Higher-order readings of *wh*-questions

Yimei Xiang, Rutgers University

Abstract In most cases, a *wh*-question expects answers naming an entity in the set denoted by the *wh*-complement. However, evidence from questions with modals or collective predicates show that sometimes a *wh*-question must be interpreted with a higher-order reading, in which this question expects answers naming a higher-order meaning such as a generalized quantifier. This paper investigates into the derivation and distribution of higher-order readings in *wh*-questions.

Keywords: *wh*-words, questions, higher-order readings, quantifiers, Boolean coordinations, number-marking, uniqueness, collectivity, reconstruction

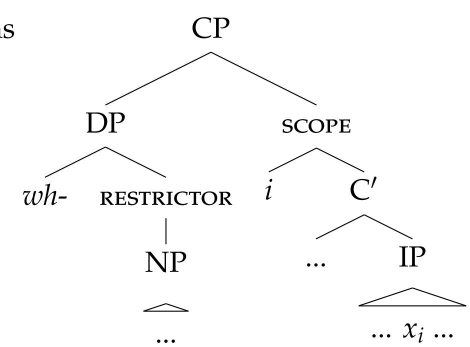
1. Introduction

A *wh*-question with *who*, *what*, or *which*-NP expects answers naming either an entity in the set denoted by the *wh*-complement or a generalized quantifier (GQ) over of a subset of this set. This requirement is especially robustly seen with short answers. For example in (1), the speaker uttering the short answer (1a) is committed to that the mentioned individual is a math professor (Jacobson 2016). Moreover, this inference projects over quantification: the most prominent reading of the disjunction (1b) yields that both mentioned individuals are math professors.[1]

(1) Which math professor left the party at midnight?
  a. Andy. $\rightsquigarrow$ *Andy is a math professor.*
  b. Andy or Billy. $\rightsquigarrow$ *Andy and Billy are math professors.*

To capture this question-answer relation, it is commonly assumed that the *wh*-determiner functions as a binder (such as an ∃-closure or a λ-operator) of *e*-type variables. An LF schema for *wh*-questions is given in (2): the *wh*-phrase binds an *e*-type variable inside the question-nucleus (namely, IP) and assigns this variable with a value in the extension of the NP-complement.

(2) LF schema of *wh*-questions



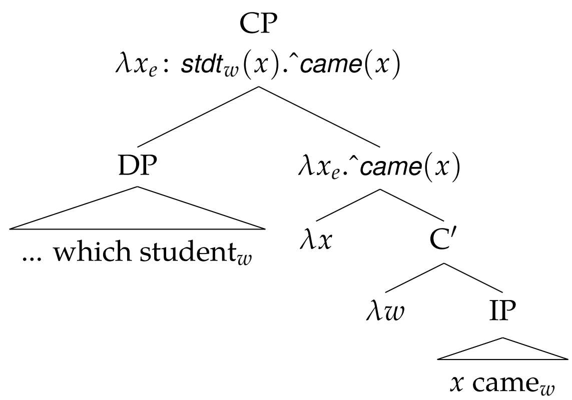
[1] Elided disjunctions are scopally ambiguous relative to this commitment, as described in (i). Here and throughout the paper, I consider only the reading (ia). The other reading can be derived by accommodating the presupposition locally.

(i) a. *Andy and Billy are math professors, and one of them left the party at midnight.*
  b. *Either Andy or Billy is math professor who left the party at midnight.*

The above assumption with *wh*-phrases predicts that a *wh*-question denotes either a function defined for values in the extension of the NP-complement (as assumed by categorial approaches) or a set of propositions naming such values (as assumed by Karttunen Semantics). For convenience in describing the relation between *wh*-phrases and *wh*-questions in meaning, I will present the story using categorial approaches to question composition. However, the core idea of this paper is independent from the assumptions of categorial approaches in defining and composing questions.

Categorial approaches define questions as functions (henceforth called **Q-functions**) and *wh*-phrases as **function domain restrictors**. As in (3), *which student* combines with a function defined for any individuals and returns a more restrictive Q-function only defined for atomic students. I henceforth call the domain of a Q-function as a **Q-domain**. Treating short answers as bare nominal, categorial approaches regard the relation between matrix questions and short answers as a simple function-argument relation. For example, in (4), applying the Q-function denoted by the question to an individual denoted by the short answer yields the inference that this individual came and the presupposition that this individual is a student.

(3) Which student came?



(4) Combining w. entity-denoting answers
$[\![\text{WH-Q}]\!]([\![\textit{Andy}]\!])$
$= (\lambda x_e\colon \textsf{stdt}_w(x).\hat{}\textsf{came}(x))(a)$
$= \textsf{stdt}_w(a).\hat{}\textsf{came}(a)$

(5) Combining w. GQ-denoting answers
$[\![\textit{Andy or Billy}]\!]([\![\text{WH-Q}]\!])$
$= (a^{\Uparrow} \sqcup b^{\Uparrow})(\lambda x_e\colon \textsf{stdt}_w(x).\hat{}\textsf{came}(x))$
$= \textsf{stdt}_w(a) \wedge \textsf{stdt}_w(b).\hat{}\textsf{came}(a) \cup \hat{}\textsf{came}(b)$

The above assumptions also automatically predict the fact seen in (1b) that GQs named by direct answers to a *wh*-question must quantify over a subset of the set denoted by the *wh*-complement in this question. Take (5) for example: since the disjunctive answer has a complex type ($\langle et, t\rangle$), the question-answer relation is flip-flopped into an argument-function relation. Applying the Boolean disjunction $a^{\Uparrow} \sqcup b^{\Uparrow}$ to the Q-function yields the presupposition that both of the disjoined individuals a and b are atomic students. (See (16) in section 2.2 for the definition of Boolean disjunctions.)

The above discussion is focused on readings where the Q-domain of a *wh*-question ranges over only entities, called **first-order readings**. Questions with first-order readings are subject to two restrictions regarding to their GQ-naming answers. First, the answer space (viz., the Hamblin set) of each such question consists of only individual-naming answers. If a propositional answer names a GQ, the derivation of this answer involves external Boolean operations applied to the entity-naming answers in the answer space. Second, the named GQs must be interpreted with wide scope relative to the scopal elements in the question-nucleus. For a simple illustration, consider (6) and assume that the answer space of this question consists of only propositions naming individuals that are atomic or plural students. With this assumption, the conjunction answer (6b) can be ruled into the answer space only if this conjunction is interpreted as a plural entity (of type e), not as a Boolean conjunction (of type $\langle et, t\rangle$). The GQ-naming answers (6b-c) involve external operations such as join and universal quantification, and the named GQs would be read with a wide scope relative to *might*.

(6) Which student or students might come?
   a. Andy.
   b. Andy and Billy.
   c. Andy or Billy.
   d. Every student.

However, as first noticed by Spector (2007, 2008), the above two predictions are overly restrictive for some *wh*-questions, which shows that these questions admit also **higher-order readings**. When having a higher-order reading, the Q-domain of a question ranges over higher-order meanings (of type $\langle et, t\rangle$) such as Boolean disjunctions and existential quantifiers. For example in (7), the elided disjunction is interpreted under the scope of the necessity modal. Spector argues that to obtain this narrow scope reading, *which books* should bind a higher-order trace across the necessity modal, so that a disjunction can be reconstructed to a scopal position under the necessity modal.

(7) a. Which books does John have to read?
   b. The French novels or the Russian novels. The choice is up to him. $\Box \gg$ *or*

This paper significantly expands on Spector's view and examines the semantics of *wh*-questions and derivation of higher-order readings. The rest sections address the following two questions:

A. What higher-order meanings are included in a higher-order Q-domain? (§3 and §4)

   A higher-order Q-domain of 'which-A f?', if it exists, consists of the positive GQs that range over a subset of A and the Boolean coordination compounds of such positive GQs.

B. What *wh*-questions admit higher-order readings, and how can we account for the distributional constraints of these readings? (§5 and §6)

   In general, higher-oder readings are unavailable in questions where the *wh*-phrase is singular-marked or numeral-modified (as in *which book does John have to read?* and *which two books does John have to read?*). To account for this distributional constraint, I argue that the derivation of a higher-order reading involves the application of a H-shifter to the root of the *wh*-complement, and that the atomicity constraint of singular nouns and the cardinality constraint of numerals block the application of the H-shifter.

   However, questions with a singular-marked or numeral-modified *wh*-phrase marginally admit narrow scope disjunctive answers, which shows that these questions can be interpreted with a higher-order reading but the Q-domain yielded in this reading has no conjunction. I will provide two ways to account for this 'conjunction-rejecting' reading in section 6.

The rest of this paper is organized as follows. Section 2 introduces the basics of Boolean coordinations and GQs. Section 3 presents evidence for cases where a question must be interpreted with a higher-order reading, drawn on facts about questions with modals or collective predicates. In particular, diagnostics based on *non-reducibility* rule in Boolean disjunctions and existential quantifiers, and diagnostics based on *stubborn distributivity* rule in Boolean conjunctions and universal quantifiers. Section 4 further shows that the higher-order Q-domain of a *wh*-question is subject to "The Positiveness Constraint", which says that only positive GQs and their coordination compounds can be included in a higher-order Q-domain. Sections 5 and 6 investigate into the compositional derivation of two types of higher-order readings and explain their distributional constraints. Section 7 concludes.

2. Coordinations and GQs

2.1. Pluralities

The ontology of individuals from Sharvy (1980) and Link (1983) assumes that both singular and plural terms denote sets of entities (of type $\langle e, t \rangle$). In particular, a singular term denotes a set of atomic elements, while a plural term denotes a set consisting of both atomic and sum elements.[2] If sums are defined in terms of part-hood relations ($\leq$) as in (8), the ontology of individuals can be represented with the mereological structure in Figure 1. Letters *abc* each denotes an atomic student. Lines indicate *part of* relations from the bottom to the top. For example, atomic entities a and b are parts of their sum $a \oplus b$.

(8) a. **Overlap**
$x \circ y =_{\text{df}} \exists z[z \leq x \wedge z \leq y]$
(x is overlapped with y if and only if they have a part in common.)

b. **Sum of a set**
For any non-empty set A:
$\bigoplus A =_{\text{df}} \iota x : \forall y[y \in A \rightarrow y \leq x] \wedge \forall z[z \leq x \rightarrow \exists z'[z' \in A \wedge z \circ z']]$
(For any non-empty set A, the sum of A is the unique x such that every member of A is a part of x and that every part of x is overlapped with a member of A.)

c. **Binary sum**
$x \oplus y =_{\text{df}} \bigoplus\{x, y\}$

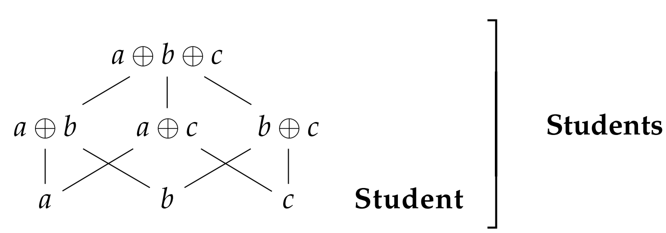


Figure 1: Ontology of individuals (Sharvy 1980, Link 1983)

Further, Link assumes that the extension of a plural term is obtained by applying a star (*)-operator to the extension of the corresponding singular term. This *-operator closes a set of entities under

[2]The view of treating plurals as sets ranging over not only sums but also atomic elements is called the "weak" theory of plurality (Sauerland et al. 2005, and among others), as opposed to the "strong" theory which defines plurals as denoting sets consisting of only non-atomic elements. Whether plurals are treated weak or strong is not crucial in this paper. The following presentation follows the weak theory.

merelogical sum.[3]

(9) **The *-operator** (Link 1983)
$^*A = \{x \mid \exists A' \subseteq A[x = \bigoplus A']\}$
($^*A$ is the set that contains any sum of things taken from A.)

2.2. Boolean coordinations

Conjunctions of entity-denoting expressions are semantically ambiguous. For example, the sentence in (10) is ambiguous between a 'one-team reading' and a 'two-teams reading'. In particular, the two-teams reading reduces a conjunction of two plural entities to a conjunction of two propositions. In the literature, this ambiguity has been thought of as a contrast between "collective" and "intersective" readings, or a contrast between "non-Boolean" and "Boolean" readings.

(10) The two boys and the two girls formed a team.
  a. One-team reading: 'the four boys and girls all together formed one team.'
  b. Two-teams reading: 'the two boys formed a team and the two girls formed a team.'

A simple way to capture the above ambiguity is to interpret the conjunctive *and* ambiguously as either a summation operator ($\oplus$) or a meet operator ($\sqcap$) (*pace* Link 1983; Hoeksema 1988; among others).[4] When interpreted as a summation operator, *and* combines with two entities and yields a complex sum entity, as computed in (11). The collective predicate *formed a team* then combines with the entire complex sum entity, yielding the one-team reading.

(11) $$\begin{aligned}[\![\textit{the two boys and the two girls}]\!] &= [\![\textit{the two boys}]\!] \oplus [\![\textit{the two girls}]\!] \\ &= (b_1 \oplus b_2) \oplus (g_1 \oplus g_2) \\ &= b_1 \oplus b_2 \oplus g_1 \oplus g_2\end{aligned}$$

[3] Another representative view is to model plural individuals as sets (Landman 1989; Schwarzschild 1996; Winter 2001; among others). In this view, a singular term ranges over atomic elements or singleton sets of these atomic elements, while a plural term ranges over all non-empty sets recursively formed out of these atomic elements. The *-operator used to form pluralities is alternatively defined as in (ia). The sum-based view and the set-based view are mostly interchangeable. The main difference is that recursive application of the set-formation *-operator yields non-flat structures, as in (ib). For example, a plurality of pluralities ***student* (of type $\langle\langle et, t\rangle, t\rangle$) ranges over non-empty subsets of **student*. In a discourse domain with two students a and b, the plural term *students* has a simple meaning $\{\{a\}, \{b\}, \{a, b\}\}$ of type $\langle et, t\rangle$ and a complex meaning $\{\{\{a\}\}, \{\{b\}\}, \{\{a, b\}\}, \{\{a\}, \{b\}\}, ...\}$ of type $\langle ett, t\rangle$.

(i) a. $^*A = \textsc{pow}(A) - \{\varnothing\}$
  ($^*A$ is the powerset of A excluding the emptyset.)
 b. $^{**}A = \textsc{pow}(^*A) - \{\varnothing\}$
  ($^{**}A$ is the powerset of $^*A$ excluding the emptyset.)

The non-flat structure has similar consequences as a Boolean conjunction and is especially helpful in analyzing collective statements. (See ...) Despite of this advantage, this non-flat structure isn't more helpful in interpreting GQ-like disjunctions. For example, the translation of *ab or cd* would still involve an extra disjunctive (as in $\lambda P[P(\{a, b\}) \vee P(\{c, d\})]$) or a choice function (as in $f_{\text{CH}}(\{\{a, b\}, \{c, d\}\})$). Hence, I follow the sum-based approach and treat higher-order conjunctions and disjunctions uniformly Boolean.

[4] Another approach is to assign the conjunctive *and* a single meaning but ascribe the ambiguity to covert operations. For example, Champollion (2016) treats *and* unambiguously an intersection operation, and uses covert type-shifting operations to derive the collective/non-Boolean reading.

The meet operator must be applied to meanings of the same conjoinable type. Conjoinable types are roughly types of the form '$\langle ...t\rangle$', as defined recursively in (12). An inductive definition of meet is given in (13). Since entities are not of a conjoinable type, to be conjoined with meet, they have to be first type-shifted into GQs of a conjoinable type $\langle et, t\rangle$) via Montague-lift. Hence, in a coordination of two definite DPs, the meet-denoting *and* combines with two Montagovian individuals and returns their Boolean conjunction (Keenan and Faltz 1984: Part 1A).

