Higher-order readings of WH-questions

Yimei Xiang, Rutgers University

Abstract In most cases, a WH-question expects an answer that names an entity in the set denoted by the extension of the WH-complement. However, evidence from questions with necessity modals and questions with collective predicates shows that sometimes a WH-question must be interpreted with a higher-order reading, in which this question expects an answer naming a generalized quantifier.

This paper investigates the distribution and compositional derivation of higher-order readings of WH-questions. First, I argue that the generalized quantifiers that can serve as direct answers to WH-questions must be homogeneously positive. Next, on the distribution of higher-order readings, I observe that questions in which the WH-complement is singular-marked or numeral-modified can be responded by elided disjunctions but not by conjunctions. I further present two ways to account for this disjunction-conjunction asymmetry. In the uniform account, these questions admit disjunctions because disjunctions (but not conjunctions) may satisfy the atomicity requirement of singular-marking and the cardinality requirement of numeral-modification. In the reconstruction-based account, the WH-complement is syntactically reconstructed, which gives rise to local uniqueness and yields a contradiction for conjunctive answers.

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 an answer that names either an entity in the set denoted by the WH-complement or a generalized quantifier (GQ) ranging over of a subset of this set. This requirement is especially robustly seen with short answers to questions. 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.*

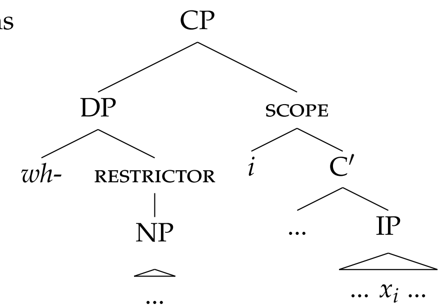
To capture this question-answer relation, it is commonly assumed that WH-phrases are functions (e.g., existential ($\exists$-)quantifiers or function domain restrictors) over first-order predicates, and that the domain for quantification or abstraction is the set denoted by the extension of the WH-complement. An LF schema for WH-questions is given in (2): the WH-phrase combines with a first-order function denoted by the scope and binds an *e*-type variable inside the question nucleus (viz., the IP).

[1] Elided disjunctions are scopally ambiguous relative to this commitment, as described in (i). This paper considers 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.*

(2) LF schema of wh-questions



In this view, the root denotation of a wh-question is either a one-place function defined for values in the extension of the NP-complement, as assumed in categorial approaches and structured meaning approaches, or a set of propositions naming such values, as assumed in propositional approaches (such as Hamblin-Karttunen Semantics, Partition Semantics, and Inquisitive Semantics). For convenience in describing the relation between wh-phrases and wh-questions in meaning, the following presentation follows categorial approaches (Hausser and Zaefferer 1979; Hausser 1983; among others). The core ideas of this paper, however, are independent from the assumptions of categorial approaches on defining and composing questions.

Categorial approaches define questions as functions and wh-phrases as function domain restrictors. In (3), for example, in forming the question *which student came?*, the wh-phrase *which student* applies to a first-order function defined for any individuals and returns a more restrictive first-order function that is only defined for atomic students. I henceforth call this functional denotation of a question a "Q-function" and the domain of a Q-function a "Q-domain".

(3) a. $[\![\textit{which student}]\!] = \lambda P_{et} \lambda x_e \colon \textsf{student}(x). P(x)$

b. $[\![\textit{which student came?}]\!] = [\![\textit{which student}]\!](\lambda x_e. \textsf{came}(x))$
$= \lambda x_e \colon \textsf{student}(x). \textsf{came}(x)$

Treating short answers as bare nominals, categorial approaches regard the relation between matrix questions and short answers as a simple function-argument relation — the Q-function serves as a function for an entity-denoting answer and an argument for a GQ-denoting answer. For example, in (4a), applying the Q-function denoted by the question to an individual denoted by the short answer yields that this individual came and the presupposition that this individual is a student. In (4b), in contrast, 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 $a^{\Uparrow} \cup b^{\Uparrow}$ (i.e., the union of two Montagovian individuals[2]) to the Q-function yields the presupposition that both of the disjoined individuals a and b are students.

[2] Disjunctions over set-denoting expressions are standardly treated as **unions** '∪'. This idea follows a more general schema defined in Partee and Rooth 1983. Since entities are not sets, to be disjoined, they have to be first type-shifted into GQs of a conjoinable type $\langle et, t \rangle$ via Montague-lift. Hence, in a disjunction of two referential DPs, *or* combines with two Montagovian individuals and returns their union (Keenan and Faltz 1985: Part 1A).

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

The conjunctive *and* is commonly treated ambiguously as either an **intersection** operator '∩' (for combining sets, in analogy to the union meaning of *or*) or a **summation** operator '⊕' (for combining entities) (Link 1983; Hoeksema 1988). Another view is to interpret *and* uniformly and attribute the ambiguity to covert operations. For example, Winter (2001) and Champollion (2016b) treat *and* unambiguously an intersection operator and use covert type-shifting operations to derive the summation-like reading.

(4) a. Combining with an entity

$$\llbracket \text{WH-Q} \rrbracket(\llbracket \textit{Andy} \rrbracket) = (\lambda x_e\colon \textsf{student}(x).\textsf{came}(x))(a)$$
$$= \textsf{student}(a).\textsf{came}(a)$$

b. Combining with a GQ

$$\llbracket \textit{Andy or Billy} \rrbracket(\llbracket \text{WH-Q} \rrbracket) = (a^{\Uparrow} \cup b^{\Uparrow})(\lambda x_e\colon \textsf{student}(x).\textsf{came}(x))$$
$$= \textsf{student}(a) \wedge \textsf{student}(b).\textsf{came}(a) \vee \textsf{came}(b)$$

The above discussion considers **first-order readings** of WH-questions. If a question has a first-order reading, the Q-function denoted by this question is a first-order function. However, as first observed by Spector (2007, 2008), in some cases a WH-question can only be properly addressed by an answer that specifies a GQ. For example in (5), the elided disjunction in the answer is interpreted under the scope of the necessity modal *have to*. Spector argues that to obtain this narrow scope reading, *which books* should bind a higher-order trace (of type $\langle et, t \rangle$) across the necessity modal, so that a disjunction can be semantically reconstructed to a scopal position under the modal.

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

Examples like (5) show that questions can also have **higher-order readings**, in which the yielded Q-functions take GQs as arguments. This paper investigates into those higher-order readings.

The rest of this paper is organized as follows. Section 2 discusses cases where a question must be interpreted with a higher-order reading, drawn on evidence from questions with modals and/or collective predicates. Section 3 examines what higher-order meanings can be members of a Q-domain and be used as semantic answers to higher-order questions. I argue that the higher-order meanings involved in a Q-domain must be "homogeneously positive". Sections 4 and 5 investigate the derivation and distributional constraints of higher-order readings. These two sections focus on a puzzling conjunction-disjunction asymmetry — questions with a singular-marked or numeral-modified *wh*-phrase reject conjunctive answers but admit disjunctive answers. I present two ways to account for this asymmetry, including a uniform account and a reconstruction-based account. Section 6 concludes.

2. Evidence for higher-order readings

Saying that a question has a first-order reading yields two predictions regarding to its GQ-naming answers. First, the named GQ must be interpreted with wide scope relative to any scopal expressions in the question nucleus. Second, the answer space (viz., the Hamblin set) of this question consists of only propositions denoted by the entity-naming answers. If an answer names a GQ, the proposition denoted by this answer is not in the answer space of this question, and the named GQ is not in the Q-domain; instead, those answers are derived by applying additional Boolean operations to propositions in the answer space.

This section presents counterexamples to both predictions, showing that first-order readings are insufficient. First, evidence from questions with necessity modals (e.g., *which books does John have to read?*) shows that sometime the Q-domain of a question must contain Boolean disjunctions and existential quantifiers (Sect. 2.1). Second, evidence from questions with collective predicates (e.g., *which children formed one team?*) shows that sometimes the Q-domain of a question must contain Boolean conjunctions and universal quantifiers (Sect.

2.2). Finally, combinations of these two diagnostics rule in the Boolean coordinations of the aforementioned GQs (Sect. 2.3).

2.1. Non-reducibility: Evidence for disjunctions and existential quantifiers

In general, to completely address a question, one needs to provide the strongest true answer to this question (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 seen in (6), in responding to a basic WH-question, a disjunctive answer is always partial/incomplete — whenever the disjunctive answer is true, it is asymmetrically entailed by another true answer, namely, a/the true disjunct.

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

Spector (2007, 2008) observes that, however, disjunctions can completely address WH-questions in which the nucleus contains a necessity modal (called "$\Box$-questions" henceforth). For example in (7), the elided disjunction is scopally ambiguous. If the disjunction takes scope over the necessity modal *have to*, the disjunctive answer has a partial answer reading. Alternatively, if interpreted under the scope of the modal, the elided disjunction can be regarded as a complete specification of John's reading obligations — there is not any specific book that John has to read, his only reading obligation is to choose between the French novels and the Russian novels. This narrow scope complete answer reading is also observed with existential quantifiers, as seen in (8).

(7) a. Which books does John have to read?
    b. The French novels or the Russian novels.
        i. 'John has to read the French novels or the Russian novels. I don't know which exactly.' (Partial: *or* $\gg \Box$)
        ii. 'John has to read the French novels or the Russian novels. The choice is up to him.' (Complete: $\Box \gg$ *or*)

(8) a. Which books does John have to read?
    b. At least two books by Balzac.
        i. 'There are at least two books by Balzac that John has to read. I don't know what they are.' (Partial: $\exists \gg \Box$)
        ii. 'John has to read at least two books by Balzac, which two (or more) to read is up to his own choice.' (Complete: $\Box \gg \exists$)

To obtain the complete answer reading (7b-ii), the elided disjunctive answer must be treated as a GQ (i.e., the Boolean disjunction $f^{\Uparrow} \cup r^{\Uparrow}$) and be reconstructed to a position under the scope of the necessity modal. Thus, Spector (2007) concludes that the $\Box$-question (7a) 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 relative to the modal, respectively. To highlight the contrast between these two readings with respect to the types of the yielded Q-functions, I instead call the two readings the **first-order reading** and the **higher-order reading**, respectively. As paraphrased in (9), the first-order reading expects answers that specify an entity, while the higher-order reading expects answers that specify a GQ.

(9) Which books does John have to read?

a. First-order reading:
'For which book(s) x is such that John has to read x?'

b. Higher-order reading:
'For which a GQ π over books is such that John has to read π?'

Spector assumes that the derivation of the higher-order reading involves **semantic reconstruction** (Cresti 1995; Rullmann 1995): the WH-phrase binds a higher-order trace π (of type $\langle et, t\rangle$) across the necessity modal. Adapting this analysis to the categorial approach, I propose the LFs and Q-functions for the two readings as follows. (The assumed Q-domain for the higher-order reading is subject to revision. '$\textsc{smlo}(\pi)$' stands for the **smallest live-on set** of π. For now, I just assume that the Q-domain is the set of GQs ranging over a set of books.[3]) Observe that, for the higher-order reading, the GQ-denoting answer is interpreted at whatever scopal position that the higher-order *wh*-trace π takes.

(10) First-order reading

a. [which-books λx_e [have-to [John read x]]]

b. $[\![\textsc{wh-Q}]\!] = \lambda x_e : \mathsf{books}_w(x).\Box[\lambda w.\mathsf{read}_w(j, x)]$

c. $[\![F\ or\ R]\!]([\![\textsc{wh-Q}]\!])$
$= (f^{\Uparrow} \cup r^{\Uparrow})(\lambda x : \mathsf{books}_w(x).\Box[\lambda w.\mathsf{read}_w(j, x)])$
$= \mathsf{books}_w(f) \wedge \mathsf{books}_w(r).\Box[\lambda w.\mathsf{read}_w(j, f)] \cup \Box[\lambda w.\mathsf{read}_w(j, r)]$

(11) Higher-order reading

a. [which-books $\lambda \pi_{\langle et,t\rangle}$ [have-to [π λx_e [John read x]]]]

b. $[\![\textsc{wh-Q}]\!] = \lambda \pi_{\langle et,t\rangle} : \textsc{smlo}(\pi) \subseteq \mathsf{books}.\Box[\lambda w.\pi(\lambda x_e.\mathsf{read}_w(j, x))]$ (To be revised)

c. $[\![\textsc{wh-Q}]\!]([\![F\ or\ R]\!])$
$= (\lambda \pi_{\langle et,t\rangle} : \textsc{smlo}(\pi) \subseteq \mathsf{books}.\Box[\lambda w.\pi(\lambda x.\mathsf{read}_w(j, x))])(f^{\Uparrow} \cup r^{\Uparrow})$
$= \textsc{smlo}(f^{\Uparrow} \cup r^{\Uparrow}) \subseteq \mathsf{books}.\Box[\lambda w.(f^{\Uparrow} \cup r^{\Uparrow})(\lambda x.\mathsf{read}_w(j, x))]$
$= \{f, r\} \subseteq \mathsf{books}.\Box[\lambda w.\mathsf{read}_w(j, f) \vee \mathsf{read}_w(j, r)]$

$\Box$-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. 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. $f(x)$ abbreviates for the proposition *John read* x. Arrows indicate entailments. The middle disjunctive symbol '$\vee$' stands for the disjunction of the two atomic sentences; in specific, it stands for $f(a) \vee f(b)$ in Figure 1 and $\Box f(a) \vee \Box f(b)$ in Figure 2.

