Higher-order readings of *wh*-questions

Yimei Xiang, Rutgers University

(Forthcoming in *Natural Language Semantics*)

Abstract In most cases, a *wh*-question calls for 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 argues that sometimes a *wh*-question must be interpreted with a higher-order reading, in which this question calls for an answer that names 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 semantic 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 answered 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 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) calls for 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 the claim 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 the inference that both mentioned individuals are math professors.[1]

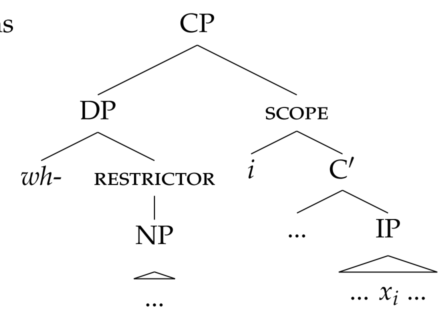
(1) Which math professor left the party at midnight?
   a. Andy. $\rightsquigarrow$ *Andy is a math professor.*
   b. Andy or Billy. $\rightsquigarrow$ *Andy and Billy are math professors.*

To capture this question–answer relation, it is commonly assumed that *wh*-phrases are functions (e.g., existential (∃-)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 a 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 the formation of 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. $[\![\textit{which student}]\!] = \lambda P_{\langle e,t \rangle} \lambda x_e \colon \textsf{student}(x).P(x)$
    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 the assertion that this individual came and the presupposition that this individual is a student. In (4b), in contrast, since the disjunctive answer has the 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 as an intersection operator and use covert type-shifting operations to derive the summation-like reading.

(4) a. Combining with an entity

$[\![wh\text{-Q}]\!]([\![Andy]\!]) = (\lambda x_e\colon \mathsf{student}(x).\mathsf{came}(x))(a)$
$\quad = \mathsf{student}(a).\mathsf{came}(a)$

b. Combining with a GQ

$[\![Andy\ or\ Billy]\!]([\![wh\text{-Q}]\!]) = (a^{\Uparrow} \cup b^{\Uparrow})(\lambda x_e\colon \mathsf{student}(x).\mathsf{came}(x))$
$\quad = \mathsf{student}(a) \wedge \mathsf{student}(b).\mathsf{came}(a) \vee \mathsf{came}(b)$

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 delves into these 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 questions with a higher-order reading. 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 disjunction–conjunction asymmetry: questions with a singular-marked or numeral-modified *wh*-phrase admit disjunctive answers but reject conjunctive answers. I present two ways to account for this asymmetry, namely a uniform account and a reconstruction account. Section 6 concludes.

2. Evidence for higher-order readings

Saying that a *wh*-question has a first-order reading yields two predictions in regards 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 the 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?*) argues that sometimes 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?*) argues that sometimes the Q-domain of a question must contain Boolean conjunctions and universal quantifiers (Sect.

2.2). Finally, combinations of these two diagnostics show that Boolean coordinations of the aforementioned GQs are also included in some Q-domain (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 response 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, however, that 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 Q-functions yielded, I instead call the two readings the 'first-order reading' and the 'higher-order reading', respectively. As paraphrased in (9), the first-order reading calls for answers that specify an entity, while the higher-order reading calls for answers that specify a GQ.

(9) Which books does John have to read?

a. First-order reading:
'Which books x is such that John has to read x?'

b. Higher-order reading:
'Which 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 following LFs and Q-functions for the two readings. (Subject-movement is ignored. '@' stands for the actual world, and '$\textsc{smlo}(\pi)$' stands for the smallest live-on set of π. The assumed Q-domain for the higher-order reading is subject to revision. For now, I just assume that this Q-domain is the set of GQs ranging over a set of books.[3] See Sect. 3 for refinements.) Observe that, for the higher-order reading, the GQ-denoting answer is interpreted at the scopal position that the higher-order *wh*-trace π takes, whichever that may be.