(12) **Conjoinable types** (Partee and Rooth 1983)

a. t is a conjoinable type;

b. if σ is a conjoinable type, then for any type τ, $\langle \tau, \sigma\rangle$ is a conjoinable type;

c. nothing else is a conjoinable type.

(13) **Binary meet** (Partee and Rooth 1983, Groenendijk and Stokhof 1989)

$$A \sqcap B =_{\text{df}} \begin{cases} A \wedge B & \text{if } A \text{ and } B \text{ are of type } t \\ \lambda x_\tau [A(x) \sqcap B(x)] & \text{if } A \text{ and } B \text{ of a relational conjoinable type } \langle \tau, \sigma\rangle \\ \text{undefined} & \text{otherwise} \end{cases}$$

(14) **Montague lift** (Partee and Rooth 1983)

For any meaning α of type τ, the Montague-lifted meaning is $\alpha^{\Uparrow}$ such that $\alpha^{\Uparrow}$ is of type $\langle\langle \tau, t\rangle, t\rangle$ and $\alpha^{\Uparrow} =_{\text{df}} \lambda m_{\langle \tau, t\rangle}.m(\alpha)$.

The Boolean reading of a DP-disjunction is computed as in (15). *The two boys* is interpreted as a set of predicates held of the plurality of the two boys $b_1 \oplus b_2$, and the entire conjunction is the set of predicates held of both the plurality of the two boys and the plurality of the two girls.

(15)
$$\begin{aligned} [\![\textit{the two boys and the two girls}]\!] &= [\![\textit{the two boys}]\!]^{\Uparrow} \sqcap [\![\textit{the two girls}]\!]^{\Uparrow} \\ &= (\lambda P'.P'(b_1 \oplus b_2)) \sqcap (\lambda P'.P'(g_1 \oplus g_2)) \\ &= \lambda P[(\lambda P'.P'(b_1 \oplus b_2))(P) \sqcap (\lambda P'.P'(g_1 \oplus g_2))(P)] \\ &= \lambda P[P(b_1 \oplus b_2) \sqcap P(g_1 \oplus g_2)] \\ &= \lambda P[P(b_1 \oplus b_2) \wedge P(g_1 \oplus g_2)] \end{aligned}$$

On a par with the meet use of *and* in deriving Boolean conjunctions, coordinating two entity-denoting expressions, the disjunctive *or* functions as a join operator ($\sqcup$) over Montagovian individuals, yielding a Boolean disjunction.

(16) **Binary join** (Partee and Rooth 1983, Groenendijk and Stokhof 1989)

$$A \sqcup B =_{\text{df}} \begin{cases} A \vee B & \text{if } A \text{ and } B \text{ are of type } t \\ \lambda x_\tau [A(x) \sqcup B(x)] & \text{if } A \text{ and } B \text{ are of a relational conjoinable type } \langle \tau, \sigma\rangle \\ \text{undefined} & \text{otherwise} \end{cases}$$

2.3. Quantificational determiners and quantified NPs

Most GQs can be decomposed into a quantificational determiner and a set-denoting term. There is a rich class of quantificational determiners in natural languages. In English, for example, quantificational determiners non-exclusively include those in (17) as well as their Boolean combinations.

(17) a. Aristotelian: *all*, *every*, *no*, *some*
b. Proportional: *most*, *at least half*, *10 percent of the*, *less than two-thirds of the*
c. Numerical: *at least two*, *less than ten*, *between six and nine*, *finitely many*, *an odd number of*
d. Exceptive: *no ... but John*, *every ... except Mary*

All of these quantificational determiners can be interpreted extensionally as relations between sets of individuals in the discourse domain. In particular, the determiners in (17a-d) express relations between two sets and are often called "type $\langle 1, 1\rangle$ quantifiers" ('1' stands for 1-ary relation). In (18), the left argument A and the right argument B serve as the restriction and the scope of the quantifier, respectively.[5] In particular, in (18d-e), following Gajewski (2008), I treat exceptives as involving subtraction from the restriction.

(18) a. $[\![every]\!](A, B) =_{\text{df}} A \subseteq B$
b. $[\![most]\!](A, B) =_{\text{df}} |A \cap B| > |A - B|$
c. $[\![at\ least\ two]\!](A, B) =_{\text{df}} |A \cap B| \geq 2$
d. $[\![no\ ...\ but\ John]\!](A, B) =_{\text{df}} (A - \{j\}) \cap B = \varnothing$
e. $[\![every\ ...\ except\ Mary]\!](A, B) =_{\text{df}} (A - \{m\}) \subseteq B$

Not every binary relation between sets of individuals can be lexicalized into determiners. Strikingly, all of the type $\langle 1, 1\rangle$ quantifiers in natural languages are conservative (Barwise and Cooper 1981; Higginbotham and May 1981; Keenan and Stavi 1986), defined as in (19a). Conservativity says that the part of the scope B that's not in the restriction A does not matter for whether $Q(A, B)$ holds. Barwise and Cooper call this property "the live-on property" and coin the notion "live-on sets", as defined in (19b).[6]

(19) **Conservativity and the live-on property**
a. A type $\langle 1, 1\rangle$ quantifier Q is conservative iff for every A and B: $Q(A, B) \Leftrightarrow Q(A, A \cap B)$.
b. A generalized quantifier π lives on a set A iff for every B: $\pi(B) \Leftrightarrow \pi(B \cap A)$.

Relatedly, a GQ is said to **range over** a set A if and only if A is the **smallest live-on set** of this GQ (Szabolcsi 1997), as illustrated in Table 1. This notion will be crucial for later discussions on defining what types of GQs should and shouldn't be ruled into a higher-order Q-domian (see the Positiveness Constraint in section 4).

[5]Here and throughout the paper, interpretations of quantifiers have an implicit requirement all of its arguments are subsets of the discourse domain.

[6]Consider *not every* and *every not* for a comparison. They both can be defined as a binary relation between two sets, as defined in (i). ($\overline{A}$ stands for the complement of A' relative to the discourse universe.)

(i) a. $[\![not\ every]\!](A, B) =_{\text{df}} A \not\subseteq B$
b. $[\![every\ not]\!](A, B) =_{\text{df}} \overline{A} \subseteq B$

However, *not every* can function as a determiner, while *every not* cannot; the best way to state the relation *every not* in English is 'everyone who isn't ___'. Conservativity captures this contrast, as shown in (ii).

(ii) a. Not every student arrived. $\Leftrightarrow$ Not every student is a student who arrived.
b. Everyone who isn't a student arrived. $\not\Leftrightarrow$ Everyone who isn't a student is a student who arrived.

π	SMLO(π)
$a^{\Uparrow}$	$\{a\}$
$a^{\Uparrow} \sqcup b^{\Uparrow}$, $a^{\Uparrow} \sqcap b^{\Uparrow}$	$\{a, b\}$
some/every/no student	*stdt*
some/two/most students	**stdt*
every student but John	*stdt* $- \{j\}$
no student except John	*stdt* $- \{j\}$

Table 1: GQs and their smallest live-on sets (SMLO)

As seen from Table 1, in most cases, the smallest live-on set of a GQ is simply the set denoted by the NP-restrictor of the determiner. For example, the smallest live-on set of *some/every/no student* is the set of atomic students *stdt*. Montagovian individuals and their Boolean coordinations can be reduced to a basic quantificational determiner and a set of individuals, as defined in (20).[7]

(20) For any two entities a and b in the discourse domain, we have:

a. $a^{\Uparrow} = \{P \mid \{a\} \subseteq P\} = [\![every]\!](\{a\}) = [\![some]\!](\{a\})$

b. $a^{\Uparrow} \sqcap b^{\Uparrow} = [\![every]\!](\{a, b\})$

c. $a^{\Uparrow} \sqcup b^{\Uparrow} = [\![some]\!](\{a, b\})$

The case of exceptives is a bit more complex: the smallest live-on set of $[\![$no/every-NP-except-$x_e]\!]$ is not $[\![$NP$]\!]$, but rather $[\![$NP$]\!] - \{y \mid y \circ x\}$. An intuitive illustration is given in (21), where 'non-John student' stands for the set of atomic students excluding John.

(21) a. Every student except John arrived.
⇔ Every student except John is a non-John student who arrived.

b. No student but John arrived.
⇔ No student but John is a non-John student who arrived.

3. Evidence for a higher-order Q-domain

This section provides empirical evidence for cases where a question must be interpreted with a higher-order reading. To get started, recall from section 1 that questions with first-order readings are subject to two constraints regarding to their GQ-naming answers: first, the named GQs must be interpreted with wide scope relative to the scopal expressions in the question nucleus; second,

[7]Note that the view of defining Boolean coordination in terms of quantification is not compositional: the connectives *and* and *or* themselves can not be viewed as quantificational determiners. Quantificational determiners (of type $\langle et, ett\rangle$) combine with a set of non-Montagovian individuals and return Boolean compounds over the corresponding Montagovian individuals. In contrast, the connectives in an NP-coordination as interpreted as Boolean operations of type $\langle (ett, ett), ett\rangle$, they combine with a sequence of Montagovian individuals and return a Boolean compound of these Montagovian individuals. If the connectives are defined as combining with non-Montagovian individuals, as in (i), the connectives would be lexically encoded with Montague-lift. In consequence, iterated application of conjunction/disjunction would yield iterated Montague-lift. For example, for two non-Montagovian individuals a and b, $a \bar{\wedge} b$ and $a \bar{\vee} b$ are of type $\langle et, t\rangle$ as a regular GQ, while a more complex compound like $(a \bar{\wedge} b) \bar{\vee} (a \bar{\vee} b)$ is of a more complex type $\langle ett, \langle ett, t\rangle\rangle$.

(i) a. $\bar{\wedge} = \lambda a \lambda b. a^{\Uparrow} \sqcap b^{\Uparrow}$

b. $\bar{\vee} = \lambda a \lambda b. a^{\Uparrow} \sqcup b^{\Uparrow}$

the Hamblin sets of these questions consist of only propositions naming entities. In the following, I provide counterexamples to these two generalizations with two diagnostics. In particular, a diagnostic based on *non-reducibility* relative to narrow scope GQs shows that *wh*-questions may expect answers naming Boolean disjunctions or existential GQs (§3.1). A diagnostic based on *stubborn collectivity* shows that the Hamblin sets of some questions must contain propositions naming Boolean conjunctions or universal GQs (§3.2). Finally, combinations of these two diagnostics rule in the Boolean coordinations of the aforementioned GQs. (§3.3)

3.1. Diagnostic based on non-reducibility

In general, to completely address a question, one must provide the strongest true answer (Dayal 1996). Hence, for an answer to be possibly complete, there must be a world in which this answer is the strongest true answer. As such, as in (22), to a basic *wh*-question, a disjunctive answer is always incomplete — whenever the disjunctive answer is true, it is asymmetrically entailed by another true answer, namely, one of its disjuncts.

(22) a. Which books did John read?
b. The French novels or the Russian novels.

However, Spector (2007, 2008) observes that disjunctions can completely address constituent questions where the nucleus has a necessity modal (called "□-questions" henceforth). For example in (23), the elided disjunction is scopally ambiguous. If the disjunction takes scope over the necessity modal, the disjunctive answer has a partial answer reading. Alternatively, the elided disjunction can also be regarded as a complete specification of John's reading obligation — there isn't any specific book that John has to read, John's only reading obligation is to choose between the French novels and the Russian novels.

(23) a. Which books does John have to read?
b. The French novels or the Russian novels.
√ 'John has to read F or R, I don't know which exactly.' (Partial: *or* $\gg$ □)
√ 'John has to read F or R, and the choice is up to him.' (Complete: □ $\gg$ *or*)

To obtain this complete answer reading, the disjunction must be treated as a GQ and be interpreted under the scope of the necessity modal. Thus, Spector (2007) concludes that the question (23a) is ambiguous between a *high reading* and a *low reading* where "high" and "low" mean that the scope of the disjunction is wide and narrow, respectively. To highlight the contrast between these two readings with respect to the type of the variable(s) bound by the *wh*-phrase, I instead call the two readings the *first-order reading* and the *higher-order reading*, respectively. As paraphrased in (24), the first-order reading expects answers that specify an entity, while the higher-order reading expects answers that specify a GQ.

(24) Which books does John have to read?
a. First-order reading: 'What is a book or books x s.t. John has to read x?'
b. Higher-order reading: 'What is a GQ π over books s.t. John has to read π?'

For any theory that assumes *wh*-movement and *wh*-binding, the higher-order reading arises only if the *wh*-phrase binds a higher-order variable of type $\langle et, t\rangle$ inside the question nucleus. This binding relation can be realized by **semantic reconstruction** (Cresti 1995; Rullmann 1995): the movement of the *wh*-phrase creates an individual trace x (of type e) and a higher-order trace π (of type $\langle et, t\rangle$), and then the compositional interpretation assigns the *wh*-phrase the logical scope corresponding to the site of π, yielding a reconstructed reading.[8] The structures, question denotations, and computations for question-answer pairs following categorial approaches are as follows. (The definitions for the higher-order Q-function and Q-domain are subject to revision. For now, I simply assume that the Q-domain is the set of GQs ranging over a subset of books.)

(25) First-order reading

a. LF: [[... which books] λx [have-to [John read x_e]]]

b. $[\![\text{WH-Q}]\!] = \lambda x\colon \textit{*book}_w(x).\Box[\lambda w.\textit{read}_w(j,x)]$

c. $[\![F \textit{ or } R]\!]([\![\text{WH-Q}]\!])$
$= (f^{\Uparrow} \sqcup r^{\Uparrow})(\lambda x\colon \textit{*book}_w(x).\Box[\lambda w.\textit{read}_w(j,x)])$
$= \textit{*book}_w(f) \wedge \textit{*book}_w(r).\Box[\lambda w.\textit{read}_w(j,f)] \cup \Box[\lambda w.\textit{read}_w(j,r)]$

(26) Higher-order reading

a. LF: [[... which books] $\lambda\pi$ [have-to [$\pi_{\langle et,t\rangle}$ λx [John read x_e]]]]

b. $[\![\text{WH-Q}]\!] = \lambda\pi_{\langle et,t\rangle}.\text{SMLO}(\pi) \subseteq \textit{*book}.\Box[\lambda w.\pi(\lambda x.\textit{read}_w(j,x))]$ (To be revised)

c. $[\![\text{WH-Q}]\!]([\![F \textit{ or } R]\!])$
$= (\lambda\pi_{\langle et,t\rangle}.\text{SMLO}(\pi) \subseteq \textit{*book}.\Box[\lambda w.\pi(\lambda x.\textit{read}_w(j,x))])(f^{\Uparrow} \sqcup r^{\Uparrow})$
$= \text{SMLO}(f^{\Uparrow} \sqcup r^{\Uparrow}) \subseteq \textit{*book}.\Box[\lambda w.(f^{\Uparrow} \sqcup r^{\Uparrow})(\lambda x.\textit{read}_w(j,x))]$
$= \{f, r\} \subseteq \textit{*book}.\Box[\lambda w.\textit{read}_w(j,f) \vee \textit{read}_w(j,r)]$

$\Box$-questions are useful in validating the existence of Boolean disjunctions in a Q-domain because the answer space of a $\Box$-question is not closed under disjunction.

(27) A proposition set Q is **closed under disjunction** iff for any two propositions p and q, if both p and q are members of Q, then the disjunction $p \vee q$ is also a member of Q.

The following figures illustrate the answer space of a plain episodic question and that of a $\Box$-question. Arrows indicate entailments. $f(x)$ abbreviates for the proposition *John read x*.

[8] There are other ways to derive the higher-order reading compositionally. In (i), for example, applying object type-raising to the transitive verb *read* allows the verb to combine with a higher-order variable. In this way, the *wh*-phrase can be moved directly from the base position.

