Testing the distribution of pair-list questions with quantifiers[1]

Thom VAN GESSEL — *University of Amsterdam*
Alexandre CREMERS — *IJN – ENS Paris*

Abstract.

Questions with quantifiers such as 'Which book did every student read?' can receive a pair-list reading, but the availability of this reading depends on the quantifier, as well as the environment of the question (matrix or embedded under various predicates), with possible interactions between these factors. The details of these interactions have been a subject of debate in the literature.

We tested the acceptability of pair-list readings with 5 quantifiers (*most*, *two*, *no*, *every*, and *fewer than three*) in 4 different environments (matrix, *find out*, *be certain*, and *wonder*). Our results confirm that the availability of pair-list interpretations for questions with quantifiers depends heavily on both the quantifier and the environment in which the question appears, and more specifically that there is a qualitative divide between responsive and rogative predicates, not between intensional and extensional predicates.

Keywords: pair-list questions, quantification, clause-embedding predicates

1. Introduction

Some questions with quantifiers seem to allow for at least two different kinds of answers (May, 1985). An example of such a question is (1). It allows an individual answer like (1a), as well as an answer in the form of a list of pairs (pair-list answer, (1b)).

(1) Which book did every student read?
  a. Every student read *War and Peace*.
  b. John read *On the Road*, Paul read *Lord of the Rings* and Mary read *War and Peace*.

It therefore seems that the question has at least two paraphrases:

(2) a. Which book x is such that every student read x?
  b. For every student x, which book did x read?

The reading that can be paraphrased as (2b) is called a *pair-list reading*.

An interesting observation about pair-list readings is that their availability depends on the quantifier that the question contains. For instance, while a pair-list reading is available with *every* as in (1), it has been argued not to be available with *most* as in (3).

(3) Which book did most students read?

According to Szabolcsi (1997), there is no reading of (3) that can be paraphrased as 'For most students x, which book did x read?'. Only the single-book reading would be available.

[1]We would like to thank three anonymous reviewers for their helpful comments. The research leading to these results has received funding from the European Research Council (ERC, grant agreement 680220), the Netherlands Organisation for Scientific Research (NWO), and the Agence Nationale de la Recherche (FrontCog Project ANR-17-EURE-0017 and ProbaSem project ANR-19-CE28-0004-01).

It is also argued that the availability of a pair-list reading depends on the environment in which the question occurs. While questions with *most* may not have a pair-list reading in matrix questions, they do seem to have such a reading when embedded under *know*: (4a) can mean that for most students, John knows what they read. However, when embedded under *wonder*, this reading again seems less available: (4b) would not mean that for most students, John wonders what they read.

(4) a. John knows which book most students read.
  b. John wonders which book most students read.

Many theories on pair-list readings of questions have been proposed, with different predictions about which quantifiers allow for pair-list readings and how the embedding verb affects this. In this paper we report the results of a series of experiments in which we tested the acceptability of pair-list readings with several quantifiers in several different environments, to see how the predictions of these theories fair.

We start by discussing the proposals in section 2. We then explain our experimental method in section 3, and present the results in section 4. The paper ends with a discussion of the results and possible directions for further research.

2. Proposals

Since the 1980s, several authors have written about questions with quantifiers and the pair-list readings they give rise to. We will not try to give a detailed or exhaustive overview in this section (see for instance Szabolcsi, 1997, Pafel, 1999, Dayal, 2016). Instead, we will focus on disagreements regarding the data.

As we will see shortly, authors disagree about some data points, but they also agree on several ones. All authors predict that *every* always allows — and *no* never allows — pair-list readings. The disagreement is about those quantifiers that are in between, for instance bare numerals (*two*), modified numerals (*fewer than three*), and *most*. Another debate is about which class of embedding verbs allows more quantifiers to give pair-list readings.

2.1. Simple scope taking

In the theory of Groenendijk and Stokhof (1984), the extension of a question is a proposition, namely the true, exhaustive answer to the question. Because these true answers can vary between worlds, the intension of a question is a (typically non-constant) function from worlds to propositions, which can be conceived as a partition of the logical space: a division into non-overlapping areas.

A compositional framework that assigns these types of semantic values to questions can capture pair-list readings more or less for free: since question extensions are of the same type as declarative sentences, the quantifier rule that accounts for wide scope readings of these sentences can also apply to questions. This is the approach taken in Groenendijk and Stokhof (1984: ch. 3). Specifically, the quantifier rule allows universal quantifiers to quantify into the question meaning of (5a) to derive the pair-list reading of (5b).

(5) a. Which book did x read?

b. Which book did every student read?

The meaning of (5a) is a function that takes a world w and returns the set of worlds in which x read the same book as in w. Such a partition of the logical space can be made for each student in the domain, and the pair-list reading of (5b) is then simply the intersection of all these partitions. It is easy to see that this too is a partition of the logical space, and thus a question meaning in Groenendijk and Stokhof's sense.

