minimality constraint have to be weakened. Note in passing that cumulative relations are not homogeneous, but this is not a problem if homogeneity is checked before applying the **-operator. I refer to Križ (2015a: chapter 2) for details, and in particular his sections 2.2.2 and 2.4.1 for these two questions.

Is *mostly* homogeneous?   Before going any further, I should point out that this definition of homogeneity does not fare well with my entry for *mostly* in (26), which takes a plurality as its restrictor. Indeed, the order which D_{et} inherits from the order on D_t is completely unrelated to the mereological order on D_e.[9] Nevertheless, we can

---

[9] Given $x \in D_e$, let P_x be the homogeneous characteristic function of the i-parts of x. For $z \in D_e$, P_x maps z to 1 if $z \preceq x$, 0 if x and z do not overlap, and # otherwise. Now, if $x \prec_e y$, we can show that P_x

give a denotation of type $\langle et, \langle et, t \rangle \rangle$ for *most* which simply states that 'Most A are B' is true if most A's are indeed B, false if they are mostly not B, and a gap otherwise. This denotation, given more formally in (39), without the non-atomicity presupposition for simplicity, is homogeneous.[10]

(39) $\llbracket \text{most} \rrbracket = \lambda A_{et}.\lambda B_{et}. \begin{cases} 1 & \text{iff } \dfrac{|\{a|AT(a) \wedge A(a) = 1 \wedge B(a) = 1\}|}{|\{a|AT(a) \wedge A(a) = 1\}|} > \alpha \\[2mm] 0 & \text{iff } \dfrac{|\{a|AT(a) \wedge A(a) = 1 \wedge B(a) = 0\}|}{|\{a|AT(a) \wedge A(a) = 1\}|} > \alpha \\[2mm] \# & \text{otherwise} \end{cases}$

This entry would fit well for sentences like *Most children sang*, where the restrictor of *most* is clearly a predicate. We can generalize it to *Most of the children sang* by assuming that the definite DP introduces a plural individual, but the preposition *of* converts it back to a predicate. For *mostly*, we can simply assume that this operation is done by the *-ly* suffix, so the denotation in (26) actually reflects a complex expression which is not itself homogeneous, but can be derived from a homogeneous quantifier and a preposition/suffix (which denotes a function that is not homogeneous but seems to be needed independently).

Note that the denotation (39) is actually close to what a trivalent theory of vagueness would assign for *most*. In particular, if $\alpha > 0.5$, the interval $[0, 1]$ can be split between clearly false situations $[0, 1-\alpha)$, borderline cases $[1-\alpha, \alpha]$, and clearly true situations $(\alpha, 1]$. Nevertheless, I will treat vagueness and homogeneity as two independent phenomena, and assume that vagueness arises not because there is a gap between $1-\alpha$ and α, but rather because there is uncertainty on the value of the free parameter α.

Application: We can now apply the proposal to (3b), repeated below, assuming the structure in (23).

(3b) Mary doesn't know which children sang.

(40) $\llbracket (3b) \rrbracket^{w_0} = \neg K(w_0)(f(Ans^S(Q)(C^{w_0})))(m)$

We obtain the meaning in (40), at which point we need to make some assumptions on the behavior of f and K. If three children sang in w_0, $Ans^S(Q)(C^{w_0})$ will simply return the mereological sum of these three children. f will map this plural individual to a proposition. A first approximation is to imagine that f simply resolves to Q, so that $f(Ans^S(Q)(C^{w_0})) = \lambda w.^{\star}(S)(w)(a \oplus b \oplus c)$. *Sing* is a simple distributive verb,

Footnote 9 continued

and P_y are generally not ordered in D_{et}: $P_x(y) = \# \not\preceq 1 = P_y(y)$, so $P_x \not\preceq P_y$. Conversely, given z such that $P_x(z) = 0$, $P_y((y-x) \oplus z) = \# \not\preceq 0 = P_x((y-x) \oplus z)$, so $P_y \not\preceq P_x$.

[10] **Proof:** Let us first show that if $A \preceq A'$, $most\,A \preceq most\,A'$. Let $B \in D_{et}$ be an arbitrary predicate; we want to show that $[\text{most } A \, B] \preceq [\text{most } A' \, B]$. Assume that $[\text{most } A' \, B] = 1$. Then an α proportion of atomic a's such that $A'(a) = 1$ satisfy B. Since $A' \succeq A$, these a satisfy $A(a) \preceq 1$, which necessarily means $A(a) = 1$. Hence $[\text{most } A \, B] = 1$. The very same reasoning can be done with 0 instead of 1, so $[\text{most } A \, B] \preceq [\text{most } A' \, B]$. Second, let's fix A and take $B \preceq B' \in D_{et}$. For all atomic a's such that $A(a) = 1$, $B'(a) \succeq B(a)$, so $B'(a) = 1 \rightarrow B(a) = 1$. If at least α of such a's are 1 for B' (i.e., if $[\text{most } A \, B']$ is 1), it follows that they are also 1 for B, so $[\text{most } A \, B]$ is 1. The same can be done for 0, so $[\text{most } A \, B] \preceq [\text{most } A \, B']$.

so this proposition will be trivalent (true if all three children sang in w, false if none did, # otherwise). As discussed in Križ (2015b), under the assumption that knowledge amounts to universal quantification over the worlds in the doxastic state of Mary, this trivalence will project as expected: (3b) will be true if all three children sang in all worlds of $Dox_m^{w_0}$, false if none of the children sang in at least one of these worlds (i.e., Mary considers it possible that none of them sang), and undefined otherwise.

To account for Williams's (2000) examples, we need to adopt a more complex value for f, which can be made homogeneous as well:

(41) $f = \lambda y.\lambda w.\begin{vmatrix} 1 & \text{iff } y \preceq \sigma z[Q(z)(w)] \\ 0 & \text{iff } y \text{ does not overlap with } \sigma z[Q(z)(w)] \\ \# & \text{otherwise} \end{vmatrix}$

3.4 Alternatives theories of plurality effects

The literature offers many theories of the various plurality effects discussed here. One may wonder whether I could have chosen different options. My approach to QVE, adapted from Xiang (2016), is quite different from the standard approach represented by Lahiri (2002). It has the advantage of offering an account of Williams's cases, although this feature is somewhat *ad hoc*. Because it relies on an answerhood operator, it still does not account for QVE with rogative predicates, as discussed in Beck and Sharvit (2002). At this point, the benefits in terms of empirical coverage compared with Lahiri's approach are not striking, but we avoid the central issue of having to define the atoms of a proposition, and in Sect. 5 we will see that the proposed account helps solve some problems with previous EXH-based approaches to exhaustivity.

My account of cumulative readings is rather standard. The crucial bit is that I need to assume, following Beck and Sauerland (2000), that the cumulative operation can apply to a complex expression containing the verb and the silent variable f.

Some other theories of homogeneity on the market may be compatible with the rest of my proposal. Löbner (1985, 2000) propose to treat homogeneity as a presupposition and Gajewski (2005) implements this idea in his theory of homogeneity with embedded questions. Such a proposal should in principle be compatible with the present theory as well, but its interactions with homogeneity removers, cumulative relations, and quantifiers are not fully worked out. Magri (2014) proposes to treat homogeneity as an implicature. But in addition to problems identified in Križ (2015a), this approach will not be compatible with the theory of exhaustivity proposed in Sect. 5 below.

Križ (2015b) proposes a theory of homogeneity in questions which is essentially the propositional version of my categorical approach. He shows that, provided some reasonable hypotheses on the projection of truth value gaps, the homogeneity of questions derives immediately from the fact that propositional answers which involve plural individuals are trivalent themselves. This theory also explains how an adverb *within* a question can remove homogeneity in languages like German or in English dialects which allow the "who all" construction (see Beck and Rullmann 1999 for a discussion of these examples). If the variable f picks Q as its value, this is explained in my version of his theory as well. If it picks (41), the trivalent

properties of Q are not preserved, therefore the homogeneity remover has no effect. This can be fixed by making the falsity conditions of the function weaker and sensitive to the falsity conditions of the propositions in the range of the question, as in (42).[11]

(42) $f = \lambda y.\lambda w.\begin{vmatrix} 1 & \text{iff } y \preceq \sigma z[Q(z)(w)] \\ 0 & \text{iff } Q(y) \text{ does not overlap with } Q(\sigma z[Q(z)(w)]) \\ \# & \text{otherwise} \end{vmatrix}$

4 Incorporating stronger exhaustivity

In this section, I will introduce some background on strongly and intermediate exhaustive readings, before implementing their derivation in the theory introduced in Sect. 3.

4.1 Different exhaustive readings of embedded questions

Different readings have been proposed for sentences with embedded questions such as (43). In the previous section, we focused on the Weakly Exhaustive (WE) reading, which corresponds to Karttunen's (1977) proposal. However, at least two other readings have been discussed in the literature: the Intermediate Exhaustive (IE) and Strongly Exhaustive (SE) readings. All three readings are paraphrased in (44).

(43)   John knows which children sang.

(44)   WE: For all children who sang, John knows that they did.
         IE: For all children who sang, John knows that they did,
               and he does not believe that any other child sang.
         SE: For all children who sang, John knows that they did,
               and he knows that no other child sang.

Both SE and IE readings were first discussed by Groenendijk and Stokhof (1982), although they quickly rejected the possibility of the IE reading. The IE reading has since been discussed several times, and theories focusing on the SE reading have sometimes assimilated it to the WE reading (Spector 2005; Spector and Egré 2015;

---

[11] **Proof:** The truth conditions are unaffected, so the desired behavior for QVE with collective predicates is maintained. Now, let's see how the effect of homogeneity removers is accounted for. Let's place ourselves in w_0, where two children, a and b, sang. The question 'which children sang' maps $a \oplus b$ to a proposition that is true in worlds where a and b sang, undefined in worlds where one of them did but not the other, and false elsewhere. When this proposition is embedded under 'Mary knows', the result is a proposition which is true when Mary knows the conjunction 'a sang and b sang', false when Mary doesn't know the disjunction 'a sang or b sang', and undefined otherwise. In particular, if one world w_a in Mary's doxastic state is such that only a sang in w_a, *Mary knows who sang* is undefined. In this case, $f(a \oplus b) = Q(a \oplus b)$. Now, if we add an homogeneity remover in Q, it will map $a \oplus b$ to a proposition which is not trivalent. In particular, this proposition will be plain false in w_a. This means that although $a \oplus b$ overlaps with $a = \sigma z[Q(z)(w_a)]$, $Q(a \oplus b)$ does not overlap with $Q(a)$.

Theiler et al. 2018) while theories focusing on the WE reading have associated it with the SE reading (Berman 1991; Preuss 2001). Cremers and Chemla (2016) established its existence as a reading independent from both the WE and the SE readings.

Klinedinst and Rothschild (2011) show it is possible to derive IE and SE readings from the WE reading for non-factive predicates, using an EXH operator and some well-chosen alternatives (for an early attempt at deriving these readings as implicatures, see Berman (1991: 180). The view that stronger exhaustivity is the result of a process independent from question embedding *per se* will make it easier to implement it on top of the structured denotations we introduced for plurality effects; this view has also received support from acquisition data in Cremers et al. (2017). I will therefore follow K&R for the implementation of the IE and SE readings. In the next section I present their theory succinctly before introducing my own implementation.

Before going any further, a quick note is in order about the nature of the different readings in (44). Following K&R, I am going to assume that all three of them are derived in the grammar. Guerzoni and Sharvit (2007) famously argued that the WE/SE contrast correlates with the licensing of weak NPIs, thereby offering evidence that these two readings are distinctly encoded in the grammar, rather than one just being a special case of the other. Unfortunately, there does not seem to be any direct evidence for the assumption that the IE reading is indeed derived in the grammar. There is however evidence that the IE reading is derived via strengthening of the WE reading through implicatures. To the extent that the derivation of implicatures can be seen as a grammatical phenomenon (as argued in Chierchia 2004; Chierchia et al. 2012), this would imply that the IE reading is grammatical, but I am not particularly attached to this conclusion.

4.2 Klinedinst and Rothschild (2011)

K&R focus on veridical non-factive predicates, such as *predict* (although see Spector and Egré 2015 for arguments against the very existence of predicates which would be veridical when embedding interrogatives but not declaratives). They propose the readings in (46) for sentence (45), which are supported by a survey of their own and experimental results from Cremers and Chemla (2016).

(45)   John predicted which children sang.

(46)   WE: For all children who sang, John predicted that they did.
       IE: For all children who sang, John predicted that they did,
                    …and he did not predict that any other child sang.
       SE: For all children who sang, John predicted that they did,
                    …and he predicted that no other child sang.

The core idea of K&R is to use a set of alternatives to the embedded question, from which they derive as implicatures the second lines in the paraphrases of the IE and SE readings in (46). Their proposal assumes that the denotation of the question in a world w is simply its Karttunen-answer in this world, as defined in (47). They further assume that the focus value of the question (the set of alternatives) is obtained by varying the world variable, as in (48). Crucially, the focus value $Alt(Q(w))$ contains false answers.

To compute the implicatures, they use Groenendijk and Stokhof's (1984) operator EXH as defined in (49), and allow it to apply directly to questions (whereas it was originally designed to apply to sentential answers to the question only). For the basic cases it is sufficient to assume that $[\![\varphi]\!]_{F+}$ is the set of alternatives which are stronger than the prejacent (this assumption will be modified in the next section).

(47) $Q(w)$ is the true (weak) complete answer in w.
       Example: $Q(w) = \lambda w'.[\![$Ann sang and Bill sang$]\!]^{w'}$

(48)   Any potential complete answer to the question is an alternative:
       $Alt(Q(w)) = \{Q(w')|w' \in W\}$
       Example: $Alt(Q(w)) = \{$Ann sang, Bill sang, Chris sang, . . . A, B, and C sang, . . . $\}$

(49) $[\![$EXH $\varphi]\!] = [\![\varphi]\!] \wedge \left(\bigwedge_{p \in [\![\varphi]\!]_{F+}} \neg p \right)$

These simple hypotheses are sufficient to derive the three readings of (45). As an example, let us see how this works in a situation w_0 where out of three children, Ann and Bill sang but Chris did not.

– The WE reading is simply the result of combining $Q(w_0)$ with the verb:

$$[\![(45)]\!]^{w_0} = \text{predict}'(j)(Q(w_0)) = \text{predict}'(j)(Sa \wedge Sb)$$

– The IE reading is derived through global (matrix) exhaustification:

$$[\![\text{EXH}(45)]\!]^{w_0} = \text{predict}'(j)(Sa \wedge Sb) \wedge \neg\text{predict}'(j)(Sa \wedge Sb \wedge Sc)$$
$$= \text{predict}'(j)(Sa \wedge Sb) \wedge \neg\text{predict}'(j)(Sc)$$

– The SE reading is derived through local exhaustification of the embedded question:

$$[\![\text{John predicted EXH[which children sang]}]\!]$$
$$= \text{predict}'(j)\big((Sa \wedge Sb) \wedge \neg(Sa \wedge Sb \wedge Sc)\big)$$
$$= \text{predict}'(j)\big(Sa \wedge Sb \wedge \neg Sc\big)$$

K&R's proposal can thus capture the WE, IE, and SE readings of non-factive veridical verbs. It can be extended to factive verbs by adopting a two-dimensional view of meaning with independent assertive and presuppositional contents (see Spector and Egré 2015; Uegaki 2015a), but I will pursue a different route. In order to capture plurality effects, we need more structure than K&R propose. The goal of the next section will be to retrieve their focus values for questions from the denotation introduced in Sect. 3 and give a more precise characterization of the EXH-operator.

5 An exhaustification theory for plural questions

The goal of this section is to update the theory of plurality effects proposed in Sect. 3 so that it can account for strong and intermediate exhaustivity. To achieve this goal,

we will need to retrieve the alternatives of K&R from the richer structure required for plurality effects.

5.1 Hypotheses

5.1.1 Generating alternatives for answers

Unlike previous work, I will not need to postulate any special focus value for the embedded question itself. The short answer being of type e (for simplex questions), we can simply assume that it has D_e as its set of alternatives. I will further assume that after moving to a higher position, the trace left by the question can also trigger alternatives. Selkirk (1995) shows that the trace of a moved focused constituent can bear focus marking too. Indeed, in example (50), the whole VP 'review t_i' is new, but it bears no stress. Selkirk assumes that the trace inherits the focus marking from its antecedent, which is stressed (see Sauerland 1998b for further discussion).

(50) Bill read an article, but [which $BOOK_F$]$_i$ did Mary [review t_i]$_F$?

The silent focus-marked trace can then trigger alternatives which will be exhaustified within the scope of its λ-binder, just as in the case of a bound pronoun (Mayr 2012). Variables of type e are simply assumed to have D_e as their focus value.[12] This means that the node $f(y)$ in (30) can be exhaustified (as we will see, this is how SE readings can be derived). To avoid over-generating implicatures, we should also add to the value of f in (42) the presupposition that its argument is in the domain of the question (so if Q is 'which children sang', $f(z)$ should be a presupposition failure when z is assigned to an individual that is not a child).

I assume that alternatives project by pointwise function application (Rooth 1985) and combination with alternatives from any other source. In particular, I put no *a priori* restriction on multiple replacements.