(10) First-order reading

a. $[_{\text{CP}}$ which-books λx_e $[_{\text{IP}}$ have-to $[_{\text{VP}}$ John read x]]]

b. $[\![wh\text{-Q}]\!] = \lambda x_e\colon \textsf{books}_{@}(x).\Box[\lambda w.\textsf{read}_w(j, x)]$

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

(11) Higher-order reading ($\Box \gg \pi$) (to be revised in (42b))

a. $[_{\text{CP}}$ which-books $\lambda \pi_{\langle et,t\rangle}$ $[_{\text{IP}}$ have-to $[\pi\ \lambda x_e$ $[_{\text{VP}}$ John read x]]]]

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

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

As we shall see, $\Box$-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 if and only if 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 the proposition $\lambda w.\textsf{read}_w(j, x)$. Arrows indicate entailments. The middle disjunctive symbol '$\vee$' stands for the disjunction of the two atomic sentences; specifically, 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 $\langle \tau, t\rangle$, where τ is an arbitrary type, we say that π *lives on* A if and only if for every set B: $\pi(B) \Leftrightarrow \pi(B \cap A)$ (Barwise and Cooper 1981), and that π *ranges over* A if and only if A is the smallest live-on set ($\textsc{smlo}$) 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 or should not be included in a Q-domain (see Sect. 3).

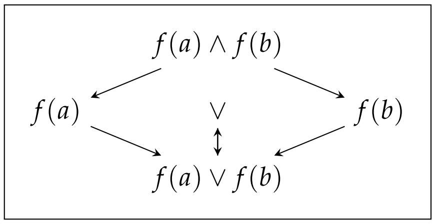


Figure 1: The answer space of *What did John read?*

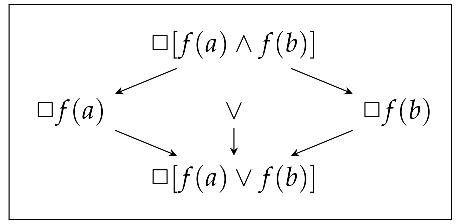


Figure 2: The answer space of *What does John have to read?*

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 $\Box f(a)$ and $\Box f(b)$. For example, if John's only reading obligation is to read either *a* or *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: □-questions may yield Q-functions that are 'non-reducible' relative to disjunctions and existential quantifiers. The following defines reducibility, where '•' stands for the combinatory operation between the function θ and a GQ:[4]

(12) For any π of type $\langle \tau t, t \rangle$, where τ is an arbitrary type, a function θ is *reducible* relative to π if and only if $\theta \bullet \pi \Leftrightarrow \pi(\lambda x_\tau . \theta \bullet x^{\Uparrow})$.

Similarly to □-questions, the following questions, with a word expressing universal quantification, also have readings where the Q-function is not reducible relative to disjunctions or to existential quantifiers. ((13) and (14) are taken from Spector 2007.)

(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?

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

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

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

b. $[\lambda\pi.\textsf{\textit{J has to read }}\pi](f^{\Uparrow} \cap r^{\Uparrow}) =$ *J has to read f* $\wedge$ *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 what I refer to here as a '*stubbornly collective predicate*' (e.g., *formed a team*, *co-authored two papers*) may have answers naming Boolean conjunctions, and furthermore, that stubborn collectivity in these questions does not trigger uniqueness.

First, to see how stubborn collectivity manifests itself, observe that the phrasal predicate *formed a/one team* admits a collective reading but not a covered/ (non-atomic) distributive reading. 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 (or cumulative/ semi-distributive) reading, and thus (17b) can be truthfully uttered.

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

a. # The children formed a/one team.

b. $\surd$ The children formed teams.

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, children $a + b$ competed against children $c + d$.] Today, the children (all) formed a/one team. ($^{\text{OK}}$collective, $^{\#}$covered/distributive)

Stubbornly collective predicates like *formed a/one team* contrast with predicates like *lifted the piano*, which admit 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 context. Sentence (19a) suffers a presupposition failure because the factive verb *know* embeds a false collective declarative. However, 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) (Context: The four relevant children *a*,*b*,*c*,*d* formed exactly two teams in total: $a + b$ formed one, and $c + d$ formed the other.)

[5] A predicate P is *quantized* if and only if 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 phrasal predicates because the extension of a quantized verbal phrase is not closed under summation formation.

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

b. √ John knows [which children formed a team].

c. ⇝ *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, or else the embedded clause in (19a) would admit a covered/distributive reading and (19a) would be felicitous, contrary to 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: 'Which GQ π over children is such that π formed a team?'

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

b. $[\![wh\text{-Q}]\!] = \lambda\pi_{\langle et,t\rangle} : \textsc{smlo}(\pi) \subseteq \textsf{\textit{children}}_{@}.\lambda w[\pi(\lambda x_e.\textsf{\textit{f.a.team}}_w(x))]$

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

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

However, Ans-H is insufficient as it cannot capture the contrast with respect to uniqueness in (22a,b). 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) (Context: The four relevant children a,b,c,d formed exactly two teams in total: $a+b$ formed one, and $c+d$ formed the other.)

a. √ John knows [which children formed a team].

