Quantifying into *wh*-dependencies:
Multiple-*wh* questions and questions with a quantifier

Yimei Xiang, Rutgers University
(Date: April 4, 2021)

Abstract *Wh*-questions with a quantificational subject have readings that seemingly involve quantification into questions (called 'QɪQ' for short). This paper unifies the derivation of QɪQ-readings and distinguish these readings from pair-list readings of multiple-*wh* questions. I propose that pair-list multiple-*wh* questions and QɪQ-questions both involve a *wh*-dependency, namely, that the *wh*-/quantificational subject stands in a dependency with the trace of the *wh*-object. In particular, in a pair-list multiple-*wh* question, the *wh*-subject binds into the trace of the *wh*-object across an identity operation; in a QɪQ-question, the quantificational subject binds into the trace of the *wh*-object across a predication operation. The differences between these two quantificational binding operations explain a contrast in domain exhaustivity between multiple-*wh* questions and questions with a universal quantifier. I further propose that the observed QɪQ-effect in a QɪQ-question is derived by extracting one of the minimal proposition sets that satisfy the aforementioned quantificational predication condition. The values of these sets determine whether the QɪQ-reading is available and whether a QɪQ-question admits pair-list answers and/or has a choice flavor.

Keywords Questions, quantifiers, multiple-*wh*, pair-list, functionality, uniqueness, domain exhaustivity, quantificational variability, categorial approaches, compositionality

1. Introduction

Pair-list readings of questions arise from two different interrogative structures: multiple-*wh* questions and single-*wh* questions with a universal quantifier (called '∀-questions' henceforth). For example, both (1a) and (1b) can be addressed by specifying a list of boy-movie pairs.

(1) a. Which boy watched which movie?
Andy watched *Ironman*, Billy watched *Spiderman*, Clark watched *Hulk*.
b. Which movie did every/each boy watch?
Andy watched *Ironman*, Billy watched *Spiderman*, Clark watched *Hulk*.

Both interrogative structures admit multiple readings. In particular, multiple-*wh* questions are ambiguous between single-pair readings and pair-list readings. For example, in (2) these two readings call for answers that specify a unique boy-movie pair and a list of boy-movie pairs, respectively.

(2) Which boy watched which movie?
a. 'Which unique boy-movie pair $\langle x, y\rangle$ is s.t. x watched y?' (Single-pair)
'Andy watched *Spiderman*.'
b. 'Which boy-movie pairs $\langle x, y\rangle$ are s.t. y is the unique movie that x watched?' (Pair-list)
'Andy watched *Ironman*, Billy watched *Spiderman*, Clark watched *Hulk*.'

In contrast, ∀-questions are ambiguous between individual readings, functional readings, and pair-list readings (Engdahl 1980, 1986). In example (3), the three readings call for answers that name an atomic movie, a Skolem function to atomic movies, and a list of boy-movie pairs, respectively.

(3) Which movie did every/each boy watch?

a. 'Which movie y is s.t. every boy watched y?' '*Spiderman*.' (Individual)

b. 'Which function f to atomic movies is s.t. every boy x_i watched $f(x_i)$?' (Functional)
'His$_i$ favorite superhero movie.'

c. 'For every boy x, [you tell me] which movie did x watch?' (Pair-list)
'Andy watched *Ironman*, Billy watched *Spiderman*, Clark watched *Hulk*.'

There are two directions to analyze pair-list readings. One direction is to give a joint analysis for pair-list readings, regardless of their origins. Accounts adopting this line of thinking either use the same LF schema to compose questions with pair-list readings, ignoring their syntactic distinctions (Engdahl 1980, 1986; Dayal 1996, 2017), or analyze these questions with different structures that nevertheless yield the same root denotation (Fox 2012a,b). However, empirical data cast doubt on the wisdom of this joint analysis for pair-list readings: pair-list readings of ∀-questions and multiple-*wh* questions contrast in domain exhaustivity, which argues that these two types of questions have different meanings and different structures.

The other direction is to assume that pair-list readings of ∀-questions involve 'quantification into questions (QiQ)' (Groenendijk and Stokhof 1984; Chierchia 1993; a.o.), which only arise in questions with a quantificational subject. An informal paraphrase for QiQ-readings is given in (4), where 'Det' stands for a determiner.

(4) Which movie did Det-boy(s) watch? (QiQ-reading)
≈ 'For Det-boy(s), [you tell me]/[I ask you] which movie did he/they watch?'

This paraphrase directly applies to *wh*-questions with an existential (∃-)quantifier (called '∃-questions'). For example, (5) exhibits a similar ambiguity between an individual reading and a QiQ-reading.[1] In contrast to the ∀-question in (3), here the QiQ-reading has a 'choice' flavor (Groenendijk and Stokhof 1984), and it doesn't call for a pair-list answer. As paraphrased in (5b), the QiQ-/choice reading asks the addressee to choose one/two of the boys and specify the unique movie he/they watched.

(5) Which movie did one/two of the boys watch?

a. 'Which movie y is s.t. one/two of the boys watched y?' '*Ironman*.' (Individual)

b. 'For one/two of the boys, [you tell me] which movie did he/they watch?' (Choice)
'Andy watched *Ironman*.'/ 'Billy and Clark watched *Spiderman*.'

However, many quantifiers cannot participate in QiQ-readings. For example, in (6) the *wh*-question with a negative quantifier (called 'no-question') cannot be responded to by silence.

(6) Which movie did {no boy, none of the boys} watch?

a. 'Which movie y is s.t. no boy watched y?' '*Revengers*.' (Individual)

b. 'Which function f to atomic movies is s.t. no boy x_i watched $f(x_i)$?' (Functional)
'The movie recommended by their$_i$ grandfather.'

[1] In ∃-questions, functional readings are only marginally acceptable. For example, the fragment functional answer (i-a) is under-informative; the identity of the boy who watched a movie has to be specified, as in (i-b). I leave this puzzle open.

(i) (Context: Among the relevant boys, only Andy watched a movie, which was his favorite superhero movie *Ironman*.)
Which movie did one of the boys watch?

a. ?? His favorite superhero movie.

b. Andy watched his favorite superhero movie.

c. # 'For no boy, [you tell me] which movie did they watch?' [Silence] (✗QɪQ)

In sum, it remains controversial whether we should treat questions with pair-list readings (abbreviated as 'pair-list questions') uniformly or questions with QɪQ-readings (abbreviated as 'QɪQ-questions') uniformly. This paper argues in favor of the latter option. The presented proposal delivers a synthesis that manages to give a distinctive analysis to pair-list readings across origins, while at the same time deriving the newly discovered subtle semantic differences w.r.t. domain exhaustivity within the descriptive umbrella category 'pair-list' from their differences in structural origin. This proposal also accounts for the distributional constraints and variations of QɪQ-readings.

I propose that QiQ-questions and pair-list multiple-*wh* questions both involve a '*wh*-dependency', namely, that the quantificational/*wh*- subject stands in a dependency with the trace of the *wh*-object. The core analysis is sketched in (7). The complex object-trace t_j^i carries a functional index j bound by the *wh*-object, as well as an argument index i bound by the quantificational/*wh*- subject (à la Chierchia 1993). In the multiple-*wh* question (7b), the *wh*-subject binds into the trace of the *wh*-object across an identity (IDENT) operation; in the QɪQ-question (7a), the quantificational subject binds into the trace of the *wh*-object across a predication (PRED) operation. Differences between these two operations explain the contrast in domain exhaustivity between these two questions (see Sect. 6).

(7) Composition schema for complex questions:

a. Which boy watched which movie? (Pair-list reading)
... [which-movie$_j$ ... which-boy$_i$ [... IDENT ... [t_i watched t_j^i]]]

b. Which movie did Dᴇᴛ-boy(s) watch? (QɪQ-reading)
... [which-movie$_j$ ... Dᴇᴛ-boy(s)$_i$ [... PRED ... [t_i watched t_j^i]]]

The rest of this paper is organized as follows. Section 2 presents evidence against the view of treating the two types of pair-list questions (i.e., ∀-questions and multiple-*wh* questions) uniformly as well as evidence that supports the view of treating QɪQ-questions uniformly. Section 3 lays out the technical challenges and relevant facts that this paper aims to account for. Section 4 reviews two influential approaches to composing pair-list questions, namely, the functionality approach of Dayal (1996, 2017) and the family-of-questions approach of Fox (2012a,b). My analysis take ingredients from these two approaches while overcoming their problems. Section 5 introduces a GB-style categorial approach to composing questions. Section 6 puts forward my central analysis for the composition of pair-list multiple-*wh* questions and QɪQ-questions. The denotations and the composition procedures of these two types of complex questions will be presented in tandem. Section 7 accounts for the quantificational (Q-)variability effects in embeddings of pair-list questions. Section 8 concludes. Appendices A and B review two additional existing accounts of QɪQ-question composition.

2. Arguments for unifying the derivation of QɪQ-readings

This section argues that pair-list ∀-questions should be composed uniformly as other QɪQ-questions, not as their multiple-*wh* counterparts. First, when ∀-questions have pair-list readings, they are subject to domain exhaustivity, whereas their multiple-*wh* counterparts are not (Sect. 2.1). This contrast argues that these two types of questions should be interpreted and composed differently. Secondly, QɪQ-questions exhibit the same subject–object/adjunct asymmetry. What's more, the distributional pattern of QɪQ-readings is preserved in questions where the subject is a coordination of quantifiers (Sect. 2.2). These facts argue that QɪQ-questions have a uniform syntax.

2.1. A contrast in domain exhaustivity

It is commonly claimed that pair-list readings of multiple-*wh* questions and ∀-questions both exhibit 'domain exhaustivity' (Dayal 1996, 2002; a.o.). For a question with a *wh*/∀-subject and a *wh*-object, the domain exhaustivity condition says that every member of the set quantified over by the *wh*/∀-subject is paired with a member of the set quantified over by the *wh*-object. For instance, in (3) and (2), repeated as (8a, b), domain exhaustivity requires that <u>every</u> boy watched a (possibly different) movie. Moreover, since the *wh*-object is singular (i.e., the *wh*-complement *movie* is singular), the two questions are also subject to 'point-wise uniqueness', which says that each boy watched <u>at most one</u> movie.

(8) a. Which movie did every/each boy watch?
b. Which boy watched which movie?

The point-wise uniqueness effect is easy to attest, but the domain exhaustivity effect is not so obvious. In the multiple-*wh* question (8b), for example, it is unclear which set of boys is quantified over by the *wh*-subject; domain exhaustivity would be trivial if the domain of quantification consisted of only the boys who did watch a movie. To remove this confound, Fox (2012a) uses the pair of examples in (9), where the quantification domain of each *wh*-phrase is explicitly specified.[2] Fox claims that (9b) rejects a pair-list reading (in contrast to (9a)), since interpreting this question with a pair-list reading would give rise to a domain exhaustivity condition that is contextually infelicitous — pairing four kids with three chairs implies that there will be multiple kids sitting on the same chair.

(9) a. Guess which one of the **three** kids will sit on which one of the **four** chairs.
b. Guess which one of the **four** kids will sit on which one of the **three** chairs.

Contrary to this widely adopted view, I argue that pair-list multiple-*wh* questions are not subject to domain exhaustivity. First, multiple-*wh* questions can be felicitously uttered in pair-list contexts where domain exhaustivity is violated. In (10), the sentence repeated from (9b) is felicitous and must be interpreted with a pair-list reading.

(10) (Context: Four kids are playing Musical Chairs and are competing for three chairs.)
Guess which one of the **four** kids will sit on which one of the **three** chairs.
$\not\rightsquigarrow$ 'Each of the four kids will sit on one of the three chairs.'

The game rules of Musical Chairs yield two conditions: (i) one of the four kids will not sit on any of the three chairs, and (ii) the remaining three kids will each sit on a different chair. Condition (ii) ensures that the embedded multiple-*wh* question has a pair-list reading, not a single-pair reading. Condition (i) contradicts the domain exhaustivity inference that each of the kids will sit on one of

[2] One might wonder whether specifying the domain of quantification explicitly can sufficiently remove the confound with domain exhaustivity — could there be additional covert domain restrictions with the *wh*-phrases? In (9) and (10), for example, the confound would remain if the quantification domain of *which one of the four kids* were covertly restricted to a subset of the four kids, excluding the kid who will not sit on a chair. I argue that such covert restrictions are not possible once the quantification domain of a *wh*-phrase has been specified explicitly. As seen in (i), uniqueness is assessed relative to a domain containing all four contextually relevant kids, as in (ia); if the phrase *which one of the four kids* could range over a subset of the four kids, the uniqueness inference would be as weak as (ib), contrary to fact.

(i) Which one of the four kids cried?
a. $\rightsquigarrow$ 'Among the four kids, only one cried.'
b. $\not\rightsquigarrow$ 'Among a certain subset of the four kids, only one cried.'

the chairs. If pair-list readings of multiple-*wh* questions were subject to domain exhaustivity, (10) would suffer a presupposition failure, contrary to fact.

Second, unlike their multiple-*wh* counterparts, pair-list ∀-questions cannot be felicitously used in contexts where domain exhaustivity is violated. In the context in (11), the quantification domain of the *wh*-/quantificational subject is greatly larger than that of the *wh*-object. The multiple-*wh* question (11a) is fully acceptable in this context, but the ∀-question (11b) is not: (11b) presupposes that each candidate will get one of the jobs, contrary to the context.

(11) (Context: **100** candidates are competing for **three** job openings.)
   a. ✓ Guess which candidate will get which job.
   b. # Guess which job every candidate will get.

Likewise, in the Musical Chairs scenario, the multiple-*wh* question is felicitous, but the corresponding ∀-question is not.

(12) (Context: Four kids are playing Musical Chairs and are competing for three chairs.)
   a. Guess which one of the four kids will sit on which one of the three chairs. =(10)
   b. # Guess which one of the three chairs each of the four kids will sit on.

One might argue that the domain exhaustivity condition of a pair-list multiple-*wh* question can be associated with any of the *wh*-phrases, including the *wh*-object. For example, in (10) and (11), it could be the case that domain exhaustivity requires every chair and every job to be taken by a kid and a candidate, respectively. This possibility is ruled out as follows: a multiple-*wh* question can be uttered in a pair-list context where neither type of domain exhaustivity is satisfied. For example, sentence (13) is felicitous, although it does not imply a domain exhaustivity inference relative to the boys or to the girls.

(13) (Context: Four boys and four girls will form four boy-girl pairs to perform in a dance competition, but only **two** of the pairs will get into the final round.)
   Guess which one of the **four** boys will dance with which one of the **four** girls in the final round.
   $\not\rightsquigarrow$ 'Each of the four boys will dance with one of the four girls in the final round.'
   $\not\rightsquigarrow$ 'Each of the four girls will dance with one of the four boys in the final round.'

In conclusion, pair-list ∀-questions are subject to domain exhaustivity, whereas pair-list multiple-*wh* questions are not. This contrast argues that these two types of questions should be interpreted differently and composed differently.

2.2. Uniform distribution of QɪQ-readings

The distribution of QɪQ-readings uniformly exhibits a subject–object/adjunct asymmetry (May 1985, 1988; Chierchia 1991, 1993). As seen in (14) and (15), pair-list readings and choice readings are available if the non-*wh* quantifier serves as the subject while the *wh*-phrase serves as the object, but not vice-versa. In (14b), the uniqueness inference triggered by the singular *wh*-subject must take wide scope relative to the ∀-object. As for the ∃-questions in (15), although (15b) marginally admits a choice reading, (15a) is preferable if the questioner seeks a choice answer.[3] The subject–adjunct

[3]The reason why (15b) and (16d) marginally admit choice readings might be that indefinites have more ways to take wide scope than ∀-quantifiers, such as through globally bound choice functions.

asymmetry is analogous, as illustrated in (16). Hence, unless there is compelling evidence to suggest otherwise, it is plausible to assume that QɪQ-readings are derived uniformly.

(14) (Context: Today the ten students cast their votes for this year's class speaker. There were three candidates. Every student had exactly one vote. The questioner wants to know all of the student-candidate pairs.)

a. Which candidate did every student vote for? (✓Pair-list)

b. # Which student voted for every candidate? (✗Pair-list)
$\rightsquigarrow$ 'Exactly one of the students voted for every candidate.'

(15) (Context: Today the ten students cast their votes for this year's class speaker. There were three candidates. Every student had exactly one vote. The questioner is only interested in knowing one of the student-candidate pairs.)

a. Which candidate did one of the students vote for? (✓Choice)
Andy voted for the first candidate.

b. ? Which student voted for one of the candidates? (?Choice)

(16) (Context: The car race passed through our town. Every driver refueled exactly once at one of our gas stations.)

a. At which station did every driver refuel? (✓Pair-list)

b. # Which driver refueled at every gas station? (✗Pair-list)

c. At which station did one of the drivers refuel? (✓Choice)

d. ? Which driver refueled at one of the nearby stations? (?Choice)

The idea of unifying the derivation of QɪQ-readings is further supported by the blocking effect of negative quantifiers. In (17a), where the subject is a conjunction of a $\forall$-quantifier and an $\exists$-quantifier, the pair-list reading associated with the $\forall$-quantifier and the choice reading associated with the $\exists$-quantifier are both preserved. This question asks the addressee to specify all of the boy-watch-movie pairs and one of the girl-watch-movie pairs. In contrast, since negative quantifiers do not participate in QɪQ-readings (as seen in (6)), coordinating a $\forall/\exists$-quantifier with a negative quantifier blocks the QɪQ-reading. For example, (17b) does not have the reading that requests the addressee to list all the boy-watch-movie pairs and not to list any teacher-watch-movie pairs. (For an explanation based on 'LF efficiency', see Sect. 6.4.3.)

(17) a. Which movie did [each of the boys and one of the girls] watch? (✓QɪQ)

b. Which movie did [each of the boys and none of the teachers] watch? (✗QɪQ)

c. Which movie did [one of the girls and none of the teachers] watch? (✗QɪQ)

3. Challenges and goals

Section 2 has laid out two goals of this paper: (i) to compose QɪQ-questions uniformly, and (ii) to compose pair-list multiple-*wh* questions and pair-list $\forall$-questions in parallel to account for their similarities in meaning and form while at the same time explaining their contrast in domain exhaustivity. It is not easy to achieve both goals: a proper solution needs to overcome several technical challenges and account for a number of semantic effects.[4]

[4] This paper does not attempt to explain effects that are more likely to be related to syntax in nature, such as constraints on extractions/movements. See Kotek 2014, 2019 and the references therein for detailed discussions.

First, for most frameworks of question semantics, the structure in (18) is ill-formed. The generalized quantifier 'Det-boy' takes arguments of type $\langle e, t \rangle$; thus it can only quantify into a t-type expression. However, the contained question *which movie did x watch* is not of type t; it is typically treated as a set of propositions as in Hamblin-Karttunen Semantics, or as a one-place predicate/property as in categorial approaches.

(18) Which movie did Det-boy(s) watch?
*[Det-boy(s) λx_e [which movie did x watch]]

There are two general strategies to solve this type-mismatch problem. One is to extract the domain of quantification of the subject via a type-shifting operation (Groenendijk and Stokhof 1984; Chierchia 1993; Dayal 1996, 2017; a.o.). For example, Dayal extracts the quantification domain of a $\forall$-quantifier via the operation of shifting a quantifier into the unique minimal witness set of this quantifier. This strategy is feasible in principle but a bit *ad hoc* (see Sect. 4.1.2 and footnote 15).

The other strategy is to create a t-type constituent in the LF that the quantifier can quantify into standardly. For example, in partition semantics (Groenendijk and Stokhof 1984), which defines the root denotation of a question as a partition of possible worlds, the formation of a partition involves a t-type node expressing the equivalence of two extensions. Alternatively, Karttunen (1977) and Krifka (2001) recast quantifying into questions as quantifying into question-embeddings. The two analyses based on partitions and question-embeddings respectively overcome the type-mismatch problem but bring other problems (reviewed in Appendices A and B). In contrast, my proposal will follow Fox (2012b) in assuming that the root of a QiQ-question contains a t-type node that expresses a predication condition (Sects. 4.2 and 6.4).[5]

Second, pair-list readings have a limited distribution in matrix QiQ-questions: only *each*/*every*-phrases can license pair-list readings for matrix questions. For example, in a choice reading, the $\exists$2-question in (19) calls for cumulative answers like (19a); the pair-list answer (19b), which distributes over the two chosen students, is over-informative (Moltmann and Szabolcsi 1994; Szabolcsi 1997b). Questions with a definite plural like (20) are analogous (Srivastav 1991; Krifka 1991; for a different view from Johnston 2019, see Sect. 6.4.5).

(19) Who did two of the students vote for?
  a. Andy and Billy voted for Mary and Jill.
  b. Andy voted for Mary, and Billy voted for Jill.

(20) Who did the students vote for?

The confound from cumulative answers can be removed by replacing *who* with a singular *wh*-phrase, which triggers a uniqueness presupposition. In the following matrix questions, distributivity taking scope over uniqueness is possible only in (21a,b), where the quantifier in the subject is lexically distributive. In contrast, for example the $\exists$2-question (21d) presupposes that two of the students voted for the same candidate and only this candidate, which contradicts the context. (In (21d-f), '~~each~~' means that the reading doesn't involve covert distributivity between the subject and the uniqueness inference triggered by the singular *wh*-object.)

(21) I know that every student voted for a different candidate. Which candidate did ...