According to this theory, pair-list readings can only be derived with universal quantifiers. If we were to quantify into (5a) with a non-universal quantifier, we would not be guaranteed to get a partition. For instance, *some student* would give us the union of all question meanings of (5a) for all students x, and the union of a set of (independent) partitions is itself not a partition.

In other words, a question like (6) with a wide scope quantifier does not have one true, exhaustive answer: a true, exhaustive answer can only be given once a single student has been chosen.

(6) Which book did some student read?

The assumption that questions always have to be partitions, and cannot have overlapping answers, results in the prediction that there is no reading of (6) that lets us pick a student and then answer for this student which book they read (this is called a *choice reading*). A theory that departs from this assumption is inquisitive semantics (Ciardelli et al., 2019). In this theory, questions and statements are analyzed as sets of sets of worlds. Since this is also a uniform approach, pair-list questions can be derived in the same way, by having the quantifier take wide scope.

Because question meanings are not limited to partitions in inquisitive semantics, it seems that in principle a pair-list reading could be derived with any quantifier. However, inquisitive semantics may offer a different explanation for why pair-list readings with some quantifiers do not arise. For instance, the lexical entry for *no* given in Ciardelli et al. (2016) always flattens the alternatives when it takes scope over a question, thereby turning the question into a statement. A challenge for inquisitive semantics is to define lexical entries for other quantifiers, such as *fewer than three* and *most*, which explain the (un)availability of pair-list readings when they occur in questions, and find independent motivation for these entries.

A second challenge is the contrast between pair-list distribution in matrix and embedded questions. In inquisitive semantics, the semantic value of a question does not depend on whether it appears embedded or not. If it is really the case that some quantifiers give pair-list readings only in embedded questions, then some other mechanism has to account for this. Ciardelli and Roelofsen (2017) suggest that quantifiers like *most* could contain a definedness condition that requires its input to be a 'flat' property, one that doesn't generate multiple alternatives. With such a condition, (7a) would be undefined, whereas (7b) would be fine.

(7) a. [most men [λx [which woman did x kiss]]]
b. [most men [Bill found out [λx [which woman did x kiss]]]]

However, such an analysis would also predict a pair-list reading for (8):

(8) [most men [Bill wonders [λx [which woman did x kiss]]]]

It is not clear how the analysis can be adapted to account for this contrast. Furthermore, for this account to work, we have to assume that quantifiers in embedded questions may leave the embedded clause.

2.2. Witness sets

Groenendijk and Stokhof (1984: ch. 6) develop an alternative derivation of pair-list questions, in which the meaning contribution of the quantifier is based on its minimal witness sets.

Witness sets of generalized quantifiers are those members of the generalized quantifier that are subsets of the set the quantifier lives on: in case of *some student*, this is the set of students. Minimal witness sets are the smallest ones that are a member of the generalized quantifier. To illustrate, *every student* has just one minimal witness set, namely the set of all students. The minimal witness sets of *some student* and *two students* are the singleton sets of students and the sets of exactly two students, respectively.

This theory is much less restrictive with respect to pair-list readings: it derives them not only for questions with universal quantifiers, but for all quantifiers that have non-empty minimal witness sets. If they have more than one, this is predicted to result in a choice reading.

The theory rules out pair-list readings of questions with *no*, because the minimal witness set of *no student* is the empty set, and we would thus derive an empty set of questions on a wide scope reading. In fact, this theory rules out pair-list readings for all downward entailing quantifiers, since such quantifiers (e.g. *fewer than three*, *at most two*) have the empty set as minimal witness set. This means that the minimal witness set theory predicts a pair-list reading for (9b), but not for (9a).

(9) a. Which book did fewer than three students read?
b. Which book did two students read?

Chierchia (1993) proposes a refinement of this theory to account for the observation that pair-list answers are only licensed when the quantifier is in subject position:

(10) a. Who does everyone like? *Pair-list*
b. Who likes everyone? *No pair-list*

This asymmetry is something that pair-list answers have in common with functional answers, as illustrated in (11).

(11) a. Who does everyone like?
Answer: his mother
b. Who likes everyone?
Answer: his mother

Individual readings are different: they are always available. For Chierchia, this is a reason to think that pair-list readings are a special case of functional readings.

In his approach, *wh*-phrases leave a functional trace. The asymmetry can then be explained by the inability of the quantifier to move past this trace. This phenomenon is called *weak crossover*, and it is a well-known phenomenon that occurs in declarative sentences as well.

However, a weakness of this theory may be that there are languages, such as German, that do have subject-object asymmetries in pair-list questions, but do not show weak crossover effects (Preuss, 2001).

Also, there seem to be counterexamples to subject-object asymmetry:

(12) Who kissed every girl?