[3] For any π of type $\langle \tau t, t\rangle$ and set A of type $\langle \tau, t\rangle$, we say that π **lives on** A iff for every set B: $\pi(B) \Leftrightarrow \pi(B \cap A)$ (Barwise and Cooper 1981), and that π **ranges over** A iff A is the **smallest live-on set** (SMLO) of π (Szabolcsi 1997). For example, the smallest live-on set of *some/every/no student* is the set of atomic students. These notions will be crucial for discussions on constraining what types of GQs should and should not be ruled into a Q-domain (see Sect. 3).

$$\begin{tikzcd}
 & f(a) \wedge f(b) \arrow[dl] \arrow[dr] & \\
f(a) \arrow[dr] & \vee \arrow[d, leftrightarrow] & f(b) \arrow[dl] \\
 & f(a) \vee f(b) &
\end{tikzcd}$$

Figure 1: The answer space of *what did John read?*

$$\begin{tikzcd}
 & \Box[f(a) \wedge f(b)] \arrow[dl] \arrow[dr] & \\
\Box f(a) \arrow[dr] & \vee \arrow[d] & \Box f(b) \arrow[dl] \\
 & \Box[f(a) \vee f(b)] &
\end{tikzcd}$$

Figure 2: The answer space of *what does John have to read?*

In Figure 1, 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 to 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 2, 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.

The diagnostic given by Spector can be generalized to the following: $\Box$-questions may yield Q-functions that are "non-reducible" relative to disjunctions and existential quantifiers. The following defines reducibility, where '$\bullet$' stands for the combinatory operation between the function θ and a GQ:[4]

(12) A function θ is **reducible** relative to a GQ π iff $\theta \bullet \pi = \pi(\lambda x.\theta \bullet x^{\Uparrow})$.

As with $\Box$-questions, the following questions ((13) and (14) are taken from Spector (2007)), with a word expressing universal quantification, also have readings where the Q-function is non-reducible relative to disjunctions or to existential quantifiers.

(13) 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?

(14) Modals
  a. Which books is it *sufficient* to read?
  b. Which books is John *required* to read?

(15) Quantifiers
  a. Which books did *all* of the students read?
  b. Which books does John *always*/*usually* read?

2.2. Stubborn collectivity: Evidence for conjunctions and universal quantifiers

Spector (2007, 2008) and Fox (2013) have assumed that a Q-domain may contain Boolean

[4]The definition of '$\bullet$' varies by the semantic type of θ. If the GQ π is of type $\langle \tau t, t\rangle$, we have the following: (i) if θ is of type $\langle \tau tt, t\rangle$, '$\bullet$' stands for the Forward Functional Application; (ii) if θ is of type $\langle \tau, t\rangle$, '$\bullet$' stands for the Backward Functional Application; (iii) if θ cannot compose with a GQ directly, then either '$\bullet$' involves a type-shifting operation or $\theta \bullet \pi$ is undefined.

conjunctions, but they have not provided empirical evidence for this assumption. Clearly, the non-reducibility diagnostic generalized in (12) does not extend to Boolean conjunctions: the Q-functions of □-questions as well as those discussed in (13) to (15) are reducible relative to Boolean conjunctions.

(16) a. $[\lambda\pi.\textsf{\textit{J has to read }}\pi](\textsf{\textit{f}}^{\Uparrow} \cup \textsf{\textit{r}}^{\Uparrow}) \neq \textsf{\textit{J has to read f}} \vee \textsf{\textit{J has to read r}}$

b. $[\lambda\pi.\textsf{\textit{J has to read }}\pi](\textsf{\textit{f}}^{\Uparrow} \cap \textsf{\textit{r}}^{\Uparrow}) = \textsf{\textit{J has to read f}} \wedge \textsf{\textit{J has to read r}}$

This section introduces a new diagnostic for ruling in Boolean conjunctions. This diagnostic draws on the fact that questions with a stubbornly collective predicate (e.g., *formed a team*, *co-authored two papers*) may have answers naming Boolean conjunctions, and especially that stubborn collectivity in these questions does not trigger uniqueness.

First, to see what is stubborn collectivity, observe that the phrasal predicate *formed a/one team* admits a collective reading but not a covered/ (non-atomic) distributive reading. The sentence (17a) 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 covered/ (non-atomic) distributive reading and thus (17b) can be truthfully uttered.

(17) (w: *The four relevant children abcd formed exactly two teams in total:* $a + b$ *formed one, and* $c + d$ *formed one.*)

a. # The children formed a/one team.

b. $\surd$ The children formed teams.

Note that the falsehood of (17a) is not improved even if the context has explicitly separated the four children into two pairs, as seen in (18).

(18) [Yesterday, the pair of children $a + b$ competed against the pair of children $c + d$.] Today, the children (all) formed a/one team.

($^{\text{OK}}$collective, $^{\#}$(non-atomic) distributive)

Hence, I call the predicate *formed a/one team* "stubbornly collective", in contrast to predicates like *lift the piano* which admits both collective and covered/distributive readings. Stubborn collectivity is widely observed with quantized phrasal predicates of the form "V + counting noun", such as *formed one committee* and *co-authored two papers*.[5]

Second, for the absence of uniqueness effects, compare the sentences in (19a-b) in the same discourse. The declarative-embedding sentence (19a) suffers a presupposition failure, because the factive verb *know* embeds a false collective declarative. However, the sentence (19b), where *know* embeds the interrogative counterpart of this collective declarative, does not suffer a presupposition failure. Moreover, intuitively, (19b) implies that John knows precisely the component members of each team formed by the considered children, which is a conjunctive inference.

(19) (w: *The four relevant children abcd formed exactly two teams in total:* $a + b$ *formed one, and* $c + d$ *formed one.*)

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

[5] A predicate P is **quantized** iff whenever P holds for x, P does not hold for any proper subpart of x (Krifka 1997). Formally: $\forall x \forall y[P(x) \wedge P(y) \rightarrow [x \leq y \rightarrow x = y]]$. Defining predicates as sets of events, Champollion (2016a) argues that distributive readings are not available with quantized predicates because the extension of a quantized verbal phrase is not closed under summation formation.

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

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

The conjunctive inference in (19c) is quite surprising — 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 (19a) would admit a covered/distributive reading and (19a) would be felicitous, contra fact. In contrast, I argue that this conjunctive closure is provided by the WH-phrase: the WH-phrase quantifies over a set of higher-order meanings including the Boolean conjunction $(a \oplus b)^{\Uparrow} \cap (c \oplus d)^{\Uparrow}$.

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

a. [which-children $\lambda\pi_{\langle et,t\rangle}$ [$_{\text{IP}}$ π λx_e [$_{\text{VP}}$ x formed a team]]]

b. $[\![\text{WH-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \text{SMLO}(\pi) \subseteq \textsf{children}.\lambda w[\pi(\lambda x.\textsf{f.a.team}_w(x))]$

c. $[\![\text{WH-Q}]\!]((a \oplus b)^{\Uparrow} \cap (c \oplus d)^{\Uparrow})$
$= \{a \oplus b, c \oplus d\} \subseteq \textsf{children}.\lambda w[\textsf{f.a.team}_w(a \oplus b) \wedge \textsf{f.a.team}_w(c \oplus d)]$

One might suggest to ascribe the conjunctive closure to an operator outside the question denotation, such as Heim's (1994) answerhood-operator ANS-H. As schematized in (21), ANS-H 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.

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

b. $\bigcap\{\lambda w.\textsf{f.a.team}_w(a \oplus b), \lambda w.\textsf{f.a.team}_w(c \oplus d)\}$
$= \lambda w.\textsf{f.a.team}_w(a \oplus b) \wedge \textsf{f.a.team}_w(c \oplus d)$

However, ANS-H is insufficient as it cannot capture the contrast with respect to uniqueness in (22). The question-embedding sentence (22b) is infelicitous because the embedded numeral-modified question (viz., the embedded question in which the WH-complement is numeral-modified) has a uniqueness presupposition which contradicts the context.

(22) (w: *The four relevant children abcd formed exactly two teams in total: $a+b$ formed one, and $c+d$ formed one.*)

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

b. # John knows [**which two children** formed a team].
$\rightsquigarrow$ *Only two of the children 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). For a question with a Hamblin set Q, its strongest true answer is the true proposition in Q entailing all the true propositions in Q. In the rest of this section, I argue that the contrast between (22a-b) is due to the following: in (22a), the embedded simple plural-marked question has a strongest true answer in the given discourse, while in (22b), the embedded numeral-modified question does not.

Dayal's presupposition is originally motivated to explain the uniqueness requirement of singular-marked WH-questions (i.e., questions in which the WH-complement is singular-marked). In Srivastav 1991, she observes that a singular-marked WH-question cannot have multiple true answers. For illustration, compare the examples in (23). The continuation in (23a) is infelicitous because the preceding singular-marked question has a uniqueness presupposition that only one of the children came. In contrast, this inconsistency disappears if the singular WH-phrase *which child* is replaced with the plural phrase *which children* or the bare *wh*-word *who*, as in (23b-c).

(23) a. "Which child came? # I heard that many children came."
b. "Which children came? I heard that many children came."
c. "[Among the children,] who came? I heard that many children came."

To capture the uniqueness presuppositions of singular-marked questions, Dayal (1996) defines a presuppositional answerhood-operator ANS-D which checks the existence of the strongest true answer. Applying ANS-D to a world w and the Hamblin set Q returns the unique strongest of the propositions in Q true in w and presupposes the existence of this strongest true proposition.

(24) $$\text{Ans-D}(w)(Q) = \exists p[w \in p \in Q \wedge \forall q[w \in q \in Q \rightarrow p \subseteq q]].\\ \iota p[w \in p \in Q \wedge \forall q[w \in q \in Q \rightarrow p \subseteq q]]$$

Adopting the ontology of individuals by Sharvy (1980) and Link (1983), Dayal assumes that the Hamblin set of a singular-marked WH-question is smaller than that of its plural-marked counterpart. The ontology of individuals assumes that both singular and plural nouns denote sets of entities. In particular, a singular noun denotes a set of atomic entities, while a plural noun denotes a set consisting of both atomic and sum entities.[6] If sums are defined in terms of part-hood relation, this ontology can be represented as in Figure 3. Letters *abc* each denotes an atomic child. Lines indicate *part of* relations from bottom to top. For example, atomic entities a and b are parts of their sum $a \oplus b$.

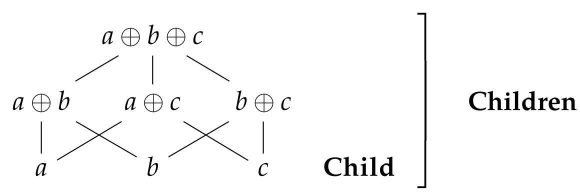


Figure 3: Ontology of individuals (Sharvy 1980; Link 1983)

Accordingly, as illustrated in (25), the Hamblin set of the singular-marked question includes only propositions naming an atomic child, while the Hamblin set of the corresponding plural-marked question includes also propositions naming a sum of children. Q_w stands for the set of propositions in Q that are true in w, namely, the Karttunen set in w. As a result, in a discourse where both Andy and Bill came, (25b) has a strongest true answer $\lambda w.\textsf{came}_w(a \oplus b)$ while (25a) does not, and then employing ANS-D in (25a) gives rise to a presupposition failure. To avoid this presupposition failure, the singular-marked question (25a) can only be felicitously uttered in a world where only one of the children came, which therefore explains its uniqueness requirement.

(25) (w: *Among the considered children, only Andy and Billy came.*)

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

a. Which child came?

  i. $Q = \{\lambda w.\textsf{came}_w(x) \mid x \in \textsf{child}\}$

  ii. $Q_w = \{\lambda w.\textsf{came}_w(a), \lambda w.\textsf{came}_w(b)\}$

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

b. Which children came?

  i. $Q = \{\lambda w.\textsf{came}_w(x) \mid x \in \textsf{children}\}$

  ii. $Q_w = \{\lambda w.\textsf{came}_w(a), \lambda w.\textsf{came}_w(b), \lambda w.\textsf{came}_w(a \oplus b)\}$

  iii. $\textsc{Ans-D}(w)(Q) = \lambda w.\textsf{came}_w(a \oplus b)$

It is also straightforward that, to account for the uniqueness presupposition, the Q-domain yielded by a singular-marked WH-phrase must exclude Boolean conjunctions such as $a^{\Uparrow} \cap b^{\Uparrow}$. Otherwise, the singular-marked question (25a) would admit conjunctive answers like $\lambda w.\textsf{came}_w(a) \wedge \textsf{came}_w(b)$ and would not be subject to uniqueness, contra fact.[7]

Numeral-modified questions also have a uniqueness presupposition. For example, the numeral-modified question in (26a) implies that only two of the children came, and the one in (26b) implies that only two or three of the children came. Both inferences contradict each of their continuations.