[5] Besides these two general strategies, Inquisitive Semantics also avoids this type-mismatch problem because it defines declaratives and interrogatives uniformly as a set of sets of propositions (of type $\langle stt, t \rangle$) and generalized quantifiers as functions of type $\langle \langle e, stt \rangle, t \rangle$. To my knowledge, this idea has not been explored extensively. For a possible direction, see Ciardelli and Roelofsen 2018: Sect. 4.3.3.

a. ... every student vote for? ($every \gg \iota$)

b. ... each student/ each of the students vote for? ($each \gg \iota$)

c. # ... all/most of the students vote for? ($all/most \gg \cancel{\text{EACH}} \gg \iota$)

d. # ... two of the students vote for? ($\exists 2 \gg \cancel{\text{EACH}} \gg \iota$)

e. # ... two or more students vote for? ($\exists 2+ \gg \cancel{\text{EACH}} \gg \iota$)

f. # ... the students vote for? ($the\text{-}NP_{\text{PL}} \gg \cancel{\text{EACH}} \gg \iota$)

To account for the limited distribution of pair-list readings, many works on question composition propose to derive pair-list readings in a way that crashes whenever the quantificational subject of the question is not universal (e.g., Dayal 1996 and Fox 2012b; for details, see Sect. 4.) This strategy, however, comes at the cost of failing to account for the choice readings of ∃-questions. In contrast, I argue that a non-interrogative DP can participate in pair-list readings only if it is lexically distributive and can productively scope out of its surface position. In my analysis, the above distributional constraints of pair-list readings follow independently observed contrasts in lexical distributivity and scoping between *every*/*each*-phrases and the other quantifiers (for details, see Sects. 6.4.2 and 6.4.4).

Third, there are several semantic effects robustly observed in QIQ-questions and pair-list multiple-*wh* questions. Section 2.1 has discussed two effects, namely, the uniqueness effect triggered by the singular *wh*-object, as seen in (22a-c), and the domain exhaustivity effect observed only in ∀-questions, as seen in (22a). These effects were not extensively considered until Srivastav 1991/Dayal 1996.

(22) a. Which movie did every/each boy watch?
⇝ 'For every boy x, x watched exactly one movie.'

b. Which boy watched which movie?
⇝ 'For every boy x s.t. x watched a movie, x watched exactly one movie.'

c. Which movie did one/two of the boys watch?
⇝ 'For some x s.t. x is one/two of the boys, x watched exactly one movie.'

Moreover, embeddings of pair-list questions exhibit 'quantificational (Q-)variability'. As first observed by Berman (1991), question-embeddings modified by a quantificational adverbial (e.g., *mostly, partly, for the most part, in part*) have a Q-variability inference. As illustrated in (23) and (24), in the paraphrase of this inference, the quantification domain of the matrix quantity adverbial *mostly* can be thought of as (a) a set of propositions (Lahiri 1991, 2002; Cremers 2016), (b) a set of sub-questions (Beck and Sharvit 2002), or (c) a set of individuals or pairs (Xiang 2016, 2019a, 2020; Cremers 2018).

(23) Jill mostly knows [which students left].

a. ⇝ 'Most p: p is a true proposition of the form ⌜*student-x left*⌝, Jill knows p.'

b. ⇝ 'Most Q: Q is a question of the form ⌜*whether student-x left*⌝, Jill knows Q.'

c. ⇝ 'Most x: x is an atomic student and x left, Jill knows that x left.'

(24) Jill mostly knows [$_{\text{PAIR-LIST}}$ $\left\{\begin{array}{l}\text{which movie every boy watched}\\ \text{which boy watched which movie}\end{array}\right\}$].

a. ⇝ 'Most p: p is a true proposition of the form ⌜*boy-x watched movie-y*⌝, Jill knows p.'

b. ⇝ 'Most Q: Q is a question of the form ⌜*which movie boy-x watched*⌝, Jill knows Q.'

c. ⇝ 'Most $\langle x, y\rangle$: $\langle x, y\rangle$ is a boy-movie pair and x watched y, Jill knows that x watched y.'

It is commonly claimed that family-of-questions approaches are advantageous in accounting for the Q-variability inference of (24): if one conceives of the embedded pair-list question as a family of sub-questions, the Q-variability inference can be defined as in (24b). In contrast, I assume a categorial approach to defining and composing questions and therefore argue that this inference can be derived as in (24c), which is compatible with a simple functionality approach (for details, see Sect. 7).

4. Two general approaches to composing complex questions

There is a rich literature on the composition of pair-list multiple-*wh* questions and questions with a quantifier. This section reviews two types of approaches that have tackled both types of questions: 'functionality approaches', which assume that these complex questions involve a *wh*-dependency, and 'family-of-questions approaches', which define each such question as a family of sub-questions.[6]

In this review section, I will focus on two influential accounts by Dayal (1996, 2017) and Fox (2012a,b), which successfully account for the domain exhaustivity and point-wise uniqueness effects in ∀-questions with singular *wh*. My analysis will take ingredients from both these accounts. For extensive literature reviews, see the appendices of this paper, as well as Xiang 2016: Chaps. 5 and 6, Dayal 2017: Chap. 4, and Ciardelli and Roelofsen 2018.

4.1. Functionality approaches

Wh-questions with a functional reading (called 'functional questions') express a dependency between the non-*wh*-subject and the *wh*-object/adjunct. In (25), the fragment answer contains a pronoun interpreted as being bound by the quantificational subject in the question.

(25) Which movie did every-boy$_i$ watch?
His$_i$ favorite superhero movie.

As for pair-list questions, functionality approaches assume that pair-list readings also involve a dependency between the ∀/*wh*-subject and the *wh*-object. For example, the pair-list answer (26a) is thought of as the specification of the 'graph' of the function (26b): it pairs elements of the set that the ∀/*wh*-subject ranges over with elements of the set that the *wh*-object ranges over.[7]

(26) Which movie did every boy watch?/ Which boy watched which movie?

a. Andy watched *Ironman*,
Billy watched *Spiderman*,
Clark watched *Hulk*.

b. $f = \begin{bmatrix} a & \rightarrow & i \\ b & \rightarrow & s \\ c & \rightarrow & h \end{bmatrix}$

Functionality approaches were originally proposed for ∀-questions only (Engdahl 1980, 1986; Chierchia 1993). The primary goal of assuming functionality was to account for the subject–object/adjunct

[6] The core assumptions of these two approaches are compatible with each other. For example, Chierchia (1993) assumes a *wh*-dependency while defining a QɪQ-question as a family of questions. For more details, see footnote 15.

[7] One might wonder why we chose to treat pair-list readings as special functional readings, not vice versa. The reason is that pair-list readings are subject to more constraints than functional readings. As seen in (i), multiple-*wh* questions are congruent with fragment answers that are lists of pairs, but not with intensional functional answers (Kang 2012; Sharvit and Kang 2017). If pair-list readings were more general than functional readings, we wouldn't expect such a gap.

(i) Which boy watched which movie?
a. # His favorite superhero movie.
b. Andy, *Ironman*, Billy, *Spiderman*, Clark, *Hulk*.

Sharvit and Kang (2017) provide an explanation as to why pair-list questions do not admit intensional functional answers. However, the syntax of multiple-*wh* questions assumed by Sharvit and Kang is quite different from mine. This paper leaves this issue open.

asymmetry uniformly observed in functional readings and pair-list readings of ∀-questions, as illustrated in the following:

(27) Which woman did every boy invite? (✓Individual, ✓Functional, ✓Pair-list)
  a. Anna.
  b. His mother.
  c. Andy invited Mary, Billy invited Susi, Clark invited Jill.

(28) Which woman invited every boy? (✓Individual, ✗Functional, ✗Pair-list)
  a. Anna.
  b. # His mother. (Intended: 'Every-boy$_i$ was invited by his$_i$ mother.')
  c. # Mary invited Andy, Susi invited Billy, Jill invited Clark.

Assuming functionality, one can explain this asymmetry in terms of constraints on dependencies/ binding (Chierchia 1993; Williams 1994; Shan and Barker 2006; a.o.). For example, Chierchia (1993) argues that weak crossover arises if the object/adjunct binds into the trace of the *wh*-subject.[8] See also Jacobson 1994 and Sharvit 1997, 1999 for functionality approaches to functional readings and pair-list readings of relative clauses with quantifiers.

Further, Dayal (1996, 2017) extends the functionality approach to pair-list multiple-*wh* questions. She observes that the corresponding relation expressed by a pair-list answer is a function: the correspondence can be one-to-one or many-to-one, but not one-to-many, as witnessed in (29). See also Caponigro and Fălăuş 2020 for an application to multiple-*wh* free relatives in Romanian.

(29) Which student talked to which professor? (Dayal 2017: 96)
  a. Alice talked to Professor Carl, and Bill talked to Professor Dan.
  b. Alice and Bill both talked to Professor Carl.
  c. # Alice talked to Professors Carl and Dan.

By assuming functionality, my proposal inherits the advantages of explaining the subject–object/adjunct asymmetry and the unavailability of one-to-many relations in terms of constraints on functionality.[9]

[8] Chierchia (1993) assumes that the *wh*-trace carries two indices, namely, a functional index i bound by the *wh*-phrase and an argument index j co-indexed with the non-interrogative quantifier. To bind the j-index carried by the *wh*-trace, the non-interrogative quantifier has to be moved to a position that c-commands this *wh*-trace. Thus in (ib), unlike (ia), when the quantifier *every boy* is moved from a position lower than the *wh*-trace, it inevitably moves across a co-indexed expression (viz., the *wh*-trace), causing weak crossover.

(i) a. Which movie did every boy watch?
  [which-movie$_i$ ... [every-boy$_j$ ... [t_j watched t_i^j]]] (No crossover)
 b. Which boy watched every movie?
  *[which-boy$_i$ ... [every-movie$_j$ ... [t_i^j watched t_j]]] (With weak crossover)

In contrast, competing accounts from Safir (1984) and May (1988) analyze the asymmetry and weak crossover in terms of syntactic constraints.

[9] It might look appealing to analyze the subject–object/adjunct asymmetry in QɪQ-questions and superiority-effects in multiple-*wh* questions uniformly. For example, Hornstein (1995) extends Chierchia's (1993) complex-trace analysis of *wh*-dependencies to superiority effects. He assumes that the in-situ *wh*-phrase contains a covert *pro* co-indexed with the fronted *wh*-phrase. Accordingly, in (i-b), moving the object *what* across the co-indexed *pro* causes weak crossover.

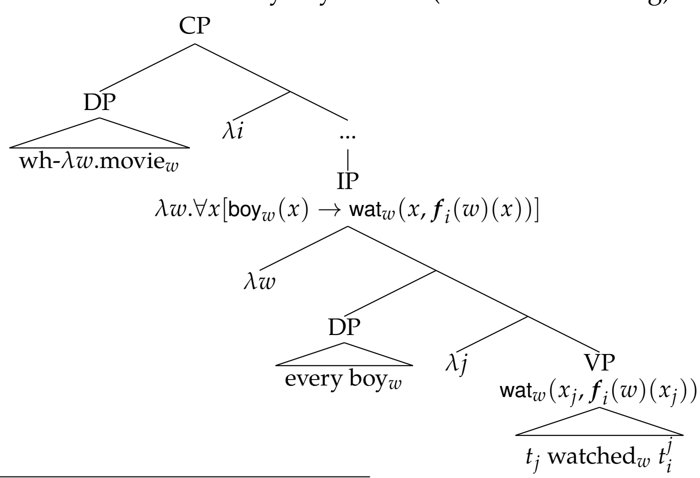
(i) a. Who bought what? (Superiority-obeying)
  [who$_i$ [t_i bought *pro*$_i$-what$_j$]] (No crossover)
 b. ?? What did who buy? (Superiority-violating)
  *[what$_i$ ... did [*pro*$_i$-who$_j$ buy t_i]]] (With weak crossover)

Moreover, in Sect. 6, I will show that *wh*-dependencies are independently needed to account for the contrast in domain exhaustivity between multiple-*wh* questions and ∀-questions.

4.1.1. *Wh*-dependency in basic functional questions

In the current dominant analysis, *wh*-dependencies in functional questions are derived by assuming a complex *wh*-trace (Groenendijk and Stokhof 1984; Chierchia 1993; a.o.).[10] The tree diagram in (30) illustrates the LF schema assumed to compose a functional ∀-question.[11] In this LF, the *wh*-trace t_i^j carries two indices, namely, an intensional functional index i (of type $\langle s, ee\rangle$) bound by the fronted *wh*-object and an argument index j (of type e) co-indexed with the trace of the quantificational subject. With such indexations, the VP denotes an open sentence expressing a dependency between the two arguments of *watched*, and the IP denotes a universal inference over this dependency, read as 'every boy x watched $\boldsymbol{f}_i(x)$'. Details of composition above IP are omitted for now because they vary with the framework of question composition. I will add more details in Sect. 5.

(30) Which movie did every boy watch? (Functional reading)



Relatedly, Shan and Barker (2006) argues that binding relations must be evaluated from left to right, and they use this single constraint to rule out crossover and superiority violations.

In contrast, I argue that superiority effects in multiple-*wh* questions and the subject–object/adjunct asymmetry in QiQ-questions have different origins: as seen in (ii), multiple-*wh* questions with *which*-phrases tolerate superiority violations and admit pair-list readings (Pesetsky 1987, 2000; Kotek 2014, 2019).

(ii) Which movie did which boy watched?
Andy watched *Ironman*, Billy watched *Spiderman*, Clark watched *Hulk*.

The analysis presented in this paper makes no prediction on the overt syntax of multiple-*wh* questions. Despite of obvious insufficiencies (see Sect. 6.3), this analysis is exempt from the under-generation problem.

[10] In contrast to the complex-trace approach, Jacobson (1999, 2014) develops a variable-free approach to functionality which does not make use of indices. In her analysis, functionality is derived by a type-shifting rule, called 'the **z**-rule', which closes off the dependency between the arguments of a predicate. (For example, $\mathbf{z}([\![\textit{watched}]\!]^w) = \lambda \boldsymbol{f}_{\langle e,e\rangle}\lambda x_e.[\![\textit{watched}]\!]^w(x, \boldsymbol{f}(x))$.) This approach is especially advantageous in tackling cases where the *wh*-dependent is in situ or inside an island. For ease of comparing with existing works on composing complex questions, this paper follows the complex-trace approach.

[11] Following Groenendijk and Stokhof (1984), I translate LF representations into the Two-sorted Type Theory (Ty2) of Gallin (1975). Ty2 differs from Montague's intensional logic in that it introduces s (the type of possible worlds) as a basic type (just like e and t), and in that it uses variables and constants of type s which can be thought of as denoting possible worlds. For example, the English common noun *boy* is translated into boy_w in Ty2, where boy is a property of type $\langle s, et\rangle$ and w a world variable of type s. With these assumptions, Ty2 can make direct reference to worlds and allows quantification and abstraction over world variables.

4.1.2. Dayal (1996, 2017) on composing pair-list questions

Dayal (1996, 2017) assumes that the two pair-list questions in (31) uniformly denote a set of conjunctive propositions, and that each of these conjunctive propositions specifies an $\langle e, e\rangle$-type function f from the quantification domain of the $\forall/wh$-subject (i.e., $\mathsf{boy}_@$) to the quantification domain of the *wh*-object (i.e., $\mathsf{mov}_@$).[12] This denotation yields domain exhaustivity since f is defined for every boy.

(31) Which movie did every boy watch?/ Which boy watched which movie?
(The discourse domain has two relevant boys b_1,b_2 and two relevant movies m_1,m_2.)

$$[\![Q_\forall]\!] = [\![Q_{\text{multiple-}wh}]\!] = \{ \bigcap\{\lambda w.\mathsf{wat}_w(x, f(x)) \mid \mathsf{boy}_@(x)\} \mid f \in [\mathsf{boy}_@ \to \mathsf{mov}_@]\}$$
$$= \left\{ \begin{array}{l} \lambda w.\mathsf{wat}_w(b_1, m_1) \wedge \mathsf{wat}_w(b_2, m_1) \\ \lambda w.\mathsf{wat}_w(b_1, m_1) \wedge \mathsf{wat}_w(b_2, m_2) \\ \lambda w.\mathsf{wat}_w(b_1, m_2) \wedge \mathsf{wat}_w(b_2, m_1) \\ \lambda w.\mathsf{wat}_w(b_1, m_2) \wedge \mathsf{wat}_w(b_2, m_2) \end{array} \right\}$$

Dayal assumes that the two pair-list questions in (31) are composed uniformly as in (32). In this LF, the quantificational/*wh*- subject and the *wh*-object are moved to the specifier of a single functional head $\mathrm{C}^0_{\text{FUNC}}$, and they each are turned into a set of entities via a type-shifting ($\uparrow_{\text{TS}}$) operation. The composition proceeds in three steps:

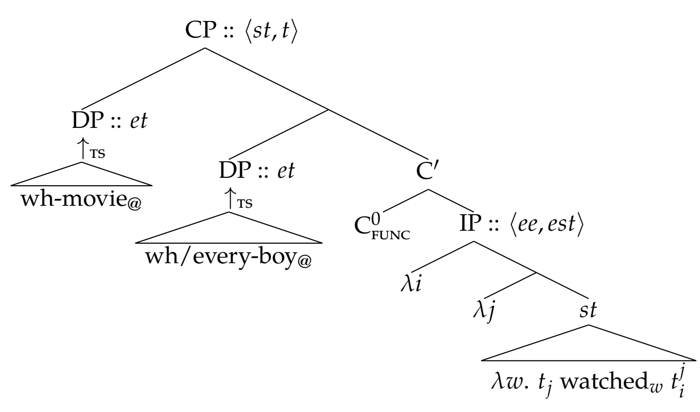
(i) The object-trace carries an extensional functional index i (of type $\langle e, e\rangle$) as well as an argument index j (of type e) that co-refers with the subject-trace. Abstracting these two indices at IP yields the property (33a), which maps an $\langle e, e\rangle$-type function and an individual to a dependency proposition (i.e., an open proposition that expresses a dependency between the two arguments of *watched*).

(ii) The functional head $\mathrm{C}^0_{\text{FUNC}}$ introduces domain and range for the function f and creates a 'graph' for f. If q (of type $\langle ee, est\rangle$) is the denotation of IP, the resulting graph of f based on q is the conjunction of the propositions of the form $\ulcorner q(f)(x)\urcorner$, where x is in the domain of f.

(iii) The sets that the $\forall/wh$-phrases range over are extracted by type-shifting operations.[13] These sets saturate the range and domain arguments introduced by $\mathrm{C}^0_{\text{FUNC}}$.

With this composition, the CP is interpreted as a set of conjunctive propositions, each of which specifies an $\langle e, e\rangle$-type function that is only defined for the set that the $wh/\forall$-subject ranges over.

(32)



[12] For simplicity, I assume that the extensions of *wh*-complements are evaluated relative to the actual world '@'.

[13] Dayal (2017) considers two ways to obtain the quantification domain of a *wh*-phrase. One way is to define a *wh*-phrase as an $\exists$-quantifier and extract out its quantification domain via the application of a Be-shifter (Partee 1986). The other way is to define a *wh*-phrase as a set of entities and derive its quantificational meaning via an $\exists$-shifter.

(33) a. $[\![\text{IP}]\!] = \lambda f_{\langle e,e\rangle}\lambda x_e \lambda w.\textsf{wat}_w(x, f(x))$

b. $[\![\text{C}^0{}_{\text{FUNC}}]\!] = \lambda q_{\langle ee,est\rangle}\lambda D\lambda R\lambda p.\exists f \in [D \to R][p = \bigcap \lambda p'.\exists x \in D[p' = q(f)(x)]]$
$\quad = \lambda q_{\langle ee,est\rangle}\lambda D\lambda R.\{\bigcap\{q(f)(x) \mid x \in D\} \mid f \in [D \to R]\}$

c. $[\![\text{C}']\!] = \lambda D\lambda R\lambda p.\{\bigcap\{\lambda w.\textsf{wat}_w(x, f(x)) \mid x \in D\} \mid f \in [D \to R]\}$

d. $[\![\text{CP}]\!] = \{\bigcap\{\lambda w.\textsf{wat}_w(x, f(x)) \mid x \in \textsf{boy}_@\} \mid f \in [\textsf{boy}_@ \to \textsf{mov}_@]\}$

To account for the uniqueness effects of singular *wh*-phrases, Dayal defines the answerhood (Ans-)operator as in (34). This definition presupposes the existence of the strongest true answer. The strongest true answer to a question is the true proposition in the Hamblin set of this question that entails all the true propositions in this Hamblin set.

(34) $\textsc{Ans}_{Dayal}(w)(Q) = \exists p[w \in p \in Q \wedge \forall q[w \in q \in Q \to p \subseteq q]].$
$\iota p[w \in p \in Q \wedge \forall q[w \in q \in Q \to p \subseteq q]]$

The following presents how the $\textsc{Ans}_{Dayal}$-operator accounts for the observed uniqueness effects. The ontology of individuals assumes that a singular noun denotes a set of atomic entities, whereas that a plural noun ranges over both atomic and sum entities (Sharvy 1980; Link 1983). Adopting this ontology, Dayal argues that the Hamblin set of the singular-*wh* question (35a) consists of only propositions naming an atomic boy, and that the Hamblin set of the corresponding plural-*wh* question (35b) includes also propositions naming a sum of boys. In a discourse where two boys Andy and Bill both watched *Hulk*, the true answers to these two questions are as in (35a') and (35b'), respectively. Crucially, the answer set (35b') has a strongest member $\lambda w.\textsf{wat}_w(a \oplus b, h)$, whereas (35a') doesn't; thus employing $\textsc{Ans}_{Dayal}(w)$ in (35a) yields a presupposition failure. Hence, question (35a) can only be felicitously uttered in worlds where only one of the boys watched *Hulk*, which explains its uniqueness effect.