5.1.2 Exhaustivity operator

I will use a version of Fox's (2007) EXH operator to derive (secondary) implicatures. Fox's notion of "innocently excludable" alternatives allows more flexibility

---

[12] Using doubly-indexed variables, we can build a formal system that lets bound variables introduce alternatives such that all alternatives introduced below a λ-binder can project beyond it, except for the alternatives that were introduced by the variable. Let us start with a set of assignments $\Gamma = \{\gamma : \mathbb{N}^2 \mapsto D_e\}$. Variables bear two indices u_{ij}, but are interpreted as expected: $[\![u_{ij}]\!]^{\gamma,w} = \gamma(i, j)$. However, we'll assume that a trace or a pronoun carries a single index i and is translated as u_{ii}. The focus value introduces alternatives using the second index: $[\![u_{ij}]\!]_f = \{[\![u_{ik}]\!] | k \in \mathbb{N}\}$. λ-abstraction, on the other hand, uses only the first index: $[\![\lambda u_i \varphi]\!]^{\gamma,w}$ is a function that maps any x of the same type as u_i to $[\![\varphi]\!]^{\gamma[i \mapsto x],w}$ where $\gamma[i \mapsto x]$ is like γ except that $\gamma[i \mapsto x](i, j) = x$ for any j. In short, λ-binding ignores the second index. This allows us to define a focus value $[\![\lambda u_i \varphi]\!]_f$ as in Rooth (1985): $[\![\lambda u_i \varphi]\!]_f = \{(\gamma, w) \mapsto [x \mapsto h(\gamma[i \mapsto x], w)] | h \in [\![\varphi]\!]_f\}$. Playing on the two indices ensures that all alternatives introduced by the variable are "flattened" when this variable is bound, while alternatives introduced by other elements in the scope of λu_i are preserved. Note that if we only look at ordinary values, everything can be reduced to standard assignments g such that $\exists \gamma \in \Gamma : \forall i, g(i) = \gamma(i, i)$.

than Groenendijk and Stokhof's (1984) operator, and this will be useful to account for interactions with other phenomena. For one, as discussed at length in the previous section, it is not clear how homogeneity interacts with exhaustivity. My assumption here will be that homogeneity of the prejacent simply projects out of EXH, but the negation of alternatives is a "weak" negation in the sense that it only states that they are not super-true. Since most question-embedding verbs come with presuppositions, I will need to specify how this operator interacts with presuppositions, which I model with partial functions (Heim and Kratzer 1998).[13, 14] Furthermore, my treatment of QVE will rely on the assumption that *most* is vague, and vagueness will also play a role in my account of the verb *surprise*. I will thus need to specify how EXH interacts with vagueness; I will follow Leffel et al. (2018) on this point.

In short, I will first assume that EXH involves a weak negation: alternatives need not be false but simply untrue (e.g., they can be presupposition failures). Second, I will require from negated alternatives not just that they be consistent with the prejacent, but also that they avoid all sorts of infelicities associated with the above-mentioned phenomena. Finally, there is one issue with K&R's approach, pointed out in Uegaki (2015a: 92), that needs to be resolved. As it stands, the prejacent for the matrix exhaustification of *John knows who sang* is the proposition that is true in every world w if John knows in w the K-answer in w. For K&R's account to make the correct prediction, the exhaustifier would have to abstract over the w argument of the verb but not that of the question. Otherwise, the prejacent would be independent from every alternative. Depending on our assumptions this would make the exhaustification vacuous or contradictory. As a consequence, Uegaki (2015a) proposes that the operator responsible for question exhaustivity is more like a quantifier than the exhaustification operator responsible for implicatures.

Mayr (2017) proposes an alternative solution stemming from the observation that the relevant notion of "strength" when comparing alternatives is not classical entailment but Strawson-entailment (independent support for this idea can be found in Sharvit 2017). If the verb is factive, alternatives may become Strawson-entailed (and this is the case with *know* in particular). Nevertheless, problematic situations arise because Strawson-entailment is not transitive and this can result in the negation of alternatives which are contextually entailed (leading to contradictions). The particular configuration that will concern us here is presented in (51).

---

[13] Gajewski and Sharvit (2012) and Spector and Sudo (2017) discuss the projection of presuppositions out of EXH. In the system I will present, Spector and Sudo's (2017) view of EXH as a presupposition hole would predict that 'which children sang' embedded under a factive verb systematically presupposes that all children sang, or that everyone is a child. My definition of EXH is therefore closer to what they call EXH[3] (it differs from EXH[3] in that I use a stronger notion of consistency). I leave for future research the possibility of a unified account.

[14] I modeled homogeneity using a trivalent logic, but I did not assume the "undefined" value # to represent presupposition failure. Križ (2015a) argues that homogeneity does not project like a presupposition (see Križ and Chemla 2015 for experimental evidence). When discussing the verb *forget* I will need to discuss potential interactions between homogeneity and presupposition (a topic that has not been explored, as far as I know), but this particular case can be handled with the simple extra assumption in (72) below.

(51)

p = Mary knows who sang
q = Mary believes that Ann sang
r = Mary knows that Ann sang
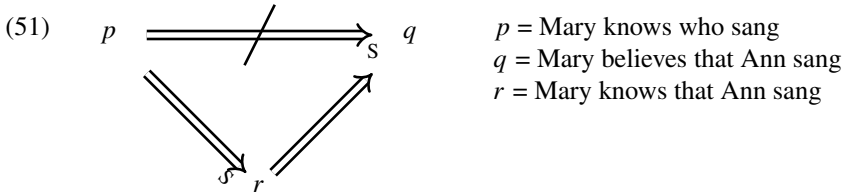
Let p be our prejacent. q would normally be negated by EXH since it is neither entailed nor Strawson-entailed. However, because r entails q, $q \neq 1$ implies $r \neq 1$. Therefore, negating the non-weaker alternative q leads to negating the Strawson-entailed alternative r. If the context satisfies the presuppositions of r, then p contextually entails r, and negating r leads to a contradiction. The solution is simply to update our notion of *innocently excludable* alternatives to make q non-excludable in such configurations. The third line in (53) takes care of this.

(52) $[\![\text{EXH } \Phi]\!]^w = 1$ iff $\left([\![\Phi]\!]^w = 1 \text{ and } \forall P \in \mathbb{IE}(\Phi, [\![\Phi]\!]_f, w), P(w) \neq 1 \right)$

(53) $\mathbb{IE}(\Phi, A, w)$
$$= \bigcap \left\{ B \subseteq A \left| \begin{array}{l} B \text{ is a maximal set in } A, \\ \{\lambda w'.p(w') \neq 1 | p \in B\} \cup \{[\![\Phi]\!]\} \text{ is satisfiable,} \\ \neg \exists r \in A : \left[(p \Rightarrow_S r) \wedge (r(w) = 1) \wedge (r \Rightarrow \bigvee B) \right] \end{array} \right. \right\}$$

(54) A set of propositions E is satisfiable if there is a world w where all propositions in E are *defined and clearly true* (i.e., no presupposition failures, borderline vague cases, or homogeneity violations).

Unlike K&R, I will assume that EXH can only combine with nodes of type $\langle s, t \rangle$, so I will not need to generalize the notions of negation, entailment, and conjunction to complex types. I will sometimes write EXH($[\![\Phi]\!]$) instead of the proper $[\![\text{EXH } \Phi]\!]$, using the same symbol for the operator in the object language and its translation.

5.1.3 Extra assumptions on specific lexical items

***Know*:**   Percus (2006) and Sauerland (2008) assume that *know* is a lexical alternative to *believe* to explain the inference in (55).

(55) John believes that I am pregnant.
       $\rightsquigarrow$ The speaker isn't pregnant.

Their explanation is as follows: Instead of (55), the speaker could have said that John *knows* that she is pregnant. If it were common ground that she is, using *believe* instead of *know* would violate Heim's (1991) maxim 'Maximize Presupposition'. From the use of *believe* we can therefore conclude that it is not common ground that the speaker is pregnant. Chemla (2008) explains how this can be strengthened into an inference that she is in fact not pregnant even when the common ground does not contain any information about this: the speaker is an authority as to whether she is pregnant or not, and for this reason the addressee could accommodate her presupposition even if it was not already common ground. Therefore, by using *believe* she is conveying that she is not pregnant.

I will assume that conversely, *believe* is a lexical alternative to *know*. This will allow me to derive the IE reading of *know* without having to assume that the assertive content of a verb can be accessed independently of its presuppositional content, as Klinedinst and Rothschild ([2011](#)), Spector and Egré ([2015](#)), and Uegaki ([2015a](#)) do.[15] One could imagine that the difference is purely a matter of aesthetics, but we will see that this assumption also makes better prediction for negative sentences, where the neg-raising behavior of *believe* plays an important role.

Most: I will argue that *most* does not exactly mean "more than 50%" but that the semantics of *most* includes a degree-predicate operating on a continuous scale, making it a vague quantifier with a context-dependent threshold.[16] This can simply be modeled as uncertainty on the value of the parameter α in ([26](#)), repeated below.

(26) $[\![\text{mostly}]\!] = \lambda P_{\langle e,t \rangle}.\lambda X_e : |At(X)| \geq 2.\big[|At(X) \cap P| > \alpha \times |At(X)|\big]$

Quantity adverbs such as *partly*, *mostly*, and *completely* are likely to form a scale, but in the cases we will discuss they do not yield any implicature that could not be derived by simply considering the structurally simpler alternatives obtained by deletion of the adverb.

5.2 Application to a simple sentence with *know*

It is now time to check that the proposed theory can derive the correct readings for the simple sentence ([43](#)), repeated below. I assume that the question must move above the verb after combining with the Ans^S operator, leaving a trace of type y.

(43) John knows which children sang.

(56) Possible parses to derive each reading:
   a. WE reading: John 1 [C Ans^S which children sang] [2 [t_1 knows f t_2]]
   b. SE reading: John 1 [C Ans^S which children sang] 2 [t_1 knows exh [f t_2]]
   c. IE reading: John 1 [C Ans^S which children sang] 2 EXH [t_1 knows [f t_2]]
   d. IE reading: EXH [John 1 [C Ans^S which children sang] 2 [t_1 knows [f t_2]]]

Again, let us imagine a situation w_0 with three children where Ann (a) and Bill (b) are the children who sang, while Chris (c) is a child who didn't sing.

---

[15] I am very grateful to Emmanuel Chemla for this suggestion.

[16] This assumption may seem controversial, as most of the literature assigns equivalent truth conditions to *most* and *more than half* (Hackl [2009](#), a.o.). Nevertheless, the threshold for *most* is typically more than just 50%, as becomes clear when one consider the following contrast from Solt ([2016](#)):

   (i)   # Most of the American population is female.

   (ii)   More than half of the American population is female.

5.2.1 WE reading

In the given situation, $Ans^S(Q)(C^{w_0}) = a \oplus b$. Then (56a) simply means that John knows the proposition that Ann and Bill are among the children who sang, which is the WE reading of (43).

5.2.2 SE reading

With the parse in (56b), the only source of alternatives for EXH is the variable denoted by t_2. We need to keep in mind that the value of this variable depends on the assignment, but the derivation is essentially the same as in K&R.

(57) $[\![\text{EXH}[\text{f } t_2]]\!]^{g,w}(w') = 1$ iff $f(g(2))(w') = 1 \wedge \forall d \in D_e, d \not\leq g(2) \rightarrow f(d)(w') \neq 1$

Any element in D_e that is not a part of $g(2)$ yields an excludable alternative. If d is an individual who is not a child in w, $f(d)(w)$ is a presupposition failure, hence the corresponding term in (57) is immediately satisfied. If d is a non-singing atomic child, $f(d)(w)$ is defined. It must therefore be false to satisfy (57). $f(d)(w) = 0$ means that d is not part of any group of singing children in w (i.e., d did not sing).

This is the SE reading. Assuming that $g(2) = a \oplus b$, which is the value assigned higher up in the tree, John knows EXH$[f \, t_2]$ if and only if he knows that Ann and Bill sang and that no other child did.

5.2.3 IE reading

The IE reading can be derived from at least two different parses. In (56c), the sources of alternatives for EXH are the trace of $Ans^S(Q)(C^{w_0})$ and the verb itself (because I assumed that $\langle believe, know \rangle$ form a scale). In (58) I informally present a subset of the alternatives for 'know$(f(t_2))$', assuming for concreteness that $g(1) = j$ and $g(2) = a \oplus b$ (though these values have not yet been fixed at this point of the derivation).

(58) Sample from the alternatives for EXH in (56c) when $g(1) = j$ and $g(2) = a \oplus b$. The prejacent is in the blue box and the excludable alternatives are in the red box:

$$K(f(a))(j) \Leftarrow \boxed{K(f(a \oplus b))(j)} \not\Rightarrow \boxed{\cancel{K(f(c))(j)} \Leftarrow \cancel{K(f(a \oplus b \oplus c))(j)}}$$
$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow$$
$$B(f(a))(j) \Leftarrow B(f(a \oplus b))(j) \not\Rightarrow \boxed{B(f(c)) \, (j) \Leftarrow B(f(a \oplus b \oplus c))(j)}$$

Here 'know(a)', 'believe(a)', and 'believe$(a \oplus b)$' are entailed. Note that 'know(c)' and 'know$(a \oplus b \oplus c)$' are innocently excludable alternatives, but they are trivially non-true in any world where $a \oplus b$ is the K-answer because of the factive presupposition of *know*. In the end, only alternatives such as 'believe(c)' and 'believe$(a \oplus b \oplus c)$' yield non-trivial implicatures.

More formally, we can show that any alternative that is not entailed is excludable. All such alternatives with *know* or with an individual that is not a child are trivially

non-true because they are presupposition failures. Only alternatives with *believe* and a child who is not an i-part of $g(2)$ yield non-trivial implicatures, hence we can reduce the result of EXH to (59).

(59) $[\![\text{EXH}[t_1 \text{ know } [f\ t_2]]]\!]^{g,w}(x)(w') = 1$ iff $K(f(g(2)))(g(1))(w') = 1$ $\wedge$
  $\forall d \in D_e, [(d \npreceq g(2) \wedge [\![\text{child}]\!]^w(d)) \rightarrow B(f(d))(g(1))(w') \neq 1]$

In particular, if d is an atomic child not part of $g(2)$ (i.e., not singing), having the property EXH[know[$f\ t_2$]] implicates not (falsely) believing that d sang. At this point, differences between the semantics of *know* and *believe*, together with my assumptions on EXH, become important. Unlike *know*, *believe* is a neg-raising predicate. I will follow Gajewski (2005) in assuming that neg-raising should be analyzed as homogeneity (see also Križ 2015a). Concretely, 'John believes p' will be true when all worlds compatible with John's beliefs support p, false when none of them does, and a truth value gap otherwise. Here is it crucial that my entry for EXH only requires the excludable alternatives to be non-true. Requiring that the excludable alternatives be false would be problematic, as it would lead to an SE reading (John believes p would be false only if John believes $\neg p$, so for a child who did not sing, John would need to have the positive belief that this child didn't sing). By contrast, (59) leads to an IE reading because in a world where Ann and Bill are the only singers, (56c) is true if and only if John knows that Ann and Bill sang and he considers it at least possible that Chris did not sing.

With (56d) the alternatives previously contributed by the trace of $Ans^S(Q)(C)$ can be derived directly from $Ans^S(Q)(C)$ itself, so the focus value is unchanged, but the prejacent is different because the denotation of $Ans^S(Q)(C)$, unlike its trace, depends on the world of evaluation through C. This is where the idea of Mayr (2017) will be important. As illustrated in (60), in a world w_0 such that $Ans^S(Q)(C^{w_0}) = a \oplus b$, alternatives such as $B(f(a))(j)$ ('John believes that Ann sang') are not excludable because they are entailed by a true S-entailed alternative ('$K(f(a))(j)$' in this case). This leaves only alternatives such as $B(f(c))(j)$ ('John believes that Chris sang') to be excluded, and we arrive at the same result.

(60) Assuming $Ans^S(Q)(C^{w_0}) = a \oplus b$, of the following alternatives, only $B(f(c))$ is excludable in w_0:
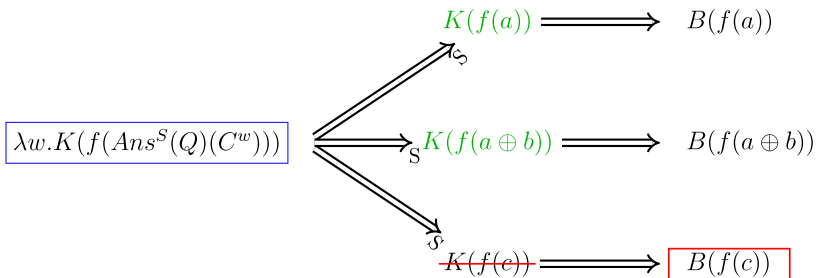
Since the subject 'John' is not quantificational and does not contribute any alternatives, it does not matter whether EXH applies to the whole sentence or is inserted anywhere between the subject and the moved question. I come back to quantificational subjects in Sect. 8.2.

5.3 False-answer sensitive readings beyond *know*

We saw how the theory derives an IE reading for *know*, but what about other factive verbs?