(26) a. "Which two children came? # I heard that three children did."

  b. "Which two or three children came? # I heard that five children did."

Dayal's account of uniqueness easily extends to numeral-modified questions. As seen in (27), for a question of the form "which n children came?" where n is a bare numeral and is read as 'exactly n', Dayal's presupposition is satisfied only if exactly n children came. If the number of children who came is smaller than n, this question has no true answer (viz., $Q_w = \emptyset$); if the number of the children who came is larger than n, the question does not have a strongest true answer.

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

  a. $Q = \{\lambda w.\textsf{came}_w(x) \mid x \in \textsf{2-children}\}$

[7] Drawing on facts from Spanish *quién* 'who.SG' which is singular-marked but does not trigger uniqueness (Maldonado 2020), Elliott et al. (2020) 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 (2019) argue against this view: 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 answer (ib) has a conjunction reading that the students formed a group and the professor formed a group. The felicity of this answer shows that the *quién*-question admits answers naming Boolean conjunctions over non-atomic elements. Alonso-Ovalle & Rouillard thus conclude that *quién* is number-neutral in meaning and is semantically ambiguous — it ranges over either a set of atomic and non-atomic individuals or a set of Boolean conjunctions and disjunctions.

b. $Q_w = \{\lambda w.\textsf{came}_w(a \oplus b), \lambda w.\textsf{came}_w(a \oplus c), \lambda w.\textsf{came}_w(b \oplus c)\}$

c. ANS-D$(w)(Q)$ is undefined

The same as in a singular-marked WH-question, the uniqueness effect shows that the Q-domain of a numeral-modified WH-question does not contain Boolean conjunctions; otherwise, (27) would have a strongest true answer based on $(a \oplus b)^{\Uparrow} \cap (c \oplus d)^{\Uparrow} \cap (b \oplus c)^{\Uparrow}$.

Return to the contrast of question-embeddings in (22), repeated below:

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

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

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

The contrast is explained if we assume that the Q-domain of a basic plural-marked question contains Boolean conjunctions, while that of a numeral-modified question does not. More specifically, in (28a), the Q-domain yielded by *which children* includes Boolean conjunctions and hence the embedded question *which children formed a team* admits conjunctive answers. In the given scenario, the Boolean conjunction $(a \oplus b)^{\Uparrow} \cap (c \oplus d)^{\Uparrow}$ yields the strongest true answer. In contrast, in (28b), the Q-domain yielded by *which two children* consists of only pluralities denoting sums of two children (e.g., $a \oplus b$ and $c \oplus d$), and hence the embedded question in (28b) has two true answers including $\lambda w.\textsf{f.a.team}_w(a \oplus b)$ and $\lambda w.\textsf{f.a.team}_w(c \oplus d)$, but neither of them counts as the strongest true answer. In conclusion, (28b) is infelicitous because the embedded question does not satisfy Dayal's presupposition, and this presupposition failure projects over the factive predicate *know*.[8]

2.3. Evidence for complex GQ-coordinations

Previous sections provide two diagnostics for simplex GQs. The diagnostic based on non-reducibility validates the existence of Boolean disjunctions and existential quantifiers in a Q-domain. The diagnostic based on stubbornly collectivity provides evidence for Boolean conjunctions and universal quantifiers. Combining these two diagnostics, the following shows that a Q-domain also contains complex GQ-coordinations:

> Context: 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:	$\{c_2, d_2\}$

[8]It is worthy noting that the argumentation for ruling in Boolean conjunctions is independent from whether using Dayal's presupposition to explain the uniqueness effects — any account of uniqueness has to explain the contrast between *which children* and *which two children* in admitting Boolean conjunctions. Recent literature has found evidence for other ways to encode or derive uniqueness. For example, to account for the projection of uniqueness presuppositions and existential presuppositions in question-embeddings, Uegaki (2018, 2020) argues to encode the uniqueness inference within the question nucleus (details omitted). Moreover, as we will see in section 5.1, Hirsch and Schwarz (2020) observe that in questions with a possibility modal, the uniqueness inference triggered by a singular-marked WH-phrase can be interpreted under the scope of the possibility modal. To derive local uniqueness, Uegaki (2020) and Hirsch and Schwarz (2020) propose that uniqueness is introduced by the lexicon of a WHICH-determiner that appears within the nucleus. For this account to explain the contrast between *which children* and *which two children*, it still has to assume that *which children* may range over Boolean conjunctions, while *which two children* cannot.

(29) 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 raised by the guest involves a necessity modal *have to* as well as a stubbornly collective predicate *present a paper together*. The answer provided by the instructor can be unpacked as follows: the disjunctive answer conveys overall a free choice inference as in (30a), and the choices are specified as in (30b-c). ('p.a.p.t.' is abbreviated for 'present a paper together'.)

(30) 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 derive the free choice inference (30a), the disjunction must be interpreted under the scope of the necessity modal. Further, since the predicate *present a paper together* is stubbornly collective, to derive the conjunctive inferences in (30b-c), each disjunct/choice must be understood as naming a Boolean conjunction over two pairs of students. In sum, the answer should be interpreted with the following scopal pattern: $\Box \gg$ *or* $\gg$ *and* $\gg$ *a paper*. To derive this scopal pattern, the nucleus of this question should contain a higher-order *wh*-trace in between the necessity modal and the collective predicate, as in (31). Instructor's answer should be read as naming a Boolean disjunction over two Boolean conjunctions, namely, $((a_1 \oplus b_1)^{\Uparrow} \cap (c_1 \oplus d_1)^{\Uparrow}) \cup ((a_2 \oplus b_2)^{\Uparrow} \cap (c_2 \oplus d_2)^{\Uparrow})$.

(31) [which-students $\lambda \pi_{\langle et,t \rangle}$ [$_{\text{IP}}$ have-to [π λx_e [x present a paper together]]]]

2.4. Interim summary

To sum up, this section discusses cases where a question must be interpreted with a higher-order reading and provides two diagnostics to rule in higher-order meanings into a Q-domain. The first diagnostic is based on narrow scope readings of GQ-naming answers to questions in which the Q-function is non-reducible relative to the named GQs. Results of this diagnostic rule in Boolean disjunctions and a class of existential quantifiers. 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 quantifiers. In addition, combining these two diagnostics, I further show that a Q-domain contains also complex GQ-coordinations.

3. Constraints on the Q-domain

The previous section has shown that the Q-domain of a *wh*-question may contain Boolean disjunctions, conjunctions, a class of existential quantifiers, universal quantifiers, as well as their Boolean coordinations. One might wonder whether we can make the following generalization:

In a higher-order reading, the Q-domain yielded by a wh-phrase consists of all GQs ranging over a subset of the set denoted by the extension of the wh-complement as well as the Boolean combinations of these GQs.

In what follows, I will show that this generalization is too strong. Spector (2007, 2008) provides some counterexamples to this generalization and argues that the GQs included in a Q-domain must be **increasing**.[9] Extending Spector's diagnostic to non-monotonic GQs and GQ-coordinations, I show that the increasing-ness requirement is too strong. In contrast, I argue that whether a higher-order meaning can be ruled into a Q-domain and be used as a semantic answer of a higher-order wh-question is determined by its **positiveness** (roughly, the property of ensuring existence with respect to certain quantification domain, see Sect. 3.2): the higher-order meanings involved in a Q-domain must be homogeneously positive.

3.1. The Completeness Test and The Increasing-ness Constraint

Whether a meaning is included in the Q-domain of a question can be examined by the **Completeness Test** generalized in (32). This test draws on a deductive relation between attitudes held towards a question and attitudes held towards the answers to this question: the question-embedding sentence *x knows Q* implies that *x* knows the complete true answer to the embedded question Q. The complete answer to 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 to Q, p is not in the Hamblin set of Q.[10][11]

(32) **The Completeness Test** (generalized from Spector 2008)
For any proposition p that names a short answer α to a question Q: if there is a world in which both p and *x knows Q* are true but *x knows p* is not true, then p is not in the Hamblin set of Q, and α is not in the Q-domain of Q.

[9] Monotonicity of GQs is defined as follows:

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

[10] This paper considers only questions with at most one complete true answer, which is the strongest true answer. For mention-some questions which can have multiple complete true answers, see Fox (2013) and Xiang (2016b: chapter 2-3).

[11] The Completeness Test does not aim to fully characterize the truth conditions of a question-embedding sentence or to exhaustively determine what can and what cannot be included in a Q-domain. First, this test is only concerned about one aspect of the truth conditions of question-embedding sentences, namely, the Completeness condition. In addition to Completeness, question-embeddings are also subject to a false-answer sensitivity condition. (Klinedinst and Rothschild 2011; George 2013; Cremers and Chemla 2016; Uegaki 2015; Xiang 2016a,b; Theiler et al. 2018; among others) For example, for the sentence (33b) being true, Sue can be ignorant about whether John should read any books by Betty, but she cannot have the false belief that John should read some book(s) by Betty.

(i) 'x knows Q' is true iff
  a. x knows a/the complete true answer of Q. (Completeness)
  b. x does not have any false belief relevant to Q. (False-answer sensitivity)

Second, the Completeness Test is only used to determine what **is not** in a Q-domain, not to determine what **is** in a Q-domain. A bi-conditional characterization could cause conflicting predictions. As I will show in Sect. 3.3, the Completeness Test shows that, for the higher-order Q-domain of the question *what does John have to read*, we should rule in simplex non-monotonic quantifiers like *exactly three books* while excluding decreasing quantifiers like *less than four books*, despite that the propositional answer yielded by the former (i.e., *John has to read exactly three books*) asymmetrically entails the propositional answer yielded by latter (i.e., *John has to read less than four books*).

For simple illustration, consider the truth conditions of the question-embedding sentence (33b) under the context described in (33a). Strikingly, the sentence (33b) implies that Sue knows John's reading obligation (a-i), but not that she knows (a-ii); Sue can be ignorant about whether John should read any books by Betty.[12]

(33) a. Context: John's reading obligations include the following:
(i) he must read at least two books by Anne; (ii) he must read no book by Betty.

b. Sue knows which books John must read.
$\rightsquigarrow$ *Sue knows (a-i).*
$\not\rightsquigarrow$ *Sue knows (a-ii).*

Given this contrast, Spector (2008) proposes that the GQs used as direct semantic answers to higher-order questions must be **increasing**. 'x knows Q' implies that x knows the complete/strongest true answer to Q; therefore, that Sue can be ignorant about the reading obligation (a-ii) excludes the decreasing quantifier *no book by Betty* and the non-monotonic GQ-coordination *at least two books by Anne and no book by Betty* from the Q-domain.[13] The higher-order reading of the question *which books John must read* is then paraphrased as follows: 'for which increasing GQ π over books, it is the case that John has to read π?'

3.2. The Positiveness Constraint

The following example applies the Completeness Test to a broader range of GQs. The game requirements listed in (34a) each name a GQ ranging over a set of cards. Among those GQs, (i-ii) are increasing, (iii-iv) are decreasing, and (v) is non-monotonic. These GQs are all interpreted with narrow scope relative to the necessity modal *require to*. The two exceptives are also read with narrow scope: (ii) implies that John is allowed not to play the largest black club in his hand, and (iv) implies that he is allowed to play the largest red heart in his hand. These exceptions give the player flexibility to determine which two Kings to play to fulfill the requirement (v). Intuitively, the question-embedding sentence (34b) implies that Sue knows about not only the game requirements (i-ii) but also (v). In particular, for the condition regarding to her knowledge towards (v), it is insufficient if Sue knows that John has to play at least two Kings but does not know that he cannot play more than two Kings. Spector's Increasing-ness Constraint incorrectly rules out the non-monotonic GQ *exactly two Kings* and fails to predict that (34b) implies that Sue knows the requirement (v).

[12] Surprisingly, in contrast to (33b), the following two sentences with a concealed question or a definite description do imply that Sue knows all of John's reading obligations list in (33a).

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

[13] The Completeness Test in (i) considers two more cases that involve GQ-disjunctions (underlined). This test further confirms that Boolean disjunctions involving a decreasing GQ-disjunct must be excluded from a Q-domain.

(i) a. Context: John's reading obligations for the summer consist of the following:
i. he must read <u>no leisure book or more than two math books</u>. (In other words, John has to read more than two math books if he reads any leisure book.)
ii. he must read <u>none or all of the Harry Potter books</u>, (because Harry Potter books must be rented in a bundle, and it would be a waste of money if he rents the entire series but only reads part of them.)

b. Sue knows which books John has to read in the summer.
$\not\rightsquigarrow$ *Sue knows (a-i)/(a-ii).*

(34) a. Context: John is playing a board game. This game requires him to play ...
   i. {at least three, more than two} black spades; ♠
   ii. every black club except the largest black club in his hand; ♣
   iii. {at most three, less than four} red diamonds; ♦
   iv. no red heart except the largest red heart in his hand; ♥
   v. exactly two Kings;

   b. Sue knows which cards John must play.
   $\rightsquigarrow$ *Sue knows that John must play (a-i), (a-ii), (a-v).*
   $\not\rightsquigarrow$ *Sue knows that John must play (a-iii)/(a-iv).*

More generally, any monotonicity-based constraint would face a dilemma — we want to rule out non-monotonic GQ-coordinations (e.g., *at least two books by Anne and no book by Betty*) while not excluding simplex non-monotonic GQs (e.g., *exactly two hearts*).