(35) (Among the considered boys, only Andy and Billy watched *Hulk* in w.)

a. Which boy watched *Hulk*?

a$'$. $\{\lambda w.\textsf{wat}_w(a, h), \lambda w.\textsf{wat}_w(b, h)\}$

b. Which boys watched *Hulk*?

b$'$. $\{\lambda w.\textsf{wat}_w(a, h), \lambda w.\textsf{wat}_w(b, h), \lambda w.\textsf{wat}_w(a \oplus b, h)\}$

In a pair-list question, if the *wh*-object is singular, the presupposition of $\textsc{Ans}_{Dayal}$ yields point-wise uniqueness. Take (31) for example: if in w_1 the boy b_1 watched only m_1 but the boy b_2 watched both m_1 and m_2, then the top two propositions in the Hamblin set (31) are both true in w_1. Since neither of the true propositions is stronger than the other, applying $\textsc{Ans}_{Dayal}(w_1)$ yields a presupposition failure.

Dayal's analysis also accounts for the domain exhaustivity and point-wise uniqueness effects in pair-list $\forall$-questions with a singular *wh*-object: domain exhaustivity is hard-wired into the meaning of C^0_{FUNC}; point-wise uniqueness comes from the conjunctive closure encoded within the meaning of C^0_{FUNC} and the presuppositional $\textsc{Ans}_{Dayal}$-operator. This account also manages to keep the semantic type of questions low (i.e., single/double-*wh* questions and $\forall$-questions are uniformly of type $\langle st, t\rangle$), reserving more elaborate tools for tackling *wh*-constructions that are more complex than pair-list questions (e.g., *wh*-triangles, multiple-*wh* echo questions).

However, this analysis faces a number of problems. On the conceptual side, the composition involves several *ad hoc* or problematic assumptions. First, in the process, index abstractions are isolated from the moved phrases. Since here the IP involves multiple abstractions, isolating these abstractions from the moved phrases renders the binding relations ambiguous. Second, C^0_{FUNC} is structure-specific and the meaning assumed for it is rather complex. It is unclear why a functional head only appears in particular structures and why it has the complex lexical entry (33b). For these reasons, Dayal is not fully satisfied with the use of the complex C^0_{FUNC}; she calls this account the 'crazy C^0 approach'. Last, for the composition of ∀-questions, it is syntactically deviant to move a non-interrogative phrase to the specifier of an interrogative CP (Heim 2012).

This analysis also makes several problematic empirical predictions. (Note that these problems are independent from the assumption of functionality.) First, by composing pair-list ∀-questions and multiple-*wh* questions with the same LF, this account predicts that these questions are semantically equivalent. However, as argued in Sect. 2.1, the two types of questions differ in domain exhaustivity. As seen in (11), repeated below, only the multiple-*wh* question can be felicitously used in a pair-list context that violates domain exhaustivity.

(36) (Context: **100** candidates are competing for **three** job openings.)
  a. Guess which candidate will get which job.
  b. # Guess which job every candidate will get.

To account for the contrast in domain exhaustivity, one might assume a twin C^0_{FUNC} that doesn't force domain exhaustivity. But even with this assumption, it would still remain puzzling why this non-exhaustive C^0_{FUNC} cannot appear in pair-list ∀-questions.

Second, this account does not extend to ∃-questions with a choice reading. As seen in Sect. 3, only *every*/*each*-phrases can license pair-list readings for matrix questions. To avoid over-generating pair-list readings for matrix ∃-questions, Dayal stipulates that the quantification domain of a non-*wh* quantifier must be obtained by extracting the 'unique' minimal witness set of this quantifier.[14] As illustrated in Table 1, among the listed quantifiers, only the ∀-quantifiers have a unique minimal witness set which is not empty. In contrast, the ∃-quantifier has multiple minimal witness sets. The negative quantifier has a unique minimal witness set, but this set is the empty set. Dayal's stipulation reins in pair-list readings as intended, but it also renders the LF schema schema (33) unavailable for questions with a non-universal quantifier, which leaves choice readings of ∃-questions unexplained.

Generalized quantifier π	Minimal witness set(s) of π
every/each boy	$\{a, b, c\}$
one of the boys	$\{a\}$, $\{b\}$, $\{c\}$
no boy	$\varnothing$

Table 1: Illustration of minimal witness sets (with three relevant boys *a*,*b*,*c*)

Moreover, without further constraints, this analysis over-predicts pair-list readings for ∃-questions. As re-illustrated in (37a,b), Dayal composes the two pair-list questions uniformly, except that she uses two distinct type-shifting operations (marked as 'TS1' and 'TS2') to extract the set of boys from *which boy* and *every/each boy*. In this analysis, nothing prevents the corresponding ∃-question from being analyzed with the LF (37c), which gives rise to an unwanted pair-list reading. In syntax, if (37b) is well-formed, (37c) should be well-formed as well. In semantics, since *one of the boys* and *which*

[14] Live-on sets and witness sets are defined as follows (Barwise and Cooper 1981): For any π of type $\langle et, t\rangle$, π lives on a set B iff $\pi(C) \Leftrightarrow \pi(C \cap B)$ for any set C; if π lives on B, then A is a witness set of π iff $A \subseteq B$ and $\pi(A)$.

boy are semantically equivalent, type-shifting operations available for *which boy* should be equally available for *one of the boys*; therefore, (37c) yields the same pair-list reading as in (37a).

(37) All the following LFs yield a pair-list reading:

a. [TS1(which-movie) [TS1(which-boy) [C^0_{FUNC} [$_{\text{IP}}$ ...]]]] Multiple-*wh* question

b. [TS1(which-movie) [TS2(every/each-boy) [C^0_{FUNC} [$_{\text{IP}}$ ...]]]] $\forall$-question

c. [TS1(which-movie) [TS1(one-of-the-boys) [C^0_{FUNC} [$_{\text{IP}}$ ...]]]] $\exists$-question

Third, as pointed out by Lahiri (2002), since it defines a pair-list question as a set of conjunctive propositions, this analysis has difficulties in accounting for the Q-variability effects in embeddings of pair-list questions. For example, sentence (38) implies a quantificational inference, which can be paraphrased as if the matrix adverbial *mostly* quantified over a set of atomic propositions. However, these atomic propositions cannot be retrieved from the question denotation assumed in (31): we cannot retrieve the atomic propositions directly from the conjunction of these propositions.

(38) Jill mostly knows [$_{\text{PAIR-LIST}}$ $\left\{\begin{array}{l}\text{which movie every boy watched}\\ \text{which boy watched which movie}\end{array}\right\}$].

$\rightsquigarrow$ 'Most *p*: *p* is a true proposition of the form $\ulcorner$*boy-x watched movie-y*$\urcorner$, Jill knows *p*.'

To account for the Q-variability effects, in an unpublished work, Dayal (2016) removes the $\bigcap$-closure from the lexical entry of C^0_{FUNC} and defines the root of a pair-list question as a family of sets of propositions. The revised account manages to keep the atomic propositions alive, but it sacrifices the advantage of keeping the semantic type of questions low.

4.2. Family-of-questions approaches

Family-of-questions approaches regard a pair-list question as a set/family of sub-questions (Hagstrom 1998; Preuss 2001; Fox 2012a,b; Nicolae 2013; Kotek 2014; Xiang 2016: Chap. 5; Dayal 2016; a.o.). As exemplified in (39), if a simplex question denotes a set of propositions, a family of questions denotes a set of sets of propositions.[15]

[15] The approaches by Groenendijk and Stokhof (1984) and Chierchia (1993) are also family-of-questions approaches. They define a QIQ-question as a family of sub-questions ranging over a minimal witness set (MWS) of the subject quantifier, as in (i). ('$\mathcal{P}_{\text{boy}_@}$' stands for a generalized quantifier ranging over the set of atomic boys. 'MWS$(\mathcal{P}_{\text{boy}_@}, A)$' means that A is a minimal witness set of $\mathcal{P}_{\text{boy}_@}$.)

(i) $[\![\textit{Which movie did } \mathcal{P}_{\textit{boy}_@} \textit{ watch?}]\!]_{\text{QiQ}} = \{[\![\textit{Which member of A watched which movie?}]\!] \mid \text{MWS}(\mathcal{P}_{\text{boy}_@}, A)\}$

However, the predictions made by these accounts are quite different from the predictions made by the non-flat semantics in (39). For example, Chierchia (1993) defines a sub-question as a set of propositions of the form $\ulcorner$*boy-x watched movie-$f(x)$*$\urcorner$, as schematized in (ii). The related $\forall/\exists$-questions are thus defined as in (iii).

(ii) $[\![Q_{\mathcal{P}}]\!] = \{\{\lambda w.\mathsf{wat}_w(x, f(x)) \mid x \in A, f \in [A \to \mathsf{boy}_@]\} \mid \text{MWS}(\mathcal{P}_{\text{boy}_@}, A)\}$

(iii) (The discourse domain has two boys b_1,b_2 and two relevant movies m_1,m_2.)

a. $[\![Q_\forall]\!] = \left\{\left\{\begin{array}{l}\lambda w.\mathsf{wat}_w(b_1, m_1), \lambda w.\mathsf{wat}_w(b_2, m_2)\\ \lambda w.\mathsf{wat}_w(b_1, m_2), \lambda w.\mathsf{wat}_w(b_2, m_2)\end{array}\right\}\right\}$

b. $[\![Q_\exists]\!] = \{\{\lambda w.\mathsf{wat}_w(b_1, m_1), \lambda w.\mathsf{wat}_w(b_1, m_2)\}, \{\lambda w.\mathsf{wat}_w(b_2, m_1), \lambda w.\mathsf{wat}_w(b_2, m_2)\}\}$

Chierchia further assumes that answering a family of sub-questions means answering *one* of the sub-questions (in contrast to Fox's assumption that answering a family of sub-questions means answering *all* of the sub-questions). Accordingly, since *one of the boys* has multiple minimal witness sets, the QIQ-reading of the $\exists$-question has a choice flavor. Although this account naturally extends to $\exists$-questions, it cannot explain the semantic effects in pair-list $\forall$-questions, such as domain exhaustivity and point-wise uniqueness.

(39) (The discourse domain has two relevant boys b_1,b_2 and two relevant movies m_1,m_2.)
Which movie did every boy watch?/ Which boy watched which movie?

$$\begin{aligned}[\![\mathrm{Q}_\forall]\!] = [\![\mathrm{Q}_{\text{multiple-}wh}]\!] &= \{[\![\textit{Which movie did x watch?}]\!] \mid x \in \mathsf{boy}_@\} \\ &= \{\{\lambda w.\mathsf{wat}_w(x,y) \mid y \in \mathsf{mov}_@\} \mid x \in \mathsf{boy}_@\} \\ &= \left\{ \begin{array}{l} \{\lambda w.\mathsf{wat}_w(b_1,m_1), \lambda w.\mathsf{wat}_w(b_1,m_2)\} \\ \{\lambda w.\mathsf{wat}_w(b_2,m_1), \lambda w.\mathsf{wat}_w(b_2,m_2)\} \end{array} \right\}\end{aligned}$$

The non-flat semantics assumed in (39) easily accounts for the Q-variability inferences of embeddings of pair-list questions. As in (40), such an inference can be defined as if the matrix adverbial *mostly* quantified over a set of sub-questions of the embedded question.

(40) Jill mostly knows $[_{\text{PAIR-LIST}} \left\{ \begin{array}{l} \text{which movie every boy watched} \\ \text{which boy watched which movie} \end{array} \right\}]$.
$\rightsquigarrow$ 'Most Q: Q is a question of the form ⌜*which movie boy-x watched*⌝, Jill knows Q.'

Fox (2012a,b) analyzes the two pair-list questions with different LFs that nevertheless yield the same root denotation. The LF of a pair-list multiple-*wh* question is illustrated in (41). Since *wh*-phrases are treated as indefinites (viz., $[\![\textit{which}]\!] = [\![\textit{some}]\!]$), this LF is read as 'the set of Q s.t. for some boy x, Q is identical to $[\![\textit{Which movie did x watch?}]\!]$', which is simply the set of questions of the form ⌜*Which movie did boy-x watch?*⌝. The composition follows the GB-style Karttunen semantics (Heim 1995) except that it treats the identity (ID-)operator as type-flexible and allows this operator to be iterated.

(41) Which boy watched which movie? (Pair-list reading)
$[_{\mathrm{CP}_2}\ \lambda Q_{\langle st,t\rangle}$ [wh-boy$_@$ λx_e $[_{\mathrm{C}'_2}$ $\mathrm{ID}(Q)$ $[_{\mathrm{CP}_1}$ λp_{st} [wh-movie$_@$ λy_e $[_{\mathrm{C}'_1}$ $\mathrm{ID}(p)$ $[_{\mathrm{IP}}$ x watched y]]]]]]]

a. $[\![\mathrm{ID}]\!] = \lambda\alpha_\tau\lambda\beta_\tau.\alpha = \beta$ (τ stands for an arbitrary type)

b. $[\![\mathrm{IP}]\!] = \lambda w.\mathsf{wat}_w(x,y)$

c. $[\![\mathrm{C}'_1]\!] = [\![\mathrm{ID}]\!](p)([\![\mathrm{IP}]\!])$
$= [p = \lambda w.\mathsf{wat}_w(x,y)]$

d. $[\![\mathrm{CP}_1]\!] = \lambda p.\exists y[\mathsf{mov}_@(y) \wedge p = \lambda w.\mathsf{wat}_w(x,y)]$
$= \{\lambda w.\mathsf{wat}_w(x,y) \mid \mathsf{mov}_@(y)\}$

e. $[\![\mathrm{C}'_2]\!] = [\![\mathrm{ID}]\!](Q)([\![\mathrm{CP}_1]\!])$
$= [Q = \{\lambda w.\mathsf{wat}_w(x,y) \mid \mathsf{mov}_@(y)\}]$

f. $[\![\mathrm{CP}_2]\!] = \lambda Q.\exists x[\mathsf{boy}_@(x) \wedge Q = \{\lambda w.\mathsf{wat}_w(x,y) \mid \mathsf{mov}_@(y)\}]$
$= \{\{\lambda w.\mathsf{wat}_w(x,y) \mid y \in \mathsf{mov}_@\} \mid x \in \mathsf{boy}_@\}$

The LF of the corresponding pair-list ∀-question is as in (42), read as 'the unique minimal set $\boldsymbol{K}$ s.t. for every boy x, $[\![\textit{Which movie did x watch?}]\!]$ is a member of $\boldsymbol{K}$'. The most important operations involved in the formation of this LF are 'quantifying into predication' and 'minimization' (à la Pafel 1999; Preuss 2001). First, the ∀-subject takes quantifier raising and quantifies into a predication operation, which is yielded by applying a predicative variable $\boldsymbol{K}$ to the open question *Which movie did x watch*. This operation yields a universal predication condition, read as 'For every boy x, $[\![\textit{Which movie did x watch?}]\!]$ is a member of $\boldsymbol{K}$'. Next, a (strong) minimization (MIN_S-)operator binds the $\boldsymbol{K}$ variable across the ∀-subject. It applies to the set of $\boldsymbol{K}$ sets that satisfy the universal predication condition and returns the unique minimal $\boldsymbol{K}$ set. This minimal $\boldsymbol{K}$ set is simply the set consisting of exactly all the sub-questions of the form ⌜*Which movie did boy-x watch?*⌝.

(42) Which movie did every boy watch? (Pair-list reading)

$[_{\text{CP}_2}$ MIN_S $\lambda \boldsymbol{K}_{\langle stt,t\rangle}$ [every-boy$_@$ λx_e [$\boldsymbol{K}$ $[_{\text{CP}_1}$ λp_{st} [wh-movie$_@$ λy_e [$\text{ID}(p)$ $[_{\text{IP}}$ x watched y]]]]]]]

a. $[\![\text{CP}_1]\!] = \{\lambda w.\mathsf{wat}_w(x,y) \mid \mathsf{mov}_@(y)\}$ (composition is the same as in (41a-d))

b. $[\![\text{CP}_2]\!] = \text{MIN}_S(\lambda \boldsymbol{K}.[\![\textit{every boy}_@]\!](\lambda x.\boldsymbol{K}(\{\lambda w.\mathsf{wat}_w(x,y) \mid \mathsf{mov}_@(y)\})))$
$= \text{MIN}_S(\lambda \boldsymbol{K}.\forall x[\mathsf{boy}_@(x) \rightarrow \boldsymbol{K}(\{\lambda w.\mathsf{wat}_w(x,y) \mid \mathsf{mov}_@(y)\})])$
$= \big\{\{\lambda w.\mathsf{wat}_w(x,y) \mid y \in \mathsf{mov}_@\} \mid x \in \mathsf{boy}_@\big\}$

(43) $\text{MIN}_S := \lambda \alpha_{\langle \sigma t,t\rangle} : \exists \boldsymbol{K}_{\langle \sigma,t\rangle}[\boldsymbol{K} \in \alpha \wedge \forall \boldsymbol{K}' \in \alpha[\boldsymbol{K} \subseteq \boldsymbol{K}']].\iota \boldsymbol{K}_{\langle \sigma,t\rangle}[\boldsymbol{K} \in \alpha \wedge \forall \boldsymbol{K}' \in \alpha[\boldsymbol{K} \subseteq \boldsymbol{K}']]$
(If α is a set of sets, $\text{MIN}_S(\alpha)$ is the unique minimal set in α which is a subset of every set in α, defined only if this minimal set exists.) (Pafel 1999)

As for the definition of answerhood, Fox (2012a,b) assumes that answering a family of sub-questions amounts to answering all of these sub-questions; in other words, answerhood is applied point-wise and exhaustively. When a point-wise answerhood-operator, defined recursively as in (44), applied to a family of sub-questions, it imposes ANS_{Dayal} onto each sub-question and returns the conjunction of propositions that are the strongest true answer to that sub-question, yielding domain exhaustivity. When the *wh*-object is singular, the presupposition that each of the sub-questions has a strongest true answer also gives rise to point-wise uniqueness.

(44) Point-wise answerhood-operator (Fox 2012a)

$$\text{ANS}_{\text{PW}} := \lambda w \lambda Q. \begin{cases} \text{ANS}_{Dayal}(w)(Q) & \text{if } Q \text{ is of type } \langle st,t\rangle \\ \bigcap\{\text{ANS}_{\text{PW}}(w)(\alpha) \mid \alpha \in Q\} & \text{otherwise} \end{cases}$$

Fox's account has two advantages over Dayal's. First, as discussed in (40), by defining a pair-list question as a family of sub-questions, this account can easily account for the Q-variability effects in embeddings. Second, the composition is quite neat; it does not use any *ad hoc* type-shifters or any complex operators. In the composition of the pair-list multiple-*wh* question, the *wh*-phrases function as $\exists$-indefinites that quantify into an identity condition. In the composition of the pair-list $\forall$-question, the $\forall$-subject standardly composes with a one-place predicate.

However, Fox's analysis has a few empirical problems similarly to Dayal's. First, since he analyzes pair-list $\forall$-questions and their multiple-*wh* counterparts as semantically equivalent, Fox as well cannot explain the contrast in domain exhaustivity.[16] Second, Fox's account does not extend to $\exists$-questions either. In the composition of a question with a quantifier, Fox uses the MIN_S-operator to obtain the unique minimal $\boldsymbol{K}$ set that satisfies a quantificational predication condition, which is unavailable if this predication condition is existential. For instance, for the $\exists$-question (45a), in a discourse with two relevant boys b_1 and b_2, $\boldsymbol{K}$ satisfies the existential predication condition (45b) as long as it is a superset of (45c) or (45d). Among these sets that $\boldsymbol{K}$ may refer to, there isn't one that is a subset of all the others.

(45) a. Which movie did one of the boys watch?
b. $\exists x[\mathsf{boy}_@(x) \wedge [\![\textit{Which movie did x watch?}]\!] \in \boldsymbol{K}]$
c. $\{[\![\textit{Which movie did } b_1 \textit{ watch?}]\!]\}$
d. $\{[\![\textit{Which movie did } b_2 \textit{ watch?}]\!]\}$

[16] One might propose to reconcile the family-of-questions approach by arguing that pair-list multiple-*wh* questions, but not pair-list $\forall$-questions, permit covert domain restriction. This possibility has been ruled out by the discussion in Sect. 2.1. First of all, the contrast between the two types of pair-list questions in domain exhaustivity remains even if the domain has been explicitly specified, as seen in (10) and (12). Moreover, as argued in footnote 2, if the quantification domain of a *wh*-phrase has been explicitly specified, it does not take further covert restrictions.

5. The formal theory

I assume a hybrid categorial approach to composing questions, developed in Xiang 2016, 2020. This approach integrates traditional categorial approaches with GB-style compositional semantics. Compared with frameworks that define questions as sets of propositions (e.g., Hamblin-Karttunen semantics), categorial approaches define questions as predicates/properties, which allows us to derive the Q-variability effects in embeddings of pair-list questions without defining pair-list questions as families of questions (Sect. 7). However, the core analysis presented in Sect. 6 with regard to the composition of the question nucleus is independent from the choice of framework.