In a series of experiments, Achimova (2011) finds that about 30% of participants consistently judges pair-list answers to (12) to be appropriate. Contrary to what Chierchia (1993) predicts, it does not matter whether the question is formulated with *who*, with *which* and a plural noun or with *which* and a singular noun.

Setting the issue of subject-object asymmetry aside, we can conclude that Chierchia's use of minimal witness sets predicts the same quantifiers to participate in pair-list readings as Groenendijk and Stokhof (1984: ch. 6).

2.3. Speech acts

Krifka (2001) posits, like Groenendijk and Stokhof (1984: ch. 3), that only universal quantifiers can scope out of questions, but gives a different explanation for this. He introduces a special type for speech acts, and adds illocutionary operators that turn propositions into speech acts, for instance, to a speech act of asking. In his theory, pair-list questions are formed by having the quantifier take scope over the speech act, rather than the sentence itself. With a universal quantifier, this results in a conjunction of speech acts:

(13) Which book did every student read?
= Which book did John read? $\wedge$ Which book did Paul read? $\wedge$ Which book did Mary read?

With a non-universal quantifier, this results in a disjunction of speech acts:

(14) Which book did two students read?
= Which book did John and Paul read? $\vee$ Which book did John and Mary read? $\vee$ Which book did Paul and Mary read?

However, what his theory relies on is the idea that speech acts can only be conjoined, not disjoined: whenever we disjoin a question, we intend to replace the former question with a new one.

The observed contrast between matrix questions and other environments is accounted for by having extensional verbs (those verbs that are sensitive only to the true answer of the question they embed, like *know*) and intensional verbs (sensitive to the intension of the question, like *wonder*) embed objects of a different type. Intensional verbs embed speech acts rather than propositions, which explains why they only have pair-list readings with a universal quantifier, while extensional verbs embed question extensions as true answers, which explains why more quantifiers can have pair-list readings when embedded under *know*.

2.4. Two kinds of pair-list

According to Szabolcsi (1997), the contrast between matrix questions and embedded questions shows that pair-list readings are in fact two separate phenomena. On the one hand, there are pair-list matrix questions and questions embedded under intensional verbs. On the other hand, there are pair-list questions embedded under extensional verbs, which are derived using lifted questions.

For Szabolcsi, question extensions are of type $\langle s,t \rangle$. For the first type of pair-list readings, Szabolcsi proposes an analysis along the lines of Groenendijk and Stokhof (1984: ch. 6) with witness sets, but she proposes that the quantifier contributes a *unique* witness set. This predicts that only universal quantifiers can give this kind of pair-list reading, because other quantifiers typically have several witness sets. Thus, under this analysis, there is no such thing as choice readings of matrix questions.

In her analysis of the second type of pair-list readings, Szabolcsi assumes that questions with any quantifier can be lifted to the type $\langle\langle\langle s,t\rangle,t\rangle,t\rangle$, while leaving behind a trace of type $\langle s,t \rangle$ that the embedding verb can take up. The observation that verbs like *wonder* do not generate pair-list readings is explained by having intensional verbs take complements of type $\langle\langle s,t\rangle,t\rangle$ rather than $\langle s,t \rangle$.

This theory predicts that pair-list readings only arise with universal quantifiers in matrix questions and questions embedded under intensional verbs, while they can arise with any quantifier embedded under extensional verbs. Although the theory does not account for it, Szabolcsi does mention that pair-list readings with *no* should not be predicted to occur.

Sharvit (2002) observes that there are intensional verbs like *be certain* and *agree on* that allows for pair-list readings with the same quantifiers as extensional verbs like *know* and *find out*. This suggests that it is in fact a different property of the verb that causes the observed difference between *know* and *wonder*. According to Sharvit, this property is the ability to take declarative and/or interrogative complements: responsive verbs like *know* and *find out* allow the question to take scope over the embedding verb, while this is not the case with rogative verbs.

The relevant difference between these verb classes is that responsive verbs allow for quantificational variability, while rogative verbs normally do not. This is connected to the observation that quantificational variability does not occur with whether-questions and declaratives, just like exceptional wide scope readings with *most* and *fewer than three*. Therefore Sharvit concludes that pair-list readings with responsive verbs are variants of quantificational variability, where quantification is over a set of relevant subquestions of the question. The division into subquestions depends on the presuppositions of the verb, and some verbs do not provide such a division, which should account for the differences between verbs.

The theory does not put forward a principled reason why verbs like *wonder* do not allow for this, or why there is no pair-list reading of matrix questions with other quantifiers if they do exist when embedded under extensional verbs. But it does make an important empirical prediction, namely that intensional verbs that can embed both declarative and interrogative complements should behave like *know* and *find out* when it comes to embedding pair-list readings.

3. Methods

As we have seen in the previous section, the theories that have been proposed to account for pair-list readings make different predictions. Not only do they differ as to which quantifiers are predicted to allow for pair-list readings, they also make different predictions about what happens with embedded questions. In order to settle these debates, we tested the acceptability of pair-list readings of questions with five different quantifiers in four different environments.