In contrast to Spector (2008), I propose that whether a simplex or complex GQ should be ruled into a Q-domain is determined by its "positiveness", not its monotonicity.

(35) **The Positiveness Constraint** (To be revised in (37))
GQs in the Q-domain of a WH-question must be positive.

A GQ being positive means that the meaning of this GQ ensures existence with respect to the set it ranges over (namely, its smallest live-on set, see definitions in footnote 3). 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 formal definition of positiveness is as follows, where $\mathbb{E}$ stands for the existential quantifier (i.e., $\mathbb{E} =_{\text{df}} \lambda Q \lambda P . Q \cap P \neq \varnothing$):

(36) For any π of type $\langle et, t \rangle$, π is **positive** iff $\pi \subseteq \lambda P_{\langle e,t \rangle} . \mathbb{E}(\textsc{smlo}(\pi))$.

Table 1 compares monotonicity and positiveness for a list of GQs that range over a set of books. (a and b are two distinct atomic books). Observe that increasing GQs are all positive, decreasing ($\downarrow_{\text{MON}}$) GQs are all non-positive, while non-monotonic (N.M.) GQs can be either positive or non-positive.

Generalized quantifier π	$\textsc{smlo}(\pi)$	Increasing?	Positive?
$a^{\Uparrow}$	$\{a\}$	Yes	Yes
$a^{\Uparrow} \cap b^{\Uparrow}$, $a^{\Uparrow} \cup b^{\Uparrow}$	$\{a, b\}$	Yes	Yes
{at least, more than} two books	*books*	Yes	Yes
every book except a	*book* $- \{a\}$	Yes	Yes
{at most, less than} two books	*books*	No ($\downarrow_{\text{MON}}$)	No
no book except a	*book* $- \{a\}$	No ($\downarrow_{\text{MON}}$)	No
less than three or more than ten books	*books*	No (N.M.)	No
every book or no book	*book*	No (N.M.)	No
exactly two books	*books*	No (N.M.)	Yes
two to four books	*books*	No (N.M.)	Yes
some but not all books	*books*	No (N.M.)	Yes
(exactly) two or four books	*books*	No (N.M.)	Yes
an even number of books	*books*	No (N.M.)	Yes

Table 1: Increasing-ness/monotonicity versus positiveness

3.3. The Homo-Positiveness Constraint

Table 1 considers only coordinations over Montagovian individuals and GQs of the simplex form 'Det+NP'. Benjamin Spector (pers. comm.) points out that, however, the Positiveness Constraint does not exclude the unwanted non-monotonic GQ-coordinations such as *every article and no book* and *some article and no book*: letting $\pi = [\![\textit{every article and no book}]\!]$ and representing π as $\{X \mid A \subseteq X \wedge B \cap X = \emptyset\}$, we have $\textsc{smlo}(\pi) = A \cup B$ and $\pi \subseteq \mathbb{E}(A \cup B)$.[14]

Basically, to include a non-monotonic GQ in a Q-domain, it is insufficient to require existence with respect to the set that the entire GQ ranges over. Instead, thinking of a non-monotonic GQ as a Boolean coordination of increasing GQs and decreasing GQs (for example, *every article and no book* is the conjunction of *every article* and *no book*, and *exactly two books* is the conjunction of *at least two books* and *no more than two books*), we should require existence relative to both (i) the set that the coordinated increasing GQs range over and (ii) the set that the coordinated decreasing GQs range over. For example, *every article and no book* should be excluded from a Q-domain as it ensures existence with respect to the set of articles but not to the set of books. Hence, I propose to strengthen the Positiveness Constraint (35) to the following:

(37) **The Homo-Positiveness Constraint** (Final)
GQs in the Q-domain of a WH-question must be homogeneously positive.

In most cases, the coordinated GQs cannot be semantically retrieved out of their coordination. First, some GQs cannot be decomposed into a simple coordination of monotonic GQs. For example, the decomposition in (38a) has to involve at least disjunctions over conjunctions of monotonic GQs. Second, even for a GQ that can be decomposed into a simple coordination, there are multiple ways to decompose it, as seen in (38b-c).

(38) a. $[\![\textit{exactly 2 A or exactly 4 B}]\!]$
$= [\![\textit{exactly 2 A}]\!] \cup [\![\textit{exactly 4 B}]\!]$
$= ([\![\textit{at least 2 A}]\!] \cap [\![\textit{at most 2 A}]\!]) \cup ([\![\textit{at least 4 B}]\!] \cap [\![\textit{at most 4 B}]\!])$

b. $[\![\textit{every A and no B}]\!]$
$= [\![\textit{every A}]\!] \cap [\![\textit{no B}]\!]$

c. $[\![\textit{every A and no B}]\!]$
$= [\![\textit{every A or some B}]\!] \cap [\![\textit{no B}]\!]$
$= ([\![\textit{every A}]\!] \cup [\![\textit{some B}]\!]) \cap [\![\textit{no B}]\!]$

However, for the purpose of determining whether a GQ-coordination is homogeneously positive, we just need to find out the involved strongest increasing and decreasing GQs

[14] The following explains why $A \cup B$ is the smallest live-on set of π, where $\pi = \{X \mid A \subseteq X \wedge B \cap X = \emptyset\}$. First, (i) shows that $A \cup B$ is a live-on set of π: replacing X with $X \cap (A \cup B)$ in the set description does not change the set.

(i) $$\left\{ X \;\middle|\; \begin{array}{l} [A \subseteq (X \cap (A \cup B))] \wedge \\ [B \cap (X \cap (A \cup B)) = \emptyset] \end{array} \right\} = \left\{ X \;\middle|\; \begin{array}{l} [A \subseteq X \wedge A \subseteq (A \cup B)] \wedge \\ [(B \cap (A \cup B)) \cap X = \emptyset] \end{array} \right\} = \pi$$

Next, (ii) shows that $A \cup B$ is the smallest live-on set: for any a, replacing X with $X \cap (A \cup B - \{a\})$ in the set description makes no change to the set being defined iff $a \notin A \cup B$.

(ii) $$\left\{ X \;\middle|\; \begin{array}{l} [A \subseteq (X \cap ((A \cup B) - \{a\}))] \wedge \\ [B \cap (X \cap ((A \cup B) - \{a\})) = \emptyset] \end{array} \right\} = \left\{ X \;\middle|\; \begin{array}{l} [A \subseteq X \wedge A \subseteq (A \cup B - \{a\})] \wedge \\ [(B \cap ((A \cup B) - \{a\})) \cap X = \emptyset] \end{array} \right\}$$
$$= \left\{ X \;\middle|\; \begin{array}{l} [A \subseteq X \wedge A \subseteq (A \cup B) \wedge a \notin A] \wedge \\ [(B - \{a\}) \cap X = \emptyset] \end{array} \right\}$$
$$= \pi \text{ iff } a \notin A \text{ and } a \notin B$$

that determine the lower and upper bounds of this GQ-coordination. I define homogenous positiveness as in (39). π^+ is the logically strongest increasing GQ entailed by π that determines the lower bound of π, and π^- is the logically strongest decreasing GQ entailed by π that determines the upper bound of π.

(39) For any π of type $\langle et, t\rangle$, π is **homogeneously positive** iff
  a. $\pi \subseteq \mathbb{E}(\textsc{smlo}(\pi^+))$, where $\pi^+ =_{\text{df}} \{P \mid \exists P' \subseteq P[\pi(P')]\}$;
  b. $\pi \subseteq \mathbb{E}(\textsc{smlo}(\pi^-))$, where $\pi^- =_{\text{df}} \{P \mid \exists P' \supseteq P[\pi(P')]\}$.

Table 2 compares the three parameters for a broader range of GQs. (A and B are two noun phrases whose extensions are non-overlapped.) Observe that whether a GQ π is homogeneously positive is independent from whether π is (non-)monotonic, upper (un-)bound, and dis/con-joined.[15]

Generalized quantifier π	π^+	π^-	Increasing?	Positive?	Homo-positive?
at most 2 B	$D_{\langle e,t\rangle}$	at most 2 B	No ($\downarrow$MON)	No	No
less than 2 or more than 5 B	$D_{\langle e,t\rangle}$	$D_{\langle e,t\rangle}$	No (N.M.)	No	No
every A or no B	every A	no B	No (N.M.)	No	No
every A or no A	every A	no A	No (N.M.)	No	No
every A and no B	every A	no B	No (N.M.)	Yes	No
at least 2 B	at least 2 B	$D_{\langle e,t\rangle}$	Yes	Yes	Yes
an even number of B	at least 2 B	$D_{\langle e,t\rangle}$	No (N.M.)	Yes	Yes
exactly 2 to 4 B	at least 2 B	at most 4 B	No (N.M.)	Yes	Yes
exactly 2 or 4 B	at least 2 B	at most 4 B	No (N.M.)	Yes	Yes
exactly 2 B	at least 2 B	at most 2 B	No (N.M.)	Yes	Yes

Table 2: Increasing-ness versus Positiveness versus Homo-positiveness

The generalizations made in Table 1 still hold here. First, a non-positive GQ also cannot be homogeneously positive. Second, for any increasing π, the retrieved π^- is trivial, and thus this increasing π being positive ensures π being homogeneously positive.[16] For non-monotonic GQs, however, the Homo-Positiveness Constraint yields a different prediction. The simplex non-monotonic GQ *exactly two books* is positive as well as homogeneously positive: *exactly two books* entails *some books*, and the retrieved π^+ *at least two books* and π^- *no more than two books* both range over the set `books`. In contrast, the complex non-monotonic GQ-coordination *every article and no book* is positive but not homogeneously positive: the retrieved π^- *no book* ranges over the set `book`, but *every article and no book* does not entail *some book*.[17]

[15] In an earlier version (Xiang 2019), treating positiveness and homogeneity as two separate conditions, I incorrectly claimed that any π of type $\langle et, t\rangle$ can be decomposed into a conjunction $\pi^+ \cap \pi^-$ and proposed that π is homogenous if π is monotonic or if π^+ and π^- range over the same set. However, as pointed out by Lucas Champollion (pers. comm.), the equation $\pi = \pi^+ \cap \pi^-$ does not hold for disjoined GQs such as *an even number of cards* and *exactly two or four cards*.

[16] If a GQ π is unbound, then one or both of the strongest GQs retrieved from π are trivial. See more details in (i). This trivial GQ (viz., $D_{\langle e,t\rangle}$) ranges over the discourse domain D_e. Increasing GQs are upper-unbound, and decreasing GQs are lower-unbound.

(i) a. If π is upper-unbound, namely, $\forall P[P \in \pi \rightarrow \exists P' \in \pi[P' \supseteq P]]$, then $\pi^- = D_{\langle e,t\rangle}$.
  Example: *at least two books*, *an even number of books*, *less than two or more than four books*
 b. If π is lower-unbound, namely, $\forall P[P \in \pi \rightarrow \exists P' \in \pi[P' \subseteq P]]$, then $\pi^+ = D_{\langle e,t\rangle}$.
  Example: *less than two or more than four books*, *at most four books*.

[17] A puzzle arises about complex non-monotonic GQ-coordinations such as *some book but no leisure book*, where

3.4. Interim summary

In summary, the Q-domain yielded by the phrase '*wh*-A' in a higher-order reading, if any, is the set consisting of the homogeneously positive GQs ranging over a subset of A. I write this set as ${}^{\text{H}}\!A$.

(40) ${}^{\text{H}}\!A = \{\pi_{\langle et,t\rangle} \mid \textsc{smlo}(\pi) \subseteq A \wedge \pi \text{ is homogeneously positive}\}$, where
π is homogeneously positive iff $\pi \subseteq \mathbb{E}(\textsc{smlo}(\pi^{+}))$ and $\pi \subseteq \mathbb{E}(\textsc{smlo}(\pi^{-}))$.

It is yet unclear where the homo-positiveness constraint comes from. It could be in the lexicon of a type-shifting operator, presupposed by the higher-order *wh*-trace, or a constraint on semantic reconstruction. For now, I just treat 'H' as a syntactically presented operator asserting homo-positiveness. (For distributional constraints of this operator, see section 4.) Then, the first-order/higher-order ambiguity of a WH-question can be attributed to the absence/presence of the H-shifter within the WH-phrase. As exemplified in (41), 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 homogenously positive GQs, and then the WH-phrase binds a higher-order trace π across the modal verb.