This section lays out only the assumptions that are central to this paper, with some simplifications and modifications. For more details and applications of this framework, see Xiang 2020.

5.1. Defining questions and answers

Wh-questions admit both short answers and full answers. In discourse, short answers are parts of speech corresponding to the *wh*-term. Following categorial approaches, I define the root denotations of matrix and embedded questions uniformly as functions that map entities (or $\langle e, e\rangle$-type functions) denoted by possible short answers to propositions denoted by corresponding full answers. As illustrated in (46), *Which boy came?* denotes a function that maps each atomic boy x to the proposition that x came. After Chierchia and Caponigro (2013), I call such denotations 'topical properties'.[17]

(46) 'Which boy came?' 'John.'

a. $[\![\textit{Which boy came?}]\!] = \lambda x_e\colon \textsf{boy}_{@}(x).\lambda w[\textsf{came}_w(x)]$

b. $[\![\textit{Which boy came?}]\!]([\![\textit{John}]\!]) = \textsf{boy}_{@}(j).\lambda w[\textsf{came}_w(j)]$

Complete true answers to questions are obtained by the application of the answerhood-operators in (47). Compared with the $\textsc{Ans}_{Dayal}$-operator (34), the main difference is that the Hamblin set Q is replaced with a topical property $\boldsymbol{P}$, which can supply both short answers and propositional answers.[18] These answerhood-operators account for uniqueness effects in the same way as $\textsc{Ans}_{Dayal}$.

(47) Answerhood-operators (modified from Xiang 2020; to be revised in (62))

a. For the complete true short answer:

$$\textsc{Ans}^S(w)(\boldsymbol{P}) = \exists\alpha \in \mathrm{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\alpha) \wedge \forall\beta \in \mathrm{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\beta) \rightarrow \boldsymbol{P}(\alpha) \subseteq \boldsymbol{P}(\beta)]].$$
$$\iota\alpha \in \mathrm{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\alpha) \wedge \forall\beta \in \mathrm{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\beta) \rightarrow \boldsymbol{P}(\alpha) \subseteq \boldsymbol{P}(\beta)]]$$

b. For the complete true propositional answer:

$$\textsc{Ans}(w)(\boldsymbol{P}) = \boldsymbol{P}(\textsc{Ans}^S(w)(\boldsymbol{P}))$$

[17] In this paper, λ-terms with presuppositions are represented in the form of $\lambda v_\tau\colon \beta.\alpha$ (where τ is the semantic type of v, β stands for the additional definedness condition or presupposition, and α stands for the value description). λ-terms without a presupposition are written in the form of $\lambda v_\tau.\alpha$ or $\lambda v_\tau[\alpha]$, whichever is easier to read.

[18] Following Fox (2013), Xiang (2016, 2020) assumes a weaker definition for complete true answers: a true answer to a question is complete as long as it is not asymmetrically entailed by any other true answer to this question. This answerhood is assumed to account for mention-some readings of questions and free relatives. Since mention-some is not the focus of this paper, for easier comparisons with competing theories of complex questions, I follow Dayal (1996, 2017) here and define the complete true answer as the unique strongest true answer. For recent accounts on solving the dilemma between uniqueness and mention-some, see Fox 2018, 2020 and Xiang 2021. Also see Dotlacil and Roelofsen 2021 for an analysis using dynamic inquisitive semantics to account for both uniqueness effects and mention-some readings.

5.2. Composing simple *wh*-questions

I define *wh*-phrases as ∃-quantifiers ranging over a polymorphic set. In questions with an extensional reading, the quantification domain of a *wh*-phrase of the form $\ulcorner wh\text{-}\mathrm{A}_w \urcorner$ contains not only elements in the extension of the *wh*-complement $[\![\mathrm{A}]\!]^w$, but also functions from a set of entities to $[\![\mathrm{A}]\!]^w$, as defined in (48a).[19] The semantics of *wh*-phrases in questions with an intensional reading is defined analogously, as schematized in (48b).

(48) a. $[\![wh\text{-}\mathrm{A}_w]\!] = \lambda P.\exists\alpha \in \bigcup\left\{ [\![\mathrm{A}]\!]^w, \{\boldsymbol{f}_{\langle e,e\rangle} \mid \mathrm{Ran}(\boldsymbol{f}) = [\![\mathrm{A}]\!]^w\} \right\}[P(\alpha)]$

b. $[\![wh\text{-}\lambda w.\mathrm{A}_w]\!] = \lambda P.\exists\alpha \in \bigcup\left\{ \begin{array}{l} \{r_{\langle s,e\rangle} \mid \forall w[r_w \in [\![\mathrm{A}]\!]^w]\}, \\ \{\boldsymbol{f}_{\langle s,ee\rangle} \mid \forall w[\mathrm{Ran}(\boldsymbol{f}_w) = [\![\mathrm{A}]\!]^w]\} \end{array} \right\}[P(\alpha)]$

c. For any function $\boldsymbol{f}$ and any set A, $\mathrm{Ran}(\boldsymbol{f}) = A$ iff $\forall x \in \textsc{Dom}(\boldsymbol{f})[\boldsymbol{f}(x) \in A]$.

In the composition of a simplex *wh*-question, the fronted *wh*-phrase is converted into a function domain restrictor via the BeDom-operator (abbreviated as 'BD' in this paper).[20] As defined in (49), if π is an ∃-quantifier, $\textsc{Be}(\pi)$ is the set that π ranges over, and $\textsc{BeDom}(\pi)$ is a function domain restrictor which combines with a function θ and returns the function that is similar to θ but is undefined for items that are not in $\textsc{Be}(\pi)$.

(49) For any π of type $\langle \sigma t, t\rangle$ where σ is an arbitrary type, we have:

a. $\textsc{Be}(\pi) = \lambda x.\pi(\lambda y.y = x)$ (Partee 1986)

b. $\textsc{BeDom}(\pi) = \lambda\theta_\tau.\iota P_\tau \left[\begin{array}{l} [\mathrm{Dom}(P) = \mathrm{Dom}(\theta) \cap \textsc{Be}(\pi)] \\ \quad \wedge \forall\alpha \in \mathrm{Dom}(P)[P(\alpha) = \theta(\alpha)] \end{array} \right]$ (Xiang 2016, 2020)

As exemplified in (50), the fronted '$^{\mathrm{BD}}$(wh-boy$_@$)' applies to the simple 'came'-function defined for all entities and returns a more restrictive 'came'-function only defined for atomic boys.

(50) Which boy came?
$[_{\mathrm{CP}}\ {}^{\mathrm{BD}}(\text{wh-boy}_@)\ [_\gamma\ \lambda i\ [_{\mathrm{IP}}\ \lambda w.\ t_i\ \text{came}_w\ ]]]$

a. $[\![\gamma]\!] = \lambda x_e \lambda w.\mathsf{came}_w(x)$

b. $[\![\mathrm{CP}]\!] = \lambda x_e : \mathsf{boy}_@(x).\lambda w[\mathsf{came}_w(x)]$

The following illustrates the derivations for individual and functional readings of *wh*-questions with a quantifier. An individual reading arises if the *wh*-phrase binds an individual trace, as in (51a); a functional reading arises if the *wh*-phrase binds an (intensional) functional trace, as in (51b).

(51) Which movie did every boy watch?

a. Individual reading:
'Which movie y is s.t. every boy watched y?'
$[_{\mathrm{CP}}\ {}^{\mathrm{BD}}(\text{wh-movie}_@)\ [_\gamma\ \lambda i\ [_{\mathrm{IP}}\ \lambda w.\ \text{every-boy}_w\ \text{watched}_w\ t_i\ ]]]$

i. $[\![\gamma]\!] = \lambda y_e \lambda w.\forall x[\mathsf{boy}_w(x) \rightarrow \mathsf{wat}_w(x,y)]$

[19] Other than assuming a polymorphic restrictor, we can alternatively assume that *wh*-phrases are semantically ambiguous between ranging over $[\![\mathrm{A}]\!]^w$ or a set of functions from individuals to $[\![\mathrm{A}]\!]^w$. For example, Engdahl (1986) assumes a type-shifter applied to the *wh*-complement that has the effect of turning a set of entities into a set of $\langle e,e\rangle$-type functions.

[20] Crucially, $\textsc{BeDom}(\pi)$ is type-flexible — it can combine with any function of a $\langle \sigma, ...\rangle$ type where σ is the type of an element in $\textsc{Be}(\pi)$. Type-flexibility makes it possible to compose a question regardless of whether the *wh*-phrase binds an individual or functional variable, and regardless of how many *wh*-phrases there are in this question. This assumption overcomes difficulties with traditional categorial approaches in composing multiple-*wh* questions with single-pair readings.

ii. $[\![\text{CP}]\!] = \lambda y_e : \mathsf{mov}_{@}(y).\lambda w[\forall x[\mathsf{boy}_w(x) \rightarrow \mathsf{wat}_w(x,y)]]$

b. (Intensional) functional reading:
'Which Skolem function f to atomic movies is s.t. for every boy x, x watched $f(x)$?'
$[_{\text{CP}}\ {}^{\text{BD}}(\text{wh-}\lambda w.\text{movie}_w)\ [_\gamma\ \lambda i\ [_{\text{IP}}\ \lambda w.\ \text{every-boy}_w\ \lambda j\ [_{\text{VP}}\ t_j \text{ watched}_w\ t_i^j\ ]]]]$

i. $[\![\gamma]\!] = \lambda f_{\langle s,ee\rangle}\lambda w.\forall x[\mathsf{boy}_w(x) \rightarrow \mathsf{wat}_w(x, f_w(x))]$

ii. $[\![\text{CP}]\!] = \lambda f_{\langle s,ee\rangle} : \forall w'[\text{Ran}(f_{w'}) = \mathsf{mov}_{w'}].\lambda w[\forall x[\mathsf{boy}_w(x) \rightarrow \mathsf{wat}_w(x, f_w(x))]]$

6. Proposal

In line with functionality approaches, I analyze pair-list readings of multiple-*wh* questions and QiQ-readings of questions with a quantifier as extensional functional readings. For both types of questions, I assume that the composition involves a quantificational binding-in operation applied to what I refer to as a '*dependency sentence*'. A dependency sentence is an open sentence with the logical form $\ulcorner x\ \text{P}\ f(x)\urcorner$, which expresses a functional dependency between the two arguments of the two-place predicate P. In particular, in the composition of a pair-list multiple-*wh* question, the *wh*-subject existentially quantifies into an identity operation (à la Karttunen semantics). In contrast, in the composition of a QiQ-question, the quantificational subject quantifies into a predication operation (à la Fox 2012b). The LF schema is as follows, repeated from (7):

(52) Composition schema for complex questions:

a. Which boy watched which movie? (Pair-list reading)
... [which-movie$_j$ ... which-boy$_i$ [... IDENT ... [t_i watched t_j^i]]]

b. Which movie did DET-boy(s) watch? (QiQ-reading)
... [which-movie$_j$ ... DET-boy(s)$_i$ [... PRED ... [t_i watched t_j^i]]]

The distinctions between IDENT and PRED give rise to a contrast in domain exhaustivity between multiple-*wh* questions and $\forall$-questions. The LF schema for QiQ-questions automatically explains why $\forall$-questions and $\exists$-questions have pair-list readings and choice readings, respectively, and why NO-questions do not have QiQ-readings. What's more, given the known contrasts among non-interrogative quantifiers in lexical distributivity and scoping, this analysis also explains why counting quantifiers do not participate in QiQ-readings.

In what follows, I will first provide the root denotation of each type of complex question (Sect. 6.1) and revisit the definition of answerhood (Sect. 6.2). Next, I will show how to derive each of these root denotations compositionally (Sects. 6.3 and 6.4). Finally, Sect. 6.5 will summarize the proposal.

6.1. Question denotations

I assume that pair-list readings and QiQ-readings of complex *wh*-questions are extensional functional readings. When a question has a pair-list/QiQ reading, it denotes a topical property that maps an $\langle e,e\rangle$-type function f to the conjunction of a proposition set that describes the graph of f. Illustrations of those topical properties are given in (53) and (54) in tandem. The (a)-denotations are represented in a way that is isomorphic to the structures of composition (for details of composition, see Sects. 6.3 and 6.4). The (b)-denotations are semantically equivalent to their (a)-counterparts, but are represented in a way that is more convenient for comparison.

(53) $[\![$*Which boy watched which movie?*$]\!]_{\text{pair-list}}$
$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle} : \mathrm{Ran}(\boldsymbol{f}) = \textsf{mov}_{@}.\bar{\bigcap}\{p \mid \exists\text{-}\textsf{boy}_{@}(\lambda x.p = \lambda w.\textsf{wat}_w(x, \boldsymbol{f}(x)))\}$ (a)
$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle} : \mathrm{Ran}(\boldsymbol{f}) = \textsf{mov}_{@}.\bar{\bigcap}\{\lambda w.\textsf{wat}_w(x, \boldsymbol{f}(x)) \mid \textsf{boy}_{@}(x)\}$ (b)

(54) $[\![$*Which movie did* Det-*boy(s) watch?*$]\!]_{\text{QiQ}}$
$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle} : \mathrm{Ran}(\boldsymbol{f}) = \textsf{mov}_{@}.\bar{\bigcap} f^{\text{MIN}}_{\text{CH}}(\{\boldsymbol{K} \mid \text{Det-}\textsf{boy}_{@}(\lambda x.\boldsymbol{K}(\lambda w.\textsf{wat}_w(x, \boldsymbol{f}(x))))\})$ (a)
$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle} : \mathrm{Ran}(\boldsymbol{f}) = \textsf{mov}_{@} \wedge \text{Det-}\textsf{boy}_{@}(\mathrm{Dom}(\boldsymbol{f}))$.
$\bar{\bigcap} f^{\text{MIN}}_{\text{CH}}(\{\boldsymbol{K} \mid \text{Det-}\textsf{boy}_{@}(\lambda x.\boldsymbol{K}(\lambda w.\textsf{wat}_w(x, \boldsymbol{f}(x))))\})$ (b)

The denotations above introduce two new operators, namely, $\bar{\bigcap}$ and $f^{\text{MIN}}_{\text{CH}}$. The $\bar{\bigcap}$-operator is like the regular intersection operator except that it ignores the undefined elements.[21]

(55) $\bar{\bigcap} := \lambda A_{\langle \tau t,t\rangle}.\bigcap\{a \mid a \in A \wedge a \text{ is defined}\}$

The $f^{\text{MIN}}_{\text{CH}}$-operator combines a weak minimization-operator MIN_W with a choice-function variable f_{CH}, which gets existentially bound at a global site.[22] The MIN_W-operator is weaker than Pafel-Fox's MIN_S-operator: for any set α, a member x of α is a minimal member of α as long as no member of α is a proper subset/subpart of x — without requiring that x be a subset/subpart of every member of α.[23] The choice between MIN_S and $f^{\text{MIN}}_{\text{CH}}$ makes no difference in $\forall$-questions, but only the latter works for $\exists$-questions (Sect. 6.4.2).

(56) $f^{\text{MIN}}_{\text{CH}} := \lambda\alpha_{\langle\sigma,t\rangle}.f_{\text{CH}}(\text{MIN}_W(\alpha))$

(57) a. $\text{MIN}_W := \lambda\alpha_{\langle\sigma,t\rangle}.\{x_\sigma \mid x \in \alpha \wedge \neg\exists y \in \alpha[y < x]\}$
b. $\text{MIN}_S := \lambda\alpha_{\langle\sigma,t\rangle}.\iota x_\sigma[x \in \alpha \wedge \forall y \in \alpha[y \geq x]]$ (generalized from (43))
['$<$' stands for the proper subset relation if α is a set of sets, and for the proper subpart relation if α is a set of non-sets; '$\geq$' is analogous.]

Notice a contrast between (53b) and (54b): (53b) only has a restriction on the range of the input functions, while (54b) also has a restriction on the domain of these functions. More concretely, in (53) the topical property of the multiple-*wh* question takes any function that maps entities to atomic movies as its input and the graph description of this function as its output. In contrast, in (54) the topical property of the QiQ-question is more restrictive — it is only defined for functions that map Det-boy(s) to atomic movies. The additional domain restriction in (54b) (i.e., $\text{Det-}\textsf{boy}_{@}(\mathrm{Dom}(\boldsymbol{f}))$) arises as a definedness condition of the value description in (54a): the quantificational predication

[21] Specifically, if a set A contains an undefined member, applying the $\bar{\bigcap}$-operator to A does not pass up the undefinedness of this member to the returned intersection, as seen in (ib), which contrasts in this respect with (ia). However, as shown in (ic), if a definedness condition is applied to the set A, not to a member of A, $\bar{\bigcap}A$ inherits this condition.

(i) Assume that $\boldsymbol{f}$ is defined for two atomic boys a and b, but not for boy c. Then:
a. $\bigcap\{\lambda w.\textsf{wat}_w(x, \boldsymbol{f}(x)) \mid \textsf{boy}_{@}(x)\}$ is undefined;
b. $\bar{\bigcap}\{\lambda w.\textsf{wat}_w(x, \boldsymbol{f}(x)) \mid \textsf{boy}_{@}(x)\} = \lambda w.\textsf{wat}_w(a, \boldsymbol{f}(a)) \wedge \textsf{wat}_w(b, \boldsymbol{f}(b))$;
c. $\bar{\bigcap}[\forall x \in \{a,b,c\}[x \in \mathrm{Dom}(\boldsymbol{f})].\{\lambda w.\textsf{wat}_w(x, \boldsymbol{f}(x)) \mid \textsf{boy}_{@}(x)\}]$ is undefined.

[22] For readers who are familiar with Boolean semantics, the $f^{\text{MIN}}_{\text{CH}}$-operator is roughly the same as the collectivity raising operator in Winter 2001.

[23] The following illustrates the contrast between the $f^{\text{MIN}}_{\text{CH}}$-operator and the MIN_S-operator:

(i) Let a and b be two distinct entities, $A = \{\varnothing, \{a\}, \{b\}\}$, and $B = \{\{a\}, \{b\}\}$. Then:
a. $\text{MIN}_S(A) = f^{\text{MIN}}_{\text{CH}}(A) = \varnothing$;
b. $\text{MIN}_S(B)$ is undefined; while $f^{\text{MIN}}_{\text{CH}}(B)$ has two possible values: $\{a\}$ and $\{b\}$.

condition $\text{Det-}\mathsf{boy}_@(\lambda x.\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x))))$, read as 'For Det-boy(s) x, the proposition ⌜*x watched* $\boldsymbol{f}(x)$⌝ is a member of $\boldsymbol{K}$', is defined only if the function $\boldsymbol{f}$ is defined for Det-boy(s).

For a concrete illustration of (54), consider the QiQ-denotation of a ∀-question:

(58) ⟦*Which movie did every/each boy watch?*⟧

$$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle}: \underbrace{\mathrm{Ran}(\boldsymbol{f}) = \mathsf{mov}_@}_{\text{from } wh\text{-obj}} . \underbrace{\bar{\bigcap} f^{\text{MIN}}_{\text{CH}}(\{\boldsymbol{K} \mid \forall\text{-}\mathsf{boy}_@(\lambda x.\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x))))\})}_{\text{(i) from question nucleus}} \quad \text{(a)}$$

$$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle}: \underbrace{\mathrm{Ran}(\boldsymbol{f}) = \mathsf{mov}_@}_{\text{from } wh\text{-obj}} . \underbrace{\bar{\bigcap}\,[\forall\text{-}\mathsf{boy}_@(\mathrm{Dom}(\boldsymbol{f})).\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x)) \mid \mathsf{boy}_@(x)\}]}_{=\text{(i)}} \quad \text{(a')}$$

$$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle}: \underbrace{\mathrm{Ran}(\boldsymbol{f}) = \mathsf{mov}_@}_{\text{from } wh\text{-obj}} \wedge \underbrace{\forall\text{-}\mathsf{boy}_@(\mathrm{Dom}(\boldsymbol{f}))}_{\text{definedness cond of (i)}} . \bar{\bigcap}\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x)) \mid \mathsf{boy}_@(x)\} \quad \text{(b)}$$

In (58a), the input $\boldsymbol{f}$ can be any $\langle e, e\rangle$-type function from entities to atomic movies. For each such $\boldsymbol{f}$, the output is the conjunction of the set that consists of exactly all the propositions of the form ⌜*boy-x watched* $\boldsymbol{f}(x)$⌝. Crucially, as represented in (58a'), the output inference carries a definedness condition: the universal predication condition w.r.t. the open sentence 'boy-x watched $\boldsymbol{f}(x)$' is defined only if $\boldsymbol{f}$ is defined for every boy. This condition projects to the entire topical property as in (58b), yielding domain exhaustivity. In short, the ∀-question (58) is semantically equivalent to its multiple-*wh* counterpart (53), except that (58) presupposes domain exhaustivity.

At this point, it is clear why I pursue a functionality approach instead of a family-of-questions approach: since I assume a *wh*-dependency, I can attribute the domain exhaustivity effect to a definedness condition arising from an operation applied within the question nucleus (viz., the universal quantificational binding-in operation w.r.t. a dependency sentence). In this approach, the contrast in domain exhaustivity between multiple-*wh* questions and ∀-questions can be explained in terms of the differences of their nuclei. In family-of-questions approaches, domain exhaustivity is instead attributed to an operation applied outside the question nucleus (e.g., the point-wise answerhood-operator of Fox 2012a,b); such accounts cannot capture the semantic contrast between ∀-questions and multiple-*wh* questions as a direct result of their structural differences.