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

Uniqueness presuppositions in *wh*-questions are standardly explained by 'Dayal's presupposition', according to which 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 subsection, I argue that the contrast between (22a) and (22b) 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 was 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 had observed 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 *wh*-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 *wh*-phrase *which children* or the bare *wh*-word *who*, as seen 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 *wh*-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) $$\textsc{Ans-D}(w)(Q) = \exists p[w \in p \in Q \land \forall q[w \in q \in Q \to p \subseteq q]].\\ \iota p[w \in p \in Q \land \forall q[w \in q \in Q \to p \subseteq q]]$$

Adopting the ontology of individuals by Sharvy (1980) and Link (1983), Dayal concludes 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 relations, this ontology can be represented as in Figure 3. Letters a,b,c each denote an atomic child. Lines indicate part-hood relations from bottom to top. For example, atomic entities a and b are parts of their sum $a \oplus b$.

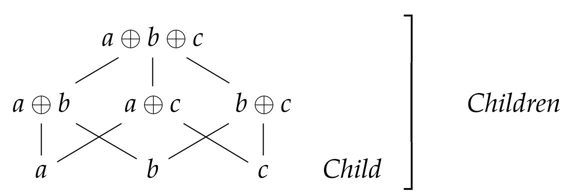


Figure 3: Ontology of individuals (Sharvy 1980; Link 1983)

Accordingly, as illustrated in (25), the Hamblin set of the singular-marked *wh*-question includes only propositions naming an atomic child, while the Hamblin set of the corresponding plural-marked *wh*-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 context where both Andy and Bill came, (25b) has a strongest true answer $\lambda w.\textsf{came}_w(a \oplus b)$ while (25a) does not; employing Ans-D in (25a) gives rise to a presupposition failure. To avoid this presupposition failure, the singular-marked *wh*-question (25a) can only be uttered in a world where only one of the children came, which therefore explains its uniqueness presupposition.

[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 as inclusive or exclusive is not crucial in this paper. The following presentation follows the inclusive theory.

(25) (Context: Among the relevant children, only Andy and Billy came.)

a. Which child came?

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

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

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

b. Which children came?

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

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

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

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 *wh*-question (25a) would admit conjunctive answers like $\lambda w.\textit{came}_w(a) \wedge \textit{came}_w(b)$ and would not be subject to uniqueness, contrary to fact.[7]

Numeral-modified *wh*-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 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 *wh*-questions. As seen in (27), for a question of the form 'Which *N*-children came?', where *N* is a cardinal numeral read as 'exactly *N*', Dayal's presupposition is satisfied only if exactly *N* of the 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 children who came is larger than *N*, the question does not have a strongest true answer.

(27) (Context: Among the relevant children, only Andy, Billy, and Clark came.)
Which two children came?

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

[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 like *formó un grupo* 'formed.SG a group', and the resulting question calls for the specification of 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.'

Answer (ib) has the conjunction reading that the students formed a group and the professors formed a group. The felicity of this answer shows that the *quién*-question admits answers naming Boolean conjunctions over non-atomic elements. Alonso-Ovalle and 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

Just as in a singular-marked *wh*-question, here 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 $(a \oplus b)^{\Uparrow} \cap (a \oplus c)^{\Uparrow} \cap (b \oplus c)^{\Uparrow}$.

Let us return to the contrast of question-embeddings in (22), repeated below:

(28) (Context: The relevant four children a,b,c,d formed two teams in total: $a + b$ formed one, and $c + d$ formed the other.)

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 *wh*-question contains Boolean conjunctions, while that of a numeral-modified *wh*-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, namely $\lambda w.\textsf{f.a.team}_w(a \oplus b)$ 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]

It is worth noting that the argumentation for ruling in Boolean conjunctions is not totally dependent on whether we use Dayal's presupposition to explain the uniqueness effects in singular-marked and numeral-modified *wh*-questions — 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 (Uegaki 2018, 2020; Hirsch and Schwarz 2020; Fox 2018, 2020). For example, to account for the projection of uniqueness presuppositions and existential presuppositions in question-embeddings, Uegaki (2018, 2020) proposes to encode the uniqueness inference within the question nucleus. Moreover, as we will see in Sect. 5.1, Hirsch and Schwarz (2020) observe that in questions with a possibility modal, the uniqueness inference triggered by

[8] In contrast to my analysis, Fox (2020) assumes higher-order pluralities to account for the data in (22)/(28). He proposes to get rid of Dayal's presupposition and accounts for the observed uniqueness effects based on his Question Partition Matching (QPM) principle. According to this principle, a singular-marked *wh*-question is only acceptable in context sets where its uniqueness presupposition is satisfied. Moreover, Fox argues that QPM can better account for the unavailability of higher-order readings in questions with a negative island, as well as for the modal obviation effects of higher-order readings in those questions.