The quantifiers we tested are *most*, bare numerals (*two*), *no*, *every* and *fewer than three*. It is uncontroversial that *every* allows a pair-list reading, so this is a good baseline. The downward entailing quantifiers *no* and *fewer than three* are generally not predicted to give pair-list readings in matrix questions. With embedded questions, however, the two quantifiers are predicted to come apart: according to Szabolcsi (1997), the only quantifier that does not allow a pair-list reading in embedded environments is *no*. There is also no consensus about *most* and *two*.

We test the acceptability of these pair-list readings in four environments: matrix questions and questions embedded under *find out*, *be certain* and *wonder*. Since there is a disagreement about which types of embedding verbs behave differently from the matrix case, we chose verbs with different combinations of properties: extensional/intensional and responsive/rogative. See Table 1 for an overview.

find out	*be certain*	*wonder*
extensional	intensional	intensional
responsive	responsive	rogative

Table 1: Properties of embedding verbs to test

It is difficult to determine in a direct manner which readings of a question are available. Therefore, for matrix questions we asked participants to judge whether a pair-list *answer* to a question is appropriate. With embedded questions, we gave a description of a situation that makes the sentence that embeds the pair-list reading of the question true, and asked for a truth value judgment.

3.1. Matrix questions

Each item started with a short description of the context. In the default context, there were three students, and each student read *exactly one* book. This ensures that any pair-list answer is exhaustive, and cannot be judged as inappropriate on that account.

This context was followed by a short dialogue: a question asked by a speaker and a response by a second speaker. The participants were asked to judge whether this response was appropriate, given the question that was asked. In order to prevent the participants from taking into account the truth or falsity of the answer when judging whether it is appropriate, we made sure that it could not be determined from the context whether the response was true. The instructions stated that the second speaker never lies.

We tested variants of the question with both *which* + *singular noun* and with *what*. The choice of *wh*-phrase is sometimes assumed to have an effect on the presupposition of the question, thereby suggesting how appropriate non-pair-list answers are in some cases.

For each quantifier, the context and question were repeated with three different types of responses, illustrated in (15): a pair-list (PL) answer (a) and two non-pair-list answers (b and c), which served as fillers.

(15) Which book / what did *Q* student(s) read?
  a. *A* read *X*, *B* read *Y*, ... *PL answer*
  b. *Q* student(s) read *X*. *Positive non-PL answer*
  c. There is no book that *Q* student(s) read. *Negative non-PL answer*

As baselines, each pair-list answer also occurred with a control question that made it very inappropriate (like (16a)) and with one that made it very appropriate (like (16b)).

(16) Answer: Anthony read *The Little Prince*, James read *On the Road* and John read *The Catcher in the Rye*.
  a. Inappropriate: Which book did no student read?
  b. Appropriate: What did Anthony read, what did James read and what did John read?

Some adaptations to the default context were made for specific quantifiers. A pair-list answer to (17) would specify for no student what they read, and thus be vacuous. Since offering this option could be confusing, we did not test pair-list answers to questions with *no* for matrix questions.

(17) What did no student read?

Another problem concerns the case of bare numerals. A question like (18) has a cumulative reading, which can be answered by (18a) (Krifka, 1992; Dayal, 1996; Szabolcsi, 1997).

(18) What did two students read?
  a. *A* and *B* read *X* and *Y* *Answer to cumulative reading*
  b. *A* read *X* and *B* read *Y* *PL answer*

In this case, the PL answer (18b) conveys the same information and, additionally, specifies which of the two students read which of the two books. In this way, the PL answer may be seen as a slightly over-informative answer to the cumulative reading. It is not necessarily an indication that participants get a PL reading if they judge this answer as appropriate. The variant with *which + singular noun* does not have this problem.

(19) Which book did two students read?
  a. # *A* and *B* read *X* and *Y* *Answer to cumulative reading*
  b. *A* read *X* and *B* read *Y* *PL answer*

Therefore, we decided to drop the *what*-variant of the bare numeral case and tested only *which*.

Questions with *most* were only tested with *what*, to avoid the existence presupposition of *which + singular noun*. On the non-PL reading, it presupposes that there is some book that most subjects read. The PL answer could then be an over-informative denial of this presupposition. However, for *what*, the pair-list answer could also be an over-informative 'nothing' answer to the non-pair-list question 'What did most students read?'. We avoided this by formulating the pair-list answer as *A read 1, B read 2, C read 3 and D read 2*, in a context of five students. From this answer, it does not follow that most students read *2*, but it does leave the option open. In this way, it could only be an appropriate answer to the pair-list reading.