6.2. Redefining answerhood

As pointed out by Floris Roelofsen (pers. comm.), the answerhood-operator assumed in (47a) overgenerates possible answers for pair-list questions. For example, the topical property of *Which boy watched which movie* is defined for any $\langle e, e\rangle$-type functions that map entities to atomic movies, not just those consisting of only boy-movie pairs. The assumed answerhood incorrectly predicts that it accepts the answer (59b), which involves an adult-movie pair.

(59) Which boy watched which movie?

a. Andy watched *Hulk*, Billy watched *Spiderman*.

b. # Andy watched *Hulk*, Billy watched *Spiderman*, Mr. White watched *Ironman*.

To solve this problem, I define answerhood in (60a) for possible short answers. The added constraint, namely, that every subset of α yields a propositional answer that is possibly true, rules out functions that allow inputs that are non-boys.

(60) Answerhood for possible answers

a. For short answers:

$$\mathbb{A}^S(\boldsymbol{P}) = \begin{cases} \mathrm{Dom}(\boldsymbol{P}) & \text{if } \boldsymbol{P} \in D_{\langle e,\tau\rangle} \\ \{\alpha \mid \alpha \in \mathrm{Dom}(\boldsymbol{P}) \wedge \forall\beta \subseteq \alpha[\exists w \in \boldsymbol{P}(\beta)]\} & \text{if } \boldsymbol{P} \in D_{\langle\langle e,e\rangle,\tau\rangle} \end{cases}$$

b. For propositional answers:

$\mathbb{A}(\boldsymbol{P}) = \{\boldsymbol{P}(\alpha) \mid \alpha \in \mathbb{A}^S(\boldsymbol{P})\}$

Next, let's consider answerhood for complete true answers. For pair-list questions like (61) with a number-neutral *wh*-subject and a semi-distributive predicate, the same pair-list propositional answer can be derived based on distinct possible short answers, such as those listed in (61a-c).

(61) Which boy or boys watched which movie?
(Context: boys b_1,b_2 both watched the movie m_1, and boy b_3 watched movie m_2.)

a. $[b_1 \to m_1, b_2 \to m_1, b_3 \to m_2]$, $[b_1 \oplus b_2 \to m_1, b_3 \to m_2]$

b. $[b_1 \to m_1, b_1 \oplus b_2 \to m_1, b_3 \to m_2]$, $[b_2 \to m_1, b_1 \oplus b_2 \to m_1, b_3 \to m_2]$

c. $[b_1 \to m_1, b_2 \to m_1, b_1 \oplus b_2 \to m_1, b_3 \to m_2]$

Given this multiple-to-one mapping from short answers to propositional answers, I redefine the answerhood for complete true short answers as follows: the complete true short answer is the maximal short answer that yields the strongest true propositional answer. Formally:

(62) Answerhood for complete true answers (final)

a. For short answers:

$\textsc{Ans}^S(w)(\boldsymbol{P}) = \exists\alpha \in \mathbb{X}(w)(\boldsymbol{P}).\textsc{Max}[\mathbb{X}(w)(\boldsymbol{P})]$, where

$\mathbb{X}(w)(\boldsymbol{P}) = \{\alpha \mid \alpha \in \mathbb{A}(\boldsymbol{P}) \wedge w \in \boldsymbol{P}(\alpha) \wedge \forall\beta \in \mathbb{A}(\boldsymbol{P})[w \in \boldsymbol{P}(\beta) \to \boldsymbol{P}(\alpha) \subseteq \boldsymbol{P}(\beta)]\}$

b. For propositional answers:

$\textsc{Ans}(w)(\boldsymbol{P}) = \boldsymbol{P}(\textsc{Ans}^S(w)(\boldsymbol{P}))$

By this definition, the complete true short answer to (61) is (61c).[24]

6.3. Composing pair-list multiple-*wh* questions

Figure 1 illustrates the composition of a pair-list multiple-*wh* question. As marked in the tree diagram, this composition proceeds in four steps.

(i) *Derive a functional dependency.* The argument index carried by the complex functional trace of the *wh*-object is co-indexed with the subject trace, yielding a dependency sentence which expresses a dependency between the two arguments of *watched*.

(ii) *Quantificational binding into an identity condition.* Employing an identity (Id-)operator yields an identity relation between a covert propositional variable p and the dependency sentence generated at IP. At node ①, the *wh*-subject, interpreted as an $\exists$-quantifier, binds the argument index inside the IP across the Id-operator, yielding an existential identity condition w.r.t. the dependency sentence.

(iii) *Derive a function graph description.* Abstracting p returns the set of propositions of the form $\ulcorner$*boy-x watched* $\boldsymbol{f}_i(x)\urcorner$. Conjoining this set yields the graph description of the function $\boldsymbol{f}_i$. Here the $\bar{\bigcap}$-closure can be perceived as a 'function graph creator' in the sense of Dayal 2017. Note

[24] For the purpose of defining answerhood, it doesn't make a big difference whether we assume that the complete true short answer is the maximal element in $\mathbb{X}(w)(\boldsymbol{P})$ as in (61c), a minimal element like those in (61a), or any other element.

that, as defined in (55), the $\bar{\bigcap}$-operator ignores the undefined members in the set it applies to; therefore, the resulting conjunction is defined even if $\boldsymbol{f}_i$ is undefined for some of the boys. In other words, the application of the $\bar{\bigcap}$-operator does not force domain exhaustivity.

(iv) *Create a topical property.* Abstracting the functional index yields a property (of type $\langle ee, st\rangle$) that maps each $\langle e, e\rangle$-type function to the graph description of this function. Further, the fronted *wh*-object '$^{\text{BD}}$(wh-movie$_@$)' restricts the domain of this property — it requests the range of each input function to be a set of atomic movies.

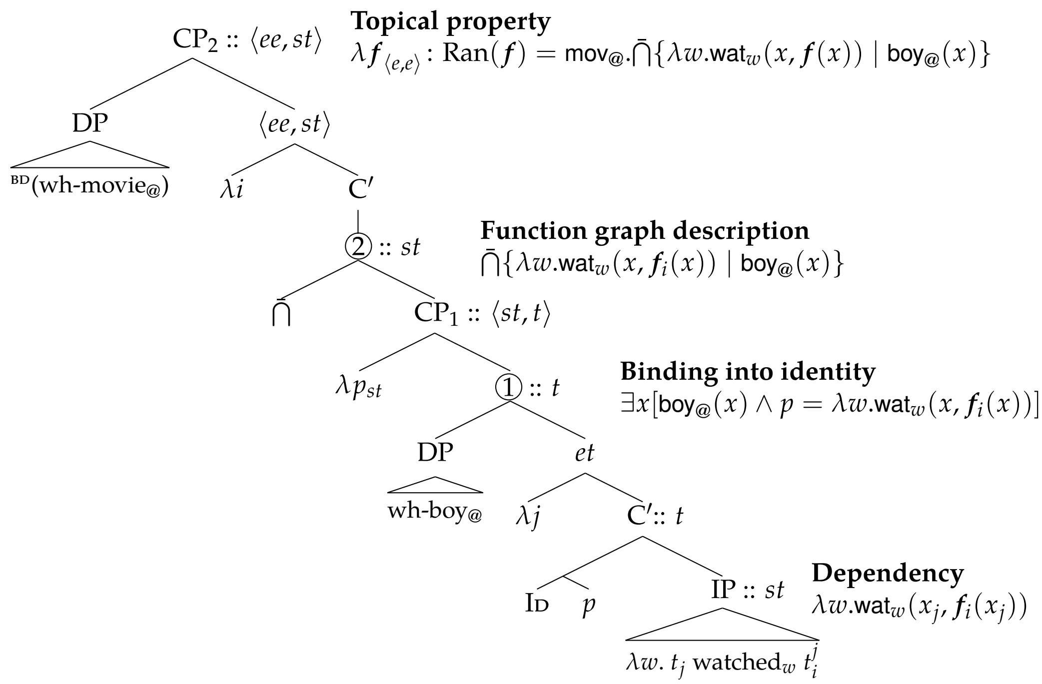


Figure 1: Composition of the pair-list multiple-*wh* question *Which boy watched which movie?*

(63) Steps (i) & (ii): Bind into an identity condition w.r.t. a dependency sentence

a. $[\![\text{IP}]\!] = \lambda w.\mathsf{wat}_w(x_j, \boldsymbol{f}_i(x_j))$

b. $[\![\text{I}\textsc{d}]\!] = \lambda \alpha_\tau \lambda \beta_\tau.\alpha = \beta$

c. $[\![\text{C}']\!] = [\![\text{I}\textsc{d}]\!](p)([\![\text{IP}]\!])$
$= [p = \lambda w.\mathsf{wat}_w(x_j, \boldsymbol{f}_i(x_j))]$

d. $[\![\textit{wh-boy}_@]\!] = \lambda P_{\langle e,t\rangle}.\exists x[\mathsf{boy}_@(x) \wedge P(x)]$

e. $[\![①]\!] = [\![\textit{wh-boy}_@]\!]([\![\text{C}']\!])$
$= \exists x[\mathsf{boy}_@(x) \wedge p = \lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))]$

(64) Step (iii): Create a function graph description

a. $[\![\text{CP}_1]\!] = \lambda p.\exists x[\mathsf{boy}_@(x) \wedge p = \lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))]$
$= \{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)) \mid \mathsf{boy}_@(x)\}$

b. $[\![②]\!] = \bar{\bigcap}\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)) \mid \mathsf{boy}_@(x)\}$

(65) Step (iv): Create a topical property
$[\![\text{CP}_2]\!] = \lambda \boldsymbol{f}_{\langle e,e\rangle} : \text{Ran}(\boldsymbol{f}) = \mathsf{mov}_@.\bar{\bigcap}\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x)) \mid \mathsf{boy}_@(x)\}$

Eagle-eyed readers might notice that here the *wh*-object is moved over the fronted *wh*-subject, which violates the generalization of 'tucking-in' (Richards 1997). Although violations of tucking-in are sometimes permitted for D-linked *wh*-phrases, it is certainly problematic to say that pair-list readings are only available in constructions that violate tucking-in. However, this problem does not stem from the specific assumptions involved in the composing of pair-list multiple-*wh* questions; it is a consequence of requiring covert/overt *wh*-fronting in question composition generally. This problem can be avoided if we assume a framework of composition that allows *wh*-in-situ. For example, in variable-free semantics (Jacobson 2014), abstractions can be passed up by type-shifting operations. Integrating my core proposal on composing pair-list questions into such frameworks allows us to create the wanted topical property without fronting the object *wh*-phrase.[25]

6.4. Composing QiQ-questions

QiQ-questions of the form *Which movie did* Det-*boy(s) watch?* are composed uniformly with the LF schema in Figure 2. The composition steps are parallel to those for the pair-list multiple-*wh* question *Which boy watched which movie?*. The following subsections will explain how this composition schema works for each type of QiQ-questions.

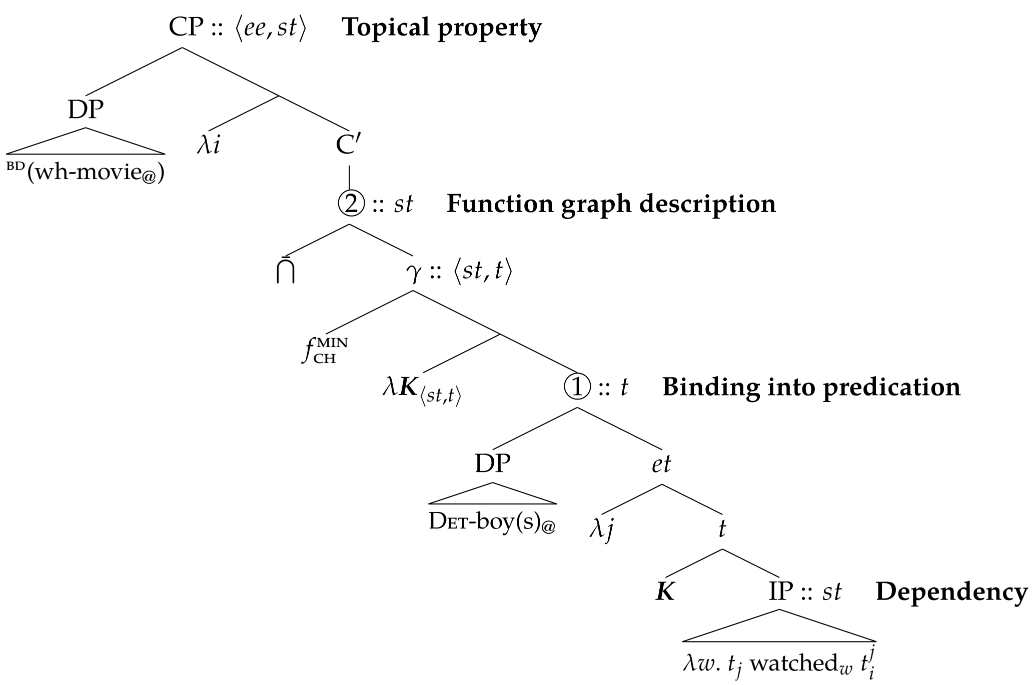


Figure 2: Composition of the QiQ-question *Which movie did* Det-*boy(s) watch?*

In denotation (66b), the condition on the range of f (i.e., that f maps to atomic movies) comes from the fronted *wh*-object. All the other conditions, including the condition on the domain of f (i.e., that f is defined for Det-boy(s)) and the output proposition which describes the graph of the input function, come from the question nucleus (viz., node ②).

[25] For an attempt to compose complex questions with a variable-free grammar, see Xiang 2019b.

(66) $[\![\textit{Which movie did } \textsc{Det}\textit{-boy(s) watch?}]\!]_{\text{QiQ}}$ (repeated from (54))

$$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle}: \underbrace{\text{Ran}(\boldsymbol{f}) = \mathsf{mov}_@}_{\text{from } wh\text{-object}} . \underbrace{\bar{\bigcap} f^{\text{MIN}}_{\text{CH}}(\{\boldsymbol{K} \mid \textsc{Det}\text{-}\mathsf{boy}_@(\lambda x.\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x))))\})}_{\text{(i) from nucleus}} \quad \text{(a)}$$

$$\Leftrightarrow \lambda \boldsymbol{f}_{\langle e,e\rangle}: \text{Ran}(\boldsymbol{f}) = \mathsf{mov}_@ \wedge \overbrace{\textsc{Det}\text{-}\mathsf{boy}_@(\text{Dom}(\boldsymbol{f}))}^{\text{definedness cond of (i)}} . \bar{\bigcap} f^{\text{MIN}}_{\text{CH}}(\{\boldsymbol{K} \mid \textsc{Det}\text{-}\mathsf{boy}_@(\lambda x.\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x))))\}) \quad \text{(b)}$$

Recall that *wh*-questions with a quantificational subject admit both functional readings and QIQ-readings. At this point, let us compare the derivations of these two readings. In both readings, the question involves a *wh*-dependency, derived by letting the subject bind into the complex functional trace of the *wh*-object. However, the composition of a pair-list reading makes use of two additional operations, i.e., quantifying into predication and minimization, which are not involved in the functional readings. These operations are similar to what Fox (2012b) assumes for composing ∀-questions (see (42)), but they depart from Fox's implementation in two respects, yielding desirable consequences in explaining the contrast in domain exhaustivity between ∀-questions and multiple-*wh* questions and in deriving the choice readings of ∃-questions. First, in the presented analysis, the predication operation is applied to a dependency sentence (not to a question). Quantificational binding into the dependency sentence is crucial for the derivation of domain exhaustivity in ∀-questions (Sect. 6.4.1). Second, the $f^{\text{MIN}}_{\text{CH}}$-operator is weaker than the $\textsc{min}_S$-operator that Fox adopts from Pafel (1999): $f^{\text{MIN}}_{\text{CH}}$ doesn't require the existence of a unique minimal member (Sect. 6.1). Replacing $\textsc{min}_S$ with $f^{\text{MIN}}_{\text{CH}}$ makes it feasible for the analysis to tackle choice ∃-questions (Sect. 6.4.2).

6.4.1. Composing ∀-questions

This section presents the composition of pair-list ∀-questions. Its primary goals are to derive the pair-list readings and to account for the domain exhaustivity effects. The LF is given in Figure 3 below. In parallel to the composition of pair-list multiple-*wh* questions (Sect. 6.3), I divide the composition into four steps:

(i) *Derive a functional dependency.* The IP is a dependency sentence, composed in the same way as the IP in the pair-list multiple-*wh* question.

(ii) *Bind into a predication condition.* A null predication-operator $\boldsymbol{K}$ is applied to IP, yielding a predication condition to the effect that the meaning of the dependency sentence generated at IP is a member of $\boldsymbol{K}$. Next, the subject *every/each-boy* quantifies into this predication condition and binds the argument index j, yielding a universal predication condition, as stated in (67b).

(iii) *Create a function graph description.* Abstracting the predicative variable $\boldsymbol{K}$ returns the set of $\boldsymbol{K}$ sets that satisfy the universal predication condition, as in (68a). These are the sets that contain all the propositions of the form ⌜*boy-x watched* $\boldsymbol{f}_i(x)$⌝. Next, applying the minimizer $f^{\text{MIN}}_{\text{CH}}$ returns a minimal $\boldsymbol{K}$ set that satisfies the universal quantification predication condition, as in (68b). Here there is only one such minimal $\boldsymbol{K}$ set, i.e., $\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)) \mid \mathsf{boy}_@(x)\}$. Finally, applying the intersection-operator $\bar{\bigcap}$ to this set returns the graph description of $\boldsymbol{f}_i$, as in (68c).

(iv) *Create a topical property.* The fronted '$^{\text{BD}}$(wh-movie$_@$)' binds the functional index j and restricts the range of any input $\boldsymbol{f}$ to the set of atomic movies. The possible inputs of this topical property are therefore functions that map each boy to an atomic movie, and the outputs are conjunctive propositions that describe the graph of each input function.

Step (ii) of this composition — quantificational binding into predication — is the heart of the analysis. First, it carries forward the advantage of Fox's analysis that the quantificational subject can standardly combine with a one-place predicate of type $\langle e, t\rangle$. In contrast to earlier accounts (e.g., Groenendijk and Stokhof 1984; Chierchia 1993; Dayal 1996, 2017), there is no need to assume a type-shifting operation or make use of a minimal witness set. What's more, since here the $\forall$-subject binds into the functional *wh*-trace, it yields a definedness condition stating that the function $\boldsymbol{f}_i$ is defined for every boy. This definedness condition projects to CP, yielding domain exhaustivity.[26]

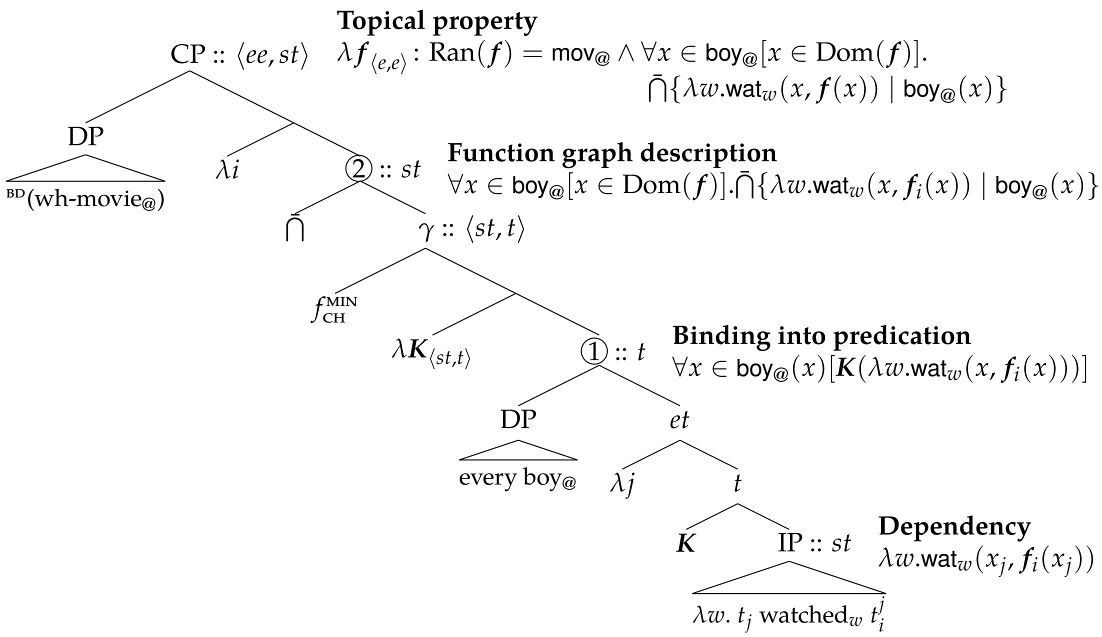


Figure 3: Composition of the $\forall$-question *Which movie did every boy watch?*

(67) Steps (i) & (ii): Bind into the predication condition w.r.t. a dependency sentence

a. $[\![\mathrm{IP}]\!] = \lambda w.\mathsf{wat}_w(x_j, \boldsymbol{f}_i(x_j))$ (equivalent to (63))

b. $[\![①]\!] = [\![\textit{every boy}_{@}]\!](\lambda x.\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))))$
$= \forall x \in \mathsf{boy}_{@}[\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)))]$ (defined only if $\forall x \in \mathsf{boy}_{@}[x \in \mathrm{Dom}(\boldsymbol{f}_i)]$)
(For every boy x, the proposition 'x *watched* $\boldsymbol{f}_i(x)$' is a member of $\boldsymbol{K}$.)