(i) LF: [[... which books] $\lambda\pi$ [[$_{\text{IP}}$ has-to [$_{\text{VP}}$ John read$^{\Uparrow\text{OBJ}}$ $\pi_{\langle et,t\rangle}$]]]]

(ii) **Object Type-Raising** (Hendriks 1993)
If R is of type $\langle e, et\rangle$, then $R^{\Uparrow\text{OBJ}} = \lambda\pi_{\langle et,t\rangle}\lambda x_e.\pi(\lambda y_e.R(x,y))$

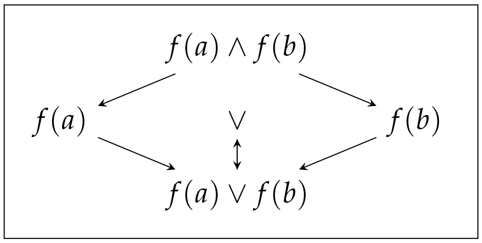


Figure 2: Answer space for 'what did John read?'

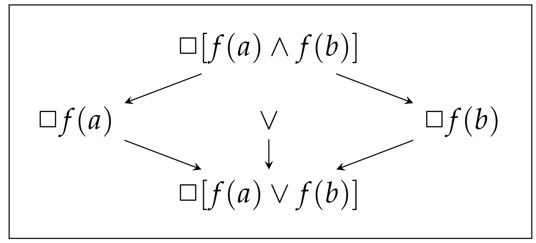


Figure 3: Answer space for 'what does John have to read?'

In Figure 2, the disjunctive answer $f(a) \vee f(b)$ is semantically equivalent to the disjunction of the two individual answers $f(a)$ and $f(b)$. Hence, the disjunctive answer can never be the strongest true answer of the question — whenever the disjunctive answer is true, there will be another true answer, $f(a)$ or $f(b)$, asymmetrically entailing it. In contrast, in Figure 3, the disjunctive answer $\Box[f(a) \vee f(b)]$ can be the strongest true answer since it is semantically weaker than the disjunction of the two individual answers. If John's only reading obligation is to choose between a and b, the individual answers are false, and the disjunctive answer is the unique true answer and hence the strongest true answer.

More generally, Spector's argumentation for ruling in Boolean disjunctions can be made based on any *wh*-question whose Q-function is not reducible relative to disjunctions. The following defines reducibility relative to disjunctions:[9]

(28) **Reducibility relative to Boolean disjunctions**
A function θ is reducible relative to Boolean disjunctions iff for any two entities a and b such that θ is defined for a and b or defined for $a^{\Uparrow}$ and $b^{\Uparrow}$: $\theta \bullet (a^{\Uparrow} \sqcup b^{\Uparrow}) \equiv (\theta \bullet a^{\Uparrow}) \sqcup (\theta \bullet b^{\Uparrow})$.

It's easy to see that functions taking non-scopal arguments are reducible relative to disjunctions.

(29) a. John read [Book A or Book B].
$\equiv$ John read Book A or John read Book B.
b. John has to read [Book A or Book B].
$\not\equiv$ John has to read Book A or John has to read Book B.

The non-reducibility diagnostic can also be made based on the following questions ((30) and (31) are taken from Spector (2007)). The same as $\Box$-questions, thanks to the presence of an attitude verb, a modal verb, or a non-existential quantifier, those questions have readings where the Q-function is non-reducible relative to disjunctions or to existential quantifiers.

(30) Attitude verbs
a. Which books did John *demand* that we read?
b. Which books is John *certain* that Mary read?
c. Which books does John *expect* Mary to read?

[9] In (28), '$\bullet$' stands for the combinatory operation between θ and a GQ. If θ is of type $\langle ett, t\rangle$, '$\bullet$' stands for left-to-right functional application; if θ is of type $\langle e, t\rangle$, '$\bullet$' stands for right-to-left functional application. If θ cannot compose with a GQ directly, '$\bullet$' would involve a type-shifting operation.

(31) Modals

a. Which books is it *sufficient* to read?

b. Which books is John *required* to read?

(32) Quantifiers

a. Which books did *most*/*all* of the students read?

b. Which books does John *always*/*usually* read?

The non-reducibility diagnostic for Boolean disjunctions easily extends to other GQs, as generalized in (33). Q-functions of □-questions are non-reducible relative to many other existential GQs, as exemplified in (34).[10]

(33) **Reducibility relative to GQs**
A function θ is reducible relative to a GQ π if and only if $\theta \bullet \pi \equiv \pi(\lambda x.\theta \bullet x^{\Uparrow})$

(34) John must read $\left\{\begin{array}{c}\text{at least two}\\ \text{more than two}\\ \text{exactly two}\end{array}\right\}$ books by Balzac.
$\not\equiv$ There are $\left\{\begin{array}{c}\text{at least two}\\ \text{more than two}\\ \text{exactly two}\end{array}\right\}$ books by Balzac that John must read.

The same as disjunctions, the above GQs can completely address a □-question. Hence, those GQs should likewise be ruled into a higher-order Q-domain.

(35) a. Which books does John have to read?

b. $\left\{\begin{array}{c}\text{At least two}\\ \text{More than two}\\ \text{Exactly two}\end{array}\right\}$ books by Balzac. ($^{\text{OK}}$partial: $\exists \gg \Box$, $^{\text{OK}}$complete: $\Box \gg \exists$)

3.2. Diagnostic based on stubborn collectivity

Spector (2007, 2008) and Fox (2013) have assumed that a higher-order Q-domain contains also Boolean conjunctions, but they have not found empirical evidence for this assumption. For two reasons, Spector's non-reducibility diagnostic does not extend to Boolean conjunctions. First, the Q-functions of □-questions as well as those discussed in (30) to (32) are reducible relative to conjunctions, as exemplified in (36).

(36) John must read the French novels and the Russian novels.
$\equiv$ John must read the French novels, and John must read the Russian novels.

[10] For now, I consider only non-decreasing GQs that can be decomposed into a type $\langle 1, 1\rangle$ determiner (e.g., *at least two*) and a set term (e.g., *books by Balzac*). Other GQs, such as decreasing GQs (e.g. *at most two books by Balzac*) and Boolean compounds of GQs (e.g. *at least two books by Balzac and no book by Shakespeare*), are more complex and will be discussed separately in section 4.

Second, conjunctions are ambiguous between Boolean and non-Boolean interpretations. For example, the conjunction of two individuals can be interpreted either as a GQ-like Boolean conjunction as in (37a), or a plural entity as in (37b). Thus, although some scopal items (e.g., the attitude verb *wonder*) yield functions that are non-reducible relative conjunctions as seen in (38), the seeming narrow scope reading of a conjunction might simply be a scope-less reading of a plural entity.

(37) ⟦*Andy and Billy*⟧ can be interpreted as a or b:
   a. $a^{\Uparrow} \sqcap b^{\Uparrow}$ the meet of Montagovian individuals
   b. $a \oplus b$ the sum of non-Montagovian individuals

(38) Mary is worried that Andy and Billy will come to the party. (They make a trouble whenever they stay together.) $\not\rightsquigarrow$ *Mary is worried that Andy will come to the party.*

Alternatively, this section provides a new diagnostic to rule in Boolean conjunctions to a higher-order Q-domain, drawn on facts that questions with a quantized collective predicate (e.g., *form a team*, *co-authored two papers*) are not necessarily subject to uniqueness and that these questions admit answers naming Boolean conjunctions. I will show that quantized collective predicates are "stubbornly collective" and hence say that this diagnostic is based on stubborn collectivity.

First, for stubborn collectivity, observe that the phrasal predicate *formed a team* admits a collective reading but not a (non-atomic) distributive reading. The sentence (39a) cannot be truthfully uttered in the given context, because it admits only a collective reading and this reading is false in the given scenario. In contrast, the plural counterpart *formed teams* admits a cumulative/ (non-atomic) distributive reading and thus (39b) can be truthfully uttered.

(39) (w: *The kids abcd formed exactly two teams in total:* $a + b$ *formed one, and* $c + d$ *formed one.*)
   a. # The kids formed a team.
   b. $\surd$ The kids formed teams.

Note that the falsehood of (39a) is not improved even if the linguistic or extra-linguistic context has explicitly separated the four kids into two pairs, as seen in the following:[11]

(40) [Yesterday, the pair *ab* competed against the pair *cd*.] Today, the kids (all) formed a team.
(^{OK}collective, #non-atomic distributive, #atomic distributive)

[11] The falsehood of (39a) is a challenge to the cover-based theory of distributivity of Schwarzschild (1996), which predicts that every verb phrase, with enough supporting context, admits a non-atomic distributive reading. Assuming a generalized distributor PART which distributes over a contextually determined cover variable *Cov*. The derived reading can be atomic distributive, non-atomic distributive, or collective, depending on the value of the cover. For example, with four relevant kids *abcd*, the sentence (i) is ambiguous between three readings: a collective reading that all the kids together participated in a single piano-lifting event (if $Cov = \{a \oplus b \oplus c \oplus d\}$), a distributive reading that each of the kids lifted the piano (if $Cov = \{a, b, c, d\}$), and a non-atomic distributive reading that the kids are separated into several possibly overlapped subgroups and that each subgroup of kids participated in a piano-lifting event (for example, if $Cov = \{a \oplus b, c \oplus d\}$).

(i) The kids lifted the piano.
   a. LF: [[The kids] PART_{Cov} [lifted the piano]]
   b. $\forall x[x \in Cov \rightarrow \textsf{lifted-the-piano}(x)]$ where Cov is a cover of *the-kids*.

Hence, I call *formed a team* a "stubbornly collective predicate", in parallel to what Schwarzschild (2011) calls "stubbornly distributive predicates" (e.g., *be intelligent*, *have blue eyes*) which admit only atomic distributive readings. Stubborn collectivity is also observed with quantized phrasal predicates of the form "V + counting noun", such as *formed one committee* and *co-authored two papers*.

(41) A predicate P is **quantized** if and only if (Krifka 1997)
$\forall x \forall y[P(x) \wedge P(y) \rightarrow [x \leq y \rightarrow x = y]]$
(Whenever x is in P, no proper part of x is also in P.)

Second, for the absence of uniqueness, compare the sentences in (42a-b) in the same discourse. The declarative-embedding sentence (42a) suffers presupposition failure, because the factive verb *know* embeds a false collective declarative. But (42b), where *know* embeds an interrogative counterpart of this collective declarative, does not suffer presupposition failure. Moreover, intuitively, (42b) expresses that John knows precisely the component members of all the teams formed by the considered kids, which is a conjunctive inference.

(42) (w: *The kids abcd formed exactly two teams in total:* $a + b$ *formed one, and* $c + d$ *formed one.*)

a. # John knows [that **the kids** formed a team].

b. $\surd$ John knows [**which kids** formed a team].

c. $\rightsquigarrow$ *John knows that* $a + b$ *formed a team* <u>*and*</u> $c + d$ *formed a team.*

The conjunctive inference (42c) is striking — where does the conjunctive closure come from? Clearly, no matter how we analyze collectivity, this conjunctive closure cannot come from the predicate *formed a team* or anywhere within the question nucleus, otherwise the embedded clause in (42a) would admit a non-atomic distributive/ covered reading and (42a) would be felicitous. In contrast, I argue that this conjunctive closure is provided by the *wh*-phrase: the *wh*-phrase binds a higher-order trace and quantifies over higher-order meanings including the Boolean conjunctions $(a \oplus b)^{\Uparrow} \sqcap (c \oplus d)^{\Uparrow}$.

(43) Which kids formed a team?
Higher-order reading: 'For which GQ π over kids is such that π formed a team?'

a. Logical Form
[[... which kids] $\lambda\pi$ [$_{\text{IP}}$ $\pi_{\langle et,t\rangle}$ λx [$_{\text{VP}}$ x_e formed a team]]]

b. Q-function (domain to be revised)
$[\![\text{WH-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \text{SMLO}(\pi) \subseteq \textit{*kid}.\hat{}\pi(\lambda x.\textit{form-a-team}(x))$

c. Combining with a Boolean conjunction
$[\![\text{WH-Q}]\!]((a \oplus b)^{\Uparrow} \sqcap (c \oplus d)^{\Uparrow})$
$= \{a \oplus b, c \oplus d\} \subseteq \textit{*kid}.\lambda w[\textit{form-a-team}_w(a \oplus b) \wedge \textit{form-a-team}_w(c \oplus d)]$

One might suggest to ascribe the conjunctive closure to an operator outside the question denotation, such as Heim's (1994) answerhood-operator. As schematized in (44), this operator contains a $\bigcap$-closure, it applies to an evaluation world w and a Hamblin set Q and returns the conjunction of all the propositions in Q that are true in w.

(44) a. $\text{Ans-H}(w)(Q) = \bigcap\{p \mid w \in p \in Q\}$

b. $\bigcap\{\hat{}\textit{f.a.team}(a \oplus b), \hat{}\textit{f.a.team}(c \oplus d)\} = \hat{}\textit{f.a.team}(a \oplus b) \cap \hat{}\textit{f.a.team}(c \oplus d)$

However, this definition of answerhood cannot capture the contrast in (45). The question-embedding sentence (45b) is infelicitous because the embedded numeral-modified question (viz., the question in which the *wh*-complement is numeral-modified) has a uniqueness presupposition which contradicts the context.

(45) (w: *The kids abcd formed two teams in total:* $a+b$ *formed one, and* $c+d$ *formed one.*)
  a. $\surd$ John knows [**which kids** formed a team].
  b. # John knows [**which two kids** formed a team].
     $\rightsquigarrow$ *Only two of the kids formed any team.*

Uniqueness presuppositions in *wh*-questions are standardly explained by "Dayal's presupposition" — a question is defined only if it has a strongest true answer (Dayal 1996). In the rest of this section, I argue that the contrast between (45a-b) is due to the following: in (45a), the embedded simple plural-marked question has a strongest true answer in the given discourse, while (45b), the embedded numeral-modified question does not.

Dayal's presupposition was originally motivated to explain the uniqueness requirement of singular-marked *wh*-questions. In Srivastav 1991, she observes that a singular-marked *wh*-question (viz., a question in which the *wh*-complement is singular-marked) can have only one true answer. Compare the examples in (46) for illustration. The continuation in (46a) is infelicitous because the singular-marked question has a uniqueness presupposition that only one of the kids came, which is inconsistent with the second clause. By contrast, this inconsistency disappears if the singular *wh*-phrase *which kid* is replaced with a plural phrase *which kids* or a bare *wh*-word *who*, as seen in (46b-c).

(46) a. "Which kid came? # I heard that many kids did."
  b. "Which kids came? I heard that many kids did."
  c. "[Among the kids,] who came? I heard that many kids did."

To capture the uniqueness presuppositions of singular-marked questions, Dayal (1996) defines a presuppositional answerhood-operator that checks the existence of the strongest true answer, as schematized in (47). Applying this Ans-D-operator returns the unique strongest of the propositions in Q true in w and presupposes the existence of this strongest true proposition.

(47) $$\textsc{Ans}\text{-D}(w)(Q) = \exists p[w \in p \in Q \wedge \forall q[w \in q \in Q \rightarrow p \subseteq q]].\ \iota p[w \in p \in Q \wedge \forall q[w \in q \in Q \rightarrow p \subseteq q]]$$
  ($\textsc{Ans}\text{-D}(w)(Q)$ is defined only if the set of answers in Q that are true in w has a strongest member; when defined, $\textsc{Ans}\text{-D}(w)(Q)$ returns this unique strongest true answer.)

Adopting the ontology of individuals by Sharvy (1980) and Link (1983) (see section 2.1), Dayal assumes that the Hamblin set of a singular-marked *wh*-question is smaller than the one of its plural-marked counterpart. As illustrated in (48), the former Hamblin set includes only propositions naming an atomic kid, while the latter also includes propositions naming a sum of kids. For easier illustration, the set of proposition that are true in w is abbreviated as Q_w. As a consequence, under a discourse where both Andy and Bill came, (48b) has a strongest true answer $\hat{}\mathit{came}(a \oplus b)$ while (48a) does not. Then employing Ans-D in (48a) gives rise to a presupposition failure. To avoid this presupposition failure, (48a) can only be uttered in a world where only one of the kids came, which therefore explains its uniqueness requirement.