For the question 'What did fewer than three students read?', we slightly changed the context. Participants may interpret the pair-list answer *A read 1, B read 2* as *only A read 1 and only B read 2*, thus answering the non-pair-list reading by specifying that both *1* and *2* are read by fewer than three students. To prevent this, the context for this question explained that students might have read more than one book. The pair-list answer is then *A read 1 and 2, B read 2*, which cannot get such a mutually exhaustive reading.

For the target items, we added a repetition in which the context was different (kids who saw animals in the zoo instead of students who read books), but otherwise the items were exactly the same. In total, the survey consisted of an instruction page, two training items and 60 items of interest.

3.2. Responsive verbs

For questions embedded under responsive verbs, items were reversed: instead of judging for each of the three answers (positive non-pair-list, negative non-pair-list and pair-list) whether they are appropriate, the context indicated that a character in the story hears one of the statements in (20).

(20) a. Q student(s) read X. *Positive non-PL*
b. There is no book that Q student(s) read. *Negative non-PL*
c. A read X, B read Y, ... *PL*

Then, the participant was asked to judge whether the statement in which the question is embedded under the verb *find out* or *be certain* (as in (21)) is true or false.

(21) Ann [found out / is certain] [which book / what] Q student(s) read.

The instructions depended on the verb tested: for *find out*, the participant was told that Ann initially knows nothing more than what is explicitly written down about the context. For *be certain*, the participant was told that Ann never doubts the information she gets. In case the sentence is judged true after the pair-list statement, we infer that a pair-list interpretation of the question is judged to be appropriate.

Compared to the set-up with matrix questions, a few other changes were made. Firstly, the quantifier *no* was added (both with *what* and with *which + singular noun*). This is generally assumed to be one of the few quantifiers that do not allow for a PL reading, even when the question is embedded under an extensional verb. Thus, we should expect to find that (22) is judged false when (22a) is the case.

(22) Ann found out which book no student read.
a. # There is no student such that Ann found which book they read.
b. Ann found out which book is such that no student read it.
c. # Ann found out that there is no book such that no student read it.

Because of the existence presupposition of *which*, we expect (22) also to be false in the negative non-PL case (22c). In fact, this goes for all quantifiers. Therefore, all negative non-PL cases with *which* were removed from this part of the experiment.

3.3. Wonder

Unlike with *find out* and *be certain*, the truth of 'Ann *wonders* which book every student read' does not depend on the information Ann has, but on the information she desires. Therefore we needed a natural story with different contexts in which different information is desired. The following story was written in the instructions:

> *Professor Smith teaches mathematics. Recently, a group of his students took an exam, and he is now in the process of grading it.*
>
> *Before the exam, each of Professor Smith's students chose a tutor: someone who helped him or her by practicing assignments together.*
>
> *Professor Smith has a tendency to blame the tutors for the malperformance of their students. However, Professor Smith does not know which tutor any of the students chose. Therefore, if a student fails the exam, Professor Smith wants to know who the responsible tutor is.*

In each item of the survey, the participant was asked to judge whether a statement of the form of (23a) or (23b) is true (where *Q* is one of the quantifiers from the previous experiment):

(23) a. Professor Smith wonders which tutor *Q* students chose.
b. Professor Smith wonders who *Q* students chose.

This statement was preceded by a short description of some relevant facts, which aim to push either toward a non-PL or PL interpretation of the statement. The participant's truth judgment should indicate whether the intended reading is available.

To illustrate, let us examine the items for one of the quantifiers, *most*. Each item started with a description of the students involved:

(24) Five students took the exam: John, Mary, Olivia, Ben and Emma.

This is followed by a description that indicates which tutor the professor wants to talk to. This description comes in two variants, (25a) and (25b). Then, the statement is followed by the target statement. All items are repeated with a true and with a false control statement, which function as baselines.

(25) a. Professor Smith heard that there is one popular tutor, who trained the majority of the students. Professor Smith thinks that this tutor had too much work, and that therefore some of these students underperformed. *(non-PL)*

Target Professor Smith wonders [wh] most students chose
False Professor Smith wonders [wh] no student chose
True Professor Smith wonders [wh] was chosen by the majority of the students

b. Four of the five students failed, namely John, Mary, Emma and Ben. Professor Smith knows that they each had a different tutor. *(PL)*

Target Professor Smith wonders [wh] most students chose

False	Professor Smith wonders [wh] Olivia chose
True	Professor Smith wonders [wh] John chose, [wh] Mary chose, [wh] Emma chose and [wh] Ben chose

3.4. Participants

Sixty participants were recruited on Amazon's Mechanical Turk for each survey (age range: 18–64). Because a few participants did not complete the survey or rushed through it answering at chance level on fillers, only 58 completed the *be certain* and matrix versions of the experiment. Participants were paid $1.20 for the matrix survey and $1.30 for the other surveys, which took slightly longer. Two participants in the *find out* survey reported native languages other than English and were therefore excluded from further analyses.