(68) Step (iii): Create a function graph description

a. $[\![\lambda \boldsymbol{K}.①]\!] = \lambda \boldsymbol{K}.\forall x \in \mathsf{boy}_{@}[\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)) \in \boldsymbol{K}]$
$= \lambda \boldsymbol{K}: \forall x \in \mathsf{boy}_{@}[x \in \mathrm{Dom}(\boldsymbol{f}_i)].\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)) \mid \mathsf{boy}_{@}(x)\} \subseteq \boldsymbol{K}$

b. $[\![\gamma]\!] = f^{\mathrm{MIN}}_{\mathrm{CH}}([\![\lambda \boldsymbol{K}.①]\!])$
$= \forall x \in \mathsf{boy}_{@}[x \in \mathrm{Dom}(\boldsymbol{f}_i)].\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)) \mid \mathsf{boy}_{@}(x)\}$

c. $[\![②]\!] = \forall x \in \mathsf{boy}_{@}[x \in \mathrm{Dom}(\boldsymbol{f}_i)].\bar{\bigcap}\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)) \mid \mathsf{boy}_{@}(x)\}$

(69) Step (iv): Create a topical property
$[\![\mathrm{CP}]\!] = \lambda \boldsymbol{f}_{\langle e,e\rangle}: \mathrm{Ran}(\boldsymbol{f}) = \mathsf{mov}_{@} \wedge \forall x \in \mathsf{boy}_{@}[x \in \mathrm{Dom}(\boldsymbol{f})].\bar{\bigcap}\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x)) \mid \mathsf{boy}_{@}(x)\}$

[26]Note that this definedness condition is not affected by the application of the $\bar{\bigcap}$-operator: as specified in the denotation of the γ-node, this definedness condition is applied to the minimal $\boldsymbol{K}$ set as a whole, not to the members of $\boldsymbol{K}$. See footnote 21 for relevant illustrations.

The explanation of domain exhaustivity crucially relies on the presence of a ∀-quantifier: domain exhaustivity comes from the binding relation between a ∀-quantifier and the argument index of the functional *wh*-trace. As a welcome effect, this analysis does not over-predict domain exhaustivity for a pair-list multiple-*wh* question: in a multiple-*wh* question, the argument variable of the functional trace of the *wh*-object is 'existentially' bound by the *wh*-subject. For comparison, the family-of-questions approach of Fox (2012a,b) attributes domain exhaustivity to an operation outside the question nucleus, namely, the point-wise answerhood-operator. Since the selection of answerhood is independent from the root structure/meaning of a question, the family-of-questions approach cannot explain the contrast in domain exhaustivity between ∀-questions and multiple-*wh* questions.

In the remaining subsections, I will describe the characteristics of the QIQ-reading of each type of questions in terms of the following three parameters:

[±D-EXH]:	whether the reading is subject to domain exhaustivity;
[±PL]:	whether the reading is a pair-list reading;
[±CH]:	whether the reading has a 'choice' flavor.

The QIQ-reading of a ∀-question is [+D-EXH,+PL,−CH]. It presupposes domain exhaustivity because the universal predication condition (from node ①) is defined only if the input function $\boldsymbol{f}$ is defined for 'every' member of the set that the subject quantifies over. It expects a pair-list answer because the yielded eligible minimal proposition set $\boldsymbol{K}$ (from node γ) that satisfies the universal predication condition is a 'non-singleton' set ranging over multiple elements in the quantification domain of the subject. It does not have a choice flavor because there is 'only one' such eligible minimal $\boldsymbol{K}$ set.

6.4.2. Composing ∃-questions

Choice readings of ∃-questions are derived in the same way as pair-list readings of ∀-questions. At node ①, the ∃-subject binds into the complex functional trace of the *wh*-object across the null predicate $\boldsymbol{K}$, yielding an existential predication condition w.r.t. a dependency sentence. At node γ, applying the $f_{\text{CH}}^{\text{MIN}}$-operator returns one of the minimal $\boldsymbol{K}$ sets that satisfy the existential predication condition. Crucially, unlike the case of the ∀-question, here there are 'multiple' minimal $\boldsymbol{K}$ sets that satisfy the quantificational predication condition, each of which is a singleton set consisting of a proposition of the form ⌜*boy-x watched* $\boldsymbol{f}_i(x)$⌝. Each such minimal $\boldsymbol{K}$ set supplies a possible topical property, which therefore gives rise to a choice flavor.

(70) Which movie did one of the boys watch?

$[_{\text{CP}}{}^{\text{BD}}(\text{wh-movie}_@)\ \lambda i\ [_{②}\ \bar{\bigcap}\ [_\gamma\ f_{\text{CH}}^{\text{MIN}}\ \lambda \boldsymbol{K}\ [_{①}\ \text{one-of-the-boys}_@\ \lambda j\ [\ \boldsymbol{K}\ (\lambda w.x_j\text{-wat}_w\text{-}\boldsymbol{f}_i(x_j))\ ]]]]]$

a. $[\![①]\!] = \exists x \in \mathsf{boy}_@[\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)))]$

b. $[\![\gamma]\!] = f_{\text{CH}}^{\text{MIN}}([\![\lambda \boldsymbol{K}.①]\!])$
$= f_{\text{CH}}(\{\{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))\} \mid x \in \mathsf{boy}_@\})$
$= \{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))\}$, where x is the chosen boy

c. $[\![②]\!] = \lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))$, where x is the chosen boy

d. $[\![\text{CP}]\!] = \lambda \boldsymbol{f}_{\langle e,e\rangle} : \text{Ran}(\boldsymbol{f}) = \mathsf{mov}_@\ .\lambda w[\mathsf{wat}_w(x, \boldsymbol{f}(x))]$, where x is the chosen boy

Note that this approach does not assume a choice-function analysis of indefinites. In (70), the $f_{\text{CH}}^{\text{MIN}}$-operator, which contains a choice-function variable f_{CH}, applies to a family of singleton sets of propositions, not to a set of boys. The subject *one of the boys* is treated standardly as an existential generalized quantifier. Therefore, more precisely, node γ should be read as 'the chosen singleton set of propositions of the form $\ulcorner\{$*boy-x watched* $\boldsymbol{f}_i(x)\}\urcorner$'. I further assume that the choice-function variable f_{CH} is existentially bound at a global site. The full paraphrase of the LF (70) is as follows:

(71) 'For some choice function f_{CH}, what is the $\langle e,e\rangle$-function $\boldsymbol{f}$ to atomic movies s.t. the conjunction of the singleton set $\{$*boy-x watched* $\boldsymbol{f}(x)\}$ chosen by f_{CH} is true?'

The QIQ-reading of an $\exists$-question yielded with the above analysis is [$-$D-EXH,$-$PL,$+$CH]. This reading is not subject to domain exhaustivity because the existential predication condition (70a) only requires the input function $\boldsymbol{f}$ to be defined for 'at least one' of the boys.[27] Possible answers to this question are single-pairs, not pair-lists, because the minimal $\boldsymbol{K}$ sets satisfying the existential predication condition are all 'singleton' sets, as seen in (70b). The yielded QIQ-reading has a choice flavor, because there can be 'multiple' minimal $\boldsymbol{K}$ sets that satisfy the existential predication condition.

The above discussion covered the $\exists$1-quantifier *one of the boys*. The rest of this section considers other indefinites of the form 'Num-(of-the-)NP'. Recall that pair-list readings are not available in matrix $\exists$-questions; for example, the $\exists$2-question (72c) cannot be interpreted with distributivity taking scope between quantification and uniqueness.

(72) I know that every student voted for a different candidate. Which candidate did ...

a. ... every/each student vote for? (*every/each* $\gg \iota$)

b. ... one of the students vote for? ($\exists 1 \gg \iota$)

c. # ... two of the students vote for? ($\exists 2 \gg \cancel{\text{EACH}} \gg \iota$)

To avoid over-generating pair-list readings, pioneering works derive these readings in ways that would crash in questions with a non-universal quantifier. In Dayal's analysis, the derivation of pair-list readings crashes because $\exists$-quantifiers have multiple minimal witness sets. In Fox's analysis, the derivation crashes because we cannot find the unique minimal set among the sets of sub-questions that satisfy an existential predication condition. Obviously, this strategy comes at the cost of failing to account for the choice readings of $\exists$-questions.

I assume that the determiner of the numeral-modified indefinite *two of the boys* is not $\exists$2 but rather the simple $\exists$; in other words, the cardinal numeral *two* is part of the restrictor of the determiner. With this assumption, the quantifier *two of the boys* ranges over the set of entities that are pluralities of two boys; in other words, it denotes a set of sets that contain at least one of such plural entities.

(73) a. $\exists 2 := \lambda P_{\langle e,t\rangle}\lambda Q_{\langle e,t\rangle}.|P \cap Q| = 2$

b. $\exists := \lambda P_{\langle e,t\rangle}\lambda Q_{\langle e,t\rangle}.P \cap Q \neq \varnothing$

This assumption is supported by the contrast between (74a) and (74b): unlike the distributive $\forall$-quantifiers *every/each boy*, the indefinite *two (of the) boys* can grammatically combine with a collective predicate such as *formed a team*. This contrast argues that the indefinite *two (of the) boys* is not intrinsically distributive; it cannot be analyzed as an existential distribution over two atomic boys.

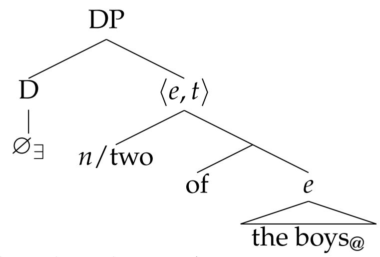
(74) a. Every/Each boy joined/*formed a team.

b. Two (of the) boys joined/formed a team.

[27] In (70c,d), there is no need to write out the domain condition that $\boldsymbol{f}$ is defined for at least one boy, because this condition is entailed by the definedness condition of the output proposition: for any x, the proposition $\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}(x))$ is defined only if $\boldsymbol{f}$ is defined for x.

The composition of *two of the boys* proceeds as in (75). First, *of* combines with a plural entity denoted by the *the*-phrase and returns a set of subparts of this entity. Next, the numeral *two*, as a predicate restrictor, combines with a set of entities and returns a subset consisting only of the entities that have exactly two atomic (AT) subparts, as in (75c,d). Finally, a covert existential determiner $\varnothing_{\exists}$ combines with this set-denoting NumP and returns an existential generalized quantifier (Link 1987).

(75) *two of the boys*



Assume that the discourse domain has three boys a,b,c:

a. $[\![\textit{the boys}_{@}]\!] = a \oplus b \oplus c$

b. $[\![\textit{of}]\!] = \lambda x_e.\{y \mid y \leq x\}$

c. $[\![\textit{two}]\!] = \lambda Q_{\langle e,t\rangle}.\{x \mid |\text{AT}(x)| = 2 \wedge Q(x)\}$

d. $[\![\textit{two of the boys}_{@}]\!] = \{a \oplus b, b \oplus c, a \oplus c\}$

e. $[\![\varnothing_{\exists}\ \textit{two of the boys}_{@}]\!] = \lambda P_{\langle e,t\rangle}.\exists x[|\text{AT}(x)| = 2 \wedge \mathsf{boys}_{@}(x) \wedge P(x)]$
$= \lambda P_{\langle e,t\rangle}.\exists x \in \{a \oplus b, b \oplus c, a \oplus c\}[P(x)]$

Let's return to the composition of the matrix $\exists$2-question. Just as in (70b), here the minimal $\boldsymbol{K}$ sets yielded by the application of the $f^{\text{MIN}}_{\text{CH}}$-operator are all 'singleton' sets. Each of these sets consists of one single proposition of the form $\ulcorner x\ \textit{watched}\ \boldsymbol{f}(x)\urcorner$, where x is the plurality of two boys, as in (76b). Hence, the derived QIQ-reading is [$-$PL], just as in the $\exists$1-question.

(76) Which movie did two of the boys watch? (QIQ-reading)

$[_{\text{CP}}{}^{\text{BD}}(\text{wh-movie}_{@})\ \lambda i\ [_{②}\ \bar{\bigcap}\ [_{\gamma}\ f^{\text{MIN}}_{\text{CH}}\ \lambda \boldsymbol{K}\ [_{①}\ \text{two-of-the-boys}_{@}\ \lambda j\ [\ \boldsymbol{K}\ (\lambda w.x_j\text{-wat}_w\text{-}\boldsymbol{f}_i(x_j))\ ]]]]]$

a. $[\![①]\!] = \exists x \in \mathsf{2\text{-}boys}_{@}[\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)))]$
('2-boys$_{@}$' abbreviates the set of entities that are pluralities of two boys in @.)

b. $[\![\gamma]\!] = \{\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))\}$, where x is the chosen two boys

c. $[\![②]\!] = \lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x))$, where x is the chosen two boys

d. $[\![\text{CP}]\!] = \lambda \boldsymbol{f}_{\langle e,e\rangle} : \text{Ran}(\boldsymbol{f}) = \mathsf{mov}_{@}.\lambda w[\mathsf{wat}_w(x, \boldsymbol{f}(x))]$, where x is the chosen two boys

In contrast to matrix $\exists$-questions, extensional embeddings of $\exists$-questions sometimes admit pair-list readings (Szabolcsi 1997b; Beghelli 1997; Appendix B). For example, the embedding sentence (77) is felicitous even if each boy watched a different movie. I assume that this sentence has the LF in (77a) and the meaning in (77b). In this LF, the indefinite moves over the embedding verb *know*. Its trace in the matrix clause is associated with a covert distributor EACH, which yields the 'EACH $\gg \iota$' reading.

(77) Susi knows [which movie two of the boys watched]. ($\exists 2 \gg$ EACH $\gg \iota$)

a. [[$\varnothing_{\exists}$ two-of-the-boys$_{@}$] λx_e [[x EACH] λy_e [Susi knows which movie y watched]]]

b. $\exists x[x \in \mathsf{2\text{-}boys}_{@} \wedge \forall y \in \text{AT}(x)[[\![\textit{Susi knows which movie y watched}]\!]]]$

This analysis is supported by the contrast between (77) and (78): adding overt *each* to the embedded question makes the pair-list reading unavailable.[28] If the matrix trace of the indefinite were associated with covert EACH, the local trace y would be atomic, and therefore could not be associated with overt *each*.

(78) Susi knows [which movie two of the boys **each** watched]. $(\exists 2 \gg \cancel{\text{EACH}} \gg \iota)$
[[$\varnothing_\exists$ two-of-the-boys$_@$] λx_e [[x EACH] λy_e [Susi knows which movie y (#**each**) watched]]]

In matrix $\exists$-questions, however, pair-list readings cannot be licensed by VP-EACH/*each*. In sentence (79), the contribution of *each* to the meaning is that the two chosen boys watched the (same) movie separately, not that they watched possibly a different movie each. The presented analysis straightforwardly explains this fact: to derive a pair-list QIQ-reading as in the $\forall$-question, the quantificational predication condition has to be distributive. Such distributivity arises only if one of the following conditions is met: (i) the quantifier itself is lexically distributive, or (ii) a separate distributor appears between the quantifier and the null predication-operator $\boldsymbol{K}$. Condition (i) is satisfied in questions with an *each*/*every*-subject. Condition (ii) does not apply to English, because VP-EACH/*each* should apply locally to VP as in (79a), not as high as in (79b).

(79) Which movie did two of the boys **each** watch? $(\exists 2 \gg \cancel{\text{EACH}} \gg \iota)$

a. ... [two-of-the-boys$_@$ λx_e [$\boldsymbol{K}$ [$_{\text{IP}}$ λw [[x **each**] λy_e [$_{\text{VP}}$ y watched$_w$ $\boldsymbol{f}(y)$]]]]]

b. * ... [two-of-the-boys$_@$ λx_e [[x **each**] λy_e [$\boldsymbol{K}$ [$_{\text{IP}}$ λw [$_{\text{VP}}$ y watched$_w$ $\boldsymbol{f}(y)$]]]]]

6.4.3. Questions with a negative quantifier

Negative quantifiers do not participate in QIQ-readings. For example, the NO-question in (80) can be responded to by specifying a single movie or a Skolem function to atomic movies, but not by silence.

(80) Which movie did no boy/ none of the boys watch? (✓Individual, ✓Functional, ✗QIQ)
Hulk./ The movie that his grandpa recommended./ #[Silence]

The proposed analysis easily explains why QIQ-reading are not available in NO-questions. The minimal set that contains none of the propositions of the form ⌜*boy-x watched* $\boldsymbol{f}_i(x)$⌝ is the empty set, whose conjunction is undefined. Hence, composing a NO-question with the proposed LF schema for QIQ-readings yields a deviant topical property, which maps any input to undefinedness.

(81) Which movie did none of the boys watch? (✗QiQ-reading)
$[^{\text{BD}}$(wh-movie$_@$) λi [$_{②}$ $\bar{\bigcap}$ [$_\gamma$ $f^{\text{MIN}}_{\text{CH}}$ $\lambda \boldsymbol{K}$ [$_{①}$ none-of-the-boys$_@$ λj [$\boldsymbol{K}$ $(\lambda w.x_j\text{-wat}_w\text{-}\boldsymbol{f}_i(x_j))$]]]]]

a. $[\![\textit{none of the boy}_@]\!] = \lambda P_{\langle e,t\rangle}.\neg\exists x[\mathsf{boys}_@(x) \wedge P(x)]$

b. $[\![①]\!] = \neg\exists x \in \mathsf{boys}_@[\boldsymbol{K}(\lambda w.\mathsf{wat}_w(x, \boldsymbol{f}_i(x)))]$

c. $[\![\gamma]\!] = f^{\text{MIN}}_{\text{CH}}([\![\lambda \boldsymbol{K}.①]\!]) = \varnothing$

d. $[\![②]\!]$ is undefined

Relatedly, recall that QIQ-readings are unavailable if the quantificational subject is a GQ-coordination involving a negative conjunct, as seen in (82a,b). Such questions, if analyzed with the LF schema for QIQ-readings, give rise to meanings equivalent to the QIQ-readings of the questions in (83a,b), which do not contain a negative conjunct. For example, for (82a), the minimal set that contains every

[28] I thank an anonymous reviewer of *L&P* for bringing this data to my attention.

proposition of the form ⌜*boy-x watched* $f(x)$⌝ and no proposition of the form ⌜*teacher-x watched* $f(x)$⌝ is simply the set of propositions of the form ⌜*boy-x watched* $f(x)$⌝.

(82) a. Which movie did [each of the boys and none of the teachers] watch? (✗QɪQ)
b. Which movie did [one of the girls and none of the teachers] watch? (✗QɪQ)

(83) a Which movie did [each of the boys] watch? (✓QɪQ)
b Which movie did [one of the girls] watch? (✓QɪQ)

Given the semantic equivalence between (82a,b) and (83a,b), I propose to explain the blocking effect of negative quantifiers in terms of the Efficiency constraint (Meyer 2013). According to this constraint, the QɪQ-structures of (82a,b) are ill-formed because of the existence of the simplifications in (83a,b).[29]

(84) Efficiency (à la Meyer 2013)
a. LF α is ill-formed if there is an LF β s.t. β is a simplification of α.
b. β is a simplification of α iff (i) $[\![\alpha]\!] = [\![\beta]\!]$, and (ii) β can be derived from α by replacing nodes in α with their subconstituents.

6.4.4. Questions with a counting quantifier

Decreasing quantifiers (e.g., *at most two boys*, *less than three boys*) do not license QɪQ-readings. In (85), the boy(s)-movie-pair answer (85b) is not a choice answer; instead, it is an individual answer, where uniqueness scopes above the quantifier.

(85) Which movie did at most two/ less than three boys watch?
\# 'For at most two/ less than three boys x, [tell me] which movie did x watch?'
a. *Hulk*. (Intended: '*Hulk* is the only movie watched by at most two/ less than three boys. The other movies were watched by more boys.')
b. Andy and Billy watched *Hulk*.
i. ✓ Individual reading: '*Hulk* is the **only** movie watched by at most two/ less than three boys, who are Andy and Billy. The other movies were watched by more boys.'
ii. ✗ Choice reading: 'Andy and Billy are two boys who both watched **only** *Hulk*.'

It is quite appealing to extend the analysis proposed for negative quantifiers to these decreasing quantifiers. Following Hackl (2000), Xiang (2019a) decomposes a decreasing quantifier into a negative determiner *no* and a set-denoting restrictor, as in (86). With this decompositional analysis, the unavailability of QɪQ-readings in (85) can be explained in the same way as in the NO-question (81).

(86) a. $[\![\textit{at most two boys}_{@}]\!] = \lambda P_{\langle e,t\rangle}.\neg\exists x[|\textsc{At}(x)| > 2 \wedge \textsf{boys}_{@}(x) \wedge P(x)]$
b. $[\![\textit{less than three boys}_{@}]\!] = \lambda P_{\langle e,t\rangle}.\neg\exists x[|\textsc{At}(x)| \geq 3 \wedge \textsf{boys}_{@}(x) \wedge P(x)]$

However, despite having a non-decreasing subject, sentence (87) below doesn't admit a QɪQ/choice-reading either. As in (85), here the uniqueness inference triggered by the singular *wh*-object must scope above the quantificational subject. This fact argues that the unavailability of QɪQ-readings in (85) and (87) has nothing to do with the monotonicity pattern of the quantificational subject.