(48) (w: *Among the considered kids, only Andy and Billy came.*)

a. Which kid came?
$Q = \{{}^\wedge\textsf{came}(x) \mid x \in \textsf{kid}_w\}$
$Q_w = \{{}^\wedge\textsf{came}(a), {}^\wedge\textsf{came}(b)\}$
$\textsc{Ans}\text{-D}(w)(Q)$ is undefined

b. Which kids came?
$Q = \{{}^\wedge\textsf{came}(x) \mid x \in {}^*\textsf{kid}_w\}$
$Q_w = \{{}^\wedge\textsf{came}(a), {}^\wedge\textsf{came}(b), {}^\wedge\textsf{came}(a \oplus b)\}$
$\textsc{Ans}\text{-D}(w)(Q) = {}^\wedge\textsf{came}(a \oplus b)$

It is also straightforward to see that, to account for the uniqueness presupposition, the Q-domain yielded by a singular-marked *wh*-phrase must exclude Boolean conjunctions such as $a^{\Uparrow} \sqcap b^{\Uparrow}$. Otherwise, the singular-marked question (48a) would admit conjunctive answers like ${}^\wedge\textsf{came}(a) \cap {}^\wedge\textsf{came}(b)$ and would not be subject to uniqueness, contra fact.[12]

A numeral-modified question also has a uniqueness presupposition. For example, the numeral-modified question in (49a) implies that only two of the kids came, and the one in (49b) implies that only two or three of the kids came. Both inferences contradict each of their follow-up clauses.

(49) a. 'Which two kids came? # I heard that three kids did.'

b. 'Which two or three kids came? # I heard that five kids did.'

Dayal's account of uniqueness easily extends to numeral-modified questions. As seen in (50), for a question of the form "which n students came?" where n is a bare numeral and is read as 'exactly n', Dayal's presupposition is satisfied only if exactly n students came. If the number of students who came is smaller than n, this question has no true answer (viz., $Q_w = \emptyset$); if the number of the students who came is larger than n, the question does not have a strongest true answer.

(50) (w: *Among the considered kids, only Andy, Billy, and Clark came.*)
Which two kids came?

a. $Q = \{{}^\wedge\textsf{came}(x) \mid x \in \textsf{2-kids}_w\}$

b. $Q_w = \{{}^\wedge\textsf{came}(a \oplus b), {}^\wedge\textsf{came}(a \oplus c), {}^\wedge\textsf{came}(b \oplus c)\}$

c. $\textsc{Ans}\text{-D}(w)(Q)$ is undefined

[12] Drawn on facts from Spanish *quién* 'who.SG' which is singular-marked but does not trigger uniqueness (Maldonado 2017), Elliott et al. (2017) by contrast propose that *quién*-questions admit also higher-order readings, in which the yielded Q-domain ranges over a set of Boolean conjunctions over atomic elements. Alonso-Ovalle and Rouillard (2018) argue against this view, drawn on facts of questions with a collective predicate. As seen in (i), *quién* 'who.SG' can be used to combine with a stubbornly collective predicate *formó un grupo* 'formed.SG a group', and the formed question expects to specify the component members of one or more groups.

(i) Quién formó un grupo?
who.SG formed.SG a group
'Who formed a group?'

a. Los estudiantes.
the students

b. Los estudiantes y los profesores.
the students and the professors.

The felicity of answer (ib) suggests that the *quién*-question admits answers Boolean conjunctions over non-atomic elements. Hence, Alonso-ovalle & Rouillard conclude that *quién* 'who.SG' is semantically number-neutral: it ranges over the set of atomic and non-atomic individuals when taking a lower-order reading, and a set of Boolean conjunctions and disjunctions over these individuals when taking a higher-order reading.

Crucially, the same as the case of singular-marked questions, the Q-domain of a numeral-modified *wh*-question must exclude Boolean conjunctions; otherwise, (50) would have a strongest true answer yielded based on $(a \oplus b)^{\Uparrow} \sqcap (c \oplus d)^{\Uparrow} \sqcap (b \oplus c)^{\Uparrow}$.

Return to the contrast of question-embeddings in (45), repeated below:

(51) (w: *The kids abcd formed two teams in total: $a + b$ formed one, and $c + d$ formed one.*)

a. $\surd$ John knows [**which kids** formed a team].

b. # John knows [**which two kids** formed a team].
$\rightsquigarrow$ *Only two of the kids formed any team.*

The contrast is explained if we assume that the Q-domain of a basic plural-marked question can range over higher-order meanings, while that of a numeral-modified question cannot. More specifically, in (51a), the Q-domain yielded by *which kids* includes Boolean conjunctions and hence the embedded question *which kids formed a team* admits conjunctive answers. In the given scenario, the Boolean conjunction $(a \oplus b)^{\Uparrow} \sqcap (c \oplus d)^{\Uparrow}$ yields the strongest true answer. In contrast, the Q-domain yielded by *which two kids* consists of only pluralities denoting sums of two kids (such as $a \oplus b$ and $c \oplus d$), and hence the embedded question in (51b) has two true answers, namely, $^\wedge f.a.team(a \oplus b)$ and $^\wedge f.a.team(c \oplus d)$, neither of which counts as the strongest true answer. As such, (51b) is infelicitous because the embedded question does not satisfy Dayal's presupposition, and this presupposition failure projects over the factive predicate *know*.

3.3. Evidence for Boolean compounds

Previous sections provide two diagnostics for simple GQs. The diagnostic based on non-reducibility validates the existence of Boolean disjunctions and existential GQs in a higher-order Q-domain. The diagnostic based on stubbornly collectivity provides evidence for Boolean conjunctions and universal GQs. Combining these two diagnostics, this section shows that a higher-order Q-domain also contains Boolean coordination compounds of GQs.

3.3.1. Disjunction over conjunctions

Assume that the 8 students enrolled in a class are separated into four pairs by year and major. As part of the course requirement, each pair of students have to co-present one paper this or next week. Moreover, the instructor requires the presentations in each week to be given by students from the same department.

junior linguists:	$\{a_1, b_1\}$	junior philosophers:	$\{a_2, b_2\}$
senior linguists:	$\{c_1, d_1\}$	senior philosophers:	$\{c_2, d_2\}$

With the above background, consider the following conversation:

(52) a. Guest: "[In your class,] which students have to present a paper together this week?"

b. Instructor: "The two junior linguists and the two senior linguists, OR, the two junior philosophers and the two senior philosophers."

The question from the guest involves a necessity modal *have to* and a stubbornly collective predicate *present a paper together*. The answer provided by the instructor can be unpacked as follows: the disjunctive answer conveys in general a free choice inference as in (53a), and the choices are specified as in (53b-c). ('p.a.p.t.' is abbreviated for 'present a paper together'.)

(53) a. *The presentations this week have to be given by either the linguists or the philosophers. They can be given by the linguists, and can be given by the philosophers.*
b. *If the presentations are given by the linguists,* $a_1 \oplus b_1$ *will p.a.p.t., and* $c_1 \oplus d_1$ *will p.a.p.t..*
c. *If the presentations are given by the philosophers,* $a_2 \oplus b_2$ *will p.a.p.t., and* $c_2 \oplus d_2$ *will p.a.p.t..*

To get the free choice inference (53a), the disjunction must be interpreted under the scope of the necessity modal. Further, since the predicate *present a paper together* is stubbornly collective, to get the conjunctive inferences in (53b-c), each disjunct/choice must be understood as naming a Boolean conjunction over two pairs of students. In sum, the strongest true answer of the question is an inference with the following scopal pattern: $\Box \gg$ *or* $\gg$ *and* $\gg$ *a paper*. To get this scopal pattern, the LF of the question should involve a higher-order *wh*-trace in between the necessity modal and the collective predicate, as in (54a). Instructor's answer should be read as naming a Boolean disjunction over two Boolean conjunctions, as in (54b).

(54) a. [[... which students] $\lambda\pi$ [$_{\text{IP}}$ have to [$\pi_{\langle et,t\rangle}$ λx [$_{\text{VP}}$ x_e p.a.p.t.]]]]
b. $((a_1 \oplus b_1)^{\Uparrow} \sqcap (c_1 \oplus d_1)^{\Uparrow}) \sqcup ((a_2 \oplus b_2)^{\Uparrow} \sqcap (c_2 \oplus d_2)^{\Uparrow})$

3.3.2. Conjunction over disjunctions

As mentioned in section 3.2, a few emotive attitude predicates (e.g., *be worried*, *be surprised*, *be happy*) are not reducible relative to conjunctions.

(55) a. Mary is worried that Andy and Billy will join the party. (They make a mess whenever they stay together.) $\not\rightsquigarrow$ *Mary is worried that Andy will join the party.*
b. Jess is worried that Andy will leave and Billy will stay. (Billy becomes a trouble-maker when Andy is not around.) $\not\rightsquigarrow$ *Jess is worried that Andy left.*

In the following example, the elided complex conjunctions completely addresse the question if read with the following scopal pattern: *be worried* $\gg \Box \gg$ *and* $\gg$ *or*/$\exists$.[13]

(56) (*w: Jack is tolerated of taking up to one course a year in syntax or semantics, but he would be worried that if he has to take one or more courses for each subfield.*)
What courses is Jack worried that he must take [this year]?
a. Semantics I or II, and, Syntax I or II.
b. At least one course in semantics and at least one course in syntax.

(57) a. [[...which courses] $\lambda\pi$ [$_{\text{IP}}$ Jack$_i$ is worried that [must [$\pi_{\langle et,t\rangle}$ λx [$_{\text{VP}}$ he$_i$ takes x_e]]]]]]
b. $(\textsf{sem}_1^{\Uparrow} \sqcup \textsf{sem}_2^{\Uparrow}) \sqcap (\textsf{syn}_1^{\Uparrow} \sqcup \textsf{syn}_2^{\Uparrow})$

[13]Since $\Box$-statements are reducible relative to conjunctions, the scope relation between the necessity modal and the conjunctive does not make a difference to the truth conditions.

3.4. Interim summary

To sum up, this section provides two diagnostics for a richer Q-domain. The first diagnostic is based on narrow scope readings of elided quantificational answers in questions where the question nucleus is not reducible relative to the named GQs. Results of this diagnostic rules in Boolean disjunctions and a class of existential GQs. The second diagnostic is based on the absence of uniqueness effects in questions with a stubbornly collective predicate. This diagnostic rules in Boolean conjunctions and universal GQs. Moreover, combining these two diagnostics, I also show that a higher-order Q-domain consists of also Boolean coordinations of GQs.

4. Defining a higher-order Q-domain: The Positiveness Constraint

Evidence from the previous section has ruled in Boolean disjunctions, conjunctions, a class of existential GQs and universal GQs, and their Boolean coordinations. One might now ask whether we can make the following generalization: the higher-order Q-domain yielded by a wh-phrase consists of all GQs over subsets of the wh-complement and the Boolean compounds of these GQs. Spector (2007, 2008) gives some counterexamples to this generalization and argues that the higher-order GQs [or GQ-compounds] involved in a higher-order Q-domain must be *increasing* (or say *upward-entailing*). Extending Spector's diagnostic to non-monotonic GQs, I show that the increasingness requirement is too strong and then argue that whether a higher-order meaning can be ruled in is determined by its *positiveness*.

4.1. The Completeness Test

Whether a meaning is included in the Q-domain of a question can be examined by the **Completeness Test**, as generalized in (58). This test draws on a deductive relation between attitudes held towards a question and attitudes held towards the answers of this question: the question-embedding sentence *x knows Q* implies that *x* knows the complete true answer of Q. The complete answer of a question is the strongest true proposition in the Hamblin set of this question (Dayal 1996); hence, if a proposition *p* is true but is not entailed by the complete true answer of Q, *p* is not in the Hamblin set of Q.[14]

(58) **The Completeness Test** (generalized from Spector (2008))
For any proposition *p* that names a short answer *x* to a question Q: if *x knows Q* does not entail *x knows p*, then: *p* is not in the Hamblin set of Q, and *x* is not in the Q-domain of Q.

Consider (59) for an illustration of the simplest case. In the given context, the true answers of the question *who came* include positive propositions like *Andy and Billy came*, negative propositions like *Cindy didn't come*, and their Boolean coordinations such as *Andy and Billy came but Cindy didn't*. However, the question-embedding sentence (59) being true only requires the belief-holder Sue to know the positive answers; it does not require her to know the negative answers. This asymmetry suggests that the Q-domain of the embedded question includes *Andy and Billy* (interpreted either as a plural individual $a \oplus b$ or a Boolean conjunction $\lambda P.P(a) \wedge P(b)$), but not the negative GQ *not*

[14] For simplicity, here I ignore mention-some questions and assume that a question has at most one complete true answer, which is its strongest true answer.

Cindy ($\lambda P.\neg P(c)$) or the conjunctive compound which involves a negative conjunct *Andy and Billy but not Cindy* ($\lambda P.P(a \oplus b) \wedge \neg P(c)$).

(59) (*w: Among the relevant individuals, Andy and Billy came, but Cindy didn't.*)
Sue knows who came.
  a. $\rightsquigarrow$ *Sue knows that Andy and Billy came.*
  b. $\not\rightsquigarrow$ *Sue knows that Cindy didn't come.*

4.2. The Increasing-ness Constraint and its problem

Next, apply the Completeness Test to case of a □-question. The background in (60a) lists out all of John's reading obligations, where each inference names a GQ over a set of books. ((a-i) and (a-iii) are taken from Spector (2008).) In particular, the inference (a-i) names an increasing GQ and the rest each names a decreasing GQ.[15] For the same reason seen in (23) in section 3.1, here the embedded □-question must be interpreted with a higher-order reading — there is no particular book that John has to read. The task is to find out the truth conditions of the question-embedding sentence (60b) in the described background.

(60) a. Assumed that John's reading obligations are as follows:
  i. John must read {at least two, more than one} novel(s) by Andy,
  ii. John must read every book by Andrew except one out of stock,
  iii. John must read no book by Billy,
  iv. John must read no book by Betty except the one in a blue cover,
  v. John must read {at most one, less than two} book(s) by Cindy,
  vi. John must read {no more than two, up to two} books by Danny.
 b. Sue knows which books John must read.
  $\rightsquigarrow$ *Sue knows (i) & (ii).*
  $\not\rightsquigarrow$ *Sue knows (iii)/(iv)/(v)/(vi).*

The question-embedding sentence (b) being true requires that Sue knows the inferences (a-i) and (a-ii) but does not require that she knows any of the rest.[16] This contrast suggests that the Q-domain of the embedded □-question does not include decreasing GQs such as *no book by Andy* or non-monotonic GQ-compounds such as the GQ-conjunction *at least two books by Andy and no book by*

[15] Monotonicity of quantifiers is defined as follows:

(i) For a quantifier π of type $\langle et, t\rangle$, π is ...
  a. *increasing* if and only if $\pi(A) \Rightarrow \pi(B)$ for any sets of entities A and B s.t. $A \subseteq B$;
  b. *decreasing* if and only if $\pi(A) \Leftarrow \pi(B)$ for any sets of entities A and B s.t. $A \subseteq B$;
  c. *non-monotonic* if and only if π is neither increasing nor decreasing.

[16] Surprisingly, in contrast to (60b), the following two sentences with a concealed question or a definite description do imply that Sue knows all the inferences in (60a).

(i) a. Sue knows what John's reading obligations are.
  b. Sue knows John's reading obligations.