(41) Which books does John have to read?
a. **First-order reading**
[which-books λx_e [have-to [John read x]]]
$[\![\text{WH-Q}]\!] = \lambda x_e \colon x \in \textit{books}_w . \Box\lambda w.\textit{read}_w(j, x)$
b. **Higher-order reading** ($\Box \gg \pi$) (Revised from (11))
[which-${}^{\text{H}}$books $\lambda \pi_{\langle et,t\rangle}$ [have-to [π λx_e [John read x]]]]
$[\![\text{WH-Q}]\!] = \lambda \pi_{\langle et,t\rangle} \colon \pi \in {}^{\text{H}}\textit{books}_w . \Box\lambda w.\pi(\lambda x.\textit{read}_w(j, x))$

4. Distributing 'conjunction-admitting' higher-order readings

As discussed in section 2.2, uniqueness effects in WH-questions show 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:

(42) a. Which child came?
$\leadsto$ *Exactly one of the children came.*

---

the extensions of the two involved noun phrases are overlapped. In (i), telling John that he has to read a book is clearly insufficient — John might incorrectly think that reading a leisure book suffices for this requirement. Hence, it is appealing to say that the embedding sentence (i) entails both (ia) and (ib), and that the complete answer to the embedded question names the GQ-coordination *some book but no leisure book*. However, this GQ-coordination is not homogeneously positive — it does not entail existence relative to the set of leisure books.

(i) (Context: John has to read a book, but he is not allowed to read any leisure books.)
Sue will tell John what he has to read.
a. *Sue will tell John that he has to read a book.*
b. *Sue will tell John that he cannot read any leisure books.*

I argue that the Homo-positiveness Constraint still holds here. In the described context, the Completeness condition of (i) should be read as (ic). The complete true short answer to *what John has to read* is the existential quantifier *some non-leisure book*, which is stronger than *some book* and is homogeneously positive.

c. *Sue will tell John that he has to read a non-leisure book.*

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

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

According to Dayal (1996), the singular-marked question (42a) presupposes uniqueness because its strongest true answer exists only when it has exactly one true answer. This analysis also extends to the numeral-modified questions (42b-c), as argued in section 2.2. Adopting this analysis of uniqueness, I have concluded 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 (44) and (45) for example. Although *students (who are) in a group of two* is semantically similar to *two students*, the embedded question in (45) 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 readings is sensitive to the internal structure of the WH-complement.

(43) I know $\left\{\begin{array}{l}\textbf{who}\\ \text{which } \textbf{students}\end{array}\right\}$ presented a paper together,

a. ... the two boys.

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

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

a. ... the two boys.

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

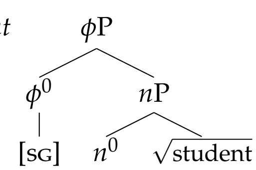
(45) I know which **students (who are) in a group of two** presented a paper together,

a. ... the two boys.

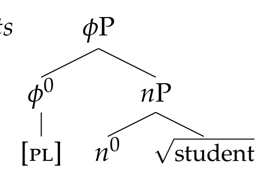
b. ... the two boys and the two girls.

To account for the above distributional constraints, I assume that the H-shifter (viz., the operator that turns a set of entities into a set of GQs) must be applied locally to the nP within the WH-complement. In what follows, I argue that the application of H is blocked in singular-marked nouns and numeral-modified nouns due to conflicts in meaning and types. First, I assume the following structure for a singular/plural bare noun:

(46) a. *student*



b. *students*



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

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

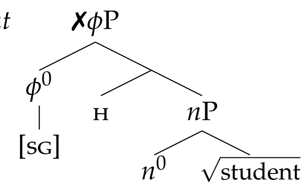
b. $[\![[\textsc{sg}]]\!] = \lambda P_{\langle e,t\rangle}\lambda x_e.\textsc{Atom}(x) \wedge P(x)$

c. $[\![[\textsc{pl}](n\mathrm{P})]\!] = [\![n\mathrm{P}]\!] = \{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$

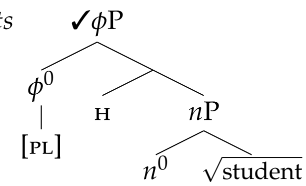
d. $[\![[\textsc{sg}](n\mathrm{P})]\!] = \{a, b, c\}$

The above assumptions straightforwardly explain why the H-shifter cannot be used in singular nouns. In (48a), applying the H-shifter to nP returns a set of GQs, which are non-atomic and conflict with the atomicity requirement of [SG]. (See a possible refinement of this view in Sect. 5.2.) Hence, the H-shifter cannot be applied in a singular-marked WH-question because it would yield an empty Q-domain. In contrast, the H-shifter can be freely used in simple plural-marked and number-neutral WH-questions because in these questions the [PL] feature carried by ϕ^0 is semantically vacuous.[18]

(48) a. *student*

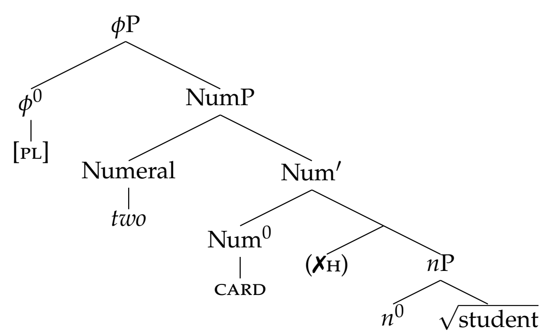


b. *students*



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 (49a), CARD combines with a predicate P and a numeral n and returns the set of individuals x such that P holds for x and x is constituted of exactly-n atoms. These assumptions easily explain why the H-shifter cannot be used in a numeral-modified NP: the CARD-predicate at Num^0 checks the cardinality of the elements in the set it combines with and hence it cannot combine with a set of GQs. (See a possible amendment of this generalization in Sect. 5.2.)

(49) *two students*



a. $[\![\textsc{card}]\!] = \lambda P \lambda n \lambda x.P(x) \wedge |x| = n$

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

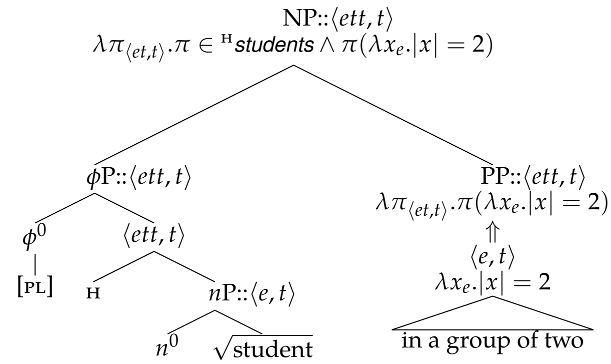
c. With the H-shifter
$[\![\mathrm{Num}']\!]$ is undefined (or Num′ has type-mismatch)

In contrast to numeral-modifiers, PP-modifiers are adjoined to the entire NP/ϕP. Hence, the H-shifter can be used in the modified NP without causing a type-mismatch. As illustrated

[18] This claim holds regardless of whether plurals are treated inclusively or exclusively. One can also treat [PL] as a predicate restrictor that asserts/presupposes non-atomicity or anti-presupposes atomicity (see also footnote 6). I also assume that the bare *wh*-words *who* and *what* have a structure similar to *which people/things*.

in (50), all we need is applying argument-lifting to the PP-modifier and shifting it into a set of GQs. Then, the lifted PP composes with the higher-order ϕP standardly via Predicate Modification. This analysis also extends to NPs modified by a relative clause.

(50) *students in a group of two*



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

5.1. The puzzles

In section 2.2, based on stubborn collectivity and uniqueness effects, I showed that singular-marked questions and numeral-modified questions do not admit answers naming Boolean conjunctions. I further concluded in section 4 that these questions do not have higher-order readings and explained this distributional constraint. The explanation attributed the unavailability of higher-order readings to that applying the H-shifter yields semantic consequences that conflict with the atomicity requirement of singular nouns and the cardinality requirement of numerals.

Surprisingly, in responding to a $\Box$-question where the WH-phrase is singular-marked or numeral-modified, narrow scope disjunctions are not as bad as conjunctions. This contrast is seen clearly in (51) and marginally in (52).[19]

(51) I know which book John has to read,
  a. # ... Book A and Book B.
  b. ? ... Book A or Book B. (#*or* $\gg \Box$, ?$\Box \gg$ *or*)

(52) 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*)

Narrow scope readings of disjunctions are even more readily available in discourse. In (53), the disjunction in the 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 of non-reducibility in section 2.1, that the disjunctive answer admits a

[19] The conjunctive continuation in (52a) is intuitively more acceptable than the conjunctive continuation in (51a), as pointed out by Gennaro Chierchia (pers. comm.). A reviewer of *Natural Language Semantics* also reported that they found no clear contrast between (52a-b). One possibility for the improvement in (52a) is that the numeral *two* alone can be reconstructed to the nucleus, which yields a simple plural-marked question roughly read as 'which books are two books that John have to read?'

narrow scope reading suggests that here the □-question admits higher-order answers, which conflicts with the aforementioned generalization that singular-marked questions do not have higher-order readings.

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

A similar fact is observed in 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; for a discussion on what is mention-some, see Xiang 2016b: chapter 2). As exemplified in (54), 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 the addressee to exhaustively list out all the feasible options. Crucially, MA-answers of ◇-questions can have either an elided conjunctive form, as in (54b), or an elided disjunctive form read as free choice, as in (54c). 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 & Grammar* for this class.

(54) What can we use [as a textbook] for this class?
   a. *Heim & Kratzer*. MS
   b. *Heim & Kratzer* and *Meaning & Grammar*. Conjunctive MA
   c. *Heim & Kratzer* or *Meaning & Grammar*. Disjunctive MA

Xiang (2016b) proposes that MS-readings are higher-order readings: in the LF of a ◇-question with a MS-reading, the WH-phrase binds a higher-order trace across the possibility modal. MA-readings of ◇-questions arise as long as one of the following conditions is met: (i) the higher-order *wh*-trace takes wide scope, or (ii) this trace is associated with an operator with a meaning akin to the Mandarin free choice licensing particle *dou*. In particular, condition (i) yields conjunctive MA, and condition (ii) yields disjunctive MA. I will give more details of this analysis in section 5.3.2.

It is commonly believed that MS-readings and multi-choice readings are unavailable in singular-marked ◇-questions because these questions presuppose uniqueness (Fox 2013; Xiang 2016b: chapter 3). The infelicity of the continuations in (55) supports this view: the continuations name multiple choices of textbooks, while the preceding question-embedding sentence implies that there is only one feasible choice.

(55) I know which textbook we can use for this class, ...
   a. # ... *Heim & Kratzer* and *Meaning & Grammar*.
   b. ? ... *Heim & Kratzer* or *Meaning & Grammar*.

However, Hirsch and Schwarz (2019) novelly observe that the matrix singular-marked ◇-question in (56) does admit a multi-choice reading.[20] They argue that the singular WH-phrase triggers uniqueness but the uniqueness presupposition can be accommodated under the scope of the modal verb *could*. The question can be read as 'for which x, it is the case that x is the unique letter missing in *fo__m*?'.

(56) Which letter could be missing in *fo__m*? (Hirsch and Schwarz 2019)

[20]Here I cite the example from the original poster presentation of Hirsch and Schwarz (2019). See updates of this work in the proceedings paper Hirsch and Schwarz 2020.

a. (The missing letter could be) *a*.
b. The missing letter could be *a* and the missing letter could be *r*.

Note that in example (56), the multi-choice answer (56b) is not a direct answer. As seen in (57a-b), in the form congruent with the question or in the form of a short answer, the conjunctive answers are greatly degraded. In contrast, the multi-choice inference can be felicitously expressed in the form of an elided free choice disjunction, as in (57c). The same pattern is seen with numeral-modified WH-questions, as shown in (58). (For reasons why (58b) is marginally acceptable, see footnote 19.)

(57) 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*.
  c. *a* or *r*. (Both are possible.)

(58) [Hearing a rhotic back vowel] Which two letters could be missing in *f_ _m*?
  a. ?? *or* and *ar*.
  b. *or* or *ar*.

To directly compare with the number-neutral ◇-question (54), I re-illustrate Hirsch and Schwarz's (2019) idea in (59). According to Hirsch and Schwarz, the uniqueness presupposition triggered by the singular-marked WH-phrase *which textbook* can be interpreted globally or locally. The global uniqueness reading says that there is only one 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 feasible option, as in a MS-reading, or all the feasible options, as in a MA-reading. In contrast to the numeral-neutral question (54) to which an elided MA-answer can be either a conjunction or a disjunction, here an elided MA-answer must be a disjunction, as seen in (54a-b). The disjunction-conjunction contrast is also seen with the universal free choice item *any book*, which is argued to be existential in lexicon (Chierchia 2006, 2013), and the basic universal quantifier *every book*.

(59) 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.

In sum, singular-marked and numeral-modified ◇-questions admit multi-choice readings if uniqueness is interpreted locally. In multi-choice readings, their MA-answers must be expressed in the form of a free choice disjunction, not a conjunction.

Gentile and Schwarz (2018) observe similar local uniqueness readings with *how many*-questions. The same as singular-marked and numeral-modified WH-phrases, *how many* presuppose uniqueness. For example, the question in (60) cannot be felicitously responded by a multi-choice answer expressed by the conjunction of two cardinal numerals. Given that the predicate *solved this problem together* is stubbornly collective, Gentile and Schwarz conjecture from the uniqueness effect that the Q-domain of this question does not include Boolean conjunctions over numerals.