Finally, we measured the error rate on controls and unambiguous fillers and excluded participants who were more than one standard deviation above the mean for their survey (this was done by survey to avoid possible differences in difficulty level). This led to the further exclusion of 16 participants, leaving a final 56 participants for the *find out* survey, 54 participants for *be certain*, 53 for *wonder*, and 57 for matrix. The average error rate for remaining participants is 4.3%.

4. Results

The results on pair-list targets and controls are presented in Figure 1. Performance on controls is high and there is little variation across quantifiers or environments. At first glance, *find out* and *be certain* seem to pattern together in that they allow pair-list readings with all but the downward entailing quantifiers (with *fewer than three* slightly less degraded than *no*), while *wonder* and matrix seem to follow a gradient from *every* to *no*. Our statistical analyses confirm this impression.

A model comparing targets to controls would not converge, so we worked under the assumption that performance on controls was uniform, allowing us to compare targets directly.[2] We leave aside the quantifier *no* for the moment, as it couldn't be tested for matrix questions.

We first fitted a mixed-effects logistic model which distinguished between all four ENVIRONMENTS (*find out, be certain, wonder*, and matrix). We included fixed effects for ENVIRONMENT, QUANTIFIER, WH-WORD and interactions between QUANTIFIER and the two others.[3] We included random intercepts by participants, but no random slopes could be fitted. We then fitted a model for each possible grouping of the four ENVIRONMENTS (into one, two, or three categories), and compared all 15 resulting models. The best model by AIC had three categories: {*find out, be certain*}, {*wonder*} and {matrix}. The best model by BIC further put *wonder* and matrix together.

We focus on the simpler BIC-optimal model. The full list of fixed effects are presented in Table 2. For responsive verbs *find out* and *be certain*, we observe no differences between the

[2]The only apparent deviations from this assumption happens with *wonder* and the *no* appropriate control for *find out*, but they seem small enough not to affect our conclusions.

[3]This model was optimal by AIC compared to models with fewer or more interaction terms between WH-WORD and other effects.

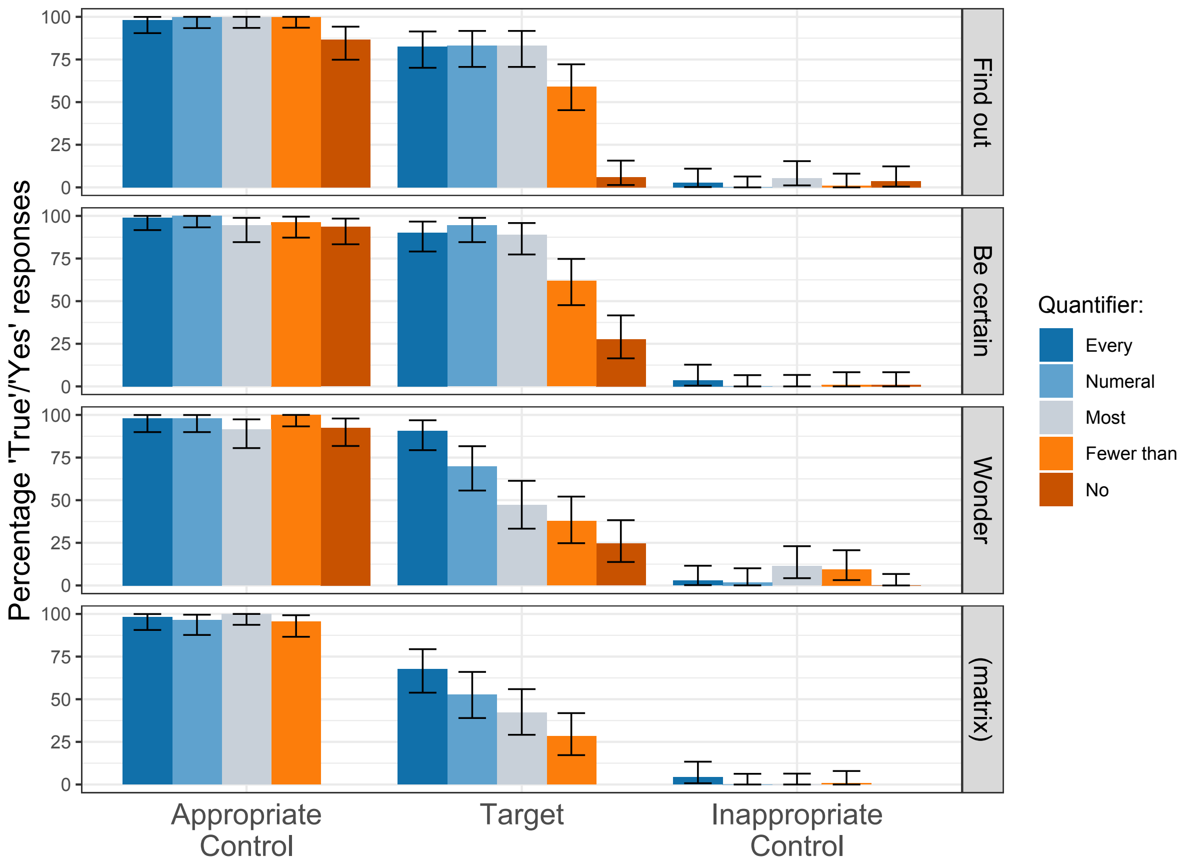

Figure 1: Mean positive response ('True' for embedded questions, 'Yes' for matrix questions) to control and target items, by embedding and quantifier. The error bars show ClopperPearson 95% confidence intervals. On targets, a positive response indicates a pair-list interpretation.