First, one should note that there is no consensus on the definition of "intermediate readings". Everyone seems to agree on the definition of the IE reading of *know*, but for other verbs most authors simply call "IE reading" the reading predicted by their theories when the mechanism they propose for *know* applies to another verb. Therefore, there are as many definitions of "IE readings" as there are ways to derive the IE reading of *know*. I will favor the term "false-answer sensitive" reading (henceforth FAS) proposed by Xiang (2016), which covers all readings making reference to false answers of the embedded question. This definition includes the IE and SE readings of *know*. To distinguish the two, I will call "global FAS reading" any FAS reading which requires exhaustification above the embedding verb.

Second, to derive the IE reading of *know* I assumed that it had a non-factive lexical alternative (*believe*). This assumption did not come out of the blue: a similar assumption can be found in the literature on anti-presupposition (Percus 2006; Sauerland 2008; Chemla 2008). For other factive verbs, however, such natural non-factive alternatives are not necessarily available, and alternatives built on false answers will always be presupposition failures in the world of evaluation. Therefore, I derive the generalization in (61). Whether this generalization is supported is an open empirical issue which will not be easily addressed, because (i) SE readings can obscure the data and (ii) determining which verbs are alternatives to each other can be challenging.

(61)   *FAS generalization:* A factive verb receives global FAS readings if and only if it has a lexical non-factive alternative.

Note that even if no global FAS reading is predicted for a verb, global exhaustification is not necessarily vacuous. Downward-entailing or non-monotonic verbs like *forget* and *surprise*, which are particularly relevant in this context, will be discussed in the next sections.

6 Application to new cases

In this section, I explore predictions of the theory for cases that are not discussed by K&R. In particular, I will examine how the exhaustification procedure interacts with the plurality effects discussed in Sect. 3.

6.1 Negative sentences

I first introduced the homogeneity effect with negative sentences such as (3b), repeated below. We can now look at how negation and homogeneity interact with exhaustivity. Depending on where the question is interpreted (below or above negation), different parses may be available, but this won't affect the overall set of readings that can be derived: one can always obtain all the alternatives below negation via

the trace of $Ans^S(Q)(C)$ and thereby derive an IE reading below negation. Conversely, even if $Ans^S(Q)(C)$ cannot escape NegP, its alternatives can still project and global exhaustification will lead to what I will take to be the default reading of (3b). For concreteness, I will adopt Pollock's (1989) standard structure for sentential negation, and assume that $Ans^S(Q)(C)$ only moves to Spec,VP (i.e., below negation).

(3b) Mary doesn't know which children sang.

(62) Possible parses and associated readings:
  a. WE reading: Mary 1 doesn't [$_C$ Ans^S which children sang] [2 [t_1 know f t_2]]
  b. SE reading: Mary 1 doesn't [$_C$ Ans^S which children sang] 2 [t_1 know exh [f t_2]]
  c. "¬IE" reading: Mary 1 doesn't EXH [[$_C$ Ans^S which children sang] 2 [t_1 know [f t_2]]]
  d. IE reading: EXH [Mary 1 doesn't [$_C$ Ans^S which children sang] 2 [t_1 know [f t_2]]]

Parse (62a), without EXH, would simply be the negation of the WE reading. Because of homogeneity, it would be true if and only if Mary doesn't know for any singing child that they sang. Of the many different parses with EXH, some may be dispreferred because local exhaustification in downward entailing environment is often considered deviant (see Chierchia 2004; Chierchia et al. 2012). This would apply to (62b,c) and any equivalent parse. Note that their exact meaning depends on assumptions regarding the falsity conditions of EXH, which I have not specified so far. A simple and natural option would be to treat EXH as a homogeneity remover (i.e., EXH(p)(A) is false whenever it's not true), which would just return the negation of the standard SE and IE readings for (62b) and (62c) respectively.

Parse (62d), or any parse where EXH takes scope over negation, will be the most interesting case. Here we need to carefully distinguish between sentential negation (which projects the truth value gap associated with presuppositions and homogeneity) and EXH (which merely states that innocently-excludable alternatives are "untrue").

In a situation where Ann and Bill are the children who sang and Chris is a child who did not sing, the prejacent is true if and only if Mary fails to know for both Ann and Bill that they sang (because of homogeneity). Compared with the affirmative case, the Strawson-entailments in (60) survive negation thanks to homogeneity (not knowing that $a \oplus b$ sang entails not knowing that a sang). However, the classical entailments from *know* to *believe* do not, because of neg-raising: not-knowing, unlike not-believing, is compatible with ignorance. As a result, all alternatives of the form "Mary doesn't believe that d sang" are innocently excludable. We therefore derive an implicature that such alternatives are not true. Given the neg-raising properties of *believe* (Gajewski 2005; Križ 2015a), this is the case if they are false (i.e., if Mary believes that d sang) or if they are undefined (i.e., if she has no opinion on whether d sang or not).

If d did not actually sing, the prejacent imposes no restriction on Mary's belief regarding d, so the implicature simply amounts to Mary not knowing that d didn't sing and maybe falsely believing that d sang. If d did in fact sing in the world of evaluation, Mary cannot believe that d sang (this would contradict the prejacent),

so the implicature can only be satisfied if Mary is agnostic as to whether *d* sang or not.[17]

In the general case, the predicted meaning for global exhaustification is the one paraphrased in (63).

(63)   ⟦EXH[Mary doesn't know which children sang]⟧ = 1
       ⇔ For any child who sang, Mary has no opinion as to whether they sang or not, and for any child who did not sing, Mary does not know that they did not sing (but possibly falsely believes that they did).

Note that the ignorance implicature about children who did sing is likely related to the Agent ignorance effect described in Paillé and Schwarz (2018) with *whether*-questions. They observe that (64) conveys that Aisha is agnostic as to whether Ben is Canadian or not. They further show that this inference is more robust than standard implicatures and conclude that it must be conventionalized to some extent. Xiang (2015) argues that the "no false belief" component of intermediate exhaustivity is not an implicature for similar reasons, a point to which I will come back in the conclusion of this paper.

(64)   Aisha doesn't know whether Ben is Canadian.

At this point, it is worth comparing the predictions of the different exhaustification theories. First, homogeneity is necessary to derive universal ignorance regarding the true answers. Second, all exhaustification theories are subject to the generalization in (61). To account for FAS readings of factive predicates, Klinedinst and Rothschild (2011) and Uegaki (2015a) build on an idea of Spector and Egré (2015): they assume separability between the assertive and the presuppositional content, so that a non-factive alternative can be defined for any factive responsive predicate. In the case of *know*, this would simply be a version of *believe* which has all the properties required for knowledge, except factivity/veridicality. Yet this "non-factive *know*" has no reason to be a neg-raising predicate, and the results of exhaustifying a sentence like (3b) would be very different. Indeed, without the excluded middle property of neg-raising predicates, the predicted inference is the negation of a non-neg-raising *Mary doesn't believe that Chris sang*, which simply amounts to an implicature that Mary falsely believes that Chris sang. More generally, these theories would predict that (3b) implicates that Mary falsely believes that every child who didn't sing did in fact sing,[18] and they would miss the agent ignorance inference regarding children who did sing.

---

[17] Note that this is a more general prediction for any account of implicatures where neg-raising can break entailments: *Mary doesn't know that Ann sang* is predicted to implicate that Mary does not falsely believe that Ann didn't sing, by competition with the non-weaker neg-raising alternative *Mary doesn't believe that Ann sang*. A related and—as far as I can see, correct—prediction is that the anti-presupposition of *believe* should be blocked under negation: Unlike the affirmative sentence *Mary believes that Ann sang*, the negative sentence *Mary doesn't believe that Ann sang* does not anti-presuppose that Ann didn't sing, because the *know*-alternative is not truth-conditionally equivalent (though some recent accounts of anti-presuppositions do not require truth-conditional equivalence; see Anvari 2018).

[18] Uegaki (2015a) assumes that exhaustification only negates stronger alternatives (rather than all non-weaker ones), but the problem arises anyway if we consider the alternative 'Mary doesn't *know*$_{\text{non-fact}}$ that Ann, Bill and Chris sang' instead of 'Mary doesn't *know*$_{\text{non-fact}}$ that Chris sang'.

6.2 QVE sentences

Let us now turn to sentence (1b), repeated below, which I used to introduce QVE. In Sect. 3.3.1, I showed how the WE reading of (1b) is derived, but there is obviously more to this sentence once we consider potential exhaustifications. The structure I proposed in (25b) provides three positions where EXH could be inserted, listed in (65), but only one of them doesn't yield a contradiction.

(1b)   John mostly knows which children sang.

(65)   a.   # John 1 [C Ans^S which children sang] mostly 2 t_1 knows EXH[f t_2]

   b.   # John 1 [C Ans^S which children sang] mostly 2 EXH[t_1 knows f t_2]

   c.   EXH$\big[$John 1 [C Ans^S which children sang] [mostly 2 t_1 knows f t_2]$\big]$

In (65a), EXH leads to a contradiction because one cannot know two or more different SE answers, while *mostly* presupposes that there are at least two true atomic answers. (65b) also leads to a contradiction, for a similar reason.

In (65c), there are three sources of alternatives: the scale ⟨*know*, *believe*⟩, deletion of *mostly* (or equivalently replacement of *mostly* with *completely*), and the short answer. Since I imposed no restriction on multiple replacements, this generates many alternatives, which makes the derivation rather technical. To keep this section readable, all technical details are moved to Appendix 1.

As it stands, the predicted implicature is too strong. Assuming that the threshold α for *most* is strictly more than $\frac{1}{2}$ but less than $\frac{2}{3}$, (65c) would convey that exactly three children actually sang, and John knows this for two of them (see Appendix A.1 for proof). This interpretation is obviously problematic. Intuitively, the oddness comes from the fact that the sentence is normally used to describe John's mental state, but it ends up imposing a restriction on which children actually sang. We might be onto something here, because in cases such as (66), the sentence is indeed used to convey information about the embedded question rather than the agent's mental state. In short, the sentence can convey information about two independent issues: which children actually sang, and someone's knowledge on this topic. In practice, it's unlikely that we would be interested in both issues at the same time. I will show that if we focus on only one of those two issues, we can always prune the set of alternatives to make (1b) sensible and relevant.

(66)   A: Who was at the party? All I know is that all of Mary's friends were present.
   B: Well, for the most part, you know who was at the party then.
   ⇝ Not many people beyond Mary's friends were present (and B may not consider their identity to be particularly relevant).

First, let us consider a situation in which the answer to *Which children sang?* is common ground, and the question under discussion is how much John knows. Pruning all alternatives obtained by replacing $Ans^S(Q)(C)$ with an a below a certain cardinality (which depends on the cardinality of the actual answer and the threshold α for *most*) yields a reading equivalent to the conjunction of the prejacent (67a) and the two implicatures (67b) and (67c) (see Appendix A.2 for proof and discussion).

(67)   a. For most children who sang, John knows that they did.

      b. He has no false belief regarding children who did not sing.

      c. John knows *significantly less* than the complete answer.

Note that (67c) is not precisely defined because it depends on which alternatives are pruned. Due to second-order vagueness effects (Dummett 1978), it is impossible to give an exact characterization of the optimal set of alternatives to prune; see the Appendix for a detailed discussion.

Preuss (2001: 145–155) proposed a strongly exhaustive reading for QVE sentences which roughly corresponds to this: John knows a significant proportion of true answers and does not have any false belief regarding false answers. She argues that this reading is particularly salient in the following examples:

(269)   a. John uncovered to a large extent who took bribes.

      b. John confessed to a large extent who he had cheated in poker.

      c. John found out to a large extent who had cheated on the exam.

Second, we can assume the opposite configuration: a situation in which it is common ground what John knows, and the question under discussion is which children actually sang. Of course, given the factivity of *know*, anything known by John must be true, hence the context provides a lower bound on what the actual answer may be. We can then prune all alternatives resulting from replacing $Ans^S(Q)(C)$ with a fixed individual (the context already fixes the truth or falsity of such alternatives since John's doxastic state is known). Of the remaining alternatives, only those obtained by deletion of *mostly* (or replacement with *completely*) play a role. The resulting interpretation is simply that what John knows constitutes most but not all of the actual true answer. If a_1 is the (known) plurality of children that John knows to have sung, and a_0 the unknown actual complete answer, an utterance of (1b) therefore informs us that the cardinality of a_0 is contained in the open interval $(|a_1|, \alpha^{-1}|a_1|)$. Again, α is taken to be a vague threshold, so the exact upper bound may vary contextually, but in simple words the conveyed interpretation is that the group of children who sang is larger than what John knows, though not by much.

6.3 Cumulative readings

Sentence (2b), repeated below, was argued to be ambiguous between a distributive and a cumulative reading. In Sect. 3, however, we only considered WE readings, which correspond to a parse without any EXH. I will now discuss the parses with EXH, listed in (68).

(2b)   The teachers know which children sang.

(68)   a. [The teachers] [C Ans^S which children sang] ** 2 1 [t_1 know EXH[f t_2]]

      b. [The teachers] [C Ans^S which children sang] ** 2 1 EXH[t_1 know [f t_2]]

      c. EXH$\big[$[The teachers] [C Ans^S which children sang] ** 2 1 [t_1 know [f t_2]]$\big]$

The parse (68a) with the most local EXH results in a SE reading as usual, but here this SE reading forces a distributive interpretation. Indeed, if we write R for the relation

of SE-knowledge between a short answer and an agent, $R(a)(x)$ presupposes the exhaustification of $f(a)$, which is incompatible with the exhaustification of $f(b)$ for any $b \neq a$. As a consequence, given two short answers a_1 and a_2 and two individuals x_1 and x_2, $R(a_1)(x_1) \wedge R(a_2)(x_2)$ presupposes $a_1 = a_2$. As a result, $^{\star\star}R(a)(X)$ can only be true if each atomic individual in X knows the SE answer associated with the complete Karttunen-answer $f(a)$.

(68b) returns a cumulative IE reading. Here $R(a)(x)$ means that x knows $f(a)$, and that x does not believe any stronger answer. This is only possible if $f(a)$ is the maximal true partial answer that x knows. The reading we obtain is that each individual teacher knows a true partial answer and does not believe any other partial answer (in particular, they have no false beliefs regarding non-singers), and these partial answers add up to the actual true complete answer. In other words, the teachers cumulatively know which children sang, and none of them has any false belief regarding the children who did not sing (some of them may however falsely believe that some singers did not sing).

For (68c), let me assume that the DP *the teachers* does not contribute any alternatives. Because of the cumulation operator, most entailments are broken. For instance, if teacher x knows $f(a)$, y knows $f(b)$, and z knows $f(c)$, then $x \oplus y \oplus z$ is in the cumulative know-f relation with $a \oplus b \oplus c$ but not with $a \oplus b$ (since z does not know any atom of $a \oplus b$). As a consequence, none of the alternatives obtained by replacement of the complete answer is Strawson-entailed by the prejacent, and all these alternatives are independent from each other. Excluding all but one gives a satisfiable exhaustified reading, while excluding all of them leads to a contradiction (the teachers would know nothing at all). The result is that no alternative is innocently excludable and EXH is vacuous in this parse. With adequate pruning, we could retrieve the reading corresponding to (68b).

(2b) can also receive distributive readings. We have already seen that local exhaustification leads to a distributive SE reading even in the presence of the **-operator. Using a distributive operator, we could derive a distributive IE reading (each teacher has complete IE knowledge), which entails the cumulative IE reading.

To summarize, we predict a WE and an IE cumulative reading, and WE, IE, and SE distributive readings. Of course, pragmatics should play an important role in selecting the appropriate reading. I did not discuss the effect of homogeneity, which does not affect the truth conditions of (2b) but imposes stronger falsity conditions. It would make its negation (69) stronger. As it turns out, (69) would always convey that no teacher has knowledge regarding any child. This is the case under a cumulative reading because of homogeneity effects on the question (no answer is cumulatively believed by the teachers), but also under a distributive reading, even if we assume the distributive operator to take scope under negation (in this case, we need to take into account homogeneity effects on both the question and the definite plural). For a more detailed discussion of cumulative relations and homogeneity, see Križ (2015a, Sect. 2.4.1).

(69) The teachers don't know which children sang.

6.4 Primary implicatures

Grammatical theories of implicatures such as Fox (2007) usually assume that primary implicatures (ignorance inferences) are derived independently from secondary implicatures, which are the result of EXH (though see Meyer 2013 for a more radical grammatical approach). It would be natural to assume that the alternatives I introduced for the derivation of stronger readings also participate in primary implicatures. The main effect of this would be to make the sentences odd in situations which only satisfy the WE reading; if the speaker knows that Mary falsely believes that all children sang, asserting that Mary knows who sang would be a blatant violation of the primary implicatures. Rather, we predict that a good situation to detect WE readings would be one in which the speaker (a) knows that Mary knows that all child singers sang and (b) does not believe that Mary falsely believes that any other child sang. Note that (b) is weaker than the secondary implicature involved in the IE reading, because the speaker need not be sure that Mary has no false beliefs.

7 Lexical restrictors: motivation and characterization

7.1 Lahiri's (2002) arguments

Most veridical responsive verbs happen to be factive, and conversely, all factive verbs that embed questions are veridical responsive. This led Berman (1991) and Lahiri (1991) to assume that the domain of quantification for QVE is obtained by accommodation of the presuppositions of the verb. Following Lahiri (2002), I departed from this view and assumed that the domain of quantification is determined by a contextual restrictor, the value of which is fixed by the embedding verb. It must therefore be encoded somewhere in the lexical entry of the verb, but may differ from the verb's presuppositions.