Billy. Hence, Spector (2008) concludes that the higher-order meanings included in the Q-domain of a *wh*-question must be **increasing**. He proposes that this increasing-ness constraint comes from the lexical meaning of the *wh*-phrase: in a higher-order reading, the *wh*-phrase ranges over a set of increasing GQs. The higher-order reading is then paraphrased as follows: 'for which increasing GQ π over books, is it the case that John must read π?' Note a caveat that this paraphrase does not take GQ-compounds into consideration, but Spector's argumentation does consider the GQ-coordination *at least two books by Andy and no book by Billy*. Moreover, section 3.3 has shown that a higher-order Q-domain contains also Boolean coordinations. Hence, the higher-order reading that Spector (2008) intends to define should be as follows:[17]

(61) Which books must John read? (Modified from Spector (2008))
$\approx$ 'For which π such that π is an increasing GQ over books or a Boolean coordination of increasing GQs over books, is it the case that John must read π?'

Insofar, the Increasing-ness Constraint has correctly excluded decreasing GQs or non-increasing (i.e., decreasing or non-monotonic) GQ-compounds. However, this constraint incorrectly excludes also non-monotonic GQs. For illustration, I add the following requirement (vii) to the list of John's reading obligations in (60a). Intuitively, the question-embedding sentence (60b) does entail that Sue knows the inference (60a-vii), which suggests that the Q-domain of the embedded $\Box$-question includes also non-monotonic GQs such as *exactly two SAT books* and *two or three SAT books*.

(60a) vii. John must read {exactly two, two to three} SAT books, but the choice is up to him. (To avoid over-preparation and to allocate time for the other subjects, he shouldn't read more than two SAT textbooks.)

In conclusion, in determining whether a higher-order meaning should or should not be ruled into a Q-domain, a monotonicity-based constraint faces a dilemma — it needs to rule out non-monotonic GQ-compounds while ruling in non-monotonic GQs.

4.3. The Positiveness Constraint

In contrast to Spector (2008), I propose that whether a GQ or a GQ-compound should be included in a higher-order Q-domain is determined by its "positiveness", not its monotonicity.

(62) **The Positiveness Constraint**
Only positive GQs and their coordination compounds can be in a Q-domain.

[17] The Completeness Test in (i) considers two more cases that involve GQ-disjunctions (underlined). This test further confirms that non-monotonic Boolean disjunctions must be excluded from the Q-domain of the embedded question.

(i) a. John's reading obligations for the summer consist of the following:
  i. John must read no leisure book or more than two math books. (In other words, John must read more than two math books if he reads any leisure book.)
  ii. John must read none or all of the Harry Potter books, (because Harry Potter books must be rented in a bundle, and his mom would blame him for wasting money if he rents the entire book series but only reads some of them.)
b. Sue knows which books John must read in the summer.
$\not\rightsquigarrow$ *Sue knows (i)/(ii).*

A GQ being positive means that the meaning of this GQ ensures existence relative to its smallest live-on set. For example, *at least two books* and *exactly two books*, while having different monotonicity patterns, both entail *some books* and are thus positive. By contrast, the decreasing quantifier *at most two books* does not entail *some books* and is thus not positive. A schematized definition is given in (63). $\textsf{some}(\textsc{smlo}(\pi))$ stands for the GQ derived by applying the basic existential determiner *some* to the smallest live-on set of π. (See (19) in section 2.3 for the definition of live-on set.)

(63) A generalized quantifier π is **positive** if and only if $\pi \subseteq \textsf{some}(\textsc{smlo}(\pi))$.

Table 2 compares increasing-ness/monotonicity and positiveness for a list of GQs over books. (a and b are atomic books). Observe that increasing GQs are all positive, while decreasing ($\downarrow_{\textsc{mon}}$) and non-monotonic (N.M.) GQs are not.

Generalized quantifier π	$\textsc{smlo}(\pi)$	Increasing?	Positive?
$a^{\Uparrow}$	$\{a\}$	Yes	Yes
$a^{\Uparrow} \sqcap b^{\Uparrow}$, $a^{\Uparrow} \sqcup b^{\Uparrow}$	$\{a, b\}$	Yes	Yes
at least two books	*books*	Yes	Yes
more than two books	*books*	Yes	Yes
every book except a	*book* $- \{a\}$	Yes	Yes
at most two books	*books*	No ($\downarrow_{\textsc{mon}}$)	No
less than two books	*books*	No ($\downarrow_{\textsc{mon}}$)	No
no book except a	*book* $- \{a\}$	No ($\downarrow_{\textsc{mon}}$)	No
exactly two books	*books*	No (N.M.)	Yes
two to ten books	*books*	No (N.M.)	Yes
less than three or more than ten books	*books*	No (N.M.)	Yes

Table 2: Increasing-ness/monotonicity versus positiveness

I define the Q-domain of a *wh*-question in a higher-order reading as follows: for a question of the form '*wh*-A f?' where the *wh*-complement 'A' is interpreted as a set A of type $\langle \tau, t \rangle$, the **higher-order Q-domain** of this question (if it exists) is $^{\text{H}}A$ such that $^{\text{H}}A$ is the minimal set of type $\langle \tau tt, t \rangle$ that includes all the positive GQs living on a subset of A and the Boolean coordination compounds of these GQs. A one-line definition and a recursive definition are given in the following:

(64) **Higher-order Q-domain** (one-line definition)

$$^{\text{H}}A = \textsc{min}\left\{ A' \,\middle|\, \begin{array}{r} \forall \pi_{\langle \tau t, t \rangle}[\textsc{smlo}(\pi) \subseteq A \wedge \pi \subseteq \textsf{some}(\textsc{smlo}(\pi)) \rightarrow \pi \in A'] \\ \wedge \forall \alpha[\varnothing \subset \alpha \subseteq A' \rightarrow \bigsqcup \alpha \in A' \wedge \bigsqcap \alpha \in A'] \end{array} \right\}$$

(65) **Higher-order Q-domain** (recursive definition)

a. For any π such that $\textsc{smlo}(\pi) \subseteq A$ and $\pi \subseteq \textsf{some}(\textsc{smlo}(\pi))$, $\pi \in {}^{\text{H}}A$;

b. If $\pi_1 \in {}^{\text{H}}A$ and $\pi_2 \in {}^{\text{H}}A$, then $\pi_1 \sqcup \pi_2 \in {}^{\text{H}}A$ and $\pi_1 \sqcap \pi_2 \in {}^{\text{H}}A$;

c. Nothing else is in $^{\text{H}}A$.

It is worthy of noting that $\textsc{smlo}(\pi)$ presupposes that π has a live-on property (see again (19) in section 2.3), and hence that π can be decomposed into a conservative type $\langle 1, 1 \rangle$ quantifier (i.e., a quantifier that can be lexicalized into a determiner) and a live-on set (Barwise and Cooper 1981; Higginbotham and May 1981; Keenan and Stavi 1986). This presupposition excludes many unwanted higher-order meanings from the Q-domain.

I assume that the H-shifter is a covert operator that can be applied to a predicative expression (see more about the distributional constraints of this operator in section 5). As such, the first-order/higher-order ambiguity of a *wh*-question can be attributed to the application of a H-shifter. As exemplified in (66), in the LF for the higher-order reading, a H-shifter is applied to the *wh*-complement, shifting the restrictor of the *wh*-determiner from a set of entities to a set of GQs and GQ-compounds, and the *wh*-phrase binds a higher-order trace π across the $\Box$-modal *have to*.

(66) Which books does John have to read?

a. **First-order reading**

i. LF: [[... which books] λx [have-to [John read x_e]]]

ii. $[\![\text{WH-Q}]\!] = \lambda x_e : x \in {}^{*}\textit{book}_w . \Box \lambda w.\textit{read}_w(j, x)$

b. **Higher-order reading** ($\Box \gg \pi$) (Revised from (26))

i. LF: [[... which $^{\text{H}}$books] $\lambda \pi$ [have-to [$\pi_{\langle et,t \rangle}$ λx [John read x_e]]]]

ii. $[\![\text{WH-Q}]\!] = \lambda \pi_{\langle et,t \rangle} : \pi \in {}^{\text{H}*}\textit{book}_w . \Box \lambda w.\pi(\lambda x.\textit{read}_w(j, x))$

5. Distributing higher-order readings

As discussed in section 3.2, facts of uniqueness effects in questions suggest that higher-order readings are unavailable in questions where the *wh*-complement is singular-marked or numeral-modified. Aforementioned examples are collected in the following:

(67) a. Which kid came? $\rightsquigarrow$ *Exactly one of the kids came.*

b. Which two kids came? $\rightsquigarrow$ *Only two of the kids came.*

c. Which two kids formed a team? $\rightsquigarrow$ *Only two of the kids formed any team.*

According to Dayal (1996), the singular-marked question (67a) presupposes uniqueness because its strongest true answer exists only when it has exactly one true answer. This analysis also extends to the numeral-modified question (67b-c), as I argued in section 3.2 and Xiang 2016. Adopting this analysis of uniqueness, I conclude that these questions cannot take answers that name Boolean conjunctions, and further that these questions do not have higher-order readings.

Strikingly, in contrast to a numeral-modifier, a PP-modifier does not block higher-order readings. Compare the following two sentences for example. Although *students (that are) in a group of two* is semantically similar to *two student*, the embedded question in (69), where the *wh*-complement is modified by a PP or a relative clause does not presuppose uniqueness, and the question-embedding sentence can be naturally followed by an answer sentence that names a Boolean conjunction. This contrast suggests that the availability of higher-order reading is somehow sensitive to the internal structure of the *wh*-complement.

(68) I know which **two students** presented a paper together,

a. ... the two boys.

b. # ... the two boys and the two girls.

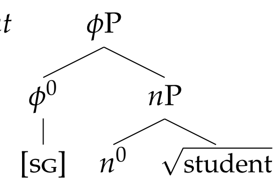
(69) I know which **students (that are) in a group of two** presented a paper together,

a. ... the two boys.

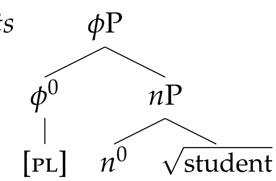
b. ... the two boys and the two girls.

To account for the above distributional constraints, I propose that the H-shifter is applied locally to nP within the *wh*-complement and argue that its application is blocked in singular-marked nouns and numeral-modified nouns. First of all, I assume the following structure of a singular/plural bare noun:

(70) a. *student*



b. *students*



At the right bottom of each tree, n^0 combines with a root $\sqrt{\text{student}}$ and returns a projection denoting a set with a complete join semi-lattice structure (Harbour 2014). For example, with three atomic students *abc*, $[\![n\text{P}]\!] = \{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$. The number feature [SG]/[PL] is evaluated at ϕ^0. Following Sauerland (2003), I interpret [SG] as a predicate restrictor that requires atomicity while [PL] as being semantically vacuous.

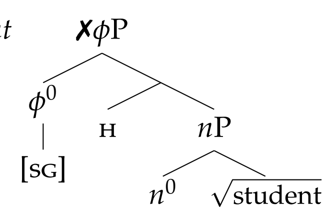
(71) a. $[\![[\text{SG}]]\!] = \lambda P_{\langle e,t\rangle} \lambda x_e.\text{ATOM}(x) \wedge P(x)$

b. $[\![n\text{P}]\!] = \{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$

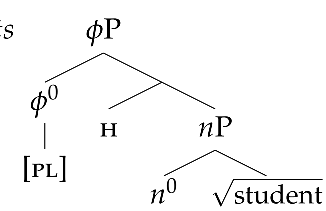
c. $[\![[\text{SG}](n\text{P})]\!] = \{a, b, c\}$

The above assumptions straightforwardly explain why the H-shifter cannot be used in singular nouns. In (72a), applying the H-shifter to nP returns a set of GQs and GQ-compounds, which are all non-atomic and are conflicting with the atomicity requirement of [SG]. As such, the H-shifter cannot be applied in a singular-marked *wh*-question because it would otherwise yield an empty Q-domain. In contrast, H-shifter can be freely used in simple plural-marked and number-neutral *wh*-questions because in these questions the ϕ^0 does not presuppose atomicity.[18]

(72) a. *student*



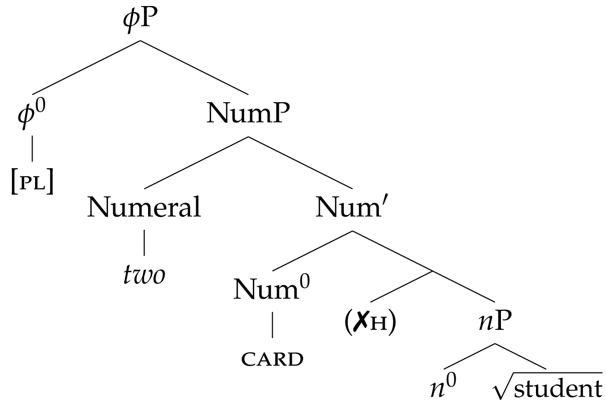
b. *students*



Next, consider numeral-modified NPs. Following Scontras (2014), I place cardinal numeral-modifiers at [Spec, NumP] and assume that Num^0 is located between n^0 and ϕ^0 and is occupied by a cardinality predicate CARD. As defined in (74a), CARD takes a predicate P and a number n and returns the set of individuals in P each of which is constituted of exactly-n atoms. These assumptions automatically explain why the H-shifter cannot be used in a numeral-modified NP — the CARD-predicate at Num^0 requires to check the cardinality of the elements in the set it combines with and hence it cannot combine with a set of GQs.

[18]Note that this claim holds regardless of whether plurals are semantically marked or unmarked. Other than One can also treat [PL] as a predicate restrictor that presupposes non-atomicity or anti-presupposes atomicity.

(73) *two students*



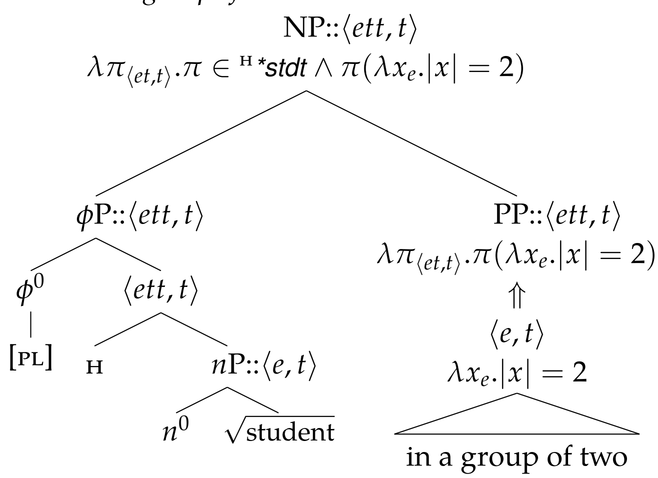
(74) a. $[\![\text{CARD}]\!] = \lambda P \lambda n \lambda x.P(x) \wedge |x| = n$

b. Without the H-shifter
$[\![\text{Num}']\!] = \lambda n \lambda x.{}^{*}\textit{stdt}(x) \wedge |x| = n$
$[\![\text{NumP}]\!] = \lambda x.{}^{*}\textit{stdt}(x) \wedge |x| = 2$
$= \{a \oplus b, b \oplus c, a \oplus c\}$

c. With the H-shfiter
$[\![\text{Num}']\!]$ is undefined (or Num′ has type-mismatch)

In contrast to numeral-modifiers, PP-modifiers are adjoined to the entire NP/ϕP. Hence, the H-shifter can be used in a PP-modified NP without causing a type-mismatch. All we need to do is to apply Montague-lift to the PP-modifier and shifts it into a set of GQs. Then, the Montague-lifted PP composes with the higher-order ϕP standardly via Predicate Modification. This analysis also extends to NPs modified with relative clauses.

(75) *students in a group of two*



6. The 'conjunction-rejecting' higher-order reading

6.1. The puzzles

Surprisingly, in responding to a □-question where the *wh*-phrase is singular-marked or numeral-modified, narrow scope disjunctions are not as bad as conjunctions.