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

Further, Gentile and Schwarz observe that possibility modals can obviate violations of uniqueness in *how many*-questions: the question in (61) admits multi-choice answers like (61a-b). In analogy to the examples in (57-59), I add that the multi-choice answer cannot be expressed by an elided conjunction, as seen in (61c).

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

In what follows, I call the reading in which a question admits a higher-order conjunctive answer the 'conjunction-admitting' higher-order reading and the higher-order reading in which a question rejects higher-order conjunctive answers the 'conjunction-rejecting' higher-order reading. Two puzzles arise from the above observations. First, why singular-marked and numeral-modified questions can be directly responded by elided disjunctions but not elided conjunctions? Second, why this 'conjunction-rejecting' higher-order reading is available despite that the WH-phrase is singular-marked or numeral-modified, in contrast to the 'conjunction-admitting' higher-order reading discussed in section 4?

The following sections provide two approaches to derive the 'conjunction-rejecting' higher-order reading and explain its distributional constraints. One approach treats the 'conjunction-rejecting' reading the very same reading as the 'conjunction-admitting' reading but gives a weaker semantics to singular and numeral-modified nouns (Sect. 5.2). The other approach considers the 'conjunction-rejecting' reading a special higher-order reading: the derivation of this reading involves reconstructing the WH-complement to the question nucleus and interpreting uniqueness locally (Sect. 5.3). Both approaches can well explain the puzzles.

5.2. A uniform approach

The uniform approach treats the 'conjunction-rejecting' the very same reading as the 'conjunction-admitting' higher-order reading. The core idea of this approach comes from a personal communication with Manuel Križ. To unify these two readings, all we need is to allow some of the Boolean disjunctions to be atomic or cardinal, just like entities.

In the following definitions, the (a)-condition on minimal witness sets ensures the atomic/cardinal GQ to be a 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.[21] For example, the Boolean conjunction $a^{\Uparrow} \cap b^{\Uparrow}$ is ruled out because its only minimal witness set $\{a, b\}$ is not singleton. The (b)-condition applies the

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

original singularity/cardinality-requirement to each element in the smallest live-on set of π. For example, the Montagovian individual $(a \oplus b)^{\Uparrow}$ and the Boolean disjunction $(a \oplus b)^{\Uparrow} \cup c^{\Uparrow}$ are ruled out by (62b) because their smallest live-on sets (viz., $\{a \oplus b\}$ and $\{a \oplus b, c\}$, respectively) include a non-atomic element $a \oplus b$.

(62) A GQ π is atomic iff
  a. the minimal witness sets of π are all singleton sets;
  b. every member in the smallest live-on set of π is atomic.

(63) A GQ π has the cardinality n iff
  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.

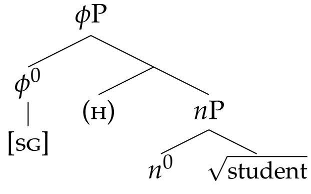
With the above assumptions, I formalize the definitions of the singular feature [SG] and the cardinality predicate CARD as in (65), where '$\textsc{mws}(A, x)$' is read as 'A is a minimal witness set of x'. Both [SG] and CARD are now treated as polymorphic restrictors which can combine with either predicates of entities or predicates of higher-order meanings.

(64) Old definitions
  a. $[\![[\textsc{sg}]]\!] = \lambda P_{\langle e,t\rangle} \lambda x_e . P(x) \wedge \textsc{Atom}(x)$
  b. $[\![\textsc{card}]\!] = \lambda P_{\langle e,t\rangle} \lambda n \lambda x_e . P(x) \wedge |x| = n$

(65) New definitions
  a. $[\![[\textsc{sg}]]\!] = \lambda P \lambda x. \begin{cases} P(x) \wedge \textsc{Atom}(x) & \text{if } P \subseteq D_e \\ P(x) \wedge \forall A[\textsc{mws}(A, x) \rightarrow |A| = 1] \wedge \\ \qquad \forall y \in \textsc{smlo}(x)[\textsc{Atom}(y)] & \text{if } P \subseteq D_{\langle et,t\rangle} \end{cases}$
  b. $[\![\textsc{card}]\!] = \lambda P \lambda n \lambda x. \begin{cases} P(x) \wedge |x| = n & \text{if } P \subseteq D_e \\ P(x) \wedge \forall A[\textsc{mws}(A, x) \rightarrow |A| = 1] \wedge \\ \qquad \forall y \in \textsc{smlo}(x)[|y| = n] & \text{if } P \subseteq D_{\langle et,t\rangle} \end{cases}$

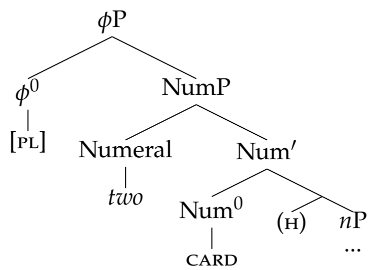
With the revised definitions, the H-shifter can be used regularly in singular nouns and numeral-modified nouns. In a discourse with three students abc, the singular noun *student* and the numeral-modified noun *two students* are interpreted as follows. Again, nP denotes a set of entities closed under summation formation: $[\![n\mathrm{P}]\!] = \{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$.

(66) *student*



  a. Without H:
  $[\![\phi\mathrm{P}]\!] = \{a, b, c\}$
  b. With H:
  $[\![\phi\mathrm{P}]\!] = \{a^{\Uparrow}, b^{\Uparrow}, c^{\Uparrow}, a^{\Uparrow} \cup b^{\Uparrow}, a^{\Uparrow} \cup c^{\Uparrow}, b^{\Uparrow} \cup c^{\Uparrow}, a^{\Uparrow} \cup b^{\Uparrow} \cup c^{\Uparrow}\}$
  $= \{\bigcup A \mid A \subseteq \{x^{\Uparrow} \mid x \in \{a, b, c\}\}\}$

(67) *two students*



a. Without H:

$[\![\phi\mathrm{P}]\!] = [\![\mathrm{NumP}]\!] = \{a \oplus b, b \oplus c, a \oplus c\}$

b. With H:

$$
\begin{aligned}
[\![\phi\mathrm{P}]\!] &= [\![\mathrm{NumP}]\!] \\
&= \left\{ \begin{array}{ccc} (a \oplus b)^{\Uparrow}, & (b \oplus c)^{\Uparrow}, & (a \oplus c)^{\Uparrow} \\ (a \oplus b)^{\Uparrow} \cup (b \oplus c)^{\Uparrow}, & (a \oplus b)^{\Uparrow} \cup (a \oplus c)^{\Uparrow}, & (a \oplus b)^{\Uparrow} \cup (a \oplus c)^{\Uparrow} \\ & (a \oplus b)^{\Uparrow} \cup (b \oplus c)^{\Uparrow} \cup (a \oplus c)^{\Uparrow} & \end{array} \right\} \\
&= \{\bigcup A \mid A \subseteq \{x^{\Uparrow} \mid x \in \{a \oplus b, b \oplus c, a \oplus c\}\}
\end{aligned}
$$

In sum, in the uniform approach, higher-order readings of WH-questions are uniformly derived as follows. First, applying an H-shifter to the *n*P within the WH-complement yields a higher-order domain. In particular, if the WH-phrase is number-neutral or bare plural, the yielded domain is the set of homogeneously positive GQs ranging over a subset of $[\![n\mathrm{P}]\!]$; if the WH-phrase is singular-marked, the yielded domain is a set of disjunctions of Montagovian individuals $x^{\Uparrow}$ where x is an atomic element in $[\![n\mathrm{P}]\!]$; if the WH-complement is modified by a numeral N, the yielded domain is a set of disjunctions of $x^{\Uparrow}$ where x is an entity in $[\![n\mathrm{P}]\!]$ with N-many atomic subparts. Second, the shifted *wh*-phrase binds a higher-order trace in the nucleus, yielding a higher-order Q-function as the root denotation of the question.[22]

5.3. A reconstruction approach

In contrast to the uniform approach, the reconstruction approach assumes that the derivation of the 'conjunction-rejecting' reading requires additional machinery — the WH-complement is syntactically reconstructed to the question nucleus. In what follows, I will present the derivation and explain the consequences of employing reconstruction, especially why it enables singular-marked $\Box$-questions to have narrow scope disjunctive answers (Sect. 5.3.1). Then I will extend this analysis to $\Diamond$-questions and show how this analysis accounts for the contrast between disjunctive and conjunctive MA-answers (Sect. 5.3.2).

5.3.1. $\Box$-questions

Let me start with a singular-marked $\Box$-question. (68) provides the rough LF structures and the yielded Q-functions for first-order and higher-order readings with local uniqueness. In both LF structures, the singular-marked WH-complement *book* is syntactically reconstructed

[22]Given that it is possible to account for the uniqueness effects with singular-marked and numeral-modified WH-phrases while assuming a higher-order reading, one might wonder whether *wh*-questions have first-order readings at all. I so far do not see any direct evidence to object to this claim. However, assuming no first-order reading would lead to a prediction that the application of the H-shifter is mandatory, which is conceptually problematic. The presence of the H-shifter should be independent from the *wh*-determiner since it is applied locally to the root *n*P; thus, if the H-shifter were mandatory, we would expect that any NP has only a higher-order reading. However, in *the student* for example, the complement of *the* has to be interpreted as a set of entities, not as a set of GQs.

to a position in the nucleus c-commanded by the necessity modal. This reconstruction has two consequences. First, it leaves a semantically unmarked variable D as the restrictor of the WH-determiner, which can be type-lifted freely by the H-shifter without causing a type-mismatch or a violation to atomicity. Thus, a higher-order reading arises if the H-shifter is applied to the D variable and if the WH-phrase binds a higher-order trace, as in (68b). Second, uniqueness is evaluated at whichever scopal position that the reconstructed noun adjoins to. In both (68a-b), uniqueness takes scope below the necessity modal.[23]

(68) Which book does John have to read?

a. First-order reading ($\Box \gg \iota$)
'For which entity x_e, it has to be the case that x is the book that John read?'

i. $[_{\text{CP}}$ which$_D$ λx_e $[_{\text{IP}}$ $\Box$ $[x$ 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 ($\Box \gg \pi \gg \iota$)
'For which (homogeneously positive) $\pi_{\langle et,t\rangle}$, it has to be the case that for π x, x is the book that John read?'

i. $[_{\text{CP}}$ which$_{^{\text{H}}D}$ $\lambda \pi_{\langle et,t\rangle}$ $[_{\text{IP}}$ $\Box$ $[\pi$ λx_e. x 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 trees illustrate the two LF structures in more details. The reconstruction of the WH-complement is realized in three steps. 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.[24] In particular, in the LF (70) for the higher-order reading, the same as what is assumed for conjunction-admitting readings, here the *wh*-restrictor (viz., the domain variable D) is type-raised by a H-shifter, and the WH-phrase binds a higher-order trace π across the necessity modal.[25]

[23] Luis Alonso-Ovalle (pers. comm.) points out that the assumed local uniqueness inference might be too strong for $\Box$-questions. For example, the question-answer pair in (i) can be felicitously uttered 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 dynamically relative to an updated context, namely, the context where the player has a bunch of cards in hand and only needs two more cards to close the game.

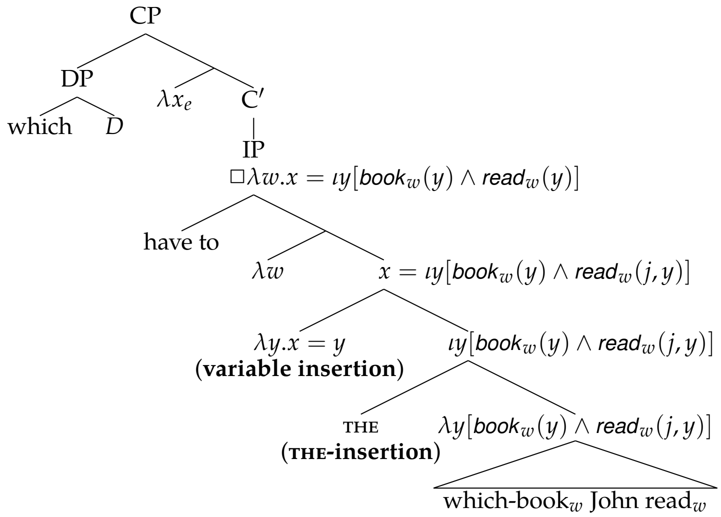
[24] One might have concerns with the assumed syntax for reconstruction. The assumed THE-insertion and variable insertion, on the one hand, are similar to the operations of determiner replacement and variable insertion used in trace conversion (Fox 2002) especially backward trace conversion (Erlewine 2014). On the other hand, in trace conversion, THE-insertion and determiner replacement are locally applied to the moved DP *which book*, while in my proposal, THE-insertion and variable insertion apply to a larger constituent DP+VP *which book John read*. I admit that the assumed syntax for reconstruction is unconventional, but this is not necessarily a problem for considering (70) as the structure that derives the 'conjunction-rejecting' reading. As seen in section 5.1, this reading itself is a bit unnatural. It is much harder to obtain than the conjunction-admitting reading, especially in question-embeddings (see (51-52) and (55)). Thus, it is 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 considered question.