One argument of Lahiri (2002) against the presupposition accommodation solution was the controversial status of so-called *intermediate accommodation* at the time he was writing (see Beaver 1995). Lahiri's second argument came from communication verbs such as *tell* or *predict*. These verbs are neither factive nor veridical when embedding propositions, but are usually veridical when embedding a question. This suggests that the restriction they impose on answers when they embed questions does not correspond to a presupposition.

Regarding the first argument, recent literature on presupposition seems to have reached a consensus in favor of the availability of a mechanism of intermediate accommodation (Geurts and van der Sandt 1999; Singh 2008, 2009). Furthermore, Spector and Egré (2015) provide arguments against the idea that communication verbs are veridical when embedding questions but not factive. In short, new data suggest that the two arguments provided by Lahiri (2002) against the intermediate accommodation solution are not valid. Since intermediate accommodation seems more natural and much less stipulative than the idea of a contextual restrictor, the value of which must be encoded in each embedding verb, it should be preferred. Nevertheless, I will present new arguments suggesting that restriction based on intermediate accommodation does

not work. The crucial observation is that factive verbs such as *forget* or *discover* have stronger presuppositions than *know*, and yet these stronger presuppositions project when these verbs embed questions instead of further restricting the set of answers. As a consequence, I will stick to Lahiri's (2002) lexical restrictors. I will now look at *forget* in greater detail, before turning to communication verbs in Sect. 7.3 and *be certain* in Sect. 7.4.

One issue I will leave aside for the moment is how exactly the verb determines the value of the restrictor. This is somewhat tricky as I will need to argue that for some verbs *C* depends not only on the verb but also on its subject.

7.2 Complex factive verbs: the case of *forget*

For concreteness I will focus on the verb *forget*, as in sentence (70), although the point against intermediate accommodation can also be made with other complex factive verbs such as *discover*, *find out*, or *remember*. *Forget* happens to be of particular interest because it was among the counter-examples to reducibility presented in George (2013). I will show that the observations of George are in fact additional evidence for homogeneity effects.

(70) Sue forgot which children sang.

The most natural entry for 'forget that *p*', given informally in (71a), introduces two presuppositions (factivity and past knowledge) and asserts that the agent no longer knows *p*. In this sense, 'forget *p*' is equivalent to 'stop knowing *p*'. According to the presupposition accommodation approach, (70) should thus receive the truth conditions in (71b), which are clearly too weak. (71c) looks like a much better paraphrase of (70). One way to maintain the presupposition accommodation approach would be to attribute different status to the two presuppositions of *forget*, so that only factivity would play a role in accommodating the restrictor.[19]

(71) a. Sue forgot *p*: <u>*p* and Sue used to know *p*</u>, Sue does not know *p* (any more).
      b. # For each child *x* who sang and for which Sue knew that *x* sang, Sue forgot that *x* sang.
      c. For each child *x* who sang, <u>Sue knew that *x* sang</u> and Sue forgot that *x* sang.

In the lexical restrictor approach, we can postulate that the restrictor is the same as *know*, that is, the simplest conceivable veridical restrictor. This makes the factivity

---

[19] Such an analysis is sketched in Theiler et al. (2016) and Uegaki (2015b), following a suggestion of Theiler (2014), although none of them apply it to factive verbs beyond *know* and emotive-factives. The idea is to derive the factivity of declarative entries from the presence of an operator which is responsible for the veridicality of the corresponding question-embedding predicates. Therefore, the factivity presupposition is not hard-coded in the semantics of the verb, but comes as a by-product of the verb being *extensional*, in Groenendijk and Stokhof's (1984) sense. Other presuppositions, such as the 'knowledge in the past' presupposition of *forget*, would remain purely lexical and would not play any role in the restriction of the question. This approach seems promising for complex factives, but it seems difficult to extend to non-factive responsive predicates. As we will see, the restrictors for verbs like *be certain* or *agree* seem arbitrary and would, if anything, correspond to lexical presuppositions of these verbs.

presupposition trivially satisfied, while the "past knowledge" presupposition remains non-trivial. Indeed, some propositions in the restrictor may satisfy the factive presupposition but not the "past knowledge" presupposition. If we want to be more specific, we need to postulate how homogeneity and presuppositions interact. Non-maximality may allow for some exceptions and in particular presupposition failure for a few atoms, but for simplicity I will assume universal projection for distributive predicates, as defined in (72).

(72) If a distributive predicate $\lambda x . P_\rho(x)$ with presupposition $\lambda x . \rho(x)$ applies to a plurality X, $P_\rho(X)$ presupposes that each atom of X satisfies ρ.

We now have everything in hand to apply the general theory to the sentence (70). The four parses in (73) are possible.

(73) a. Sue 1 [C Ans^S which children sang] 2 [t_1 forgot [f t_2]]
          <u>Sue used to know which children sang (weakly)</u>, and for each child who sang, Sue forgot that they sang.

      b. Sue 1 [C Ans^S which children sang] 2 [t_1 forgot EXH[f t_2]]
          <u>Sue used to know which children sang (strongly)</u>, and for at least one child, Sue forgot whether they sang.

      c. Sue 1 [C Ans^S which children sang] 2 EXH$\big[t_1$ forgot [f t_2]$\big]$
          EXH *is vacuous, result equivalent to (a)*

      d. EXH$\big[$Sue 1 [C Ans^S which children sang] 2 [t_1 forgot [f t_2]]$\big]$
          EXH *is again vacuous, result equivalent to (a)*

As usual in the absence of an EXH operator, (73a) yields a WE reading. However, homogeneity ensures that Sue had exhaustive knowledge in the past *and* has exhaustive oblivion in the present. George (2013) noticed that these truth conditions cannot be reduced to anything of the form 'Sue forgot that p', where p is a classical proposition. Yet we see that this reading can be derived from the usual denotation for declarative-embedding *forget* once homogeneity is taken into account, because the answer $p = f(y)$ to a question is typically a trivalent proposition when y is not atomic (as pointed out by Križ 2015b). In the end, this example does not challenge reducibility.

In (73b), local exhaustification leads to a stronger presupposition: Sue must have known which children sang in an SE sense. As discussed in Sect. 6.1, a natural assumption for how exhaustification interacts with trivalent prejacents would be to treat EXH as a homogeneity remover. Together with the fact that *forget* is Strawson-downward-entailing, this makes the assertion very weak (there is at least one child such that Sue forgot whether she sang). Again, theories of exhaustification normally block EXH-insertion in Strawson-downward-entailing environments (Chierchia 2004; Chierchia et al. 2012). The availability of this very weak reading is essentially an empirical question.

In (73c), the only alternatives are contributed by the trace of the short answer. Since the verb *forget* does not contribute any alternatives (and in particular, no non-factive alternatives), it is subject to generalization (61): matrix exhaustification does not yield a FAS reading. Since *forget* is Strawson-downward-entailing, we could expect non-trivial implicatures from alternatives corresponding to i-parts of the actual answer.

Nevertheless, homogeneity kicks in and these alternatives end up being entailed by the prejacent: when its presuppositions are satisfied, $[\![forget]\!](p)(s)$ is true if there is a world w' in Dox_s such that $p(w') = 0$. If $a \succeq b$, $f(a) = 0 \rightarrow f(b) = 0$, so forgetting a short answer entails forgetting all the weaker short answers. To sum up, the stronger alternatives which could contribute implicatures are necessarily presupposition failures in the world of evaluation (hence their implicatures are trivially satisfied), and all alternatives which would not be presupposition failures are entailed, hence not excludable.

In (73d), EXH is equally vacuous because all alternatives are Strawson-entailed: if d is a child who sang in w and Sue forgot the answer to the question in w, it follows (by homogeneity) that Sue forgot that d sang.

To summarize, we wanted to derive the presupposition that Sue used to know which children sang for sentence (70). With intermediate accommodation, the sentence would have no presupposition and very weak truth conditions (roughly: Sue forgot what she used to know, whatever this was). *Forget* and other complex factive verbs thus provide a strong new argument against intermediate accommodation as the source of the domain of quantification for QVE. Along the way, we saw that matrix exhaustification is vacuous for (70), and that this sentence does not pose much threat to the reducibility hypothesis once homogeneity is taken into account.

7.3 Communication verbs

Communication verbs have long been considered to be veridical when embedding questions, and this was surprising given that they are not factive when embedding declarative complements. Spector and Egré (2015) recently proposed that they are ambiguous between a veridical and a non-veridical reading when embedding questions, and between a factive and a non-factive reading when embedding declaratives. The latter is supported by data from Schlenker (2007), repeated in (74). From (74a) we infer that Sue is indeed pregnant, and this inference projects out of negation or questions, a projection pattern typical of presuppositions. In support of the former point, Spector and Egré (2015) provide examples such as (75). Furthermore, they argue that the veridical readings correspond exactly to the factive entries of communication verbs and they provide data from Hungarian supporting this view; Hungarian marks the distinction between factive and non-factive *tell* morphologically, and Spector and Egré show that questions embedded under the factive entry of *tell* always give rise to a veridical interpretation.

(74)   a. Sue told someone that she is pregnant.

       b. Sue didn't tell anyone that she is pregnant.

       c. Did Sue tell anyone that she is pregnant?

(75)   Every day, the meteorologists tell the population where it will rain the following day, but they are often wrong.

Given these new data, I will follow Spector and Egré (2015) in assuming that *tell* and other communication verbs come with two lexical entries: a factive/veridical one and a non-factive/non-veridical one. The factive/veridical entry is easily accounted for

if we assume that its restrictor is the usual $C = \lambda p.p(w_0)$, which yields the veridical reading in (76a). The non-veridical entry requires a bit more work.

(76) John told me which children sang.

    a. For each child x who sang, John told me that x sang.
       = John told me the true complete answer to Q

    b. For each child x such that John believes that x sang, John told me that x sang.
       = John told me what he believes to be the complete answer to Q

Spector and Egré's (2015) lexical rule involves existential quantification, hence the non-veridical reading they derive for (76) is simply 'John told me some answer to Q', without any restriction on what that answer may be (most of the recent proposals make the same prediction for non-veridical predicates). The Ans^S operator requires a restrictor as its second argument, but one could imagine picking the trivial restrictor consisting of all possible propositions. My theory, by contrast, is based on universal quantification, so less restriction would lead to a *stronger* reading. Indeed, the trivial restrictor would predict the meaning of (76) to be 'John told me that every child sang', which is clearly wrong. Non-veridical entries require a non-trivial restrictor as well, and a reasonable option for non-veridical *tell* and other communication verbs is to use the set of propositions that John believes: $C = \lambda p.B_j(p)$. This yields the reading (76b). Assuming that this restrictor is shared by all communication verbs, we derive an interesting prediction. When used non-veridically, communication verbs should at least receive an "honest" reading when they embed questions: the agent is communicating what she takes to be the true answer to the question. This prediction is supported by the fact that communication verbs which convey some degree of dishonesty do not seem to do well with questions, as exemplified by (77) and the French examples (78) and (79).[20]

(77) a.   John insinuated that Mary cheated on the exam.

      b.   * John insinuated who cheated on the exam.

(78) a.   Jean prétend que Marie a triché.
           Jean claims　that Marie cheated

      b.   * Jean prétend qui　a triché.
             Jean claims　who cheated

(79) a.   Jean insinue　que Marie a triché.
           Jean insinuates that Marie cheated

      b.   * Jean insinue　qui　a triché.
             Jean insinuates who cheated

---

[20] An anonymous reviewer points out that *John lied about who cheated* seems fine. To begin with, note that *lie* is intransitive for most speakers, and *John lied that Mary cheated on the exam* is deviant. If it embedded questions, *lie* would therefore not be responsive. I think that this example falls under what Rawlins (2013) calls the "overproductivity of 'about'". *About* allows interrogative complements to be embedded under verbs which do not normally accept them (e.g., *think about who cheated*).

Finally, another notable implication of the idea that the default non-veridical restrictor is the set of propositions believed by the agent would be the existence of systematic tautologies for questions embedded under *believe*. While this is an interesting new way to look at why *believe* does not embed questions, recent proposals elaborating on Zuber's (1982) idea that this property stems from the neg-raising behavior of *believe* seem far more promising (see Theiler et al. 2017 and Mayr 2017).

7.4 *Be certain*

As with communication verbs, a non-trivial lexical restrictor must be postulated for *be certain*. Lahiri (2002) proposes the restrictor in (80), consisting of all propositions that the agent considers possible.

(80) $[\![x \text{ is certain that } p]\!] = \lambda w.Dox_x^w \subseteq p$
      $C_x = \lambda p_{st}.\exists w \in Dox_x^{w_0} : p(w)$ ('x considers it possible that p')

This restrictor is not necessarily closed under conjunction, and this would lead my theory to predict occasional presupposition failures when the restrictor is fed to the Ans^S operator. Let us look at a concrete scenario illustrating this. Imagine that Sue considers it possible that Ann sang and that Bill sang, but she is sure that they did not both sing. The set of answers which satisfy Lahiri's C is simply $\{f(a), f(b)\}$, which notably excludes $f(a \oplus b)$ and thus lacks a maximal element. This would lead to a presupposition failure when computing $Ans^S(Q)(C)$; and yet, in such a situation, (81) is intuitively plain false since its negation (82) is clearly true.[21]

(81)   Sue is certain about which children sang.

(82)   Sue isn't certain (about) which children sang.

A simple solution is to assume that the restrictor is the closure under conjunction of Lahiri's restrictor. In the scenario under discussion, this makes $a \oplus b$ the output of Ans^S, even though Sue considers this short answer impossible. If Sue considers that a could have sung, b could have sung, and $a \oplus b$ couldn't, she cannot be certain that a sang, nor be certain that b sang. Therefore, (81) is predicted to be clearly false (Sue isn't certain of any of the atomic parts of $Ans^S(Q)(C)$), and (82) is correctly predicted to be clearly true.

Local exhaustification of (81) is vacuous. (81) already conveys that Sue is certain about all the answers she considers possible. By contrapositive, if she isn't certain that p, it must be that she doesn't consider p possible (i.e., she is certain that $\neg p$). In other words, the WE reading of (81) is already equivalent to its SE reading.

As it stands, global exhaustification of (81) would lead to the odd implicature that Sue considers it impossible that more than one child sang. Indeed, all alternatives are independent of the prejacent, and only those about an atomic child are not excludable, because they would contradict the presupposition of the question (see Uegaki 2018 for

---

[21] I am ignoring possible subtleties regarding the role of *about* here, which is mandatory in (81) but not (82). See Rawlins (2013) and Mayr (2017) for discussion. Unlike verbs which cannot embed questions without the mediation of *about*, such as *think* and *lie* discussed in footnote 20, *be certain* doesn't always require *about*, and when it is optional it does not seem to affect the meaning of this predicate.

a discussion of this presupposition). This is easily fixed if we assume being certain that p to presuppose considering p possible (this seems to be correct given the behavior of *be certain* under negation or in questions). Then all alternatives become Strawson-entailed, and no implicature is derived.

Global exhaustification of (82) is also vacuous: as in previous cases, homogeneity ensures that Strawson-entailment is maintained under negation.

Local exhaustification of (82) is not vacuous and yields a rather weak reading: namely, that there is at least one child d such that Sue thinks that d possibly sang and possibly didn't sing (again, I take EXH to be a homogeneity remover). As discussed previously, the availability of local exhaustification in downward-entailing environments is debated, and the availability of this reading is not warranted.

7.5 Conclusion

Assuming that each responsive predicate comes with a lexically specified restrictor just for the purpose of question embedding sounds like a superfluous fancy. However, when we look at a wider range of responsive predicates than the few usually studied in the literature, it does turn out impossible to simply use presuppositions as a way to restrict the set of answers. In particular, such an approach would predict that the presuppositions a verb has when embedding declarative complements systematically disappear when it embeds questions, contrary to the facts: *Sue forgot who sang* does presuppose that she used to know who sang. An additional benefit of using lexical restrictors is to allow stronger truth conditions for non-veridical predicates, in comparison to the existential meaning usually assumed in the literature. In particular, *Sue told me who sang* requires a bit more than Sue mentioning a random potential singer, even under a non-veridical interpretation. I have assumed that this sentence is true only if Sue utters what she believes to be the *complete* answer.

8 Handling more complex cases

We saw that the theory makes correct predictions in most cases and provides a reducible solution to George's (2013) challenge regarding *forget*. In this section, I will show that two cases which have been argued to be problematic for exhaustification-based theories can, under reasonable assumptions, be accounted for. First, I will discuss the case of *surprise*, which has received considerable attention but still poses a number of specific challenges for the present theory. Second, I will address an issue with questions embedded in quantified sentences, which has been around since Klinedinst and Rothschild (2011). I will argue that the problem has nothing to do with questions and comes from an over-simplistic view of the mechanisms involved in the computation of implicatures.

The verb *agree* also comes with a number of challenges, but these would take us too far from the focus of this paper. An in-depth discussion can be found in Appendix B, where I go back to Lahiri's (2002) discussion of *agree* and propose a new analysis

in light of recent experimental data by Chemla and George (2016) as well as new introspective data on *disagree*.