(76) I know which book John has to read,

a. # ... Book A and Book B.

b. ?? ... Book A or Book B. (*or* $\gg$ □, ?□ $\gg$ *or*)

(77) I know which two books John has to read ...

a. # ... the two French books and the two Russian books.

b. ?? ... the two French books or the two Russian books. (*or* $\gg \Box$, $?\Box \gg$ *or*)

Such narrow scope readings are more readily available with elided answers to matrix questions. In (78), the disjunction in the elided answer is interpreted under the scope of *should*, conveying a free choice inference that the questioner is free to use any one of the two mentioned textbooks. By the diagnostic based on non-reducibility in section 3.1, the narrow scope reading of the elided disjunctive answer suggests that here the □-question does admit higher-order answers, conflicting with the aforementioned generalization that singular-marked questions do not have higher-order readings.

(78) Which textbook should I use for this class?
Heim&Kratzer or *Meaning&Grammar*, the choice is up to you.

A similar observation is seen with questions with possibility modal (called ◇-questions henceforth). ◇-questions are known to be ambiguous between mention-some (MS-)readings and mention-all (MA-)readings (Groenendijk and Stokhof 1984). As exemplified in (79), if interpreted with a MS-reading, the question can be naturally addressed by an answer that specifies only one feasible option; while in MA-readings, the question requires an answer that exhaustively lists out all the feasible options. Crucially, mention-all (MA-)answers of ◇-questions can have either an elided conjunctive form, as in (79b), or an elided disjunctive form, as in (79c). While having different forms, both of the MA-answers convey the same conjunctive inference that we can use *Heim&Kratzer* for this class <u>and</u> we can use *Meaning and Grammar* for this class.

(79) What can we use [as a textbook] for this class?

a. *Heim&Kratzer*. MS

b. *Heim&Kratzer* and *Meaning and Grammar*. Conjunctive MA

c. *Heim&Kratzer* or *Meaning and Grammar*. Disjunctive MA

In Xiang (2016: chapter 2), I propose that MS-readings are higher-order readings, namely, the *wh*-trace binds a higher-order trace across the ◇-modal, and that MA-readings arise if the higher-order *wh*-trace takes wide scope or if it is associated with an operator that evokes a free choice inference. I will return to the details in section 6.2.2.

It is previously thought that MS-readings are unavailable in singular-marked *wh*-questions: the MS/MA contrast collapses in a singular-marked *wh*-question, because these questions presuppose uniqueness and can have at most one true answer (Fox 2013; Xiang 2016: chapter 3). The following example supports this prediction. Native speakers reported that all the four continuations are infelicitous: they each names multiple choices of textbooks for this class, while the question-embedding sentence implies that the speaker is committed that there is only one choice.

(80) I know which textbook we can use for this class, ...

a. # *Heim&Kratzer* and *Meaning and Grammar*.

b. ? *Heim&Kratzer* or *Meaning and Grammar*.

c. # Every book that teaches compositionality.

d. ?? Any book that teaches compositionality.

It is worthy noting that the disjunctive continuation (80b) is slightly more acceptable than the conjunctive one (80a), in analogous to the contrasts seen in (76) and (77). This contrast is also seen with the universal free choice item (FCI) *any book*, which is argued to be existential in lexicon (Chierchia 2006, 2013), and the basic universal quantifier *every book*.

However, Hirsch and Schwarz (2019) novelly observe that the matrix singular-marked $\diamond$-question in (81) does admit a multi-choice reading. They argue that here uniqueness is interpreted below the $\diamond$-modal. The question is roughly read as 'for which x, is it the case that x is the unique/only letter missing in *fo__m*?'.[19]

(81) Which letter could be missing in *fo__m*?
  a. (The missing letter could be) *a*.
  b. The missing letter could be *a* and the missing letter could be *r*.

Gentile and Schwarz (2018) make a similar observation with *how many*-questions. The same as *wh*-questions with a singular-marked and numeral-modified *wh*-phrase, *how many*-questions presuppose uniqueness. For example, the question in (82) has only one true answer and cannot be felicitously responded by a conjunction of two cardinal numerals. Moreover, since the predicate of this question is stubbornly collective, the uniqueness effect suggests that the Q-domain of this question does not include Boolean conjunctions over numerals.

(82) How many students solved this problem together? #Two and three.
(Intended meaning: 'Two students solved this problem together, and (another) three students solved this problem together.')

Gentile and Schwarz further observe that $\diamond$-modals obviate the uniqueness presuppositions of *how many*-questions. The question in (83) can have a multi-choice reading and does not seem to presuppose uniqueness.

(83) How many students are allowed to solve this problem together?
  a. Two are OK and three are OK.
  b. Two or three.

For a direct comparison with the number-neutral $\diamond$-question in (79), consider the singular-marked $\diamond$-question in (84) and its elided MA-answers. According to Hirsch and Schwarz (2019), the uniqueness inference triggered by the singular *wh*-phrase *which textbook* can be interpreted globally or locally. The global uniqueness interpretation says that there is only textbook that we can use for this class and the questioner asks to specify this book. The local uniqueness reading says that we will only use one textbook for this class and the questioner asks to list out one option, as in a MS-reading, or all the options, as in a MA-reading. Strikingly, in contrast to the numeral-neutral question in

[19] Note that in Hirsch and Schwarz's original example, the multi-choice answer (81b) is not a direct answer of the given question. As seen in (i), conjunctive answers in a form congruent with the question are greatly degraded.

(i) Which letter could be missing in *fo__m*?
  a. ?? *a* could be missing in *fo__m* and *r* could be missing in *fo__m*.
  b. # *a* and *r*.

(79) where an elided MA-answer can be a conjunction or a disjunction, here an elided MA-answer must be a disjunction, as seen in (79a-b). Moreover, the disjunction/conjunction contrast is also seen with the universal free choice item *any book* and the basic universal quantifier *every book*.

(84) Which textbook can I use for this class?

a. *Heim&Kratzer* or *Meaning and Grammar*. Disjunctive MA

b. # *Heim&Kratzer* and *Meaning and Grammar*. Conjunctive MA

c. Any book that teaches compositionality.

d. # Every book that teaches compositionality.

The same pattern is seen with Hirsch & Schwarz's example and a similar example with a numeral-modified *wh*-question, as shown in (85) and (86).

(85) Which letter could be missing in *fo_m*?

a. Letter *a* or letter *r*.

b. # Letter *a* and letter *r*.

(86) Which two letters could be missing in *f_ _m*?

a. Letters *oa* or letters *or*.

b. # Letters *oa* and letters *or*.

Three puzzles arise from the observations in this section. First, why these singular-marked or numeral-modified □/◇-questions admit disjunctive answers but not conjunctive answers? Second, why this 'conjunction-rejecting' higher-order reading is available even though the *wh*-phrase is singular-marked or numeral-modified, in contrast to the regular (viz., 'conjunction-admitting') reading discussed in section 5? Last, why this 'conjunction-rejecting' higher-order reading is more easily attested in matrix questions than in embedded questions?

The following sections provide two approaches to derive the 'conjunction-rejecting' higher-order reading and explain its distributional constraints. One approach assumes that the derivation of this reading involves reconstructing the *wh*-complement to the question nucleus and interpreting uniqueness locally. In this approach, conjunctive answers are out because applying conjunction directly over uniqueness yields a contradiction. The other approach treats the 'conjunction-rejecting' reading the very same reading as the regular higher-order reading but gives a weaker definition to singular and numeral-modified nouns. In this approach, the distributional difference between conjunctive and disjunctive answers comes from that atomicity and cardinality restrictors remove Boolean conjunctions but not disjunctions. Both approaches have desired predictions.

6.2. A reconstruction-based approach

6.2.1. □-questions

Let us start with a singular-marked □-question. (87) provides the rough LF structures and the yielded Q-functions in local uniqueness readings. In both structures, the singular noun *book* is reconstructed to a position c-commanded by the □-modal. This reconstruction has two consequences. First, it leaves a simple variable *D* as the restrictor of the *wh*-phrase. This *D* variable is semantically unmarked; it can be type-lifted freely by the H-shifter without causing a type-mismatch or a conflict with respect to atomicity. A higher-order reading arises if the H-shifter is applied and if *wh*-phrase binds a higher-order trace, as in (87b). Second, uniqueness is evaluated at whichever scopal position that the reconstructed noun adjoins to. In both (87a-b), uniqueness takes scope below the $\Box$-modal.[20]

(87) Which book does John have to read?

a. First-order reading ($\Box \gg \iota$)
'Which x is s.t. it has to be the case that x is the book that John read?'

i. [$_{\text{CP}}$ which$_D$ λx $\Box$ [x is the book John read]]

ii. $[\![\text{WH-}Q]\!] = \lambda x_e : x \in D.\Box\lambda w[x = \iota y[\textsf{book}_w(y) \wedge \textsf{read}_w(y)]]$

b. Higher-order reading ($\Box \gg \pi \gg \iota$)
'Which π is s.t. it has to be the case that π is the book that John read?'

i. [$_{\text{CP}}$ which$_{{}^{\text{H}}D}$ $\lambda\pi$ $\Box$ [$\pi_{\langle et,t\rangle}$ λx. x_e is the book John read]]

ii. $[\![\text{WH-}Q]\!] = \lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}D.\Box\lambda w[\pi(\lambda x_e.x = \iota y[\textsf{book}_w(y) \wedge \textsf{read}_w(j,y)])]$

The following tree diagrams give the details for the two LF structures. Both structures involve reconstructing the *wh*-complement. This reconstruction is realized via three operations. First, a copy of *which book* is interpreted within the nucleus. As assumed in categorial approaches, *which book John read* denotes a one-place predicate. Second and third, THE-insertion introduces uniqueness, and variable insertion introduces a variable bound by the *wh*-phrase.[21][22] In particular, in the LF in

[20] Luis Alonso-Ovalle (p.c.) points out that the assumed local uniqueness inference might be too strong for $\Box$-questions. For example, the question-answer in (i) is fine in a context where it is taken for granted that to win the game, one needs a group of two cards and also other cards.

(i) Which two cards do you need to win the game?
The two red aces or the two black aces.

I argue that the local uniqueness inference in (i) is assessed relative to a local context, namely, the context where the player has a bunch of cards in hand and only needs two more cards to close the game.

[21] I assume that the variable introduced by variable assertion has to be the variable directly bound by the *wh*-phrase. With this assumption, in the LF of the higher-order reading, variable insertion introduces a higher-order variable π, instead of as the following where it introduces an individual variable x bound by the higher-order *wh*-trace:

(i) * [*wh*P $\lambda\pi_{\langle et,t\rangle}$ [have to [π λx_e [$\underline{\lambda y.x = y}$ [THE [which book John read]]]]]

This assumption avoids unattested split scope readings of conjunctive answers to questions with an existential quantifier. Observe that the question in (ii) cannot be felicitously responded by a conjunction. The infelicity of the conjunctive answer suggests that this answer cannot be interpreted with a split scope reading as follows: 'for a math problem x_1, Andy is the unique student who solved x_1, and for a math problem x_2, Billy is the unique student who solved x_2' (*and* $\gg \exists \gg \iota$). The unavailability of this reading rules out the LF in (iib) where the existential quantifier *a math problem* takes scope between the higher-order trace π and the inserted THE.

(ii) Which student solved a math problem?
\# Andy and Billy. (*and* $\gg \iota \gg \exists$)

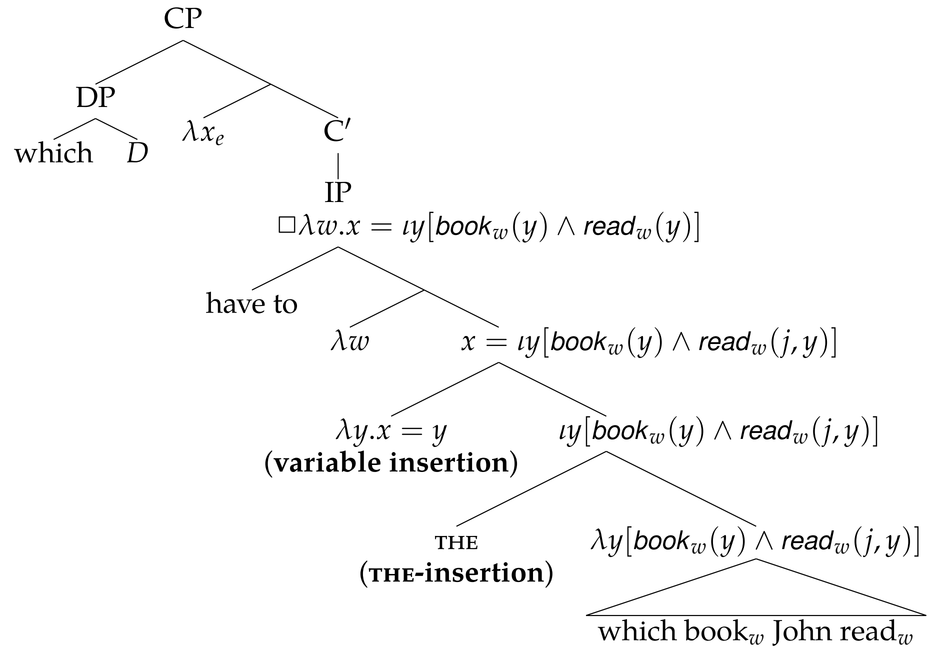
a. [*wh*P $\lambda\pi_{\langle et,t\rangle}$ [$\underline{\lambda y.\pi(\lambda x.x = y)}$ [THE [which student solved a math problem]]]]

b. * [*wh*P $\lambda\pi_{\langle et,t\rangle}$ [π λx_e [[a math problem] λz [$\underline{\lambda y.x = y}$ [THE [which student solved z]]]]]]

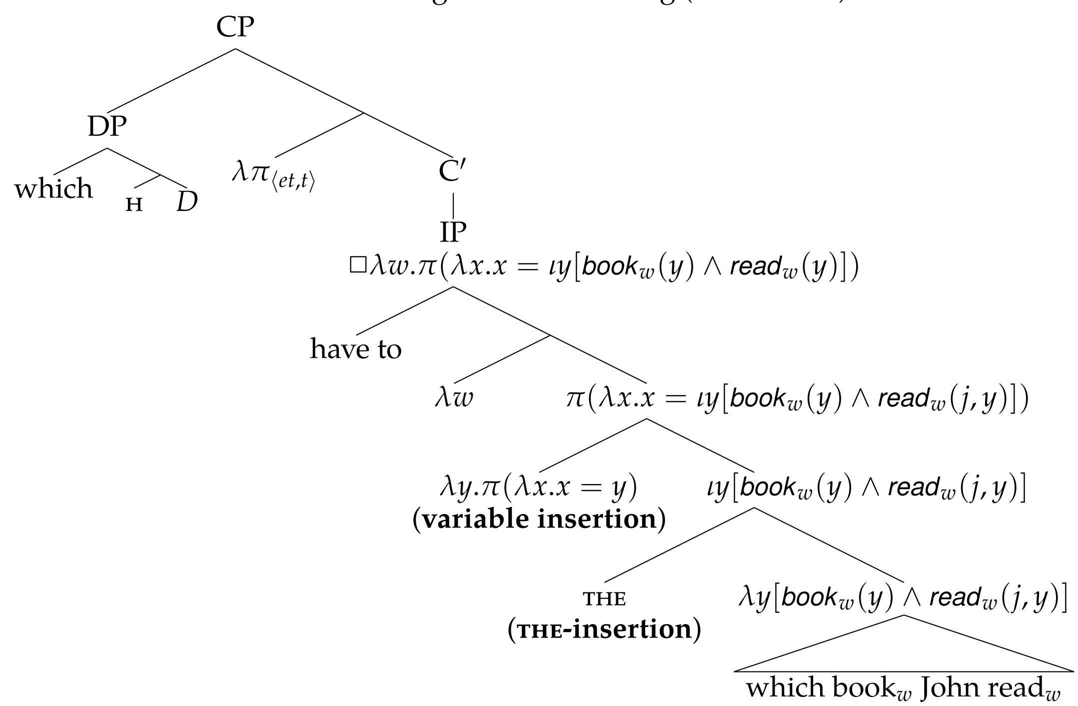
[22] One might have concerns with the operations used for reconstruction. The assumed THE-insertion and variable insertion, on the one hand, are similar to the operations of determiner replacement and variable insertion used in trace conversion (Fox 2002) especially backward trace conversion (Erlewine 2014). On the other hand, in trace conversion, THE-insertion and determiner replacement are locally applied to the moved DP (e.g. *which book*), while in my proposal, THE-insertion and variable insertion apply to a larger expression 'DP+VP' (e.g., *which book John read*). I admit that the

(89) for the higher-order reading, the same as assumed for regular (viz., conjunction-admitting) higher-order readings, the *wh*-restrictor (viz., a null domain variable D) is type-raised by a H-shifter, and the *wh*-phrase binds a higher-order trace π across the $\Box$-modal.