(i) *Question Partition Matching* (Fox 2018, 2020)
For any question with a Hamblin set Q, if it induces a partition P in a context set A, this question is acceptable in A if and only if: (i) every cell in P is identical to the exhaustification of a proposition in Q; and conversely (ii) every proposition p in Q is such that the exhaustification of p is identical to a cell in P.

The QPM principle, however, predicts that the *wh*-phrase in any non-modalized question cannot range over GQs. For example, if the question *Who left?* admits a higher-order reading, its Hamblin set should contain plain disjunctions such as $\lambda w.\textsf{left}_w(a) \vee \textsf{left}_w(b)$, which cannot be paired with any partition cell by exhaustification, violating QPM. To account for the absence of uniqueness in *Which children formed a team?*, Fox assumes that here *which children* ranges over a set of higher-order pluralities, not over a set of GQs. For example, in Fox's account, the conjunctive answer $a + b$ *and* $c + d$ is interpreted as $\{\{a, b\}, \{c, d\}\}$, not as $(a \oplus b)^{\Uparrow} \cap (c \oplus d)^{\Uparrow}$.

a singular-marked *wh*-phrase can be interpreted under the scope of the possibility modal. In line with Rullmann and Beck (1998), Uegaki and Hirsch and Schwarz propose that uniqueness is introduced as a presupposition of an unpronounced WHICH-determiner that appears within the question 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. In Sect. 5.3, I will present a way to account for the unavailability of Boolean conjunctions in singular-marked and numeral-modified *wh*-questions while encoding uniqueness locally.

2.3. Evidence for complex GQ-coordinations

Previous sections have provided 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 stubborn collectivity provides evidence for ruling in Boolean conjunctions and universal quantifiers. Combining these two diagnostics, the following shows that a Q-domain also contains complex GQ-coordinations:

(29) Context: The eight students enrolled in a class are separated into four pairs by year and major. As part of the course requirements, each pair of students has to co-present one paper this or next week. Moreover, the instructor requires the presentations in each week to be given by students having the same major.

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

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* (abbreviated as '*p.a.p.t.*'). The instructor's disjunctive answer can be analyzed as consisting of a free-choice inference, as shown in (30a), with the choices specified as in (30b,c).

(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 they can be given by the philosophers.*

b. *If the presentations are given by the linguists,* $a_1 + b_1$ *will p.a.p.t., and* $c_1 + d_1$ *will p.a.p.t.*

c. *If the presentations are given by the philosophers,* $a_2 + b_2$ *will p.a.p.t., and* $c_2 + d_2$ *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 the question should contain a higher-order *wh*-trace between the necessity modal and the stubbornly collective predicate, as in (31). The answer of the instructor should be interpreted as 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) [$_{\text{CP}}$ which-students $\lambda\pi_{\langle et,t\rangle}$ [$_{\text{IP}}$ have-to [π λx_e [$_{\text{VP}}$ x present a paper together]]]]

2.4. Interim summary

To sum up, this section discussed cases where a *wh*-question must be interpreted with a higher-order reading and provided two diagnostics to determine what higher-order meanings should be included in the Q-domain of a *wh*-question. One 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. Using this diagnostic, I argued to rule in Boolean disjunctions and a class of existential quantifiers. The other diagnostic is based on the absence of uniqueness effects in questions with a stubbornly collective predicate. Using this diagnostic, I argued to rule in Boolean conjunctions and universal quantifiers. Combining these two diagnostics, I further showed that a Q-domain may also contain 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 Boolean coordinations among all of these. One might wonder whether we can characterize higher-order Q-domains more generally as follows:

(32) *An unconstrained characterization of higher-order Q-domain*:
In a higher-order reading, the Q-domain yielded by a *wh*-phrase consists of all the 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 this section, I will show that this characterization is too inclusive. Spector (2007, 2008) provides some counterexamples and argues that the GQs included in a Q-domain must be *increasing*.[9] Furthermore, extending Spector's diagnostic to a broad range of non-monotonic GQs and GQ-coordinations, I show that the 'increasing-GQ requirement' is too strong/exclusive: in some higher-order readings, the Q-domain of a *wh*-questions includes not only increasing GQs but also some non-monotonic GQs. I also find that not all non-monotonic GQs can serve as semantic answers. To capture these observations, I argue that whether a GQ can serve as a semantic answer to a question is determined by the property which I refer to as '*positiveness*' (roughly, the property of ensuring existence with respect to a certain quantification domain; see Sect. 3.2): the GQs involved in a Q-domain must be homogeneously positive.[10]

[9]Monotonicity of GQs is defined as follows:

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

[10]The issues and proposals in Sect. 4 and Sect. 5 are independent from the findings in this section. Readers may choose to jump to the interim summary in Sect. 3.4.

3.1. Completeness-based tests and the increasing-GQ constraint

Whether a meaning can serve as a semantic answer to a question can be examined by the *completeness-based test* stated in (33). This test draws on a deductive relation between attitudes held towards a question and attitudes held towards the complete true answer to this question: the question-embedding sentence '*x knows* Q' implies a *completeness* condition that *x* knows the complete true answer to 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.[11,12]

(33) *Completeness-based 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.

For simple illustration, consider the truth conditions of the question-embedding sentence (34b) under the context described in (34a). Strikingly, there is a reading of (34b) in which this sentence implies that Sue knows John's reading obligation (a-i), but not that she knows (a-ii); Sue can be ignorant about whether John has to read any Russian novels.[13] (Note that Sue cannot have a false belief that John has to read some Russian novels, due to a separate *false-answer sensitivity* condition. See details in footnote 12.)

(34) a. Context: John's summer reading obligations include the following:
i. he has to read at least two French novels;
ii. he has to read no Russian novel (since he has already read too many Russian novels).

b. Sue knows which books John has to read this summer.
$\rightsquigarrow$ *Sue knows (a-i).*

[11] The completeness-based test here 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, 2018) and Xiang (2016b: Chap. 2).

[12] The completeness-based test does not aim to fully characterize the truth conditions of a question-embedding sentence or to exhaustively determine what can or cannot be included in a Q-domain. First, this test is only concerned with 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 sentence (34b) to be true, Sue can be ignorant about whether John should read any Russian novels, but she cannot have the false belief that John should read some Russian novel(s).

(i) '*x* knows Q' is true if and only if
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-based test can only be used to determine what meanings should be *excluded* from a Q-domain, not what meanings should be *included* in a Q-domain. A bi-conditional characterization, namely, ruling in all the short answers that are not filtered out by the completeness-based test, could yield conflicting predictions. As I will show in Sect. 3.3, the completeness-based test shows that the higher-order Q-domain of the question *What does John have to read?* contains the non-monotonic quantifier *exactly three books* but not the decreasing quantifier *less than four books*, even though 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*).

[13] Strikingly, in contrast to (34b), the following two sentences with a concealed question or a definite description do imply that Sue knows both of John's summer reading obligations listed in (34a).

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

$\not\rightsquigarrow$ *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 John's reading obligation (a-ii) excludes the decreasing GQ *no Russian novel* and the non-monotonic GQ-coordination *at least two French novels and no Russian novel* from the Q-domain.[14] The higher-order reading of the embedded question *which books John has to read* is then paraphrased as follows: 'Which increasing GQ π over books is such that John has to read π?'

3.2. The Positive-GQ Constraint

The following example makes a minimal change to the context of the example in (34): John also has to read exactly two Chinese novels. With this added reading obligation, the embedding sentence (34′b) implies that Sue knows that John has to read exactly two Chinese novels, not just that Sue knows that John has to read at least two Chinese novels.

(34′) a. Context: John's summer reading obligations include the following:
  i. he has to read at least two French novels;
  ii. he has to read no Russian novel;
  iii. he has to read exactly two Chinese novels.

b. Sue knows which books John has to read.
  $\rightsquigarrow$ *Sue knows (a-i) and (a-iii).*
  $\not\rightsquigarrow$ *Sue knows (a-ii).*

Example (35) below applies the completeness-based test to a broader range of GQs. The game requirements listed in (35a) each name a GQ ranging over a set of cards. Among these GQs, (i) and (ii) are increasing, (iii) and (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 leave the player the flexibility to determine which two kings to play to fulfill requirement (v). Intuitively, the question-embedding sentence (35b) implies that Sue knows about not only the game requirements (i) and (ii) but also about (v). In particular, for the condition regarding 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.[15]

---

[14]The illustration of the completeness-based 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 summer reading obligations consist of the following:
  i. He has to read no leisure book or more than two math books. (In other words, John has to read more than two math books if he reads any leisure book.)
  ii. He has to read none or all of the Harry Potter books (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 it).