8.1 *Surprise*

Several aspects of the behavior of questions embedded under *surprise*, as in (83), do not seem to be immediately captured by my theory. Nevertheless, I will show that under closer scrutiny facts that may seem problematic at first are either predicted or can be accommodated under standard assumptions.

(83) It surprised Chris which children sang.

First, (83) does not seem to give rise to homogeneity effects. Indeed, it seems sufficient that one specific child singing surprised Chris for (83) to be true, whereas distributivity and homogeneity would predict that the sentence is true only when Chris expected none of the children who did sing to sing (similar to what I correctly predict for *forget*). This may not be a problem though, because *surprise* may not be distributive to begin with (Lahiri 2002; Sharvit 2002, 2004). Indeed, the entailments in (84) do not hold, as becomes evident when one considers a scenario in which Peter does not take John's and Mary's respective presence at the party to be independent events.

(84) It surprised Peter that John and Mary came to the party.
      $\nRightarrow$ It surprised Peter that John came to the party
      $\nRightarrow$ It surprised Peter that Mary came to the party

For instance, if John and Mary have a baby together, Peter may expect that at least one of them will have to stay home, without specific expectations as to which of them will. In this case, he has no specific expectation regarding either Mary's coming to the party or Peter's, but he has a strong expectation that they won't show up together.

Non-distributive denotations have been proposed by Villalta (2008), Romero (2015), and Uegaki (2015a), among others. Križ (2015a, Sect. 2.2.2) discusses how non-distributive predicates (in particular collective predicates) can be accounted for in a theory of homogeneity, but I won't get into the details here. For concreteness, I will adopt a degree semantics for *surprise* inspired by Villalta (2008) and Romero (2015). This semantics relies on the idea that *surprise* is focus-sensitive, which can be modelled with Rooth's (1992) squiggle operator '$\sim$' and a variable Γ (distinct from the lexical restrictor C), as in (85), which receives the denotation in (86).

(85) $S_1 = [[[\text{that } \varphi] \sim \Gamma] \text{ surprise}_\Gamma \text{ Chris}]$
      $[\![[[\text{that } \varphi] \sim \Gamma]\!] = \underline{p \in \Gamma \subseteq [\![\varphi]\!]_f} : p$ with $p = [\![\varphi]\!]$

(86) $[\![S_1]\!]^{w_0} = \underline{p \in \Gamma \subseteq [\![\varphi]\!]_f \wedge p(w_0) \wedge Dox_c^{w_0} \subseteq p} : \mu_{c,w_0}(p) \geq \theta_\Gamma$
      where $\mu_{c,w_0}(p)$ is the degree to which the p was unexpected for Chris in w_0

(86) presupposes that φ is true and known by Chris in the world of evaluation, and that the variable Γ is a subset of the focus value of φ which contains the ordinary semantic value of φ. It asserts that φ was unexpected for Chris compared to other elements in Γ.

Interrogative complements undergo IR as usual, leaving only a trace as the complement of *surprise*. The comparison class is necessarily a set of answers (as argued in Romero 2015):

(87) $S_2 = [C \, Ans^S \, Q] \, 1 \, [[[f \, t_1] \sim \Gamma] \, \text{surprise}_\Gamma \, \text{Chris}]$

(88) $[\![S_2]\!]^{w_0} = f(a_0) \in \Gamma \subseteq \text{range}(f) \wedge Dox_c^{w_0} \subseteq f(a_0) : \mu_{c,w_0}(f(a_0)) \geq \theta_\Gamma$
       where $a_0 = Ans^S(Q)(C^{w_0})$ and C^{w_0} is the usual veridical restrictor

Although *surprise* does not give rise to homogeneity, it does allow quantificational variability effects, as illustrated in (89) below.

(89) For the most part, it surprised Chris which children sang.
       $\rightsquigarrow$ For most children who sang, Chris was surprised that they did.

Nothing in my account states that a verb must be distributive to allow QVE, and we can account for this sentence straightforwardly by assuming the structure in (90).

(90) $[C \, Ans^S \, Q] \, \text{mostly} \, 1 \, [[[f \, t_1] \sim \Gamma] \, \text{surprise}_\Gamma \, \text{Chris}]$

Crucially, Γ can contain false answers, so (90) conveys that most atomic true answers are surprising in comparison to the whole set of possible answers (not just other true answers, which would probably be a contradiction).

Finally, let us look at possible exhaustifications of (83), listed in (91).

(91) a. $[C \, Ans^S \, Q] \, 1 \, [[\text{EXH}[f \, t_1] \sim \Gamma] \, \text{surprise}_\Gamma \, \text{Chris}]$
       b. $[C \, Ans^S \, Q] \, 1 \, [\text{EXH}[[f \, t_1] \sim \Gamma] \, \text{surprise}_\Gamma \, \text{Chris}]$
       c. $[C \, Ans^S \, Q] \, 1 \, \text{EXH}\big[[[f \, t_1] \sim \Gamma] \, \text{surprise}_\Gamma \, \text{Chris}\big]$
       d. $\text{EXH}\big[[C \, Ans^S \, Q] \, 1 \, [[[f \, t_1] \sim \Gamma] \, \text{surprise}_\Gamma \, \text{Chris}]\big]$

Romero (2015) argues that (91a) results in a presupposition failure, because the exhaustified answer is not a member of the comparison class Γ. However, this depends on how EXH projects focus alternatives. We could imagine that it projects a set of exhaustified alternatives, for instance. We could also imagine that EXH applies above the squiggle operator, thereby not interfering with its presupposition (in which case we would assume that both EXH and *surprise* can access the variable Γ). Data from Cremers and Chemla (2017) suggest that SE readings are possible for *surprise*, so at least one of (91a) and (91b) should be possible.

Turning to (91c), we need to make some assumptions on how the variable Γ behaves under quantification. The simplest assumption would be that Γ remains fixed and does not covary with the bound variable. If $\Gamma = \text{range}(f)$, the result is a very strong reading: Chris is surprised at the particular combination of children who sang, but there is no subgroup of children whose singing clearly surprises him.[22] This interpretation

---

[22] **Proof:** Unlike *forget*, *surprise* does not make the inference in (i) valid. By contrapositive, we can deny *It surprised Chris that Ann sang* and *It surprised Chris that Bill sang* without denying *It surprised Chris that Ann and Bill sang* (if we see degrees of unexpectedness in terms of probabilities, this is simply a correlate of the fact that a conjunction can be less likely than each of its conjuncts). Therefore, all alternatives built on answers which are proper subparts of the actual answer could be negated.

(i) It surprised Chris that Ann and Bill sang.
     $\not\Rightarrow$ It surprised Chris that Ann sang or it surprised Chris that Bill sang.

may be possible in some very specific situations (e.g., the couple with a baby), but in general it seems too strong. One particular issue is that as a gradable predicate *surprise* is clearly vague, since there does not seem to be any obvious threshold for a proposition to count as "surprising" (even with full knowledge of the comparison class). (91c) would convey that the complete answer was just above the threshold of unexpectedness, but since we don't know precisely where this threshold is, this doesn't tell us much. Leffel et al. (2018) investigated the interaction between vagueness and implicatures and argue that in such configurations the implicature is typically blocked. Here this would mean that alternatives that are "too close" to the complete answer are pruned. The more alternatives are pruned, the weaker the implicature, i.e., the more unexpected the complete answer may be.

Let us conclude with matrix exhaustification, as in (91d). A consequence of the non-distributivity of *surprise* is that no alternative is Strawson-entailed. Without further restrictions on the context, we can show that no alternative is innocently excludable.[23] Here, pruning alternatives would lead to the implicature that none of the remaining alternatives is the complete answer, and if one of them is a part of the complete answer, then Chris is not surprised by it. In particular, pruning only the alternative corresponding to the complete answer (assuming it is known) results in the same interpretation as (91c).

8.2 Quantificational subjects

In all the examples I have treated so far, the subject of the sentence was a proper noun or a definite plural. K&R note that their theory makes too strong predictions for sentences like (92) if EXH is allowed to apply globally, and there is no reason my theory wouldn't inherit this prediction. In particular, they predict an (unattested) implicature that no child made any false prediction.

(92)   EXH[At least one child predicted who sang]
       ⇒ For any false answer p, it is not the case that at least one child predicted p
       ⇒ No child made any false prediction regarding who sang

K&R propose that the scope of EXH is limited to the VP level, thus preventing it from interacting with quantificational subjects. Of course, this would mean that the exhaustification operator responsible for the strong readings of questions is different from the operator responsible for scalar implicatures. If this were the case, it would be a formal argument against the exhaustification approach (see Xiang 2015 for a development of this argument).

---

[23] **Proof:** First let us note that not all alternatives can be excluded together, as this would lead to a presupposition failure. Let $a \in D_e$, possibly non-atomic. Let $E_a = [\![(83)]\!]_f \setminus \{\varphi_a\}$ where φ_a denotes the proposition that Peter was surprised that a sang. E_a is a set of excludable alternatives because we can find a world in which Peter is surprised at the combination a without being surprised by any combination of parts of a singing (think about the couple-with-baby situation, but with an arbitrary plurality instead of a couple), and in which no one beyond a sang (making other alternatives, including those Strawson-entailed by φ_a, presupposition failures). E_a is also maximal, since the only alternative that could be added is φ_a and this would make the prejacent a presupposition failure. By definition, the set of innocently excludable alternatives is the intersection of all such maximal sets, but $\bigcap_{a \in D_e} E_a = [\![(83)]\!]_f \setminus \bigcup_{a \in D_e} \{\varphi_a\} = \emptyset$.

However, their analysis of (92) may be over-simplistic. The quantifier *at least one* presumably generates its own alternatives, and it has in fact received quite a lot of attention in the literature on so-called *modified numerals* (following Geurts and Nouwen 2007). In fact, in light of recent work on modified numerals, the predicted implicature should be weaker than what K&R assume. For example, Coppock et al. (2016) would predict that the implicature of *At least one student talked to five professors* is 'The speaker does not believe that more than one student talked to six or more professors, or if she believes that one did, then she only knows a lower bound to the number of professors that this student talked to'.[24] Without going into further detail—(92) is far more complex than the examples usually discussed in current theories of modified numerals—it seems hard to tell what the predicted implicatures of (92) should be and whether they really challenge the theory.

Instead of trying to derive what the implicatures are for quantified sentences with embedded questions, a more theory-neutral approach could simply compare them to similar sentences involving known implicature triggers. My prediction is that they should generate similar implicatures, independently of the theory we adopt to derive them. If one looks at the few examples in (94) and (95), there do not seem to be any obvious discrepancies.

(93)   Context: Each child received a bag with jelly beans of various colors.

(94)   a.   At least one child ate the [blue]$_F$ jelly beans.
            $\not\rightsquigarrow$ No child ate any jelly beans of any other color

       b.   No child ate the [blue]$_F$ jelly beans.
            $\rightsquigarrow$ For each other color, at least one child ate jelly beans of this color
            $\rightsquigarrow$ For at least one other color, at least one child ate jelly beans of this color

       c.   Every child ate the [blue]$_F$ jelly beans.
            $\rightsquigarrow$ For each other color, at least one child didn't eat jelly beans of this color
            $\rightsquigarrow$ For each other color, no child ate jelly beans of this color

(95)   a.   At least one child predicted who sang. *[repeated from (92)]*
            $\not\rightsquigarrow$ No child made any false prediction

       b.   No child predicted who sang.
            $\rightsquigarrow$ For each x who did not sing, at least one child predicted that x sang

---

[24] **Proof:** Consider the alternative $\psi_{1,6}$ = 'At least one student talked to six professors', which is more informative than the assertion. The speaker could have uttered $\psi_{1,6}$ if she knew that at least one student talked to six or more professors and if she didn't know whether there was one such student or more than one. The primary implicature derived by their system is thus the following disjunction: either the speaker does not believe that $\psi_{1,6}$ is true or she knows whether only one student or at least two students talked to more than six professors. If the latter, she would not have been able to sincerely assert the prejacent (she would have known that more than one student talked to at least five professors). Thus, either she does not know whether any student talked to six or more professors or she knows that only one did. Alternatives of the form $\psi_{1,n}$ with $n \geq 7$ give rise to similar disjunctive primary implicatures. If we write α_n for the set of states which do not support $!\psi_{1,n}$ and β_n for those which support 'Exactly one student talked to n or more professors', we can observe the following: $\alpha_n \subseteq \alpha_{n+1}$ and $\beta_{n+1} \subseteq \beta_n$ (the latter is true only because we restrict our set to states that satisfy **sincere**(ϕ)). We can show that $\bigcap_{n\geq 6}(\alpha_n \cup \beta_n) = \alpha_6 \cup \bigcup_{n\geq 6}(\alpha_{n+1} \cap \beta_n)$. This can be paraphrased with the following disjunction: either the speaker doesn't know whether any students talked to more than five students (α_6) or she knows that exactly one student talked to at least n professors for some $n \geq 6$ and doesn't know whether this student talked to more than n professors.

    c. Every child predicted who sang.
       ⤳ For each x who did not sing, at least one child didn't predict that x sang
       ⤳̸ For each x who did not sing, no child predicted that x sang

In particular, it seems clear that (94a) also lacks the implicature one would derive by considering only the alternatives of the adjective *blue*, thus suggesting that the issue with (92/95a) is more likely a general puzzle for the theory of implicatures than some unexplained restriction on the scope of EXH for embedded questions. Testing the implicatures of these sentences seems to be an important step towards an experimental validation of my theory (and exhaustification-based theories of embedded questions more generally).

9 Conclusion

Following previous literature, I proposed a treatment of embedded questions as definite descriptions of their answers, in order to account for plurality effects. This idea had originally been proposed to account for the uniqueness presupposition of singular questions and multiple *wh*-questions (Dayal 1996; Fox 2012) and was later shown to be important for accounts of plurality effects (Lahiri 2002; Gajewski 2005). However, I diverged from most of this literature and, following Xiang (2016), adopted a categorial approach, in which what is described is a short answer instead of a propositional answer.

I showed that such a theory can be complemented with a module that derives false-answer sensitive readings through exhaustification. This extension turns out to be very powerful, in that, starting from very general hypotheses, it makes correct predictions for a wide range of sentences and accounts for phenomena for which an *ad hoc* explanation had to be postulated so far (e.g., intermediate QVE readings, non-reducibility of *forget*, agent ignorance of negated *know-wh*). The theory also has the advantage of being supported by psycholinguistic data on questions embedded under *know*. In particular, it explains why young children seem to prefer a WE reading and why the most prominent reading for adults is the IE reading, which corresponds to regular matrix exhaustification, whereas the SE reading, in this framework, is derived through local exhaustification, the availability of which is presumably more constrained.

There is a strong intuition, even among proponents of the exhaustification approach to embedded questions, that the EXH operating on questions is distinct from the operator responsible for other implicatures (in particular, scalar implicatures). Xiang (2015) argues against the exhaustification approach by pointing to cases where the behavior of the no-false-beliefs inference differs from run-of-the-mill scalar implicatures. In particular, the no-false-beliefs inference is harder to cancel: it persists in downward entailing environments, and it does not seem to be computed above matrix negation. I showed how a careful account of the neg-raising properties of *believe* and of the projection of homogeneity through EXH solves issues related to negative sentences, and that the issue with quantified sentences has more to do with the theory of implicatures than with embedded questions specifically. Regarding the other issues, I would like to point out that exact readings of numerals share most of the properties that distinguish

strengthened readings of embedded questions from the usual scalar implicatures: they are hard to cancel and easy to embed (even in downward entailing environments). Yet they are often analyzed in an exhaustification-based theory (see Spector 2013). We could thus imagine that the different properties we discussed are correlates of one underlying property (e.g., easily accessible alternatives) but do not necessarily reflect the presence of distinct operators. Variation along this dimension would explain the difference with scalar triggers like *some*, just as some underlying properties of scales may explain the differences observed by van Tiel et al. (2016).[25]

It will be hard to settle this debate without quantitative data (for instance, on correlations between the derivation of implicatures, strengthened readings of questions, and exact readings of numerals). This may be an interesting topic for future research on the psycholinguistics of embedded questions. What I have tried to show here is that an exhaustification-based approach is perfectly compatible with the data currently available and offers clear advantages when it comes to complex sentences which, for alternative theories, usually require a lot of *ad hoc* hypotheses.

To conclude, I would like to mention a few topics I have not addressed in this paper. First, I only looked at responsive predicates and ignored rogative predicates like *wonder* entirely. Beck and Sharvit (2002) suggest that these predicates also give rise to QVE and possibly other plurality effects. Given the importance of the *Ans* operator in my theory, I cannot account for any plurality effect with rogative predicates. This brings us to another issue I have only skimmed over: that of selectional restrictions. I briefly mentioned a hypothetical reason why *believe* and some communication verbs do not embed questions, but one could also wonder why rogative predicates do not embed declarative complements, or why emotive-factives such as *surprise* do not embed *whether*-questions. The theory presented in this paper offers little insight into this question, but recent work by Mayr (2017) builds on Klinedinst and Rothschild (2011) and Gajewski's (2005) theory of neg-raising to offer a new explanation within an exhaustification-based approach to exhaustivity. Finally, Phillips and George (2018) show that mention-some questions such as *Where can I get gas?* are also sensitive to false beliefs. Accounting for mention-some questions would require relatively important changes in Sect. 3 (see Xiang 2016 and Fox 2018 for recent accounts), but there is good hope that applying the exhaustification procedure described in Sect. 5 to the denotations generated by a proper account of mention-some would yield a no-false-belief reading.