quantifiers *every*, *most* and the numeral, while the pair list reading is significantly less likely with *fewer than*. For rogative predicates, pair-list readings are less likely with *every*, and even more so with each of the other quantifiers. Finally, we observe a significant main effect of WH-WORD, indicating that pair-list readings are slightly less likely with *which* than with *who/what*.

	β	z-value	p-value
(Intercept)	2.49	11.16	$< .001^{***}$
QUANTIFIER:numeral	0.10	0.28	.779
QUANTIFIER:*most*	-0.24	-0.74	.459
QUANTIFIER:*fewer than*	-1.88	-9.33	$< .001^{***}$
Rogative	-0.85	-2.83	$.005^{**}$
WH-WORD	-0.46	-2.15	$.032^{*}$
[QUANTIFIER:numeral]×[Rogative]	-1.19	-3.21	$.001^{**}$
[QUANTIFIER:*most*]×[Rogative]	-1.83	-5.02	$< .001^{***}$
[QUANTIFIER:*fewer than*]×[Rogative]	-0.78	-2.65	$.008^{**}$
[QUANTIFIER:numeral]×[WH-WORD]	0.71	1.56	.119
[QUANTIFIER:*most*]×[WH-WORD]	0.05	0.12	.902
[QUANTIFIER:*fewer than*]×[WH-WORD]	-0.19	-0.69	.488

Table 2: Results from the BIC-optimal model on targets. *Every* is the baseline QUANTIFIER. Environments are separated into responsive (*find out, be certain*; baseline) and rogative (*wonder, matrix*). WH-WORD is sum-coded (-0.5 for *who* and *what*, and $+0.5$ for *which*).

The same procedure was repeated including the quantifier *no* but excluding matrix questions. The optimal BIC model placed *find out* and *be certain* in one category separate from *wonder*, in line with the model without *no*. It further revealed that pair-list readings with *no* are as unacceptable under *wonder* as under responsive predicates ($z = 0.3, p = .76$).

A reviewer pointed out that acceptance of items with the wh-words *who/what* may not reflect true pair-list reading due to the possibility of a cumulative reading, so we may be overestimating pair-list readings, and this may bias our results if this effect interacts with quantifiers or environments. To address this concern, we first compared models with and without the triple interaction between quantifier, environment, and wh-word, and found no evidence for any interaction ($\chi^2(3) = .05, p = .997$). We then re-ran the comparison between models representing all possible groupings of the four environments, this time keeping only the data for *which*-questions. We obtained qualitatively similar results, despite the reduced statistical power: the BIC optimal model didn't distinguish between any environments, but the AIC-optimal model still distinguished between {*found out*, *be certain*} and {matrix, *wonder*}.

Anonymized data are available in the OSF folder `https://osf.io/2cxnv/`, together with our R script.

5. Discussion

We found a divide between two categories of predicates. On the one hand, responsive predicates *find out* and *be certain* clearly allow pair-list readings with quantifiers *every*, *most* and bare numerals, but less so with *fewer than three* and not at all with *no*. On the other hand, matrix questions and questions embedded under *wonder* display a full gradient from quantifier *every*, which clearly licenses pair-list readings, to quantifier *no*, which clearly does not.

Many authors indeed predict a difference between two classes of environments. However, they disagree on the boundary. The results suggest that this boundary is not between matrix questions and embedded questions, as would be expected in the theory of Ciardelli and Roelofsen, and also not between intensional verbs and extensional verbs, as predicted by Szabolcsi and Krifka. Instead, the divide seems to be between responsive and rogative verbs, as argued by Sharvit. Her analysis does not rely on an essential difference between responsives and rogatives, but on differences in presuppositions of the verb. This is not worked out into a full theory of which verbs pair with which, but at least the observation that (in general) we should expect responsives to behave like *find out* seems to be correct.

Nevertheless, *wonder* may not be representative of the whole class of rogative predicates: questions embedded under *wonder* behave like matrix questions in other respects. For instance, in some cases they can have matrix word order (pseudo-coordination) as in (26) (McCloskey, 2006).

(26) John wonders will Mary leave.