(88) LF with reconstruction for the first-order reading ($\Box \gg \iota$)



(89) LF with reconstruction for the higher-order reading ($\Box \gg \pi \gg \iota$)



assumed syntax of reconstruction is unconventional, but this is not necessarily a problem for considering (89) the structure that derives the 'conjunction-rejecting' higher-order reading. As seen in section 6.1, this reading is very hard to get, especially in question-embedding sentences of the form 'I know Q' (see examples in (76-77) and (80)). Thus, it is highly likely that the derivation of this reading requires abnormal operations, and it is possible that the structure used for deriving this reading is not the real LF of the given question.

The above derivation predicts that the higher-order trace π immediately scopes over uniqueness. This prediction explains why a question in this reading rejects conjunctive answers: if π is a Boolean conjunction, combining π with a predicate of uniqueness yields a contradiction. As seen in (90b), unless Book A and Book B refer to the same book, combining the Q-function with the Boolean conjunction $a^{\Uparrow} \sqcap b^{\Uparrow}$ yields a contradiction.

(90) Which book does John have to read?
$[\![\text{WH-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}D.\Box\lambda w[\pi(\lambda x_e.x = \iota y[\textsf{book}_w(y) \wedge \textsf{read}_w(j,y)])]$

a. Book A or Book B.
$[\![\text{WH-Q}]\!](a^{\Uparrow} \sqcup b^{\Uparrow})$
$= \Box\lambda w[a = \iota y[\textsf{book}_w(y) \wedge \textsf{read}_w(j,y)]] \vee [b = \iota y[\textsf{book}_w(y) \wedge \textsf{read}_w(j,y)]]$
(It has to be the case that the book that John read is Book A or is Book B.)

b. # Book A and Book B.
$[\![\text{WH-Q}]\!](a^{\Uparrow} \sqcap b^{\Uparrow})$
$= \Box\lambda w[a = \iota y[\textsf{book}_w(y) \wedge \textsf{read}_w(j,y)]] \wedge [b = \iota y[\textsf{book}_w(y) \wedge \textsf{read}_w(j,y)]]$
$= \bot$ (unless $a = b$)
(#It has to be the case that the book that John read is Book A and is Book B.)

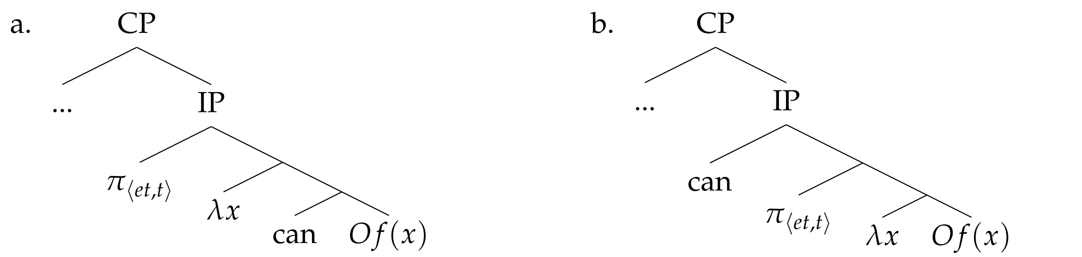
6.2.2. $\Diamond$-questions

The MA-answer of a question is the true answer that entails all the true answers to this question. In Xiang (2016: chapter 2), I argue that in the reading that expects disjunctive MA-answers and the reading that expects conjunctive MA-answers, a $\Diamond$-question has different LF structures and denotes different Q-functions.

In the conjunctive MA-reading, the *wh*-phrase binds a higher-order trace which takes scope above the $\Diamond$-modal. The following illustrates the question nucleus and the yielded Q-function and answer space when the higher-order trace π takes scope below and above the $\Diamond$-modal. In both LF structures, the local O-operator ($\approx$ *only*) is associated with the individual trace x.[23] In the illustrations of the answer space, arrows indicate entailment relation, and shades mark the true answers in w. The included propositions are derived by applying the Q-function to the Boolean conjunction $a^{\Uparrow} \sqcap b^{\Uparrow}$, the Montagovian individuals $a^{\Uparrow}$ and $b^{\Uparrow}$, and the Boolean disjunction $a^{\Uparrow} \sqcup b^{\Uparrow}$. f stands for the predicate *use as a textbook for this class*. For example, the proposition $\Diamond Of(a)$ is read as 'Book A can be the only textbook of this class.'

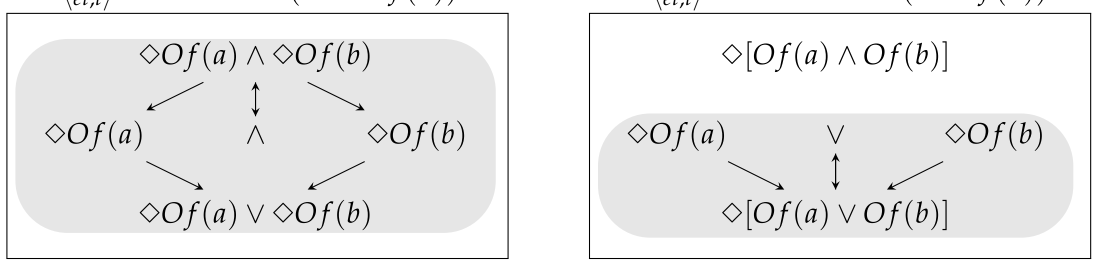
[23] In Xiang (2016: chapter 2), I argue that the narrow scope π yields a MS-reading. The local O-operator ($\approx$ *only*) is assumed for predicting the facts that MS-answers are always mention-one answers, and that any answer that names one feasible option is a possible MS-answer. These issues are beyond the scope of this paper.

(91) (*w: We can use Book A as the textbook of this class, and we can use Book B as the textbook of this class.*) What can we use [as a textbook] for this class? Book A and Book B.



a. $\lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}\textit{hmn}.\pi(\lambda x.\Diamond Of(x))$

b. $\lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}\textit{hmn}.\Diamond\pi(\lambda x.Of(x))$



As seen in (91a), if π scopes above the $\Diamond$-modal, the conjunctive inference derived by combining the Q-function with the Boolean conjunction entails all the true answers, and thus it is the MA-answer. This conjunctive answer is read as 'it is possible that *a* is the only textbook for this class, and it is possible that *b* is the only textbook of this class.' In contrast, as shown in (91b), if π scopes under the $\Diamond$-modal, the derived inference is a contradiction, read as 'it is possible that only *a* is a textbook for this class and only *b* is a textbook for this class.' To sum up, the take-away point is that conjunctive MA answers are used only if the LF of the question has the $\pi \gg \Diamond$ scopal pattern.

Now consider the singular-marked $\Diamond$-question in (92). Again, the puzzle is that although this question is compatible with a multi-choice reading, it rejects a conjunctive MA-answer. Reconstructing the *wh*-complement and letting π take scope above the $\Diamond$-modal yields the following scopal pattern: $\pi \gg \iota \gg \Diamond$. As seen in (92b), unless A and B are the same book, combining the derived higher-order Q-function with the Boolean conjunction $a^{\Uparrow} \sqcap b^{\Uparrow}$ yields a contradiction.

(92) Which textbook can we use [for this class]? # Book A and Book B.

a. $[\![\text{WH-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}D.\lambda w[\pi(\lambda x_e.x = \iota y[\textsf{book}_w(y) \wedge \Diamond_w \textsf{chair}(y)])]$

b. $[\![\text{WH-Q}]\!](a^{\Uparrow} \sqcap b^{\Uparrow})$
$= \lambda w.[a = \iota y[\textsf{book}_w(y) \wedge \Diamond_w O\textsf{chair}(y)]] \wedge [b = \iota y[\textsf{book}_w(y) \wedge \Diamond_w O\textsf{chair}(y)]]$
(#*a* is the unique prof who can chair the committee alone, and *b* is the unique prof who can chair the committee alone.)

In contrast, disjunctive MA arises if the higher-order *wh*-trace is associated with an DOU-operator, regardless of whether this trace scopes below or above the $\Diamond$-modal. The DOU-operator is the covert counterpart of the Mandarin particle *dou*. In questions, the *dou* particle forces a MA-reading when associated with the *wh*-phrase, as seen in (93a). In declaratives, this particle yields a free choice (FC) inference when associated with a pre-verbal disjunction, as shown in (93b).

(93) a. **Dou** [shei] keyi jiao jichu hanyu?
DOU who can teach Intro Chinese
'Who can teach Intro Chinese?' (MA only)

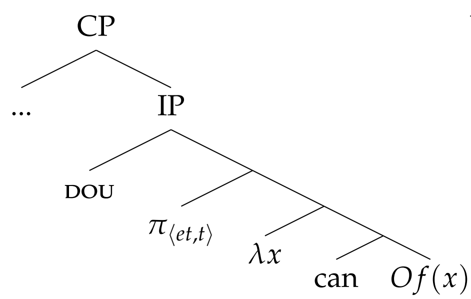
b. [Yuehan huozhe Mali] **dou** keyi jiao jichu hanyu
John or Mary DOU can teach intro Chinese
Intended: 'Both John and Mary can teach Intro Chinese.'

I define *dou* as a pre-exhaustification exhaustifier over sub-alternatives. As schematized in (94), *duo* affirms its propositional argument and negates the exhaustification of each of the sub-alternatives of its propositional argument (Xiang 2016: chapter 7; Xiang 2019). With minimal variations on the semantics of sub-alternatives, this definition predicts various functional uses of *dou* (details omitted). In particular, the sub-alternatives for disjunctive propositions of the form $\diamond(p \vee q)$ or $\diamond p \vee \diamond q$ are $\diamond p$ and $\diamond q$.

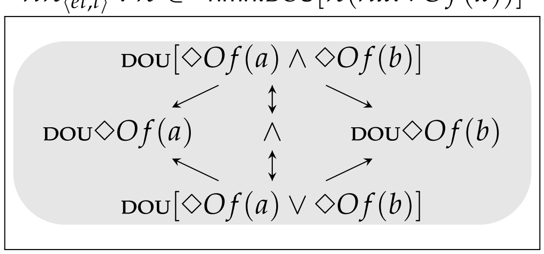
(94) $[\![dou_C]\!] = \lambda p \lambda w : \exists q \in \textsc{Sub}(p, C).p(w) = 1 \wedge \forall q \in \textsc{Sub}(p, C)[O_C(q)(w) = 0]$

The following illustrates two possible LF structures for the disjunctive MA-readings and the Q-function and the answer space derived by each LF. In both LF structures, a covert DOU-operator is presented and is associated with the higher-order trace π. The two structures differ only with respect to the scopal pattern between π and the $\diamond$-modal. As computed in (96), no matter whether π scopes above or below $\diamond$-modal, DOU strengthens the disjunctive answer into a free choice statement which is semantically equivalent to the conjunction of the two individual answers.

(95) a.

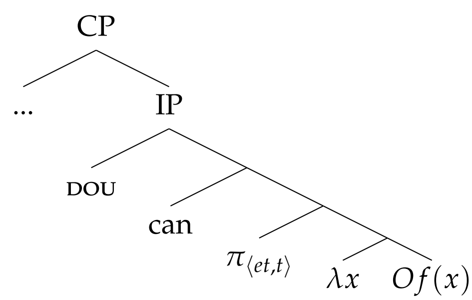


$\lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}hmn.\textsc{dou}[\pi(\lambda x.\diamond Of(x))]$

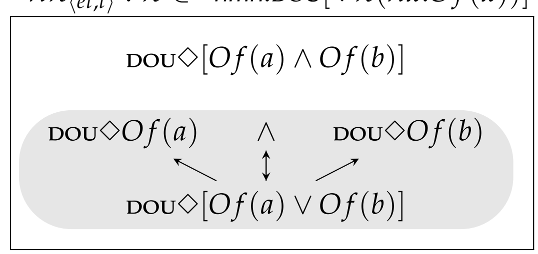


With DOU ($\pi \gg \diamond$): disjunctive/conjunctive MA

b.



$\lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}hmn.\textsc{dou}[\diamond\pi(\lambda x.Of(x))]$



With DOU ($\diamond \gg \pi$): disjunctive MA

(96) a. If $\pi \gg \diamond$
$\textsc{dou}[\diamond Of(a) \vee \diamond Of(b)]$
$= [\diamond Of(a) \vee \diamond Of(b)] \wedge \neg O\diamond Of(a) \wedge \neg O\diamond Of(b)$
$= [\diamond Of(a) \vee \diamond Of(b)] \wedge [\diamond Of(a) \rightarrow \diamond Of(b)] \wedge [\diamond Of(b) \rightarrow \diamond Of(a)]$
$= [\diamond Of(a) \vee \diamond Of(b)] \wedge [\diamond Of(a) \leftrightarrow \diamond Of(b)]$
$= \diamond Of(a) \wedge \diamond Of(b)$

b. If $\Diamond \gg \pi$
$\textsc{dou}\Diamond[Of(a) \vee Of(b)]$
$= \Diamond[Of(a) \vee Of(b)] \wedge \neg O\Diamond Of(a) \wedge \neg O\Diamond Of(b)$
$= \Diamond[Of(a) \vee Of(b)] \wedge [\Diamond Of(a) \rightarrow \Diamond Of(b)] \wedge [\Diamond Of(b) \rightarrow \Diamond Of(a)]$
$= \Diamond[Of(a) \vee Of(b)] \wedge [\Diamond Of(a) \leftrightarrow \Diamond Of(b)]$
$= \Diamond Of(a) \wedge \Diamond Of(b)$

Now return to the case of a singular-marked $\Diamond$-question. While rejecting a conjunctive MA-answer, this question admits an elided disjunction as its MA-answer. The following considers the two possibilities where a covert DOU-operator is associated with a higher-order trace. For the numeral-neutral question in (95), the Q-functions yielded by the two possible LFs have the same output (i.e., free choice statements) when combining with a Boolean disjunction. In the singular-marked question, however, whether π takes scope below or above the $\Diamond$-modal yields a crucial difference with respect to the interpretation of the disjunctive answer. If π takes wide scope, as seen in (97a), the derived free choice inference is a contradiction, just like the case of the wide scope conjunctive answer in (92). In contrast, as seen in (97b), if π takes a narrow scope relative to the $\Diamond$-modal, the derived free choice inference is not contradictory and is a desired MA-answer.