---

[25] Barner et al. (2011) propose that children's specific difficulty with scalar implicatures stems from a difficulty in accessing lexical alternatives (in contrast with other implicatures, which do not require accessing lexical alternatives). Note that my theory predicts that lexical alternatives are still involved in the computation of the IE reading of *know*. The fact that children need to access the lexical alternative *believe* in order to compute this reading would explain their difficulty in Cremers et al.'s (2017) experiment.

Appendix: Global-FAS reading of QV sentences

This section contains proofs of results discussed in Sect. 6.2 for the global exhaustification of the simple QVE sentence (1b), repeated below, with the assumed parse (65c).

(1b)   John mostly knows which children sang.

(65c)   EXH$\left[\text{John 1 [C } Ans^S \text{ which children sang] [mostly 2 } t_1 \text{ knows } f \ t_2]\right]$

I first show how keeping all alternatives yields a very peculiar interpretation, usually too strong. I then show how pruning a certain set of alternatives can yield a more suitable interpretation. The proof is for a standard *wh*-question with a distributive upward-entailing predicate.

A.1 With all alternatives

As discussed in the main text, there are three sources of alternatives for (65c): the scale $\langle know, believe \rangle$, deletion of *mostly* (or equivalently replacement of *mostly* with *completely*), and the short answer. Keeping the short answer as is, we have four alternatives, and those without *mostly* are innocently excludable and give us the implicature that John doesn't know the complete answer (only most of it). I assume that the threshold for *mostly* is more than $\frac{1}{2}$, but less than $\frac{2}{3}$. The lower bound seem uncontroversial. Adopting a higher upper bound would only increase the implicated upper bound on the number of actual singers.

Alternatives involving replacement of the short answer can be ordered by their cardinality. Those with an atomic child (cardinality 1) are not innocently excludable, because excluding them all would contradict the prejacent (John would not know anything). Alternatives involving a pair of children (cardinality 2) are not innocently excludable either, since that would mean that John only believes that one child sang. This is in contradiction with the prejacent which presupposes that (a) there are at least two children who sang and (b) John knows of a majority of these at least two children that they sang. Assuming that $\alpha > \frac{1}{2}$, this means that he knows of at least two children that they sang. For a similar reason, the alternatives with a triplet of children and the adverb *mostly* cannot all be excluded, as that would entail that John knows about at most one child (if he knew about two, then he would know about most of any triplet containing these two).

By contrast, alternatives with a triplet of children (cardinality 3) but without *mostly* and all alternatives with cardinality 4 or more are innocently excludable since they can all be excluded together with the "completely knows" alternative without making the prejacent unsatisfiable. Indeed, assuming there are three children who actually sang

and John only knows about two of them, the prejacent is true (since $\alpha < \frac{2}{3}$), and none of the mentioned alternatives is satisfied.

In the end, without any restriction on the set of alternatives, (65c) conveys (a) that exactly three children sang and (b) that John only knows about two of them.

A.2 Pruning alternatives up to a certain cardinality

Here we work under the assumption that the actual true answer to the question *Which children sang?* is known and common ground. Let us write a_0 for the plurality of children who did sing and $n_0 = |a_0|$. Let β be a number between α and 1 (strictly). I will discuss the interpretation obtained from parse (65c) when, among alternatives obtained by replacement of $Ans^S(Q)(C)$ with a plurality a, only alternatives such that a is a plurality in which the proportion of true answers doesn't exceed β (i.e., $|At(a) \cap At(a_0)| \leq \beta |At(a)|$) are kept. I will show that the result then is the conjunction of (67a), (67b), and (67c).

(65c)   EXH$\big[$John 1 [C Ans^S which children sang] [mostly 2 t_1 knows f t_2]$\big]$

(96)   a.   John knows at least $\lceil \alpha n_0 \rceil$ true answers

     b.   John has no false belief regarding children who did not sing

     c.   John knows less than $\min(\alpha \lceil \beta^{-1} n_0 \rceil, n_0)$ true atomic answers.

In Sect. A.2.1, I will give the proof, and in Sect. A.2.2 I will explain what it means exactly and what value we should choose for β.

A.2.1 Proof of the equivalence

Let n be the number of true answers believed by John.

(67a) simply is the prejacent. (67b) derives from the negation of alternatives containing a single false answer and *completely* (or deletion of *mostly*). (67c) derives from the negation of all alternatives with *believe* and *mostly* built with $a_0 \oplus b$ such that b contains only false answers and $\frac{n_0}{n_0 + |b|} \leq \beta$. By denying for any such answer that John believes most of it, we can put an upper bound on n: John knows n atoms of a_0 and, given (67b), none of b. We therefore conclude that for any b such that $\frac{n_0}{n_0 + |b|} \leq \beta$, $n < \alpha(n_0 + |b|)$. From this, it follows that:

(97) $$\begin{aligned}
n &< \min_{\{b \mid n_0 + |b| \geq \beta^{-1} n_0\}} \alpha(n_0 + |b|) \\
&< \alpha \lceil \beta^{-1} n_0 \rceil \\
&\leq \lfloor \alpha \lceil \beta^{-1} n_0 \rceil \rfloor \leq n_0
\end{aligned}$$

Note that $n = n_0$ is not excluded in principle (pick $\alpha = 0.72$, $\beta = 0.75$, $n_0 = 5$ for a counter-example), but will be if β is sufficiently larger than α and n_0 sufficiently large. Otherwise, the alternative obtained by deletion of *mostly* is excludable and yields the implicature that John knows less than n_0 true answers.

Now, let us show that all the alternatives we have excluded are indeed innocently excludable, and that nothing else follows. We have excluded the alternatives with the

complete answer, which only yield the implicature that John knows less than n_0 true answers. Of all the alternatives derived by replacement of the complete answer:

– Those with more than β true answers (in proportion) have been pruned.
– Those with at least $1-\beta$ false answers (in proportion) and *know* are excludable but yield trivial implicatures since they're all presupposition failures in the context.
– Those with only false answers, *believe* but not *mostly* give us the no-false-belief implicature.
– Those with at least $1-\beta$ false answers (in proportion), *believe*, and *mostly* give us the implicature that John knows less than $\alpha \lceil \beta^{-1} n_0 \rceil$ true atomic answers.
– The rest (alternatives with *believe* and without *mostly*) are excludable but do not contribute anything beyond the previous two points.

As we can see, all the alternatives that haven't been pruned are excludable and yield a satisfiable result, hence they are all innocently excludable, and the result is the conjunction of the three terms in (96).

A.2.2 Discussion

First, let me explain what effect β has concretely. If we pick β very close to α, the upper bound on the number of true answers known by John will approach n_0, which amounts to admitting all proportions between α and 1. Note that this does not mean that EXH is vacuous, as we still derive the "not completely" implicatures, as well as no-false-belief. If, on the contrary, we pick β very close to 1, the result becomes unsatisfiable. Indeed, the upper bound $\alpha \lceil \beta^{-1} n_0 \rceil$ gets close to αn_0 itself. The interpretation would then be that the proportion of answers known by John is *exactly* α. However, since *most* is vague, (a) we don't know what α is exactly, and (b) knowing exactly α would put us in the borderline area of the vague predicate. Note that the definition of EXH I adopted in (52) relies on a notion of *satisfiability* defined in (54) which rules out borderline vague cases, but it does not require contextual satisfiability, hence does not block the problematic implicatures here.

A naive solution would be to pick the β that maximizes informativity without making the result unsatisfiable in the context. This would mean getting $\alpha \lceil \beta^{-1} n_0 \rceil$ to match the limit between borderline and true cases of the prejacent. Of course, the existence of such a limit is dubious. As noted by Dummett (1978), vague predicates give rise to *higher-order vagueness*: one cannot precisely characterize the extension of borderline cases. One recent solution to second-order vagueness has been to attribute a probability distribution over [0, 1] to α, and to reason in terms of probabilities rather than binary truth values (for probabilistic accounts of vagueness, see Franke 2011; Lassiter and Goodman 2014; Qing and Franke 2014). For instance, we could assume that the probability to derive an implicature from an alternative ψ is proportional to the probability of $\neg \psi$ being consistent with the prejacent S. The next step would be to derive a probability that John knows n true answers for each n in $[0, n_0]$ from the probability that n/n_0 is above the threshold and the probability that n satisfies any given implicature, weighted by the probability that this implicature be derived. This would obviously take us too far from the topic of this paper, but see Franke and Jäger (2016) for an introduction to Bayesian pragmatics and Leffel et al. (2018) for a discussion of

interactions between vagueness and implicatures. Keeping things suitably vague here, I will simply conclude that an implicature that John knows *significantly less* than the complete answer is derived.

B *Agree* and *disagree*

As we discussed in Sect. 7, every responsive verb must come with a non-trivial lexical restrictor. In the case of non-veridical predicates, there often is no obvious candidate and the restrictor must be postulated. *Agree* and *disagree* are no exception, but they pose numerous additional problems. I will first summarize Lahiri's (2002) extensive discussion of *agree* and then present my own analysis, which is motivated by new data (mostly on *disagree*) as well as theoretical considerations (e.g., adding homogeneity to the picture).

B.1 Lahiri's proposal

Lahiri (2002: 111–109) proposes the denotation in (98), where Com is the predicate of common belief (Fagin et al. 1995). The free variable G is filled by the *with*-phrase if one is present, or else is determined contextually. If the subject is plural, Lahiri proposes that the reciprocal reading results from a possibly covert 'with each other'. This phrase does two things: it introduces a distributivity operator D and it fills the G variable with the plural subject itself, as illustrated in (99). When *agree (on)* embeds a question, Lahiri proposes a similar structure, based on a lexical restrictor which is possibly different for each agent: $C_y^w = \lambda p.B(w)(y)(p)$. As it stands, this yields very weak truth conditions for (100): for the sentence to be true, each teacher has to believe that every other teacher agrees with him or her, but nothing guarantees that any two teachers share any beliefs regarding the question. In short, it simply predicts that each teacher imagines that all other teachers agree with them.[26]

(98) $[\![\text{agree}]\!]^w = \lambda p.\lambda x.Dox_x^w \subseteq p \wedge Dox_x^w \subseteq \left(\lambda w'.Com_G^{w'}(p)\right)$
where G is a free variable

(99) $[\![\text{The teachers agree (with each other) that } p]\!]^w$
$= D(\sigma x[{}^\star Tx])(\lambda y.[\![\text{agree}]\!]^{w,G\to\sigma x[{}^\star Tx]}(p)(y))$
$= \forall y.Ty \to [Dox_y^w \subseteq p \wedge Dox_y^w \subseteq (\lambda w'.Com_{\sigma x[{}^\star Tx]}^{w'}(p)]$

(100) $[\![\text{The teachers agree (with each other) on } Q]\!]^w$
$= D(\sigma x[{}^\star Tx])(\lambda y.[\![\text{agree}]\!]^{w,G\to\sigma x[{}^\star Tx]}(f(Ans^S(Q)(C_y^w)))(y))$
$= \forall y.Ty \to [\underbrace{Dox_y^w \subseteq f(Ans^S(Q)(C_y^w))}_{\text{trivial}} \wedge Dox_y^w \subseteq (\lambda w'.Com_{\sigma x[{}^\star Tx]}^{w'}$
$(f(Ans^S(Q)(C_y^w)))]$
$= \forall y.Ty \to [Dox_y^w \subseteq (\lambda w'.Com_{\sigma x[{}^\star Tx]}^{w'}(f(Ans^S(Q)(C_y^w)))]$

---

[26] The reason for the weakness of the resulting truth conditions seems to come from Lahiri's generalization of the formula in his example (110) to a proposition that covaries with the variable y in his example (119) (because C makes reference to y). Replacing φ with a proposition that covaries with the index i in (108) results in an invalid application of Fagin et al.'s (1995) Fixed-Point Axiom.

A natural solution would be to assume that the restrictor is determined by the G variable rather than covarying with the subject, so that the first conjunct in (100) would not be trivial anymore. In a preliminary proposal, (Lahiri, 2002, p30) proposes the restrictor in (101): the set of propositions believed by at least one individual part of G.

(101) $C_G^w = \lambda p.\exists y \preceq G : Dox_y^w \subseteq p$

Lahiri argues that this does not work when the subject consists of more than two individuals.[27,28] An independent issue is that, just like with *be certain*, this restrictor is not closed under conjunction and, combined with Lahiri's denotation for *agree*, would result in unexpected presupposition failures in my theory.[29] As an example, consider sentence (102) below in a situation where Mary believes that only Ann sang, while John believes that only Bill sang. In such a situation, 'Ann sang' and 'Bill sang' would be in the restrictor, but 'Ann and Bill sang' would not, because neither Mary nor John believes that both Ann and Bill sang. Ans^S would then return a presupposition failure. This seems wrong, because in this context (103) below sounds clearly true, suggesting that (102) must be false and not just a presupposition failure (assuming that *agree* and *disagree* share the same restrictor).

(102)   John and Mary agree on which children sang.

(103)   John and Mary disagree on which children sang.

---

[27] As an example, in the situation described in (i), the restrictor (101) would contain 24 individuals while the FBI agents would only agree on four individuals. Yet sentence (ii) is intuitively true.

  (i)   Context: There are twenty FBI agents observing a meeting of five alleged subversives in an otherwise empty building. They agree completely on who four of those five are, but disagree about the fifth, about whose identity they each have a different hypothesis.

  (ii)   The FBI agents agree on who the people in the building are, for the most part.

While the judgment is clear, the effect is intuitively linked to the fact that the FBI agents agree that there are five people in the building and only disagree on the identity of *one* person. It seems that a better model of the complement in (ii) would be the family of questions {⟦who person 1 is⟧, ⟦who person 2 is⟧, ...}. With this denotation for the questions, the naive proposal correctly predicts (ii) to be true in situation (i). Interestingly, it predicts the sentence to be false in a context where the FBI agents agree that a, b, c, d, and e are the five people in the building but each agent has a different opinion on which of the people they see is a, which is b, and so on. I won't try to give a proper analysis of identity questions and how their QV readings should be analyzed.

[28] An independent issue with the solution Lahiri proposes is that it does not predict the correct truth conditions for *agree-with* constructions. Indeed, his semantics requires that for each individual part x in the subject, most of what x believes is agreed on by the members of the free variable G. Yet, in context (i) below, (ii) seems to be much worse than (iii). This is not expected since it is true that FBI agents all agree that Ann and Bill are in the building. The fact that (iii) is better than (ii) in this context suggests that the relation between the subject of *agree* and the complement of *with* is the reverse of what Lahiri proposes.

  (i)   Context: John believes that Ann and Bill are in the building and is agnostic as to who else may be present. The FBI agents believe Ann, Bill, and ten other individuals to be hiding in the building.

  (ii)   John agrees with the FBI agents on who is in the building.

  (iii)   The FBI agents agree with John on who is in the building.

[29] Thanks to Wataru Uegaki for pointing this to me.

As with *be certain*, we could try to fix this by assuming that the actual restrictor is the closure under conjunction of (101), but doing so would be neither necessary nor sufficient this time. Rather, I will stick to this simple restrictor but assume a different structure, which will ensure that C_g^w is always closed under conjunction anyway.

B.2 A new solution

I will assume the denotation in (104a) for *agree*, and that *agree with each other* is interpreted as in (104b,c). The main novelties here are (i) a first presupposition which solves the weakness identified in Lahiri's proposal (as well as the issue discussed in footnote 26) and will strongly reduce the possible uses of *disagree*, (ii) the treatment of the '$Dox \subseteq (\lambda w.Com^w \dots)$' component as a presupposition here (otherwise the truth conditions for *disagree* would be very weak), and (iii) a different analysis of *each other*. I propose to interpret this expression very literally and assume that *each* applies to *other* and not to the subject. This means that an additional covert distributive operator is needed for the subject, and so in the general case the *agree*-relation will apply to pairs of atomic individuals. We will see how this better captures new facts about *disagree*.

(104) a. $[\![\text{agree}]\!]^{w,G} = \lambda p.\lambda x : \underline{Com_G^w(p) \wedge Dox_x^w \subseteq (\lambda w'.Com_G^{w'}(p)).Dox_x^w \subseteq p}$

b. $[\![X \text{ agree (with each other) that } p]\!]^w$
$$= D(X)(\lambda x.D(X-x)(\lambda y.[\![\text{agree}]\!]^{w,G \to y}(p)(x)))$$

c. $[\![X \text{ agree (with each other) on } Q]\!]^w$
$$= D(X)\left[\lambda x.D(X-x)\left[\lambda y.[\![\text{agree}]\!]^{w,G \to y}\left(f(Ans^S(Q)(C_y^w))\right)(x)\right]\right]$$

Unless the parts of X are not atomic individuals (e.g., in a context where there are salient subgroups which may agree or disagree on an issue), the variable y ends up denoting atomic individuals. In this case we can replace all occurrences of Com with simple belief.[30]

A consequence of these assumptions is that the lexical restrictor covaries with the *with*-phrase, which itself covaries with the atoms of the subject, so IR must target a position below the *with*-phrase.