Dayal (2019) recently proposed that an intermediate syntactic layer between CP and the speech-act level, ForceP, is responsible for matrix syntax and can be embedded under *wonder* and *ask* but not under other intensional/rogative predicates such as *depend on* and *investigate*. While she does not discuss questions with quantifiers, ForceP could affect which pair-list readings are available, and this would predict a slightly different division, equally compatible with our results. We leave for future research the question of pair-list readings under predicates such as *depend on* and *investigate*, which Dayal argues cannot embed ForceP despite being rogative.

The data for *find out* roughly matches the predictions of the theories we discussed, setting aside the fact that *be certain* follows the same pattern. However, it is not immediately clear why pair-list readings with *fewer than three* are not judged as good as pair-list with *two*, *most*, and *every*.

Note also that while all authors agree that pair-list readings are not available for *no* under *find out*, the only account to actually capture this is that of Groenendijk and Stokhof — although they do not distinguish embedded and matrix questions.

The pattern of matrix questions and questions under *wonder* does not show a clear cut-off point between acceptable and unacceptable pair-list readings. The acceptability of pair-list readings with *most* and *two* is not as high as with *every* and not as low as with *fewer than three*. This suggests that they are not completely fine, but also that they are probably not underivable, as many theories do suggest. Instead, the results seem to call for a pragmatic account in which such a reading of the question would be derivable, but for some reason dispreferred.

If pair-list readings with *fewer than three* are derivable, then they would allow for a silent answer, which could explain why they are dispreferred. Such an explanation is not immediately available for questions with *most*. Perhaps this can be explained by the unlikelihood of the QUD/speaker goal that would call for such a question — that is, a situation in which someone needs to know something about an arbitrary majority, which can be picked by the person answering. Another possibility is that in all such situations, there is a different question that can be asked that is for some reason less costly and thus competes with the pair-list reading of questions with *most*. We leave this as an open question. A challenge for any pragmatic explanation may be that it is not immediately clear how it can also apply to questions embedded under *wonder*. However, Dayal suggests that the semantic effect of ForceP is to presuppose that the question is active for some agent, which may be a starting point for a theory about pragmatic effects in such embedded questions.

To conclude, our results confirm the classic observation that the availability of pair-list interpretations for questions with quantifiers depends heavily on both the quantifier and the environment in which the question appears. Importantly, we offer a clarification on the shape of this dependency. We observe two clearly distinct patterns: under responsive predicates *find out* and *be certain* there is a clear divide between acceptable upward-entailing quantifiers and degraded downward entailing quantifiers. Under *wonder* and in matrix position, we observe a full gradient from *every* to *no*. This suggests that the division between environments lies somewhere between responsives and rogatives (or a subclass of rogatives), but our survey leaves open *why* each category of environment follows the pattern it does.

References

Achimova, A. (2011). Resolving wh/quantifier ambiguities: a psycholinguistic approach. MSc thesis, Rutgers University.

Chierchia, G. (1993). Questions with quantifiers. *Natural Language Semantics 1*(2), 181–234.

Ciardelli, I., J. Groenendijk, and F. Roelofsen (2019). *Inquisitive Semantics*. Oxford University Press.

Ciardelli, I. and F. Roelofsen (2017). An inquisitive perspective on modals and quantifiers. *Annual Review of Linguistics 4*, 1–22.

Ciardelli, I., F. Roelofsen, and N. Theiler (2016). Composing alternatives. *Linguistics and Philosophy 40*, 1–36.

Dayal, V. (1996). *Locality in Wh Quantification*. Kluwer, Dordrecht.

Dayal, V. (2016). *Questions*. Oxford University Press.

Dayal, V. (2019). The Fine Structure of the Interrogative Left Periphery. Slides at `https://sites.google.com/site/inquisitivesemantics/workshops/inqbnb3`.

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

Krifka, M. (1992). Definite NPs Aren't Quantifiers. *Linguistic Inquiry 23*(1), 156–163.

Krifka, M. (2001). Quantifying into Question Acts. *Natural Language Semantics 9*, 1–40.

May, R. (1985). *Logical Form: Its Structure and Derivation*. MIT Press.

McCloskey, J. (2006). Questions and questioning in a local English. In R. Zanuttini, H. Campos, E. Herburger, and P. H. Portner (Eds.), *Crosslinguistic research in syntax and semantics: Negation, tense, and clausal architecture*, pp. 87–126. Washington, D.C.: Georgetown University Press.

Pafel, J. (1999). Interrogative Quantifiers within Scope. *Linguistics and Philosophy 22*(3), 255–310.

Preuss, S. (2001). *Issues in the Semantics of Questions with Quantifiers*. Ph. D. thesis, Rutgers University.

Sharvit, Y. (2002). Embedded Quantifiers in Which- and Whether-Questions. In *Proceedings of SALT XII*.

Szabolcsi, A. (1997). Quantifiers in pair-list readings. In A. Szabolcsi (Ed.), *Ways of Scope Taking*. Springer.