[25] I assume a locality constraint that the variable introduced by variable insertion has to be the variable directly bound by the WH-phrase. With this assumption, in the LF for the higher-order reading, variable insertion introduces a higher-order variable π; it cannot be as follows where it introduces an individual variable x bound by the higher-order *wh*-trace:

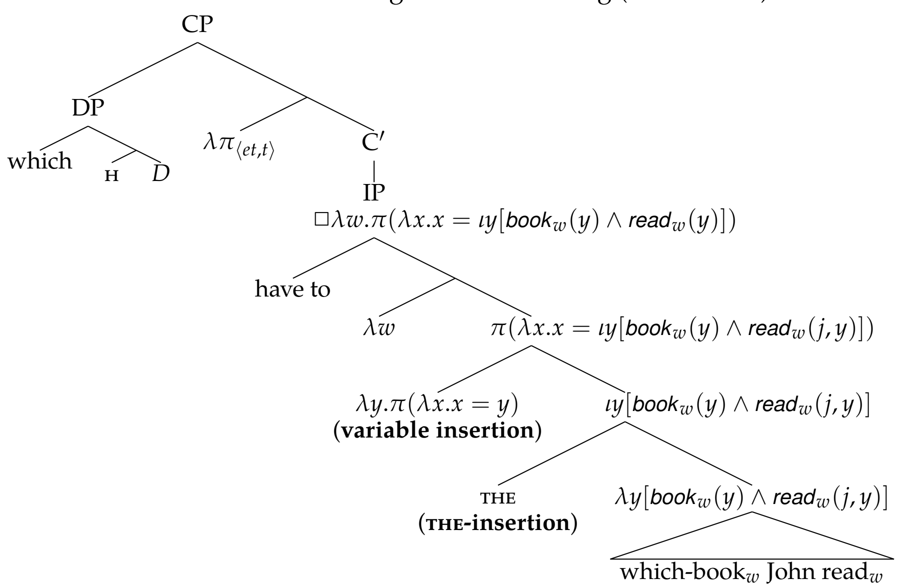
(i) * [*wh*P $\lambda \pi_{\langle et,t\rangle}$ [have to [π λx_e [$\underline{\lambda y.x = y}$ [THE [which book John read]]]]]

This constraint 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

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



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



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 shown in (71b), unless Book A and B are the same book, combining the Q-function with the conjunction $a^{\Uparrow} \cap b^{\Uparrow}$ yields a contradiction.

---

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 requests to rule 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]]]]]]

(71) Which book does John have to read?

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

a. Book A or Book B.

$$[\![\text{WH-Q}]\!](a^{\Uparrow} \cup b^{\Uparrow}) = \Box\lambda w[[a = \iota y[\mathsf{book}_w(y) \wedge \mathsf{read}_w(j,y)]]\vee [b = \iota y[\mathsf{book}_w(y) \wedge \mathsf{read}_w(j,y)]]]$$

(It has to be the case that the unique book that John read is Book A or that the unique book that John read is Book B.)

b. # Book A and Book B.

$$[\![\text{WH-Q}]\!](a^{\Uparrow} \cap b^{\Uparrow}) = \Box\lambda w[[a = \iota y[\mathsf{book}_w(y) \wedge \mathsf{read}_w(j,y)]]\wedge [b = \iota y[\mathsf{book}_w(y) \wedge \mathsf{read}_w(j,y)]]]$$

(#It has to be the case that the unique book that John read is Book A and that the unique book that John read is Book B.)

5.3.2. $\Diamond$-questions

The MA-answer to a question is the true answer that entails all the true answers to this question. As seen in section 5.1, in responding to a $\Diamond$-question, the MA-answer can be expressed in the form of a conjunction or a free choice disjunction. However, if the WH-phrase in the $\Diamond$-question is singular-marked of numeral-modified, the MA-answer can only be expressed in the form of a free choice disjunction. As argued in Xiang 2016b: chapter 2, the MA-reading expecting a conjunctive answer and the MA-reading expecting a disjunctive answer are derived via different LF structures. With the assumed reconstruction, the contrast in LF naturally predicts the conjunction-disjunction asymmetry in conjunction-rejecting readings of singular-marked $\Diamond$-questions.

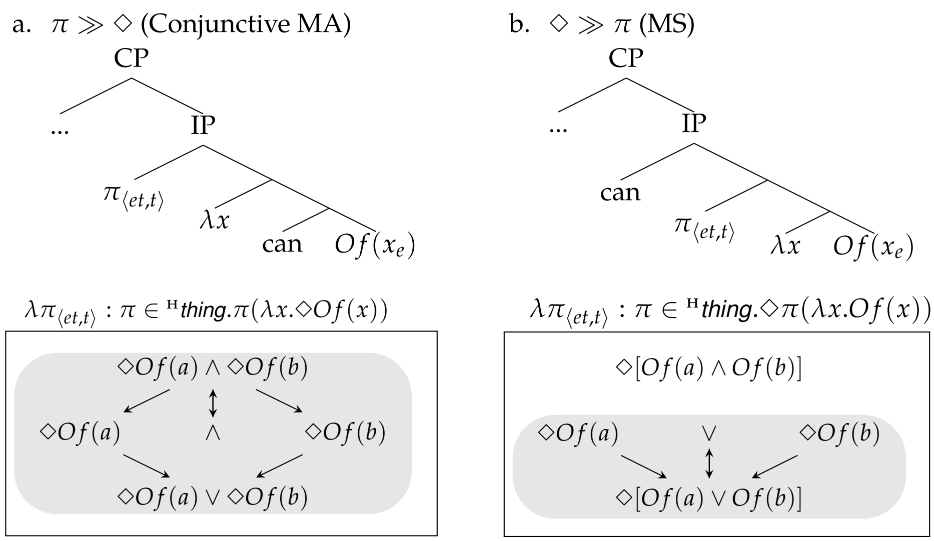
In the conjunctive MA-reading, the WH-phrase binds a higher-order trace which takes scope above the possibility modal. The following considers the interpretations of a number-neutral $\Diamond$-question in cases where the higher-order *wh*-trace (π) takes scope below and above the possibility modal. For each case, (72a/b) illustrates the structure of the question nucleus and the yielded Q-function and answer space.[26] The illustration of the answer space considers only the propositions derived by applying the Q-function to the conjunction $a^{\Uparrow} \cap b^{\Uparrow}$, the Montagovian individuals $a^{\Uparrow}$ and $b^{\Uparrow}$, and the disjunction $a^{\Uparrow} \cup b^{\Uparrow}$. Arrows indicate entailment relations among the propositions, and shades mark the answers that are true in the described world. f stands for the predicate $\mathsf{use\ (as\ a\ textbook)\ for\ this\ class}$ (e.g., $\Diamond Of(a)$ is read as 'Book A can be used as the only textbook for this class.')

(72) (*w: Book A and Book B each can be used as the only textbook for this class; no other book can be used as a textbook for this class.*)
What can we use [as a textbook] for this class? Book A and Book B.

[26] In both structures, a local exhaustivity operator 'O' ($\approx$ *only*) is associated with the individual trace x.

(i) $O_C(p) = \lambda w.p(w) = 1 \wedge \forall q \in C[p \not\subseteq q \rightarrow p(w) = 0]$ (Chierchia et al. 2012)

Xiang (2016b: chapter 2) argues that MS-reading arises if the higher-order trace π scopes below the possibility modal, and the local O-operator 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.



In (72a) where the *wh*-trace π scopes above the possibility modal, the conjunctive answer derived by combining the Q-function with the Boolean conjunction $a^{\Uparrow} \cap b^{\Uparrow}$ entails all the true answers, and thus it is the complete/MA- answer to the $\Diamond$-question. This conjunctive answer is read as 'it is possible that we use Book A as the only textbook for this class, and it is possible that we use Book B as the only textbook for this class.' In contrast, in (72b) where π scopes under the possibility modal, the inference derived based on $a^{\Uparrow} \cap b^{\Uparrow}$ is a contradiction (and therefore is not shaded), read as '# it is possible that we use Book A as the only textbook for this class and we use Book B as the only textbook for this class.' In short, the take-away point is that conjunctive MA-answers are available only if the higher-order WH-trace π scopes above the possibility modal ($\pi \gg \Diamond$).

Next, consider the correponding singular-marked $\Diamond$-question in (73). Again, the puzzle is that multi-choice answers to this question cannot have an elided conjunctive form. As assumed in section 5.3.1, the derivation of the higher-order reading of a singular-marked WH-question involves syntactically reconstructing the WH-complement. Reconstructing the singular noun *book* and letting the higher-order WH-trace π take scope above the possibility modal yield the following scopal pattern: $\pi \gg \iota \gg \Diamond$. As shown in (73b), unless A and B are the same book, combining the derived Q-function with the Boolean conjunction $a^{\Uparrow} \cap b^{\Uparrow}$ yields a contradiction.

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

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

b. $[\![\text{WH-Q}]\!](a^{\Uparrow} \cap b^{\Uparrow}) = \lambda w[[a = \iota y[\textit{book}_w(y) \wedge \Diamond_w Of(y)]] \wedge$
$[b = \iota y[\textit{book}_w(y) \wedge \Diamond_w Of(y)]]]$

(#*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.)

In contrast to conjunctive MA, disjunctive MA arises only if the higher-order *wh*-trace is associated with an DOU-operator, regardless of whether this trace takes scope below or above the possibility modal. The DOU-operator is the covert counterpart of the Mandarin particle *dou*. Despite of many different uses of *dou*, the most important uses for the purpose of interpreting $\Diamond$-questions are the following: in $\Diamond$-questions, associating *dou* with a WH-phrase blocks the MS-reading, as seen in (74a); in $\Diamond$-declaratives, associating *dou* with a pre-verbal disjunction yields a free choice (FC) inference, as shown in (74b). (For other uses of *dou* and a unified analysis, see Xiang 2016b: chapter 7 and Xiang 2020.) It is thus appealing to unify the derivation of free choice disjunction in $\Diamond$-declaratives and the derivation of disjunctive MA-readings of $\Diamond$-questions.

(74) 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.'

Xiang (2016b, 2020) defines *dou* as a pre-exhaustification exhaustifier over sub-alternatives: *dou* affirms its propositional argument and negates the exhaustification of each of the sub-alternatives of its propositional argument. The alternations in function of *dou* come from minimal variations with the semantics of sub-alternatives (details omitted). In particular, for a disjunctive sentence of the form $\Diamond(\phi \vee \psi)$ or the form $\Diamond\phi \vee \Diamond\psi$, the sub-alternatives are $\Diamond\phi$ and $\Diamond\psi$. The covert DOU is semantically identical to *dou* except that it does not presuppose non-vacuity. With this semantics, applying *dou*/DOU to a disjunctive sentence yields a universal free choice inference.

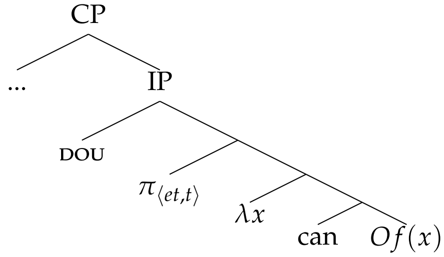
(75) $[\![dou_C]\!] = \lambda p \lambda w : \exists q \in \textsc{Sub}(p, C).p(w) = 1 \wedge \forall q \in \textsc{Sub}(p, C)[O_C(q)(w) = 0]$
(For any proposition p and world w, $[\![dou_C]\!](p)(w)$ is defined only if C contains a sub-alternative of p. When defined, $[\![dou_C]\!](p)(w)$ asserts that p is true in w, and that for any q that is a sub-alternative of p, the exhaustification of q is false in w.)

(76) $[\![\textsc{dou}_C]\!] = \lambda p \lambda w : p(w) = 1 \wedge \forall q \in \textsc{Sub}(p, C)[O_C(q)(w) = 0]$

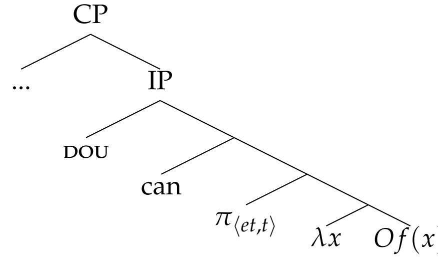
The following illustrates two possible structures of the question nucleus for the disjunctive MA-reading as well as the Q-function and answer space yielded by each structure. In both structures, a covert DOU-operator is presented at the left edge of the question nucleus and is associated with the higher-order trace π. The two structures differ only with respect to the scopal pattern between the trace π and the possibility modal *can*. As computed in (78), no matter whether π scopes above or below *can*, DOU strengthens the disjunctive answer into a free choice statement that is semantically equivalent to the conjunction of the two individual answers.