B.3 Predictions for *agree*

Let's look at a declarative complement embedded under *agree* with two agents, as in (105). The top-most distributivity operator splits the subject into its atomic parts, yielding the proposition in (105a) for Mary and the symmetric one in (105b) for John. Before going further, we need to specify how presuppositions project from the distributivity operator.

(105) Mary and John agree (with each other) that Ann sang.

---

[30] *Com* still plays a non-trivial role when explicit plural complements are used. For instance, 'Mary agrees with the teachers that Ann sang' presupposes that the teachers are in *Com*-agreement that Ann sang. Whether *Com* may be replaced with regular belief in such cases is an open empirical question.

$$\text{a. } D(m \oplus j - m)\big[\lambda y.\underline{Com_y^w(p) \wedge Dox_m^w \subseteq (\lambda w'.Com_y^{w'}(p))}.Dox_m^w \subseteq p\big]$$
$$= \underline{Com_j^w(p) \wedge Dox_m^w \subseteq (\lambda w'.Com_j^{w'}(p))}.Dox_m^w \subseteq p$$
$$= \underline{Dox_j^w \subseteq p \wedge Dox_m^w \subseteq (\lambda w'.Dox_j^{w'} \subseteq p)}.Dox_m^w \subseteq p$$

$$\text{b. } \underline{Dox_m^w \subseteq p \wedge Dox_j^w \subseteq (\lambda w'.Dox_m^{w'} \subseteq p)}.Dox_j^w \subseteq p$$

One solution is to see the distributive operator as a big symmetric conjunction, and to generalize the projection pattern of simple symmetric conjunctions (e.g., using local contexts from Schlenker 2009).

(106) Projection patterns for conjunctions:

a. $\underline{p}.S \wedge \underline{p'}.S'$ presupposes $p \wedge (S \rightarrow p')$

b. $\bigwedge_i [\underline{p_i}.S_i]$ presupposes $\forall i, \Big[\bigwedge_{j \neq i} S_j\Big] \rightarrow p_i$

c. $D(X)\Big(\lambda y.\underline{\rho(y)}.P(y)\Big)$ presupposes $\forall x \in At(X), [D(X - x)(\lambda y.P(y))]$ $\rightarrow \rho(x)$

The result is the following:

(107) $[\![105]\!]^w = \underline{[Dox_m^w \subseteq p]} \rightarrow \big[Dox_m^w \subseteq p \wedge Dox_j^w \subseteq (\lambda w'.Dox_m^{w'} \subseteq p)\big]$
$$\wedge \underline{[Dox_j^w \subseteq p]} \rightarrow \big[Dox_j^w \subseteq p \wedge Dox_m^w \subseteq (\lambda w'.Dox_j^{w'} \subseteq p)\big]$$
$$\wedge \, Dox_m^w \subseteq p \wedge Dox_j^w \subseteq p$$
$$= [Dox_m^w \subseteq p] \rightarrow \big[Dox_j^w \subseteq (\lambda w'.Dox_m^{w'} \subseteq p)\big]$$
$$\wedge \underline{[Dox_j^w \subseteq p]} \rightarrow \big[Dox_m^w \subseteq (\lambda w'.Dox_j^{w'} \subseteq p)\big]$$
$$\wedge \, Dox_m^w \cap Dox_j^w \subseteq p$$

The assertive component of (105a) entails the first presupposition of (105b), so only the second presupposition is non-trivial, and the result is that if Mary believes that Ann sang, John knows that she does (and vice versa). Simplifying the conditional presuppositions, (105) asserts that John and Mary both believe p and, given the assertion, presupposes that they each know that the other believes p. Note that this is weaker than what Lahiri intended with the *Com* operator. In particular, the sentence can be true even if John or Mary is not aware that the other knows that they share the belief that Ann sang. In such a situation, Mary would be able to say "I agree with John that Ann sang", but she would not be able to say "John agrees with me that Ann sang" or "John and I agree that Ann sang".

(108) Mary, Nancy, and Sue agree (with each other) that Ann sang.

Generalizing this to more than two agents, as in (108), results in the same kind of pairwise weak agreement: each of Mary, Nancy, and Sue must believe that Ann sang and is presupposed to know that the others believe that Ann sang, but not necessarily know that the others know that she does.

Let's now come back to questions embedded under *agree*, as in (102), repeated below. The predicted truth conditions are developed in (109). The denotation is equivalent to the conjunction in (109a). (109b) details the meaning of the first conjunct. The first conjunct in the presupposition is trivial: it states that John believes whatever he believes. The second conjunct states that Mary is aware of what John believes.[31] For simplicity, I ignore presuppositions arising from the definite description in Ans^S.[32] (109c) gives the final result, with the conditional presuppositions resulting from the symmetric conjunction. The assertion is simply that John believes every atomic answer that Mary believes, and vice versa. The presuppositions are quite hard to parse, but they state that if John believes what Mary believes, then Mary is aware of what he believes, and vice versa. Given the assertion, the presuppositions can be greatly simplified and simply state that each of John and Mary is aware of what the other believes. In case John and Mary have overlapping but different beliefs, homogeneity kicks in and the sentence is neither true nor false. If they have non-overlapping beliefs, the conditional presuppositions are trivially satisfied and the sentence is plain false.

(102)   John and Mary agree on which children sang.

(109)   a. $[\![(102)]\!]^{w_0} = [\![\text{agree}]\!]^{G \to j}(f(Ans^S(Q)(C_j^{w_0})))(m) \wedge [\![\text{agree}]\!]^{G \to m}$
$$(f(Ans^S(Q)(C_m^{w_0})))(j)$$

   b. $[\![\text{agree}]\!]^{G \to j}(f(Ans^S(Q)(C_j^{w_0})))(m) = \underline{Com_j^{w_0}(f(Ans^S(Q)(C_j^{w_0})))}$

$\wedge \underline{Dox_m^{w_0} \subseteq (\lambda w.Com_j^w(f(Ans^S(Q)(C_j^{w_0}))))}$

$\wedge Dox_m^{w_0} \subseteq f(Ans^S(Q)(C_j^{w_0}))$

$= \underline{Dox_m^{w_0} \subseteq (\lambda w.Dox_j^w \subseteq f(Ans^S(Q)(C_j^{w_0})))}$

$\wedge Dox_m^{w_0} \subseteq f(Ans^S(Q)(C_j^{w_0}))$

   c. $[\![(102)]\!]^{w_0} = \Big[\Big[Dox_j^{w_0} \subseteq f(Ans^S(Q)(C_m^{w_0})) \Big] \to$

$$\Big[Dox_m^{w_0} \subseteq (\lambda w.Dox_j^w \subseteq f(Ans^S(Q)(C_j^{w_0}))) \Big] \Big]$$

$\wedge \Big[\Big[Dox_m^{w_0} \subseteq f(Ans^S(Q)(C_j^{w_0})) \Big] \to$

$$\Big[Dox_j^{w_0} \subseteq (\lambda w.Dox_m^w \subseteq f(Ans^S(Q)(C_m^{w_0}))) \Big] \Big]$$

$\wedge Dox_m^{w_0} \subseteq f(Ans^S(Q)(C_j^{w_0})) \wedge Dox_j^{w_0} \subseteq f(Ans^S(Q)(C_m^{w_0}))$

In a situation with more than two agents, such as (110), the predicted meaning can ultimately be reduced to pairwise agreement: (110) simply means 'Each teacher agrees with each other teacher on which children sang'. Homogeneity ensures that the

---

[31] Thanks to an anonymous reviewer for catching an error in an earlier version of this manuscript which made this presupposition trivial as well.

[32] This would translate as a presupposition that $C_{j/m}^{w_0}$ is non-empty, which seems to project existentially and lead to the overall presupposition that one of John or Mary must believe that at least one child sang (see Chemla and George 2016). Similarly, there would be a presupposition that at least one of John or Mary believes the other to have non-trivial beliefs regarding Q.

sentence is plain false only if no two teachers have overlapping beliefs. On top of this, the presupposition, given the assertion, can be simplified to 'Each teacher is aware of what each other teacher believes'.

(110)   The teachers agree (with each other) on which children sang.

If John or Mary [or any teacher in the case of (110)] believes that a given child sang, the other one (resp. all other teachers) must also believe so. Nevertheless, this leaves open the possibility that one be ignorant as to whether a child sang or not, as long as the other one (resp. all other teachers) is ignorant too or believes that this particular child did not sing. This is in line with the results of Chemla and George (2016). They show that in a situation where Mary believes a, b and is agnostic regarding c, (102) is judged false if John believes a, b, and c to be true, but true if John believes a and b to be true but c to be false. While they interpret this result as an indication that *agree-wh* is not reducible to *agree-that*, I show that one can account for this fact in a reducible fashion (except for the lexical restrictor).[33]

B.4 Predictions for *disagree*

The theory imposes strong restrictions on possible uses of *disagree*. Indeed, the conditional presuppositions of (111) are of the form (112). Since the consequent is incompatible with the antecedent, the conditional is equivalent to the negation of the antecedent. As a result, (111) is predicted to presuppose that both Mary and John believe that Ann sang, but it makes an assertion which contradicts this presupposition. It is thus deviant (necessarily false or a presupposition failure), and this is reflected in grammaticality judgments (Gajewski 2002). More generally, the expression *disagree with each other* is predicted never to embed declarative complements.[34]

(111)   # Mary and John disagree (with each other) that Ann sang.

(112) $\left[Dox_m^w \not\subseteq p \rightarrow Dox_m^w \subseteq p \land Dox_j^w \subseteq (\lambda w'.Dox_m^{w'} \subseteq p) \right] \equiv Dox_m^w \subseteq p$

With embedded questions no contradictions arise, since each individual is related to a different proposition (the answers believed by another one). (114) below gives the derivation of the meaning of (113). (114b) corresponds to the first conjunct in (114a): Mary disagrees with John on which children sang. Because of homogeneity, this is only true if Mary believes none of the atomic answers believed by John. The presupposition is simply that Mary is aware of what John believes. Putting everything together in (114c), we can simplify the result a bit (because the non-overlap condition of*disagree* is symmetric, unlike the inclusion condition of *agree*). We obtain a rather simple assertion: John and Mary's beliefs about Q do not overlap. The conditional

---

[33] Note that this may not have been an ideal test case of non-reducibility to begin with, because the two situations can be distinguished on the basis of 'agree that' facts, provided that we distinguish presupposition failures from falsity: in the first case 'Mary agrees with John that c' is plain false, whereas in the second case it results in a presupposition failure.

[34] As a reviewer points out, replacing *disagree* with *do not agree* may alleviate the issue by allowing the negation to scope above the distributive operator. In that case, we obtain the negation of (105): at least one of John and Mary does not believe that Ann sang, and it is presupposed that if one of them believes that she did, the other one is aware of this.

presupposition states that if John and Mary disagree, then they're each aware of what the other believes.

(113) Mary and John disagree with each other on which children sang.

(114)   a. $[\![(113)]\!] = [\![\text{disagree}]\!]^{G\to j}(f(Ans^S(Q)(C_j^{w_0})))(m)$
$\qquad\qquad\qquad \wedge [\![\text{disagree}]\!]^{G\to m}(f(Ans^S(Q)(C_m^{w_0})))(j)$

   b. $[\![\text{disagree}]\!]^{G\to j}(f(Ans^S(Q)(C_j^{w_0})))(m) =$
$\qquad\qquad\qquad \underline{Dox_m^{w_0} \subseteq (\lambda w. Dox_j^w \subseteq f(Ans^S(Q)(C_j^{w_0}))) \wedge Dox_m^{w_0} \cap}$
$\qquad \underline{f(Ans^S(Q)(C_j^{w_0})) = \emptyset}$

   c. $[\![(113)]\!]^{w_0} = \left[\left[\underline{Dox_j^{w_0} \cap f(Ans^S(Q)(C_m^{w_0})) = \emptyset}\right] \to \right.$
$\qquad\qquad\qquad\qquad \underline{\left[Dox_m^{w_0} \subseteq (\lambda w. Dox_j^w \subseteq f(Ans^S(Q)(C_j^{w_0})))\right]}\left.\right]$
$\qquad \wedge \left[\left[\underline{Dox_m^{w_0} \cap f(Ans^S(Q)(C_j^{w_0})) = \emptyset}\right] \to \right.$
$\qquad\qquad\qquad\qquad \underline{\left[Dox_j^{w_0} \subseteq (\lambda w. Dox_m^w \subseteq f(Ans^S(Q)(C_m^{w_0})))\right]}\left.\right]$
$\qquad \wedge \underline{Dox_m^{w_0} \cap f(Ans^S(Q)(C_j^{w_0})) = \emptyset \wedge Dox_j^{w_0} \cap f(Ans^S(Q)(C_m^{w_0})) = \emptyset}$
$\qquad = \left[\underline{f(Ans^S(Q)(C_j^{w_0})) \cap f(Ans^S(Q)(C_m^{w_0})) = \emptyset}\right] \to$
$\qquad\qquad\qquad\qquad \left[\left[Dox_j^{w_0} \subseteq (\lambda w. Dox_m^w \subseteq f(Ans^S(Q)(C_m^{w_0})))\right]\right.$
$\qquad\qquad\qquad \underline{\wedge \left[Dox_j^{w_0} \subseteq (\lambda w. Dox_m^w \subseteq f(Ans^S(Q)(C_m^{w_0})))\right]}\left.\right]$
$\qquad \wedge \left[f(Ans^S(Q)(C_j^{w_0})) \cap f(Ans^S(Q)(C_m^{w_0})) = \emptyset\right]$

This result may seem too strong, because it predicts (113) to be a truth value gap in context (115), whereas intuitively it feels true. However, we will see that an independent strongly exhaustive reading may account for this case.

(115) Context: Mary believes that only Ann and Bill sang, John believes that only Ann and Chris sang.

This proposal also predicts that any situation in which there are more agents than possible atomic answers to the question should be disallowed, since it would force at least two agents to have overlapping beliefs. This would rule out sentences like (116) and (117), which in actual fact seem quite acceptable.

(116) The twenty teachers disagree with each other on which of the three children sang.

(117) Mary, John, and Sue disagree with each other about whether Ann sang.

One way out of this problem would be to allow more flexibility in the ways a plural subject can be split into different parts. We could assume, for instance, that the twenty teachers can be split into three groups—those who believe that Ann sang, those who believe that Bill sang, and those who believe that Chris did—and that *each* does not

distribute down to atomic teachers but only to these subgroups. Since each subgroup would disagree with every other subgroup, (116) could be true even though there are pairs of teachers who do agree with each other. While this may be a good explanation for the acceptability of (116), it seems less tenable for an example like (117), where the atoms are explicitly mentioned.

B.5 Strengthening

Before discussing possible exhaustifications of (102), repeated below, I need to spell out the full structure I assume for this sentence. It is shown in (118) with all dependencies. This structure provides many possible positions for EXH, listed in (119), which I will now review.

(102) John and Mary agree on which children sang.

(118) $[D$ [John and Mary]] 1 [(with each other$_1$)] 2 $[C_2$ Ans Q] 3 t_1 agree$_2$ $[f\ t_3]$

(119) a. $[D$ [John and Mary]] 1 [(with each other$_1$)] 2 $[C_2$ Ans Q] 3 t_1 agree$_2$ EXH$[f\ t_3]$

   b. $[D$ [John and Mary]] 1 [(with each other$_1$)] 2 $[C_2$ Ans Q] 3 EXH$[t_1$ agree$_2$ $[f\ t_3]]$

   c. $[D$ [John and Mary]] 1 [(with each other$_1$)] 2 EXH$[[C_2$ Ans Q] 3 t_1 agree$_2$ $[f\ t_3]]$

   d. $[D$ [John and Mary]] 1 EXH$[[$(with each other$_1$)] 2 $[C_2$ Ans Q] 3 t_1 agree$_2$ $[f\ t_3]]$

   e. EXH$[[D$ [John and Mary]] 1 [(with each other$_1$)] 2 $[C_2$ Ans Q] 3 t_1 agree$_2$ $[f\ t_3]]$

The most local exhaustification, in (119a), leads to the usual strongly exhaustive reading for *agree*: agents must all have exhaustive opinions on the situation and their beliefs must match perfectly. This reading was not detected by Chemla and George (2016).

The exhaustification in (119b) turns out to be vacuous (all alternatives stronger than the prejacent end up being presupposition failures when the values for 2 and 3 are filled in).

(119c,d) are vacuous as well, because all alternatives are Strawson-entailed.

In the case of (119e), the alternatives are essentially of the form (105), where 'Ann' is replaced with an arbitrary, possibly plural, individual d. Because the alternatives only have very weak conditional presuppositions, they are not Strawson-entailed anymore (in fact, their presupposition is entailed by the prejacent). They cannot all be excluded, as this would violate the presupposition that John and Mary believe at least one answer, but it is possible to exclude all non-atomic alternatives without contradiction. This would just mean that John and Mary believe that only one child sang (and agree on which child it is). In short, we find ourselves in a situation close to what we had with QVE in Sect. 6.2. Pruning the alternative obtained from d simply allows the answer on which John and Mary agree to be d, but does not add a useful implicature.