(97) Which book can we use [as a textbook] for this class? Book A or Book B.

a. If $\pi \gg \iota \gg \Diamond$:
$[\![\text{WH-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}D.\lambda w[\textsc{dou}[\pi(\lambda x_e.x = \iota y[\textsf{book}_w(y) \wedge \Diamond_w Of(y)])]]$
$[\![\text{WH-Q}]\!](a^{\Uparrow} \sqcup b^{\Uparrow})$
$= \lambda w[\textsc{dou}[(a^{\Uparrow} \sqcup b^{\Uparrow})(\lambda x_e.x = \iota y[\textsf{book}_w(y) \wedge \Diamond_w Of(y)])]]$
$= \lambda w.\textsc{dou}[a = \iota y[\textsf{book}_w(y) \vee \Diamond_w Of(y)]] \wedge [b = \iota y[\textsf{book}_w(y) \wedge \Diamond_w Of(y)]]$
$= \lambda w.[a = \iota y[\textsf{book}_w(y) \wedge \Diamond_w Of(y)]] \wedge [b = \iota y[\textsf{book}_w(y) \wedge \Diamond_w Of(y)]]$
$= \bot$ (unless $a = b$)
(#*a* is the unique book that we can use as the only textbook for this class, and *b* is the unique book that we can use as the only textbook for this class.)

b. If $\Diamond \gg \pi \gg \iota$:
$[\![\text{WH-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}D.\textsc{dou}\Diamond[\lambda w.\pi(\lambda x_e.x = \iota y[\textsf{book}_w(y) \wedge Of_w(y)])]$
$[\![\text{WH-Q}]\!](a^{\Uparrow} \sqcup b^{\Uparrow})$
$= \textsc{dou}\Diamond[\lambda w.(a^{\Uparrow} \sqcup b^{\Uparrow})(\lambda x_e.x = \iota y[\textsf{book}_w(y) \wedge Of_w(y)])]$
$= \Diamond\lambda w.[a = \iota y[\textsf{book}_w(y) \wedge \Diamond_w Of(y)]] \cap \Diamond\lambda w.[b = \iota y[\textsf{book}_w(y) \wedge \Diamond_w Of(y)]]$
(*a* can be the unique book that we can use as the only textbook for this class, and *b* can be the unique book that we can use as the only textbook for this class.)

To sum up, in a number-neutral $\Diamond$-question in (95), a disjunction can serve as its MA-answer regardless of whether this disjunction is interpreted below or above the $\Diamond$-modal. However, in responding to the singular-marked $\Diamond$-question in (97), the disjunction can have a MA-answer reading but must be interpreted with a narrow scope.

6.3. A unified approach

As suggested by Manuel Križ (p.c.), there is no need to treat 'conjunction-rejecting' higher-order readings differently — all we need is to allow Boolean disjunctions to be atomic or cardinal. In the following definitions, the condition (a) on minimal witness sets ensures the atomic/cardinal GQ to be a Boolean disjunction, an existential quantifier, or a Montagovian individual. In comparison, if π is a universal quantifier or a Boolean conjunction, its minimal witness set is not singleton; if π is a decreasing quantifier, its minimal witness set is the empty set.[24]

(98) A GQ π is atomic if and only if

a. the minimal witness sets of π are all singleton sets;

b. every member in the smallest live-on set of π is atomic.

(99) A GQ π has the cardinality n if and only if

a. the minimal witness sets of π are all singleton sets;

b. every member in the smallest live-on set of π has the cardinality n.

Based on the above assumptions, I re-define the singularity feature [SG] and the cardinality predicate CARD polymorphically as in (101). '$\text{MWS}(A, x)$' is read as 'A is a minimal witness set of x'.

(100) Old definitions

a. $[\![[\text{SG}]]\!] = \lambda P_{\langle e,t \rangle} \lambda x_e . P(x) \wedge \text{ATOM}(x)$

b. $[\![\text{CARD}]\!] = \lambda P \lambda n \lambda x . P(x) \wedge |x| = n$

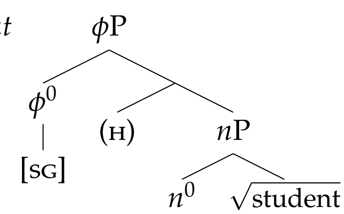
(101) New definitions

a. $[\![[\text{SG}]]\!] = \lambda P \lambda x . \begin{cases} P(x) \wedge \text{ATOM}(x) & \text{if } P \subseteq D_e \\ P(x) \wedge \forall A[\text{MWS}(A, x) \rightarrow |A| = 1] \wedge \forall y \in \text{SMLO}(x)[\text{ATOM}(y)] & \text{if } P \subseteq D_{\langle et,t \rangle} \end{cases}$

b. $[\![\text{CARD}]\!] = \lambda P \lambda n \lambda x . \begin{cases} P(x) \wedge |x| = n & \text{if } P \subseteq D_e \\ P(x) \wedge \forall A[\text{MWS}(A, x) \rightarrow |A| = 1] \wedge \forall y \in \text{SMLO}(x)[|y| = n] & \text{if } P \subseteq D_{\langle et,t \rangle} \end{cases}$

With the revised definitions, the H-shifter can be used regularly in singular nouns and numeral-modified nouns. In a discourse with three students *abc*, the singular noun *student* and the numeral-modified noun *two students* are interpreted as follows. Then the conjunction-rejecting higher-order reading is derived in the same way as regular higher-order readings.

(102) *student*



a. Without H: $[\![\phi\text{P}]\!] = \{a, b, c\}$

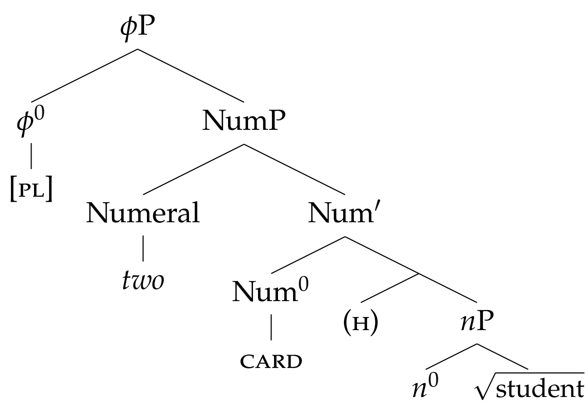
b. With H: $[\![\phi\text{P}]\!] = \{a^{\Uparrow}, b^{\Uparrow}, c^{\Uparrow}, a^{\Uparrow} \sqcup b^{\Uparrow}, a^{\Uparrow} \sqcup c^{\Uparrow}, a^{\Uparrow} \sqcup c^{\Uparrow}, a^{\Uparrow} \sqcup b^{\Uparrow} \sqcup c^{\Uparrow}\}$

[24]Witness sets are defined in terms of the living-on property as follows (Barwise and Cooper 1981):

(i) if a GQ π lives on a set B, then A is a **witness set** of π iff $A \subseteq B$ and $\pi(A)$.

For example, given a discourse domain including three students *abc*, the universal quantifier *every student* has a unique minimal witness set $\{a, b, c\}$, while the singular existential quantifier *some student* has three minimal witness sets $\{a\}$, $\{b\}$, and $\{c\}$, each of which consists of one atomic student.

(103) *two students*



a. Without H: $[\![\phi\mathrm{P}]\!] = [\![\mathrm{NumP}]\!] = \{a \oplus b, b \oplus c, a \oplus c\}$

b. With H:

$$[\![\phi\mathrm{P}]\!] = [\![\mathrm{NumP}]\!] = \left\{ \begin{array}{c} (a \oplus b)^{\Uparrow}, \quad (b \oplus c)^{\Uparrow}, \quad (a \oplus c)^{\Uparrow} \\ (a \oplus b)^{\Uparrow} \sqcup (b \oplus c)^{\Uparrow},\ (a \oplus b)^{\Uparrow} \sqcup (a \oplus c)^{\Uparrow},\ (a \oplus b)^{\Uparrow} \sqcup (a \oplus c)^{\Uparrow} \\ (a \oplus b)^{\Uparrow} \sqcup (b \oplus c)^{\Uparrow} \sqcup (a \oplus c)^{\Uparrow} \end{array} \right\}$$

6.4. Comparing the two approaches

Both the reconstruction approach and the unified approach can properly account for 'conjunction-rejecting' readings. First, both approaches explain why singular-marked and numeral-modified questions admit higher-order readings. In the reconstruction approach, the atomicity/cardinality restrictor in the *wh*-complement can block the application of the H-shifter, but this blocking effect disappears once the *wh*-complement is reconstructed to the question nucleus. In the unified account, since disjunctions can be singular/cardinal, the atomicity/cardinality restrictor in the *wh*-complement does not block the application of H, allowing the Q-domain of a singular-marked/numeral-modified question to range over a set of Boolean disjunctions (and Montagovian individuals).

Second, both approaches explain why questions in these readings reject conjunctive answers. In the reconstruction approach, conjunctive answers are not acceptable because conjoining two uniqueness inferences yields a contradiction. In the unified approach, Boolean conjunctions are not atomic or cardinal, and hence are ruled out immediately by the atomicity/cardinality restrictor within the WH-complement.

Last, both approaches capture the local uniqueness effects. In the reconstruction approach, reconstruction involves THE-assertion which introduces uniqueness. In the unified approach, disjunctions that are considered singular range over a set of atomic entities, and likewise, disjunctions having the cardinality *b* range over a set of entities each of which has the cardinality *n*.

These two approaches, however, are not notational equivalence of each other. First, they attribute the deviance of conjunctive answers to different reasons and thus can make different predictions in some cases. In the reconstruction approach, disjunctive answers are acceptable because disjoining two uniqueness inference does not yield a contradiction. However, the computation in (97a) shows an exception: if disjunctions are interpreted as wides scope FCIs, they would yield contradictions the same as conjunctions. In contrast, the unified approach does not predict disjunctions to be deviant in any case. Unfortunately, it is very hard to check the predictions with real data. Second, the unified approach derives the 'conjunction-rejecting' reading in the very same way as regular higher-order readings, while the reconstruction approach uses a salvaging strategy. As such, one the one hand, the unified approach is technically neater, and on the other hand, the reconstruction approach predicts the general difficulty in interpreting singular-marked and numeral-modified questions with higher-order readings.

7. Conclusion

This paper investigates into the derivation and distribution of higher-order readings of *wh*-questions. First, using diagnostics based on non-reducibility and stubbornly collectivity, I provided three sets of evidence showing that sometimes a *wh*-question must be interpreted with a higher-order reading and have a higher-oder Q-domain. Next, I argued that the meanings involved in a higher-order Q-domain are subject to The Positiveness Constraint — a higher-order Q-domain consists of only positive GQs and their Boolean coordinations. Incorporating this constraint into the meaning of a H-shifter, I proposed that a *wh*-question has a higher-order reading if the H-shifter converts the *wh*-restrictor into a set of higher-order meanings, and if the *wh*-phrase binds a higher-order trace. Finally, I explained the distributional constraints of higher-order readings, including cases where a question admits all types of positive answers, and cases where a question admits only non-conjunctive answers.

Acknowledgement [To be added ...]

References

Alonso-Ovalle, Luis, and Vincent Rouillard. 2018. Number inflection, spanish bare interrogatives, and higher-order quantification. In *Proceedings of NELS 49*.

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

Champollion, Lucas. 2016. Ten men and women got married today: Noun coordination and the intersective theory of conjunction. *Journal of Semantics* 33:561–622.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic inquiry* 37:535–590.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention*. Oxford: Oxford University Press.

Cresti, Diana. 1995. Extraction and reconstruction. *Natural Language Semantics* 3:79–122.

Dayal, Veneeta. 1996. *Locality in Wh Quantification: Questions and Relative Clauses in Hindi*. Dordrecht: Kluwer.

Elliott, Patrick D, Andreea C Nicolae, and Uli Sauerland. 2017. Who and what do who and what range over cross-linguistically. In *Conference: Theoretical and experimental approaches to presuppositions, Genoa, Italy*.

Erlewine, Michael Yoshitaka. 2014. Movement out of focus. Doctoral Dissertation, Massachusetts Institute of Technology.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.

Fox, Danny. 2013. Mention-some readings of questions. *MIT seminar notes* .

Gajewski, Jon. 2008. Npi *any* and connected exceptive phrases. *Natural Language Semantics* 16:69–110.

Gentile, Francesco, and Bernhard Schwarz. 2018. A uniqueness puzzle: *how many*-questions and non-distributive predication. In *Proceedings of Sinn und Bedeutung 21*.

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

Groenendijk, Jeroen, and Martin Stokhof. 1989. Type-shifting rules and the semantics of interrogatives. In *Properties, types and meaning*, 21–68. Springer.

Harbour, Daniel. 2014. Paucity, abundance, and the theory of number. *Language* 90:185–229.

Heim, Irene. 1994. Interrogative semantics and Karttunen's semantics for *know*. In *Proceedings of IATOML1*, volume 1, 128–144.

Hendriks, Herman. 1993. Studied flexibility: categories and types in syntax and semantics. Doctoral Dissertation, ILLC, University van Amsterdam.

Higginbotham, James, and Robert May. 1981. Questions, quantifiers and crossing. *The linguistic review* 1:41–80.

Hirsch, Aron, and Bernhard Schwarz. 2019. Singular which, mention-some, and variable scope uniqueness. In *Proceedings of Semantics and Linguistic Theory 29*.

Hoeksema, Jack. 1988. The semantics of non-boolean "and". *Journal of Semantics* 6:19–40.

Jacobson, Pauline. 2016. The short answer: implications for direct compositionality (and vice versa). *Language* 92:331–375.

Keenan, Edward L, and LM Faltz. 1984. *Boolean semantics for natural language*, volume 23. Springer Science & Business Media.

Keenan, Edward L, and Jonathan Stavi. 1986. A semantic characterization of natural language determiners. *Linguistics and philosophy* 9:253–326.

Krifka, Manfred. 1997. The expression of quantization (boundedness). In *Workshop on Cross-Linguistic Variation in Semantics. LSA Summer Institute. Cornell*.

Landman, Fred. 1989. Groups, i. *linguistics and philosophy* 12:559–605.

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. De Gruyter.

Maldonado, Mora. 2017. Plural marking and d-linking in Spanish interrogatives. URL `https://semanticsarchive.net/Archive/TBiNGExN/Maldonado-PluralSpanish.pdf`.

Partee, Barbara, and Mats Rooth. 1983. Generalized conjunction and type ambiguity. In *Meaning, use, and interpretation of language*, ed. Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow, 334–356. Blackwell Publishers Ltd.

Rullmann, Hotze. 1995. Maximality in the semantics of *wh*-constructions. Doctoral Dissertation, University of Massachusetts at Amherst.

Sauerland, Uli. 2003. A new semantics for number. In *Semantics and Linguistic Theory*, volume 13, 258–275.

Sauerland, Uli, Jan Anderssen, and Kazuko Yatsushiro. 2005. The plural is semantically unmarked. *Linguistic evidence* 413–434.

Schwarzschild, Roger. 1996. *Pluralities*. Springer Science & Business Media.

Schwarzschild, Roger. 2011. Stubborn distributivity, multiparticipant nouns and the count/mass distinction. In *Proceedings of NELS*, volume 39, 661–678.

Scontras, Gregory. 2014. The semantics of measurement: Harvard university dissertation. Doctoral Dissertation.

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

Spector, Benjamin. 2007. Modalized questions and exhaustivity. In *Proceedings of SALT 17*.

Spector, Benjamin. 2008. An unnoticed reading for wh-questions: Elided answers and weak islands. *Linguistic Inquiry* 39:677–686.

Srivastav, Veneeta. 1991. Wh dependencies in hindi and the theory of grammar. Doctoral Dissertation, Cornell University Ithaca, New York.

Szabolcsi, Anna. 1997. *Background notions in lattice theory and generalized quantifiers*, 1–27. Dordrecht: Springer Netherlands.

Winter, Yoad. 2001. Flexibility principles in boolean semantics: The interpretation of coordination, plurality, and scope in natural language. *MA: MIT Press, Cambridge* .

Xiang, Yimei. 2016. Interpreting questions with non-exhaustive answers. Doctoral Dissertation, Harvard University Cambridge, Massachusetts.

Xiang, Yimei. 2019. Function alternations of the mandarin particle dou: Distributor, free choice licensor, and 'even'. *Journal of Semantics* .