[29] I thank an anonymous reviewer of *L&P* for pointing me in this direction for an explanation.

(87) Which movie did at least/ exactly two boys watch? (✓Individual, ✓Functional, ✗QIQ)
'For at least/ exactly two boys x, [tell me] which movie did x watch?'

In contrast to Xiang 2019a, this paper attributes the unavailability of QIQ-readings in (85) and (87) to a syntactic constraint stating that counting quantifiers are scopally unproductive (Szabolcsi 1997a; Beghelli and Stowell 1997). Beghelli and Stowell (1997) classify non-interrogative quantifiers into the following categories and argue that they have different landing sites. In particular, counting quantifiers have very local scope (take scope essentially in situ) and resist specific interpretations.

(88) Types of non-interrogative quantifiers (Beghelli and Stowell 1997)
   a. Negative quantifiers: *no*-NP
   b. Universal-distributive quantifiers: *every/each*-NP
   c. Grouping quantifiers: indefinites like *a/some/several*-NP, bare-numeral quantifiers (e.g., *one student*, *three students*), and *the*-phrases
   d. Counting quantifiers: decreasing quantifiers headed with determiners like *few*, *fewer than five*, and *at most six* and generally cardinality expressions with a modified numeral (e.g., *more than five*, *between six and nine*)

To derive a QIQ-reading, the quantifier must escape IP and move across a null predication-operator $\boldsymbol{K}$. Counting quantifiers cannot have such global scope and hence do not participate in QIQ-readings.

6.4.5. Questions with a non-quantificational subject

For questions with a non-quantificational subject, the difference between their individual reading and the reading generated from the LF schema for QIQ-readings is trivial. For example, the QIQ-answer to (89), if available, is the conjunction of the minimal set containing the proposition '*The boys watched $\boldsymbol{f}$(the-boys)*', which is simply this proposition itself. Although it is hard to tell whether QIQ-readings are truly available in these questions, composing such questions with the proposed LF schema for QIQ-readings does not over-generate any unwanted meanings.

(89) Which movie did the boys watch?
   a. $[_{\text{CP}}{}^{\text{BD}}(\text{wh-movie}_@)\ \lambda i\ [_{②}\ \bar{\bigcap}\ [_{\gamma}\ f^{\text{MIN}}_{\text{CH}}\ \lambda \boldsymbol{K}\ [_{①}\ \textsc{lift}(\text{the-boys}_@)\ \lambda j\ [\ \boldsymbol{K}\ (\lambda w.x_j\text{-wat}_w\text{-}\boldsymbol{f}_i(x_j))\ ]]]]]$
   b. $\lambda \boldsymbol{f}_{\langle e,e\rangle} : \text{Ran}(\boldsymbol{f}) = \mathsf{mov}_@.\lambda w[\mathsf{wat}_w(\mathsf{the\text{-}boys}_@, \boldsymbol{f}(\mathsf{the\text{-}boys}_@))]$

Since non-quantificational expressions are not distributive in the lexicon, the readings derived from the LF schema of QIQ-readings are not pair-list. Just as with the matrix ∃2-question in (79), distributivity from VP-*each* is applied locally within the IP as in (90a), i.e., not as high as in (90b,c). Hence, the QIQ-readings of questions with a non-quantificational subject are [+D-EXH,−PL,−CH].[30]

(90) Which movie did the boys **each** watch?
   a. ... [$\textsc{lift}(\text{the-boys}_@)\ \lambda x_e$ [$\boldsymbol{K}$ [$_{\text{IP}}$ λw [[x **each**] λy_e [y watched$_w$ $\boldsymbol{f}(y)$]]]]]
   b. * ... [$\textsc{lift}(\text{the-boys}_@)\ \lambda x_e$ [[x **each**] λy_e [$\boldsymbol{K}$ [$_{\text{IP}}$ λw [y watched$_w$ $\boldsymbol{f}(y)$]]]]]
   c. * ... [[the-boys$_@$ **each**] λy_e [$\boldsymbol{K}$ [$_{\text{IP}}$ λw [y watched$_w$ $\boldsymbol{f}(y)$]]]]

[30] Here domain exhaustivity is trivially satisfied. For example, the set that the Montagovian individual 'LIFT(the-boys)' ranges over is a singleton set containing only the plural entity denoted by *the boys*.

Strikingly, Johnston (2019) observes cases like (91a), where it appears that the definite plural *the players* licenses a pair-list reading. In this example, a cumulative answer that does not specify the player-number correspondence is too weak to address the question. Johnston further notices that such pair-list readings exhibit a subject–object asymmetry, similar to what is observed in ∀-questions.[31] To account for these observations, Johnston assumes that the definite plural carries a covert DP-internal EACH, which turns this definite plural into a universal distributive quantifier.

(91) (Context: In a basketball team, each of the five players got to choose a jersey, numbered from 1 to 5.)

a. Which numbers did the players pick?
Ann picked 1, Ben picked 2, Chris picked 3, Dan picked 4, Emma picked 5.

b. Which players picked the numbers?
#Ann picked 1, Ben picked 2, Chris picked 3, Dan picked 4, Emma picked 5.

However, I would like to argue here that the seeming pair-list reading in (91a) is not a QIQ-reading; instead, it is a (non-QIQ) functional reading involving '*respective* distributivity'. First of all, to see why it is not a QIQ-reading, compare the following questions in the same pair-list context:

(92) (Context: In a basketball team, each of the five players got to choose a jersey, numbered from 1 to 5.)

a. Which {#numbers, number} did each of the players pick?

b. Which {numbers, #number} did the players pick?

In the ∀-question (92a), the *wh*-object must be singular because each player picked only one number; in (92b), the *wh*-object must be plural because multiple numbers were picked collectively. This contrast argues that (92a) and (92b) have different question nuclei; if these questions had the same nucleus, they would allow for the same *wh*-phrases in the given context.

Why does (91a) admit a pair-list answer? The discussion above has excluded the possibility of applying a DP-internal EACH to the definite plural: if it were available, *the players* would function in the same way as *each of the players*, which leaves the contrast in (92) unexplained. The licensing of pair-list cannot be ascribed to a (covert) VP-*each* either: as seen in (93), adding overt *each* to the question makes it infelicitous, since it gives rise to a false inference that each player picked more than one number.

(93) (Context: In a basketball team, each of the five players got to choose a jersey, numbered from 1 to 5.)
Which numbers did the players (??each) pick?

I argue that the seeming pair-list reading of (91a) is a functional reading with *respective* distributivity. The question–answer pair is paraphrased as follows:

(94) 'Which numbers did the players pick, respectively?'
'The players Ann,Ben,Chris,Dan,Emma picked the numbers 1-to-5, respectively.'

Formally, *respective* distributivity is derived via the application of a covert operator RESP_g (Gawron and Kehler 2004; Chaves 2012; Law 2019): RESP_g combines with two pluralities (i.e., a plural predicate P and a plural individual x), breaks them into parts, pairs the parts using a pragmatically available sequencing function g, and performs a pair-wise evaluation facilitated by g.

[31] I thank Bernhard Schwarz (pers. comm.) for bringing this issue to my attention.

(95) $\textsc{Resp}_g := \lambda P \lambda x. \forall n[1 \leq n \leq |g| \rightarrow [g(P)(n)](g(x)(n))]$
(The n-th part of the property P holds for the n-th part of the individual x.)

Question (91a) can thus be composed as follows. Just as with any functional reading, the subject *the players* binds into the complex functional trace of the *wh*-object, yielding a *wh*-dependency. However, unlike other functional readings, here a $\textsc{Resp}_g$-operator is applied between the predicate $picked_w$-$\boldsymbol{f}_i(x_j)$ and the trace of the definite subject, yielding *respective* distributivity.

(96) Which numbers did the players pick?
$[_{\text{CP}}\ {}^{\text{BD}}(\text{wh-numbers}_@)\ \lambda i\ [_{\text{IP}}\ \lambda w.\ \text{the-players}_@\ \lambda j\ [_{\text{VP}}\ x_j\ \textsc{Resp}_g\ \text{picked}_w\text{-}\boldsymbol{f}_i(x_j)\ ]]]$

This analysis can account for the constraints observed in (91) and (92). First, since *respective* distributivity involves a dependency between the two arguments of $\textsc{Resp}_g$, the subject–object asymmetry seen in (91) can be explained in terms of constraints on dependencies. Second, this analysis explains why in (92b) the *wh*-object must be plural: $\textsc{Resp}_g$ requires the predicate $picked_w$-$\boldsymbol{f}_i(x_j)$ to be plural, which in turn requires the range of $\boldsymbol{f}_i$ to be a set of plurals.

What's more, this analysis explains why such pair-list-like readings are only available in particular contexts. In Johnston's example, it is straightforward to break the numbers 1-to-5 and the players A-to-E into two sequences and match them pair-wise. When sequencing or matching are pragmatically difficult, *respective* distributivity is not available. For example, question (97a) doesn't have a pair-list-like reading (cf. (97b)): it is pragmatically difficult to come up with a sequencing function to match a sequence of tables with a sequences of pluralities of workers. When the pair-list-like reading is unavailable, interpreting (97a) with a cumulative reading yields infelicity since the cumulative answer is part of the mutual knowledge.

(97) (Context: The department hired three workers to move ten tables. Each table were handled by two workers. Who the workers were and which tables were moved are mutual knowledge.)
a. # Which tables did the workers move?
b. Which table or tables did which workers move?

6.5. Interim summary

To sum up the core analysis: I argued that pair-list readings of multiple-*wh* questions and QiQ-readings of questions with a quantificational subject are extensional functional readings. As schematized in (98) and described in (99), the composition of these questions proceeds in four steps.

(98) a. Which boy watched which movie? (Pair-list reading)
$[_{\text{D}}\ {}^{\text{BD}}(\text{wh-movie}_@)\ \lambda i\ [_{\text{C}}\ \bar{\bigcap}\ \lambda p_{\langle s,t\rangle}\ [_{\text{B}}\ \text{wh-boy}_@\ \lambda j\ [\ \textsc{Id}(p)\ [_{\text{A}}\ \lambda w.x_j\text{-watched}_w\text{-}\boldsymbol{f}_i(x_j)\ ]]]]$
b. Which movie did Det-boy(s) watch? (QiQ-reading)
$[_{\text{D}}\ {}^{\text{BD}}(\text{wh-movie}_@)\ \lambda i\ [_{\text{C}}\ \bar{\bigcap}\ f^{\text{MIN}}_{\text{CH}}\ \lambda \boldsymbol{K}_{\langle st,t\rangle}\ [_{\text{B}}\ \textsc{Det}\text{-boy(s)}_@\ \lambda j\ [\ \boldsymbol{K}\ [_{\text{A}}\ \lambda w.x_j\text{-watched}_w\text{-}\boldsymbol{f}_i(x_j)\ ]]]]$

(99) (A) Indexations with the two traces yield a *wh*-dependency.
(B) The quantificational/*wh*- subject binds into the dependency sentence across an identity/predication operator.
(C) Conjoining a set of propositions with the dependency form (A) yields a graph description.
(D) The fronted *wh*-object restricts the range of the input functions.

Table 2 compares the nuclear denotations of the multiple-*wh* question (98a) and four corresponding QiQ-questions of the form (98b). In all of these questions, the asserted component of the nuclear denotation is the conjunction of a set of propositions representing the graph of the input function f. In the four questions with a non-*wh*-subject, the quantificational predication condition generated at (B) yields a definedness condition that restricts the domain of f.

Subject type	Domain condition of f	Graph description of f	D-EXH	PL	CH
which boy		$\bar{\bigcap}\{\lambda w.\mathsf{W}_w(x, f(x)) \mid \mathsf{B}_@(x)\}$	−	+	−
every/each boy	$\forall x \in \mathsf{B}_@[x \in \mathrm{Dom}(f)]$	$\bar{\bigcap}\{\lambda w.\mathsf{W}_w(x, f(x)) \mid \mathsf{B}_@(x)\}$	+	+	−
n of the boys	$\exists x \in n\text{-}\mathsf{Bs}_@[x \in \mathrm{Dom}(f)]$	$\bar{\bigcap}\{\lambda w.\mathsf{W}_w(x, f(x))\}$ where $x \in n\text{-}\mathsf{Bs}_@$	−	−	+
LIFT(*the boys*)	$\text{the-}\mathsf{Bs}_@ \in \mathrm{Dom}(f)$	$\bar{\bigcap}\{\lambda w.\mathsf{W}_w(x, f(x))\}$ where $x = \text{the-}\mathsf{Bs}_@$	+	−	−
none of the boys	$\neg\exists x \in \mathsf{B}_@[x \in \mathrm{Dom}(f)]$	$\bar{\bigcap}\, \varnothing$	−	−	−

Table 2: Denotations of question nuclei

In questions with a quantificational subject, the QiQ-effect is derived by extracting one of the minimal proposition sets that satisfy the quantificational predication condition yielded at (B). This analysis explains the properties of ∀-questions and ∃-questions w.r.t. the following parameters:

- [±D-EXH]: As in a ∀-question, the resulting QiQ-reading presupposes domain exhaustivity if the quantificational predication condition yielded at (B) is subject to a definedness condition stating that the input f is defined for 'every' element in the quantification domain of the subject.
- [±PL]: As in a ∀-question, with other conditions being equal, the resulting QiQ-reading admits pair-list answers only if there is a 'non-singleton' set of propositions that minimally satisfies the quantificational predication condition yielded at (B). To derive such a non-singleton minimal set, the quantificational subject must be lexically distributive.
- [±CH]: As in an ∃-question, with other conditions being equal, the resulting QiQ-reading has a choice flavor if there are 'multiple' minimal proposition sets that satisfy the quantificational predication condition yielded at (B).

I further demonstrated why QiQ-readings are unavailable in many cases. In NO-questions, QiQ-readings are semantically deviant because the only minimal proposition set that satisfies a negative quantificational predication condition is the empty set. In questions with a counting quantifier, the LF schema for QiQ-readings is infeasible because counting quantifiers are unproductive in scoping.

Lastly, I discussed another source of pair-list readings in questions with a plural definite subject: although plural definites are not distributive lexically, pair-list readings might arise through a locally applied *respective* distributor.

7. Quantificational variability effects

As seen in Sect. 4.1.2, because it defines pair-list questions as sets of conjunctive propositions, the analysis of Dayal (1996, 2017) cannot account for the Q-variability effects in the embeddings of pair-list questions. Dayal defines simplex and pair-list questions uniformly as sets of propositions. For embeddings of simplex questions, the most natural way for her to derive the Q-variability inference is to let the matrix adverbial quantify over a set of atomic propositions, as exemplified in (100).

(100) Jill mostly knows [which students left].
$\rightsquigarrow$ 'Most p: p is a true proposition of the form ⌜*student-x left*⌝, Jill knows p.'

This proposition-based definition, however, is infeasible for embeddings of pair-list questions if a pair-list question denotes a set of conjunctive propositions (Lahiri 2002). For example, in a scenario where the three relevant boys b_1,b_2,b_3 watched and only watched the movies m_1,m_2,m_3, respectively, the strongest true propositional answer to the embedded pair-list question in (101) is $\lambda w.\mathsf{wat}_w(m_1, b_1) \wedge \mathsf{wat}_w(m_2, b_2) \wedge \mathsf{wat}_w(m_3, b_3)$, and the Q-variability inference is true if Jill knows at least two of the three atomic conjuncts, as in (101a); however, these conjuncts cannot be semantically retrieved out of their conjunction. In contrast, family-of-questions approaches such as Fox 2012a,b can derive this inference by letting the matrix adverb quantify over a set of sub-questions, as paraphrased in (101b).

(101) Jill mostly knows [$_{\text{PAIR-LIST}}$ $\left\{ \begin{array}{l} \text{which movie every boy watched} \\ \text{which boy watched which movie} \end{array} \right\}$].

a. $\rightsquigarrow$ 'Most p: p is a true proposition of the form ⌜*boy-x watched movie-y*⌝, Jill knows p.'
b. $\rightsquigarrow$ 'Most Q: Q is a question of the form ⌜*which movie boy-x watched*⌝, Jill knows Q.'
c. $\rightsquigarrow$ 'Most $\langle x, y \rangle$: $\langle x, y \rangle$ is a boy-movie pair and x watched y, Jill knows that x watched y.'

Although this paper does not pursue a family-of-questions approach, the assumed categorial approach to question composition unlocks the option in (101c), where the quantification domain of *mostly* is a set of atomic functions. In my proposal, a pair-list question denotes a topical property that maps each input $\langle e, e \rangle$-type function to a conjunctive proposition. From this topical property, we can extract the function that yields the strongest true answer to this question and define the quantification domain of *mostly* as a set of atomic subparts of this function. For example in (103), the strongest true answer is the function in (103a), and its atomic subparts are those in (103b).

(102) a. A function f is atomic iff $\bigoplus \mathrm{Dom}(f)$ is atomic.
b. $\textsc{At}(f) = \{f' \mid f' \subseteq f \text{ and } f' \text{ is atomic}\}$

(103) Which boy watched which movie?/ Which movie did every boy watch?
(The discourse domain includes three boys b_1,b_2,b_3 and three movies m_1,m_2,m_3. In a world w, b_1 watched only m_1, b_2 watched only m_2, and b_3 watched only m_3.)

a. $\textsc{Ans}^S(w)(\llbracket \mathrm{Q} \rrbracket) = \begin{bmatrix} b_1 \rightarrow m_1 \\ b_2 \rightarrow m_2 \\ b_3 \rightarrow m_3 \end{bmatrix}$

b. $\textsc{At}(\textsc{Ans}^S(w)(\llbracket \mathrm{Q} \rrbracket)) = \left\{ \begin{array}{c} [b_1 \rightarrow m_1] \\ [b_2 \rightarrow m_2] \\ [b_3 \rightarrow m_3] \end{array} \right\}$

Xiang 2020 provides two ways to define a Q-variability inference based on short answers. Ignoring the complications needed for accounting for mention-some readings, I schematize these two definitions as in (104a,b).[32] (For a compositional derivation, see Cremers 2018.) In both definitions, the quantification domain of the matrix adverbial *mostly* is a set of atomic entities or a set of atomic $\langle e, e \rangle$-type functions.

(104) The Q-variability inference of 'Jill mostly knows Q':

a. $\lambda w.\textsc{Most}\ x[x \in \textsc{At}(\textsc{Ans}^S(w)(\llbracket \mathrm{Q} \rrbracket))][\mathsf{know}_w(j, \llbracket \mathrm{Q} \rrbracket(x)]$
(For most x s.t. x is an atomic subpart of the strongest true short answer to Q, Jill knows the inference $\llbracket \mathrm{Q} \rrbracket(x)$.)

[32] Xiang 2020 also considers mention-some readings of questions, where a question can have multiple complete true answers. Once mention-some readings enter the picture, $\textsc{Ans}^S(w)(\llbracket \mathrm{Q} \rrbracket)$ needs to be defined as a set of entities/functions, not as a single entity/function.

b. $\lambda w.\textsc{Most}\ x[x \in \textsc{At}(\textsc{Ans}^S(w)([\![\mathrm{Q}]\!]))][\mathsf{know}_w(j, \lambda w'.x \leq \textsc{Ans}^S(w')([\![\mathrm{Q}]\!]))]$
(For most x s.t. x is an atomic subpart of the strongest true short answer to Q, Jill knows that x is a subpart of the strongest true short answer to Q.)

In (104a), the scope of the adverbial *mostly* says that Jill knows an atomic proposition, which is derived by applying the topical property of the embedded question to an entity or an $\langle e,e\rangle$-type function x, where x is an atomic subpart of the strongest true answer to the embedded question. This definition works for embeddings of multiple-*wh* questions, but not for embeddings of $\forall$-questions: the topical property of the pair-list $\forall$-question *which movie every boy watched* is only defined for $\langle e,e\rangle$-type functions that are defined for every boy, not for atomic functions such as $[b_1 \to m_1]$.

Alternatively, in (104b), the scope of *mostly* says that Jill knows a sub-divisive inference, which is semantically equivalent to the inference that Jill correctly identifies most of the boy-watched-movie pairs. This definitions works also for pair-list $\forall$-questions. In the context described in (103), this sub-divisive inference is true iff in every world w' s.t. w' is compatible with Jill's belief, the strongest true short answer to the embedded $\forall$-question in w' is among the seven functions in Figure 4. This figure illustrates a partition of possible worlds based on which movie each of the three boys watched. The world w described in (103) is located in the middle cell. In the other cells, correspondences conflicting with w are colored in light gray. It is straightforward to see that the union of the seven cells represents the following proposition: 'For most (or all) of the pairs $\langle b_n, m_n\rangle$ in $\{\langle b_1, m_1\rangle, \langle b_2, m_2\rangle, \langle b_3, m_3\rangle\}$, m_n is the unique movie watched by b_n.' Knowing this proposition means correctly identifying most of the three correspondences from a boy to the unique movie that this boy watched.