To summarize, local exhaustification yields an unattested strong exhaustive reading. Exhaustification in any higher position but above the subject's distributive operator is vacuous, and global exhaustification yields an extremely strong reading.

For *disagree*, the predicted strongly exhaustive reading is logically independent from the weakly exhaustive one: it presupposes exhaustive opinionatedness from each agent, but removes homogeneity, thus making it possible for sentences with *disagree* to be judged true even in cases of partial agreement, such as (115). Exhaustification anywhere above the verb is vacuous. This includes global exhaustification since the alternatives are similar to (111), and as such contradictory.

References

Aloni, M., D. Beaver, B. Clark, and R. Van Rooij. 2007. The dynamics of topics and focus. In *Questions in Dynamic Semantics*, vol. 17, ed. M. Aloni, A. Butler, and P. Dekker, 123–145. CRiSPI. Amsterdam: Elsevier.

Anvari, A. 2018. *Logical integrity: from Maximize Presupposition! to mismatching implicatures*. Ms., ENS.

Barker, C. 2007. Parasitic scope. *Linguistics and Philosophy* 30 (4): 407–444.

Barner, D., N. Brooks, and A. Bale. 2011. Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition* 118: 87–96.

Beaver, D.I. 1995. Presupposition and assertion in dynamic semantics. PhD dissertation, University of Edinburgh.

Beck, S., and H. Rullmann. 1999. A flexible approach to exhaustivity in questions. *Natural Language Semantics* 7 (3): 249–298.

Beck, S., and U. Sauerland. 2000. Cumulation is needed: A reply to Winter (2000). *Natural Language Semantics* 8 (4): 349–371.

Beck, S., and Y. Sharvit. 2002. Pluralities of questions. *Journal of Semantics* 19 (2): 105–157.

Berman, S.R. 1991. On the semantics and logical form of wh-clauses. PhD dissertation, University of Massachusetts at Amherst.

Champollion, L. 2010. Parts of a whole: Distributivity as a bridge between aspect and measurement. PhD dissertation, University of Pennsylvania.

Champollion, L. 2014. *Each* vs. *all*: distributivity, cumulativity, and two types of collectivity. Talk at the 4th Cornell Workshop in Linguistics and Philosophy. Ms., NYU.

Champollion, L. 2017. Distributivity, collectivity and cumulativity. In *Companion to semantics*, ed. L. Matthewson, H.R. Meier, and T.E. Zimmermann. Hoboken: Wiley.

Chemla, E. 2008. An epistemic step for anti-presuppositions. *Journal of Semantics* 25: 141–173.

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

Chierchia, G. 2004. Scalar implicatures, polarity phenomena, and the syntax-pragmatics interface. In *Structures and beyond*, vol. 3, ed. A. Belletti, 39–103. Oxford: Oxford University Press.

Chierchia, G., D. Fox, and B. Spector. 2012. Scalar implicature as a grammatical phenomenon. In *Semantics: An international handbook of natural language meaning*, vol. 3, ed. C. Maienborn, K. von Heusinger, and P. Portner, 2297–2331. Berlin: Mouton de Gruyter.

Ciardelli, I., J. Groenendijk, and F. Roelofsen. 2013. Inquisitive semantics: a new notion of meaning. *Language and Linguistics Compass* 7 (9): 459–476.

Cooper, R. 1979. The interpretation of pronouns. *Syntax and Semantics* 10: 61–92.

Coppock, E., I. Ciardelli, and F. Roelofsen. 2016. Implicatures of modified numerals: quality or quantity? Presentation at Sinn und Bedeutung 21, Edinburgh.

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

Cremers, A., and E. Chemla. 2017. Experiments on the acceptability and possible readings of questions embedded under emotive-factives. *Natural Language Semantics* 25 (3): 223–261. https://doi.org/10.1007/s11050-017-9135-x.

Cremers, A., L. Tieu, and E. Chemla. 2017. Children's exhaustive readings of questions. *Language Acquisition* 24 (4): 343–360. https://doi.org/10.1080/10489223.2017.1366207.

Dayal, V. 1996. *Locality in WH quantification: Questions and relative clauses in Hindi*. Dordrecht: Kluwer.

Dummett, M. 1978. *Truth and other enigmas*. Cambridge, MA: Harvard University Press.

Fagin, R., J.Y. Halpern, Y. Moses, and M. Vardi. 1995. *Reasoning about knowledge*, vol. 4. Cambridge, MA: MIT Press.

Fox, D. 2007. Free choice disjunction and the theory of scalar implicature. In *Presupposition and implicature in compositional semantics*, ed. U. Sauerland and P. Stateva, 71–120. New York: Palgrave Macmillan.

Fox, D. 2012. Lecture notes on the semantics of questions. Class notes (MIT Seminars).

Fox, D. 2018. Partition by exhaustification: Comments on Dayal 1996. Ms., MIT.

Franke, M. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4 (1): 1–82. https://doi.org/10.3765/sp.4.1.

Franke, M., and G. Jäger. 2016. Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft* 35 (1): 3–44.

Gajewski, J.R. 2002. On analyticity in natural language. Ms., MIT.

Gajewski, J.R. 2005. Neg-raising: Polarity and presupposition. PhD dissertation, MIT.

Gajewski, J.R., and Y. Sharvit. 2012. In defense of the grammatical approach to local implicatures. *Natural Language Semantics* 20 (1): 31–57.

George, B.R. 2011. Question embedding and the semantics of answers. PhD dissertation, UCLA.

George, B.R. 2013. Knowing-'wh', mention-some readings, and non-reducibility. *Thought: A Journal of Philosophy* 2 (2): 166–177.

Geurts, B., and R. Nouwen. 2007. *At least* et al.: the semantics of scalar modifiers. *Language* 83 (3): 533–559.

Geurts, B., and R. van der Sandt. 1999. Focus: Linguistic, cognitive, and computational perspectives. In *Domain restriction*, ed. P. Bosch and R. van der Sandt, 268–292. Cambridge: Cambridge University Press.

Groenendijk, J., and M. Stokhof. 1982. Semantic analysis of wh-complements. *Linguistics and Philosophy* 5 (2): 175–233.

Groenendijk, J., and M. Stokhof. 1984. Studies on the semantics of questions and the pragmatics of answers. PhD dissertation, University of Amsterdam.

Guerzoni, E., and Y. Sharvit. 2007. A question of strength: on NPIs in interrogative clauses. *Linguistics and Philosophy* 30 (3): 361–391.

Hackl, M. 2009. On the grammar and processing of proportional quantifiers: *most* versus *more than half*. *Natural Language Semantics* 17 (1): 63–98.

Hamblin, C.L. 1973. Questions in Montague English. *Foundations of Language* 10 (1): 41–53.

Heim, I. 1991. Artikel und Definitheit. In *Handbuch der Semantik*, ed. A. von Stechow and D. Wunderlich, 487–535. Berlin: de Gruyter.

Heim, I. 1994. Interrogative semantics and Karttunen's semantics for "know". In *Proceedings of the 9th Annual meeting of the IATL*, ed. R. Buchalla and A. Mittwoch, 128–144. Jerusalem: Hebrew University.

Heim, I., and A. Kratzer. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

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

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

Kotek, H. 2014. Composing questions. PhD dissertation, MIT.

Kotek, H. 2015. On the semantics of wh-questions. In *Proceedings of Sinn und Bedeutung 20*, ed. N. Bade, P. Berezovskaya, and A. Scholler, 430–447. University of Tübingen.

Kratzer, A. 2008. On the plurality of verbs. In *Event structures in linguistic form and interpretation: language, context and cognition*, ed. J. Dölling, T. Heyde-Zybatow, and M. Schäfer, 269–300. Berlin: de Gruyter.

Krifka, M. 1986. Nominalreferenz und Zeitkonstitution. Zur Semantik von Massentermen, Pluraltermen und Aspektklassen. PhD dissertation, Universität München. (Published by Wilhelm Finck, Munich, 1989).

Krifka, M. 1996. Pragmatic strengthening in plural predications and donkey sentences. In *Proceedings of SALT 6*, ed. T. Galloway and J. Spence, 136–153. Ithaca, NY: CLC Publications.

Križ, M. 2015. Aspects of homogeneity in the semantics of natural language. PhD dissertation, University of Vienna.

Križ, M. 2015. Homogeneity, trivalence and embedded questions. In *Proceedings of the 20th Amsterdam Colloquium*, ed. T. Brochhagen, F. Roelofsen, and N. Theiler, 207–216. Amsterdam: University of Amsterdam.

Križ, M., and E. Chemla. 2015. Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics* 23 (3): 205–248.

Lahiri, U. 1991. Embedded interrogatives and predicates that embed them. PhD dissertation, MIT.

Lahiri, U. 2002. *Questions and answers in embedded contexts*. Oxford Studies in Theoretical Linguistics 2. New York: Oxford University Press.

Lassiter, D., and N.D. Goodman. 2014. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of SALT 23*, ed. T. Snider, 587–610. Washington, D.C.: LSA.

Leffel, T., A. Cremers, J. Romoli, and N. Gotzner. 2018. Vagueness in implicature: The case of modified adjectives. *Journal of Semantics*. (in press).

Link, G. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Meaning, use, and interpretation of language*, ed. R. Bäuerle, C. Schwarze, and A. von Stechow, 302–323. Berlin: de Gruyter.

Löbner, S. 1985. Definites. *Journal of Semantics* 4 (4): 279–326.

Löbner, S. 2000. Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy* 23 (3): 213–308.

Magri, G. 2014. An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening. In *Pragmatics, Semantics and the Case of Scalar Implicatures*, ed. S.P. Reda, 99–145. London: Palgrave Macmillan.

Mayr, C. 2012. Focusing bound pronouns. *Natural Language Semantics* 20 (3): 299–348.

Mayr, C. 2017. Predicting polar question embedding. In *Proceedings of Sinn und Bedeutung 21*, ed. R. Truswell, 1–18. Edinburgh: University of Edinburgh.

Meyer, M.-C. 2013. Ignorance and grammar. PhD dissertation, MIT.

Paillé, M., and B. Schwarz. 2018. *Knowing whether* and ignorance inferences. In *Proceedings of WCCFL 36*, ed. R. Stockwell. Somerville, MA: Cascadilla Proceedings Project.

Percus, O. 2006. Antipresuppositions. In *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science*, ed. A. Uyema, 52–73. Tokyo: Japan Society for the Promotion of Science.

Phillips, J., and B.R. George. 2018. Knowledge wh and false beliefs: Experimental investigations. *Journal of Semantics* 35 (3): 467–494. https://doi.org/10.1093/semant/ffy004.

Pollock, J.-Y. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20 (3): 365–424.

Preuss, S. M.-L. 2001. Issues in the semantics of questions with quantifiers. PhD dissertation, Rutgers University.

Qing, C., and M. Franke. 2014. Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In *Proceedings of SALT 24*, ed. T. Snider, S. D'Antonio, and M. Weigand, 23–41. Washington, D.C.: LSA.

Rawlins, K. 2013. About 'about'. In *Proceedings of SALT 23*, vol. 23, ed. T. Snider, 336–357. Washington, D.C.: LSA.

Romero, M. 2015. Surprise-predicates, strong exhaustivity and alternative questions. In *Proceedings of SALT 25*, ed. S. D'Antonio, M. Moroney, and C.R. Little, 225–245. Washington, D.C.: LSA.

Rooth, M. 1985. Association with focus. PhD dissertation, University of Massachusetts at Amherst.

Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics* 1: 75–116.

Sauerland, U. 1998a. Plurals, derived predicates and reciprocals. In *The Interpretive Tract, MITWP 25*, ed. U. Sauerland and O. Percus, 177–204. Cambridge, MA: MIT Press.

Sauerland, U. 1998b. The meaning of chains. PhD dissertation, MIT.

Sauerland, U. 2008. Implicated presuppositions. In *Sentence and context: language, context, and cognition*, ed. A. Steube, 581–600. Berlin: Mouton de Gruyter.

Scha, R. 1981. Distributive, collective and cumulative quantification. In In J. Groenendijk, M. Stokhof, and T. M. V. Janssen, eds., *Formal methods in the study of language*, vol. 2, 131–158. Mathematisch Centrum, Amsterdam. Reprinted in Groenendijk et al. (eds), *Truth, Interpretation and Information*, Foris Publications, Dordrecht (1984).

Schlenker, P. 2007. Transparency: An incremental theory of presupposition projection. In *Presupposition and implicature in compositional semantics*, ed. U. Sauerland and P. Stateva, 214–242. London: Palgrave Macmillan.

Schlenker, P. 2009. Local contexts. *Semantics and Pragmatics* 2 (3): 1–78. https://doi.org/10.3765/sp.2.3.

Selkirk, E. 1995. Sentence prosody: Intonation, stress, and phrasing. In *The handbook of phonological theory 1*, ed. J. Goldsmith, 550–569. London: Blackwell.

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

Sharvit, Y. 2004. Some remarks on quantificational variability and question-embedding predicates (U. Lahiri, *Questions and answers in embedded contexts*). *English Linguistics*, 21(2): 541–569.

Sharvit, Y. 2017. A note on (Strawson) entailment. *Semantics and Pragmatics* 10 (1): 1–38. https://doi.org/10.3765/sp.10.1.

Singh, R. 2008. Modularity and locality in interpretation. PhD dissertation, MIT.

Singh, R. 2009. *Maximize Presupposition!* and informationally encapsulated implicatures. In *Proceedings of Sinn und Bedeutung 13*, ed. A. Riester and T. Solstad, 513–526. Stuttgart: Universität Stuttgart.

Solt, S. 2016. On measurement and quantification: The case of *most* and *more than half*. *Language* 92 (1): 65–100.

Spector, B. 2005. Exhaustive interpretations: What to say and what not to say. Presentation at LSA Workshop 'Context and Content'.

Spector, B. 2013. Bare numerals and scalar implicatures. *Language and Linguistics Compass* 7 (5): 273–294.

Spector, B., and P. Egré. 2015. A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese* 192 (6): 1729–1784.

Spector, B., and Y. Sudo. 2017. Presupposed ignorance and exhaustification: how scalar implicatures and presuppositions interact. *Linguistics and Philosophy* 40 (5): 473–517.

Sternefeld, W. 1998. Reciprocity and cumulative predication. *Natural Language Semantics* 6 (3): 303–337.

Theiler, N. 2014. A multitude of answers: Embedded questions in typed inquisitive semantics. Master's thesis, ILLC, University of Amsterdam.

Theiler, N., F. Roelofsen, and M. Aloni. 2016. A truthful resolution semantics for declarative and interrogative complements. Manuscript, ILLC, University of Amsterdam.

Theiler, N., F. Roelofsen, and M. Aloni. 2017. What's wrong with *believing whether*. In *Proceedings of SALT 27*, ed. D. Burgdorf, J. Collard, S. Maspong, and B. Stefánsdóttir, 248–265. Washington, D.C.: LSA.

Theiler, N., F. Roelofsen, and M. Aloni. 2018. A uniform semantics for declarative and interrogative complements. *Journal of Semantics* 35 (3): 409–466. https://doi.org/10.1093/jos/ffy003.

Uegaki, W. 2015a. Interpreting questions under attitudes. PhD dissertation, MIT.

Uegaki, W. 2015b. Predicting the variation in exhaustivity of embedded questions. Manuscript, MIT.

Uegaki, W. 2018. On the projection of the presupposition of embedded questions. In *Proceedings of SALT 28*, ed. S. Maspong and B. Stefánsdóttir. Washington, D.C.: LSA.

van Tiel, B., E. van Miltenburg, N. Zevakhina, and B. Geurts. 2016. Scalar diversity. *Journal of Semantics* 33 (1): 137–175.

Villalta, E. 2008. Mood and gradability: an investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy* 31 (4): 467–522.

von Fintel, K., D. Fox, and S. Iatridou. 2014. Definiteness as Maximal Informativeness. In *The Art and Craft of Semantics: A Festschrift for Irene Heim, MITWPL 70*, ed. L. Crnič and U. Sauerland, 165–174. Cambridge, MA: MIT Press.

Williams, A. 2000. Adverbial quantification over (interrogative) complements. In *Proceedings of WCCFL 19*, ed. R. Billery and B.D. Lillehaugen, 574–587. Somerville, MA: Cascadilla Press.

Winter, Y. 2000. Distributivity and dependency. *Natural Language Semantics* 8 (1): 27–69.

Winter, Y., and R. Scha. 2015. Plurals. In *Handbook of contemporary semantic theory*, 2nd ed, ed. S. Lappin and C. Fox, 77–113. Oxford: Wiley-Blackwell.

Xiang, Y. 2015. Complete and true: A uniform analysis for mention-some and mention-all. In *Proceedings of Sinn und 20*, ed. N. Bade, P. Berezovskaya and A. Schöller, 815–832. Universität Tübingen.

Xiang, Y. 2016. Interpreting questions with non-exhaustive answers. PhD dissertation, Harvard.

Xiang, Y. 2018. A hybrid categorial approach of question composition. Ms., Rutgers University.

Zuber, R. 1982. Semantic restrictions on certain complementizers. In S. Hattori and K. Inoue, eds., *Proceedings of the 13th International Congress of Linguists*, ed. S. Hattori and K. Inoue, 434–436, Tokyo.