(77) (*w: Book A and Book B each can be used as the only textbook for this class; no other book can be used as a textbook for this class.*)
What can we use [as a textbook] for this class? Book A or Book B.

a. DOU $\gg \pi \gg \Diamond$



b. DOU $\gg \Diamond \gg \pi$



$\lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\mathrm{H}}\textit{thing}.\textsc{dou}[\pi(\lambda x.\Diamond Of(x))]$

$$\begin{tikzcd}[column sep=small]
 & \textsc{dou}[\Diamond Of(a) \wedge \Diamond Of(b)] \arrow[dl] \arrow[dr] \arrow[d, leftrightarrow] & \\
\textsc{dou}\Diamond Of(a) & \wedge \arrow[d, leftrightarrow] & \textsc{dou}\Diamond Of(b) \\
 & \textsc{dou}[\Diamond Of(a) \vee \Diamond Of(b)] \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

With DOU ($\pi \gg \Diamond$): disj/conj-unctive MA

$\lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\mathrm{H}}\textit{thing}.\textsc{dou}[\Diamond\pi(\lambda x.Of(x))]$

$$\begin{tikzcd}[column sep=small]
 & \textsc{dou}\Diamond[Of(a) \wedge Of(b)] & \\
\textsc{dou}\Diamond Of(a) & \wedge \arrow[d, leftrightarrow] & \textsc{dou}\Diamond Of(b) \\
 & \textsc{dou}\Diamond[Of(a) \vee Of(b)] \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

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

(78) 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 $\Diamond \gg \pi$

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

Next, return to singular-marked $\Diamond$-questions. Recall that, while rejecting conjunctive answers, singular-marked $\Diamond$-questions admit elided disjunctions as their MA-answers. The following considers the two discussed possibilities where a covert DOU-operator is presented in the nucleus and is associated with a higher-order trace. For the numeral-neutral question in (77), the Q-functions yielded by the two possible LFs have the same output when combining with a Boolean disjunction — they both yield a free choice inference. In the corresponding singular-marked $\Diamond$-question, however, whether π takes scope below or above the possibility modal yields a crucial difference with free choice disjunctive answers. If π takes a wide scope, as seen in (79a), the derived free choice inference is a contradiction, just like the case of the wide scope conjunctive answer in (73). In contrast, as seen in (79b), if π takes a narrow scope relative to the possibility modal, the derived free choice inference is not contradictory and is a desired MA-answer.

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

a. If DOU $\gg \pi \gg \iota \gg \Diamond$:

$[\![\textsc{wh}\text{-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\mathrm{H}}D.\textsc{dou}[\lambda w.\pi(\lambda x_e.x = \iota y[\textit{book}_w(y) \wedge \Diamond_w Of(y)])]$
$[\![\textsc{wh}\text{-Q}]\!](a^{\Uparrow} \cup b^{\Uparrow}) = \textsc{dou}[\lambda w.[(a^{\Uparrow} \cup b^{\Uparrow})(\lambda x_e.x = \iota y[\textit{book}_w(y) \wedge \Diamond_w Of(y)])]]$
$= \textsc{dou}[\lambda w.[a = \iota y[\textit{book}_w(y) \vee \Diamond_w Of(y)]] \wedge$
$[b = \iota y[\textit{book}_w(y) \wedge \Diamond_w Of(y)]]]$
$= \lambda w.[a = \iota y[\textit{book}_w(y) \wedge \Diamond_w Of(y)]] \wedge$
$[b = \iota y[\textit{book}_w(y) \wedge \Diamond_w Of(y)]]$

(#*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 DOU $\gg \Diamond \gg \pi \gg \iota$:

$[\![\textsc{wh}\text{-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\mathrm{H}}D.\textsc{dou}\Diamond[\lambda w.\pi(\lambda x_e.x = \iota y[\textit{book}_w(y) \wedge Of_w(y)])]$
$[\![\textsc{wh}\text{-Q}]\!](a^{\Uparrow} \cup b^{\Uparrow}) = \textsc{dou}\Diamond[\lambda w.(a^{\Uparrow} \cup b^{\Uparrow})(\lambda x_e.x = \iota y[\textit{book}_w(y) \wedge Of_w(y)])]$
$= [\Diamond\lambda w.a = \iota y[\textit{book}_w(y) \wedge Of_w(y)]] \cap$

$$[\Diamond \lambda w.b = \iota y[\textsf{book}_w(y) \wedge Of_w(y)]]$$

(*a* can be the unique book that we use as the only textbook for this class, and *b* can be the unique book that we use as the only textbook for this class.)

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

5.4. Comparing the two approaches

Both the uniform approach and the reconstruction approach can properly derive and account for the distributional constraints of the 'conjunction-rejecting' higher-order reading.

First, both approaches explain why singular-marked and numeral-modified questions admit 'conjunction-rejecting' higher-order readings. In the uniform approach, assuming that disjunctions can be singular/cardinal, the atomicity/cardinality restrictor in the WH-complement does not block the application of the H-shifter, this approach allows the Q-domain of a singular-marked/numeral-modified question to range over a set of Boolean disjunctions (and Montagovian individuals). 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 syntactically reconstructed to the question nucleus.

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

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

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 certain cases. In the reconstruction approach, disjunctive answers are acceptable because disjoining two uniqueness inference does not yield a contradiction. However, the computation in (79a) shows an exception: if disjunctions are interpreted as wide scope free choice, they would yield contradictions the same as conjunctions. In contrast, the uniform approach does not predict disjunctions to be deviant in any case. Unfortunately, it is hard to check the predictions with real data. Second, the uniform approach derives the 'conjunction-rejecting' reading in the very same way as the 'conjunction-admitting' reading, while the reconstruction approach uses a salvaging strategy. Therefore, on the one hand, the uniform 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.

6. Conclusion

This paper investigates the higher-order readings of WH-questions. First, drawing on evidence from questions with necessity modals or collective predicates, I showed that sometimes a WH-question can only be completely addressed by a GQ and must be interpreted with a higher-order reading. Next, I argued that the GQs that can serve as complete answers to questions must be homogeneously positive. Incorporating this constraint into the meaning of a H-shifter, I proposed that higher-order readings arise 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. Accordingly, higher-order readings are unavailable if the application of the H-shifter is blocked, either by the atomicity constraint of the singular feature [SG] in singular nouns, or by the cardinality constraint of numerals in numeral-modified nouns.

Further, a puzzle arose that singular-marked and numeral-modified questions admit disjunctive answers but not conjunctive answers. I provided two explanations to this asymmetry. In the uniform approach, these questions admit disjunctions because some disjunctions (but no conjunction) may satisfy the atomicity/cardinality requirement. In the reconstruction approach, the WH-complement is reconstructed, which gives rise to local uniqueness and yields contradictions for conjunctive answers.

Acknowledgement This paper significantly expands on Xiang 2019. For helpful discussions, I thank Luis Alonso-Ovalle, Lucas Champollion, Gennaro Chierchia, Danny Fox, Manuel Križ, Floris Roelofsen, Vincent Rouillard, Benjamin Spector, Bernhard Schwarz, the audiences at Georg-August-Universitüt Göttingen and Ecole Normale Supérieure, the audience and abstract reviewers of the 22nd Amsterdam Colloquium, and two anonymous reviewers of *Natural Language Semantics*. All errors are mine.

References

Alonso-Ovalle, Luis, and Vincent Rouillard. 2019. Number inflection, Spanish bare interrogatives, and higher-order quantification. In *Proceedings of North East Linguistic Society*, ed. Maggie Baird and Jonathan Pesetsky, volume 49.

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

Beck, Sigrid, and Hotze Rullmann. 1999. A flexible approach to exhaustivity in questions. *Natural Language Semantics* 7:249–298.

Champollion, Lucas. 2016a. Covert distributivity in algebraic event semantics. *Semantics and Pragmatics* 9:1–65.

Champollion, Lucas. 2016b. 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.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector. 2012. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In *An international handbook of natural language meaning*, ed. Claudia Maienborn, Klaus von Heusinger, and Paul Portner, 2297–2332. Mouton de Gruyter.

Cremers, Alexandre, and Emmanuel Chemla. 2016. A psycholinguistic study of the exhaustive readings of embedded questions. *Journal of Semantics* 33:49–85.

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. 2020. Who and what do *who* and *what* range over cross-linguistically. URL https://ling.auf.net/lingbuzz/003959, manuscript, ZAS Berlin (Leibniz-Center of General Linguistics).

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* .

Gentile, Francesco, and Bernhard Schwarz. 2018. A uniqueness puzzle: *How many*-questions and non-distributive predication. In *Proceedings of Sinn und Bedeutung*, ed. Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern, and Hannah Rohde, volume 21, 445–462. URL https://semanticsarchive.net/Archive/DRjNjViN/SuB21.pdf.

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

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

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

Hausser, Roland, and Dietmar Zaefferer. 1979. Questions and answers in a context-dependent Montague grammar. In *Formal semantics and pragmatics for natural languages*, ed. Franz Guenthner and Siegfried J. Schmidt, 339–358. Dordrecht: Reidel.

Hausser, Roland R. 1983. The syntax and semantics of English mood. In *Questions and answers*, ed. Ferenc Kiefer, 97–158. Dordrecht: Springer.

Heim, Irene. 1994. Interrogative semantics and Karttunen's semantics for *know*. In *The Proceedings of the 9th Annual Conference of The Israel Association for Theoretical Linguistics (IATL) and of the Workshop on Discourse*, ed. Rhonna Buchalla and Anita Mitwoch, volume 1, 128–144. URL https://semanticsarchive.net/Archive/jUzYjk1O/Interrogative%2094.pdf.

Hirsch, Aron, and Bernhard Schwarz. 2019. Singular *which*, mention-some, and variable scope uniqueness. Poster presented at Semantics and Linguistic Theory 29, University of California, Los Angeles, May 2019.

Hirsch, Aron, and Bernhard Schwarz. 2020. Singular which, mention-some, and variable scope uniqueness. In *Proceedings of Semantics and Linguistic Theory*, ed.

Katherine Blake, Forrest Davis, Kaelyn Lamp, and Joseph Rhyne, volume 29, 748–767. URL https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/29.748.

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 L.M. Faltz. 1985. *Boolean Semantics for Natural Language*. Dordrecht: D. Reidel Publishing Company.

Klinedinst, Nathan, and Daniel Rothschild. 2011. Exhaustivity in questions with non-factives. *Semantics and Pragmatics* 4:1–23.

Krifka, Manfred. 1997. The expression of quantization (boundedness). Presentation at the Workshop on Cross-linguistic variation in semantics. LSA Summer Institute. Cornell.

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. 2020. Plural marking and d-linking in Spanish interrogatives. *Journal of Semantics* 37:145–170.

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 *Proceedings of Semantics and Linguistic Theory*, ed. Robert B. Young and Yuping Zhou, volume 13, 258–275. URL https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/2898.

Sauerland, Uli, Jan Anderssen, and Kazuko Yatsushiro. 2005. The plural is semantically unmarked. In *Linguistic evidence: Empirical, theoretical, and computational perspectives*, ed. Stephan Kepser and Marga Reis, 413–434. Berlin: Mouton de Gruyter.

Scontras, Gregory. 2014. The semantics of measurement. Doctoral Dissertation, Harvard University.

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 Semantics and Linguistic Theory*, volume 17, 282–299. URL https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/2962.

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.

Szabolcsi, Anna. 1997. Background notions in lattice theory and generalized quantifiers. In *Ways of scope taking*, ed. Anna Szabolcsi, 1–27. Dordrecht: Springer.

Theiler, Nadine, Floris Roelofsen, and Maria Aloni. 2018. A uniform semantics for declarative and interrogative complements. *Journal of Semantics* .

Uegaki, Wataru. 2015. Interpreting questions under attitudes. Doctoral Dissertation, Massachusetts Institute of Technology.

Uegaki, Wataru. 2018. On the projection of the presupposition of embedded questions. In *Proceedings of Semantics and Linguistic Theory*, ed. Sireemas Maspong, Brynhildur Stefánsdóttir, Katherine Blake, and Forrest Davis, volume 28, 789–808. URL https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/28.789.

Uegaki, Wataru. 2020. The existential/uniqueness presupposition of wh-complements projects from the answers. *Linguistics and Philosophy* .

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

Xiang, Yimei. 2016a. Complete and true: A uniform analysis for mention some and mention-all. In *Proceedings of Sinn und Bedeutung (SuB) 20*, ed. Polina Berezovskaya Nadine Bade and Anthea Schöller, 815–830.

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

Xiang, Yimei. 2019. Two types of higher-order readings of wh-questions. In *Proceedings of the 22nd Amsterdam Colloquium*, ed. Julian J. Schlöder, Dean McHugh, and Floris Roelofsen, 417–426. URL http://events.illc.uva.nl/AC/AC2019/uploaded_files/inlineitem/Xiang_Two_types_of_higher-order_readings_of_wh-ques.pdf.

Xiang, Yimei. 2020. Function alternations of the Mandarin particle *dou*: Distributor, free choice licensor, and 'even'. *Journal of Semantics* .