	$\begin{bmatrix} b_1 \to m_2 \\ b_2 \to m_2 \\ b_3 \to m_3 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_3 \\ b_2 \to m_2 \\ b_3 \to m_3 \end{bmatrix}$
$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_1 \\ b_3 \to m_3 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_2 \\ b_3 \to m_3 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_3 \\ b_3 \to m_3 \end{bmatrix}$
$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_2 \\ b_3 \to m_1 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_2 \\ b_3 \to m_2 \end{bmatrix}$	

Figure 4: Illustration of the sub-divisive inference in the quantification scope of (104b)

8. Conclusions

This paper started with the novel observation that pair-list $\forall$-questions and their multiple-*wh* counterparts are semantically different — only the $\forall$-questions are subject to domain exhaustivity. Given this contrast, I argued that the composition structure of a pair-list $\forall$-question must be distinct from that of its multiple-*wh* counterpart. Furthermore, drawing on the uniform syntactic constraints on distributing QɪQ-readings, I concluded that the QɪQ-readings of matrix questions should be derived uniformly.

Influential accounts such as Dayal 1996, 2017 and Fox 2012a,b do not reflect awareness of the contrast in domain exhaustivity between $\forall$-questions and multiple-*wh* questions. These accounts treat pair-list questions uniformly and compose these questions either with the same LF or with different LFs that yield the same root denotation. In addition, to explain why only subject *every/each*-phrases license pair-list readings, these accounts derive pair-list readings in a way that crashes in questions with a non-universal quantifier. In consequence, they over-predict domain exhaustivity effects for multiple-*wh* questions and fail to account for the choice readings of ∃-questions.

This paper presented a novel analysis of the composition of complex questions. This analysis has three main ingredients. First, in line with functionality approaches, I proposed that QiQ-questions and pair-list multiple-*wh* questions both involve *wh*-dependencies — the quantificational/*wh*- subject binds the argument index of the functional trace of the *wh*-object. In particular, in a pair-list multiple-*wh* question, the *wh*-subject quantifies into an identity operation w.r.t. a proposition expressing this dependency; in a QiQ-question, the quantificational subject quantifies into a predication operation w.r.t. this dependency. The subtle differences between the two quantificational binding-into-dependency operations are responsible for the contrast in domain exhaustivity between multiple-*wh* questions and ∀-questions. Second, for questions with a quantifier, inspired by Fox (2012b), I assumed that the seeming QiQ-effect is derived by extracting one of the minimal sets of propositions that satisfy the quantificational predication condition. This analysis naturally predicts which questions admit QiQ-readings; it also predicts whether the QiQ-reading of a question is subject to domain exhaustivity, admits pair-list answers, and has a choice flavor. Finally, by assuming a categorial approach, I showed that the presented analysis overcomes the difficulty in accounting for the Q-variability effects that Dayal's (1996) functionality analysis encountered.

Appendix A. A partition-based approach

Section 3 mentioned that the following LF, repeated from (18), suffers type-mismatch for most frameworks of question semantics:

(105) Which movie did Det-boy watch?
*[Det-boy λx_e [Which movie did x watch]]

Partition semantics is exempt from this type-mismatch problem. Groenendijk and Stokhof (1984: Chap. 3) first analyze the pair-list ∀-question (106) as a partition of possible worlds grouped in terms of which boy watched which movie. In the derivation of this denotation, the quantifier *every boy* quantifies into an identity operation (of type t), which says that x watched the same movies in w and in w'.

(106) Which movie did every boy watch?
$\lambda w \lambda w'.\forall x[\mathsf{boy}_@(x) \rightarrow \{y \mid \mathsf{mov}_@(y) \wedge \mathsf{wat}_w(x,y)\} = \{y \mid \mathsf{mov}_@(y) \wedge \mathsf{wat}_{w'}(x,y)\}]$
(w and w' are in the same partition cell iff for every boy x, x watched the same movies in w and in w'.)

However, Groenendijk and Stokhof themselves are not satisfied with this account since it does not extend to questions with a non-universal quantifier. For example, the predicted meaning for the corresponding ∃-question (107) is not a partition (see also Krifka 2001). Thus, Groenendijk and Stokhof ultimately pursue another family-of-questions approach using witness sets (footnote 15).

(107) Which movie did one of the boys watch?
$\lambda w \lambda w'.\exists x[\mathsf{boy}_@(x) \wedge \{y \mid \mathsf{mov}_@(y) \wedge \mathsf{wat}_w(x,y)\} = \{y \mid \mathsf{mov}_@(y) \wedge \mathsf{wat}_{w'}(x,y)\}]$
(w and w' are in the same partition cell iff for one of the boys x, x watched the same movies in w and in w'.)

For a concrete illustration, consider a discourse with two boys a,b and two movies m_1,m_2. The four worlds vary by which boy watched which movie. w_1,w_2,w_3 are grouped in one shaded cell C_1: a watched the same movie in w_1 and w_2 (and b watched the same movie in w_1 and w_3). Likewise, w_2,w_3,w_4 all belong to the shaded cell C_2: b watched the same movie in w_2 and w_4 (and a watched the same movie in w_3 and w_4). In addition, C_1 and C_2 are distinct cells because neither boy watched the same movie in w_1 and w_4. The world grouping in Figure 5 is clearly not a partition: C_1 overlaps with C_2 — they both contain w_2 and w_3. Moreover, from this world grouping, we cannot identify which movie any of the boys watched. For example, if w_1 is the actual world, then C_1 is the cell which the actual world belongs to; however, based on C_1, we cannot decide on whether a watched m_1 (as in w_1 and w_2) or he watched m_2 (as in w_3).

C_1:
$w_1 : \{\langle a, m_1\rangle, \langle b, m_2\rangle\}$
$w_2 : \{\langle a, m_1\rangle, \langle b, m_1\rangle\}$
$w_3 : \{\langle a, m_2\rangle, \langle b, m_2\rangle\}$
$w_4 : \{\langle a, m_2\rangle, \langle b, m_1\rangle\}$

C_2 :
$w_1 : \{\langle a, m_1\rangle, \langle b, m_2\rangle\}$
$w_2 : \{\langle a, m_1\rangle, \langle b, m_1\rangle\}$
$w_3 : \{\langle a, m_2\rangle, \langle b, m_2\rangle\}$
$w_4 : \{\langle a, m_2\rangle, \langle b, m_1\rangle\}$

......

Figure 5: World grouping yielded by (107)

In addition, this analysis inherits the theory-internal problems with partition semantics. For instance, since partition semantics cannot explain the uniqueness effects of singular-*wh* questions (Xiang 2020), a partition-based account cannot explain the point-wise uniqueness effects in pair-list ∀-questions.

Appendix B. A question-embedding approach

Another intuitive and framework-independent way to solve the type-mismatch problem in quantifying into questions is to analyze matrix questions as covertly embedded questions (Karttunen 1977; Krifka 2001). The LF assumed by Karttunen (1977) is given in (108). Basically, whatever the overt question denotes, it is embedded into a t-type expression which can be quantified into.

(108) Which movie did Det-boy(s) watch?
[Det-boy(s) λx_e [I-ask-you [Which movie did x watch]]]

This analysis crucially requires the quantifier in the embedded question to scope over the intensional embedding predicate ask. However, drawing on the limited distribution of pair-list readings in matrix questions and intensional question-embeddings, I will now argue that this scoping pattern is not available.[33]

As discussed in Sect. 3 and explained in Sect. 6.4, only *every*/*each*-phrases can license pair-list readings for matrix questions. As for question-embeddings, Szabolcsi (1997b) observes a contrast between intensional complements and extensional complements. In particular, in embeddings with an extensional predicate (e.g., *know*, *find out*), numeral-modified indefinites such as *two of the boys* may also license a pair-list reading. For example, in a pair-list context where each boy watched a

[33] Krifka (2001) assumes the structure in (i), where the quantifier scopes over a speech act operator quest. This analysis is exempt from the over-generation problem since Krifka assumes that speech acts cannot be disjoined. However, it also leaves the choice readings of ∃-questions unexplained.

(i) Which movie did every boy watch?
[every-boy λx_e [quest [Which movie did x watch]]]

different movie, (109b) can be uttered felicitously and interpreted with the following scopal pattern: '$\exists 2 \gg$ EACH $\gg$ V $\gg \iota$' where 'V' stands for the embedding predicate.[34] As argued in Sect. 6.4.2, this reading can be derived from the LF in (110) (see also (77)): the indefinite takes wide scope relative to the embedding predicate, and its trace in the matrix clause is associated with a covert EACH.[35]

(109) Susi knew that each boy watched a different movie. In addition, ...
  a. Susi knew/ found out which movie **each** of the boys watched.
  b. Susi knew/ found out which movie **two** of the boys watched.

(110) Susi V-ed which movie two of the boys watched.
  [[two-of-the-boys λx [[x EACH] λy [Susi V-ed which movie y watched]]]

However, embeddings with an intensional predicate (e.g., *ask*, *wonder*) behave the same as matrix questions — only *every*/*each*-phrases may license pair-list readings in these embeddings. For example, in (111a,b) the uniqueness inference triggered by *which movie* must be interpreted between the embedding predicate and the quantifier: ASK $\gg \iota \gg \exists 2$.

(111) Susi knew that every boy watched a different movie. ...
  a. Susi wondered/ asked me which movie **each** of the boys watched.
  b. # Susi wondered/ asked me which movie **two** of the boys watched.

The lack of pair-list readings argues that the LF (110) is not available for (111a,b). Szabolcsi (1997b) argues that intensional predicates create weak islands, which prevent the quantifiers in the embedded questions from taking wide scope. If this explanation is on the right track, the embedding structure (108), which requires the quantifier in the embedded question to scope over ASK, should be infeasible.

Acknowledgement [To be added ...]

[34] In an informal survey, 7 out of 14 speakers judged (109b) as contradictory to the context. They reported that the use of *which movie* gave rise to the inference that two of the boys watched the same movie. For these speakers, it seems that neither *wonder*-type nor *find out*-type embeddings allow a quantifier inside the embedded question to scope over the embedding-predicate. This judgement is also consistent with my claim that quantifying into questions cannot be analyzed as quantification into question-embeddings.

[35] Rather than assuming covert movement of the quantifier, Szabolcsi (1997b) derives the wide scope reading by type-lifting the interrogative complements of extensional predicates. Combining the type-lifted question denotation (i) with an embedding predicate P yields a wide scope reading of the generalized quantifier π relative to P. Further, Szabolcsi argues that *wonder*-type predicates cannot select for lifted questions, and hence that quantifiers in intensional complements cannot take wide scope.

(i) Denotation of questions embedded under *find out*-type predicates:
  $\lambda P.\pi(\lambda x.P([\![\textit{wh-movie}]\!](\lambda y.[\![\textit{watched}]\!](x,y))))$

References

Barwise, Jon, and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4(2):159–219.

Beck, Sigrid, and Yael Sharvit. 2002. Pluralities of questions. *Journal of Semantics* 19:105–157.

Beghelli, Filippo. 1997. The syntax of distributivity and pair-list readings. In *Ways of scope taking*, ed. Anna Szabolcsi, 349–408. Dordrecht: Kluwer.

Beghelli, Filippo, and Tim Stowell. 1997. Distributivity and negation: The syntax of *each* and *every*. In *Ways of scope taking*, ed. Anna Szabolcsi, 71–107. Dordrecht: Kluwer.

Berman, Stephen Robert. 1991. On the semantics and logical form of *wh*-clauses. Ph.D. dissertation, University of Massachusetts at Amherst.

Caponigro, Ivano, and Anamaria Fălăuş. 2020. Unveiling multiple *wh*- free relative clauses and their functional *wh*-words. *Semantics and Pragmatics*. DOI:http://dx.doi.org/10.3765/sp.13.2.

Chaves, Rui P. 2012. Conjunction, cumulation and respectively readings. *Journal of Linguistics* 48(2): 297–344.

Chierchia, Gennaro. 1991. Functional WH and weak crossover. In *Proceedings of the West Coast Conference on Formal Linguistics 10*, ed. Dawn Bates, 75–90. Stanford: CSLI Publications.

Chierchia, Gennaro. 1993. Questions with quantifiers. *Natural Language Semantics* 1:181–234.

Chierchia, Gennaro, and Ivano Caponigro. 2013. Questions on questions and free relatives. Presentation at the 18th Sinn und Bedeutung, University of the Basque Country (UPV/EHU) in Vitoria-Gasteiz, September 2013. https://scholar.harvard.edu/files/chierchia/files/frs_and_qs_iii_2013-9-11.pdf

Ciardelli, Ivano, and Floris Roelofsen. 2018. An inquisitive perspective on modals and quantifiers. *Annual Review of Linguistics* 4:129–149.

Dotlacil, Jakub and Floris Roelofsen. 2021. A dynamic semantics of single-*wh* and multiple-*wh* questions. In *Proceedings of Semantics and Linguistic Theory 30*, ed. Joseph Rhyne, Kaelyn Lamp, Nicole Dreier, and Chloe Kwon, 376–395. Washington, D.C.: LSA. DOI:https://doi.org/10.3765/salt.v30i0.4839

Cremers, Alexandre. 2016. On the semantics of embedded questions. Ph.D. dissertation, École normale supérieure, Paris.

Cremers, Alexandre. 2018. Plurality effects in an exhaustification-based theory of embedded questions. *Natural Language Semantics* 26:193–251.

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

Dayal, Veneeta. 2002. Single-pair versus multiple-pair answers: Wh-in-situ and scope. *Linguistic Inquiry* 33:512–520.

Dayal, Veneeta. 2016. List answers through higher order questions. Colloquium talk at MIT, February 2016.

Dayal, Veneeta. 2017. *Questions*. Oxford: Oxford University Press.

Engdahl, Elisabet. 1980. The syntax and semantics of questions in Swedish. Ph.D. dissertation, University of Massachusetts at Amherst.

Engdahl, Elisabet. 1986. *Constituent questions*. Dordrecht: Reidel.

Fox, Danny. 2012a. Multiple *wh*-questions: Uniqueness, pair-list and second order questions. Class notes for MIT seminars.

Fox, Danny. 2012b. Pair-list with universal quantifiers. Class notes for MIT seminars.

Fox, Danny. 2013. Mention-some readings of questions. Class notes for MIT seminars.

Fox, Danny. 2018. Partition by exhaustification: Comments on Dayal 1996. In *Proceedings of Sinn und Bedeutung 22*, ed. Uli Sauerland and Stephanie Solt, 403–434. Berlin: Leibniz-Centre General Linguistics. https://semanticsarchive.net/Archive/jQ5MTcyZ/.

Fox, Danny. 2020. Partition by exhaustification: towards a solution to Gentile and Schwarz's puzzle. Manuscript, MIT. https://semanticsarchive.net/Archive/TljZGNjZ/.

Gallin, Daniel. 1975. *Intensional and higher-order modal logic*. Amsterdam: North-Holland Publishing Company.

Gawron, Jean Mark, and Andrew Kehler. 2004. The semantics of respective readings, conjunction, and filler-gap dependencies. *Linguistics and Philosophy* 27:169–207.

Groenendijk, Jeroen, and Martin Stokhof. 1984. On the semantics of questions and the pragmatics of answers. *Varieties of formal semantics* 3:143–170.

Hackl, Martin. 2000. Comparative quantifiers. Ph.D. dissertation, MIT.

Hagstrom, Paul Alan. 1998. Decomposing questions. Ph.D. dissertation, MIT.

Heim, Irene. 1995. Notes on questions. Class notes for MIT seminars.

Heim, Irene. 2012. Notes on questions. Class notes for MIT semantics.

Hornstein, Norbert. 1995. *Logical form: From GB to minimalism*. Oxford: Blackwell.

Jacobson, Pauline. 1994. Binding connectivity in copular sentences. In *Proceedings of Semantics and Linguistic Theory 4*, ed. M. Harvey and L. Santelmann, 161–178. Ithaca, NY: CLC Publications. https://doi.org/10.3765/salt.v4i0.2456.

Jacobson, Pauline. 1999. Towards a variable-free semantics. *Linguistics and Philosophy* 22:117–185.

Jacobson, Pauline. 2014. *Compositional semantics: An introduction to the syntax/semantics interface*. Oxford: Oxford University Press.

Johnston, William. 2019. Pair-list questions from dependent plurals. In *Proceedings of the 22nd Amsterdam Colloquium*, ed. Julian J. Schlöder, Dean McHugh, and Floris Roelofsen, 201–210. University of Amsterdam. https://archive.illc.uva.nl/AC/AC2019/uploaded_files/inlineitem/Johnston_Pair-list_questions_from_dependent_plurals.pdf.

Kang, Jungmin. 2012. Why short-form functional reading answers are not possible in multiple *wh*-questions. In *University of Pennsylvania Working Papers in Linguistics*, volume 18. https://repository.upenn.edu/pwpl/vol18/iss1/15.

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

Kotek, Hadas. 2014. Composing questions. Ph.D. dissertation, MIT.

Kotek, Hadas. 2019. *Composing questions*. Cambridge, MA: MIT Press.

Krifka, Manfred. 1991. Definite noun-phrases are not quantifiers. *Linguistic Inquiry* 23:156–163.

Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9:1–40.

Lahiri, Utpal. 1991. Embedded interrogatives and predicates that embed them. Ph.D. dissertation, MIT.

Lahiri, Utpal. 2002. *Questions and answers in embedded contexts*. Oxford: Oxford University Press.

Law, Hoi-Ki. 2019. Constraints on distributivity. Ph.D. dissertation, Rutgers University.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Meaning, use, and interpretation of language*, ed. Christoph Schwarze, Rainer Bäuerle, and Arnim von Stechow, 302–323. Berlin: De Gruyter.

Link, Godehard. 1987. Generalized quantifiers and plurals. In *Generalized quantifiers*, ed. P. Gärdenfors, 151–180. Dordrecht: Reidel.

May, Robert. 1985. *Logical form: Its structure and derivation*. Cambridge, MA: MIT Press.

May, Robert. 1988. Ambiguities of quantification and *wh*: A reply to Williams. *Linguistic Inquiry* 19:118–135.

Meyer, Marie-Christine. 2013. Ignorance and grammar. Ph.D. dissertation, MIT.

Moltmann, Friederike, and Anna Szabolcsi. 1994. Scope interactions with pair-list quantifiers. In *Proceedings of the North East Linguistic Society 24*, ed. Mercé González, volume 2, 381–395. Amherst, MA: GLSA Publications.

Nicolae, Andreea Cristina. 2013. *Any* questions? Polarity as a window into the structure of questions. Ph.D. dissertation, Harvard University.

Pafel, Jürgen. 1999. Interrogative quantifiers within scope. *Linguistics and Philosophy* 22:255–310.

Partee, Barbara H. 1986. Noun phrase interpretation and type-shifting principles. In *Studies in discourse representation theory and the theory of generalized quantifiers*, ed. J. Groenendijk, D. de Jongh, and M. Stokhof, 357–381. Foris, Dordrecht.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In *The representation of (in) definiteness*, ed. Eric Reuland and Alice ter Meulen. 98–129. Cambridge, MA: MIT Press.

Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge, MA: MIT Press.

Preuss, Susanne Marie-Luise. 2001. Issues in the semantics of questions with quantifiers. Ph.D. dissertation, Rutgers University.

Richards, Norvin Waldemar. 1997. What moves where when in which languages? Ph.D. dissertation, MIT.

Safir, Ken. 1984. Multiple variable binding. *Linguistics Inquiry* 15:603–638.

Shan, Chung-chieh and Chris Barker. 2006. Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy* 29:91–134.

Sharvit, Yael. 1997. The syntax and semantics of functional relative clauses. Ph.D. dissertation, Rutgers University.

Sharvit, Yael. 1999. Functional relative clauses. *Linguistics and Philosophy* 22:447–478.

Sharvit, Yael, and Jungmin Kang. 2017. Fragment functional answers. In *Proceedings of Semantics and Linguistic Theory 26*, ed. Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, 1099–1118. Washington, D.C.: LSA. https://doi.org/10.3765/salt.v26i0.3792.

Sharvy, Richard. 1980. A more general theory of definite descriptions. *The Philosophical Review* 89:607–624.

Srivastav, Veneeta. 1991. WH dependencies in Hindi and the theory of grammar. Ph.D. dissertation, Cornell University.

Szabolcsi, Anna. 1997a. Strategies for scope taking. In *Ways of scope taking*, ed. Anna Szabolcsi, 109–154. Dordrecht: Kluwer.

Szabolcsi, Anna. 1997b. Quantifiers in pair-list readings. In *Ways of scope taking*, ed. Anna Szabolcsi, 311–342. Dordrecht: Kluwer.

Williams, Edwin Samuel. 1994. *Thematic structure in syntax*. Cambridge, MA: MIT Press.

Winter, Yoad. 2001. *Flexibility principles in Boolean semantics*. Cambridge, MA: MIT Press.

Xiang, Yimei. 2016. Interpreting questions with non-exhaustive answers. Ph.D. dissertation, Harvard University.

Xiang, Yimei. 2019a. Getting quantifying into questions uniformly: Functionality, domain exhaustivity, and quantificational variability. In *Proceedings of Semantics and Linguistic Theory 29*, ed. Katherine Blake, Forrest Davis, Kaelyn Lamp, and Joseph Rhyne, 160–179. Washington, D.C.: LSA. `https://doi.org/10.3765/salt.v29i0.4610`.

Xiang, Yimei. 2019b. A variable-free approach to *wh*-dependency. Presentation at the Workshop on Complex *Wh*-questions, Nantes, November 2019.

Xiang, Yimei. 2020. A hybrid categorial approach to question composition. *Linguistics and Philosophy*. Online-first articles. `https://doi.org/10.1007/s10988-020-09294-8`.

Xiang, Yimei. 2021. Relativized exhaustivity: uniqueness and mention-some. Manuscript, Rutgers University. `https://ling.auf.net/lingbuzz/005322`.