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

[15]The embedded question *which cards John has to play* also has a reading in which it admits all of the listed GQs, regardless of whether they are increasing, non-monotonic, or even decreasing. This reading is similar to what I observe with concealed questions and definite descriptions, as witnessed in footnote 13.

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

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

In the above two examples, Spector's increasing-GQ requirement over rules out the non-monotonic GQs *exactly two Chinese novels* and *exactly two kings*; thus it fails to predict that (34′b) implies that Sue knows the reading obligation (34′a-iii), and that (35b) implies that Sue knows the game requirement (35a-v).

More generally, any monotonicity-based constraint faces a dilemma: we need a characterization that rules out non-monotonic GQ-coordinations such as *at least two French novels and no Russian novel* while not excluding simplex non-monotonic GQs like *exactly two Chinese novels*.[16] Moreover, note that it is not just a matter of differentiating between complex and simplex non-monotonic GQs. For example, extending the application of the completeness-based test in (34) to other GQs, we can find that the Q-domain of the embedded question *which books John has to read* includes a few complex non-monotonic GQs such as the disjunction *exactly three French novels or exactly four Russian novels* and the conjunction *at least three French novels but no more than 20 novels (of any kind)*, but not the simplex non-monotonic GQ *less than three or more than ten novels*.

In contrast to Spector (2008), I propose that whether a GQ can serve as a direct semantic answer to a higher-order *wh*-question and be a member of the Q-domain of this question is determined by a property which I refer to here as 'positiveness', not its monotonicity. A preliminary constraint is as follows:

(36) *Positive-GQ Constraint* (preliminary)
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 (i.e., its smallest live-on set (SMLO); 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 GQ *at most two books* and the non-monotonic GQ *less than three or more than ten books* do not entail *some books* and are thus not positive. A formal definition of positiveness is as follows, where $\mathbb{E}$ stands for the existential quantifier (i.e., $\mathbb{E} =_{\text{df}} \lambda Q \lambda P.Q \cap P \neq \varnothing$):

(37) For any π of type $\langle et, t\rangle$, π is *positive* if and only if $\pi \subseteq \mathbb{E}(\textsc{smlo}(\pi))$.

Table 1 compares monotonicity and positiveness for a list of coordinations over Montagovian individuals and simplex GQs that range over a set of books. The letters *a* and *b* stand for two distinct atomic books. Observe that increasing ($\uparrow_{\textsc{mon}}$) GQs are all positive,

[16] I call a GQ 'simplex' if it can be expressed as a single 'D+NP' phrase and 'complex' otherwise. For example, the GQ-coordination *at least two books but no more than five books* is simplex because it can be equivalently expressed as *two to five books*.

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	SMLO(π)	Increasing?	Positive?
$a^{\Uparrow}$	$\{a\}$	Yes ($\uparrow_{\text{MON}}$)	Yes
$a^{\Uparrow} \cap b^{\Uparrow}$, $a^{\Uparrow} \cup b^{\Uparrow}$	$\{a, b\}$	Yes ($\uparrow_{\text{MON}}$)	Yes
{at least, more than} two books	books	Yes ($\uparrow_{\text{MON}}$)	Yes
every book except a	book $- \{a\}$	Yes ($\uparrow_{\text{MON}}$)	Yes
{at most, less than, no more 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 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: Monotonicity versus positiveness

3.3. The Homo-Positive-GQ Constraint

Table 1 considers only coordinations over Montagovian individuals and simplex GQs. However, Benjamin Spector (pers. comm.) points out that the Positive-GQ Constraint does not exclude many 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 SMLO$(\pi) = A \cup B$ and $\pi \subseteq \mathbb{E}(A \cup B)$.[17]

More generally, to determine whether a non-monotonic GQ can be included in a Q-domain, it is insufficient to examine 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 need to examine whether existence is ensured homogeneously with respect to both the following two sets:

(A) the set that the coordinated increasing GQs range over;
(B) the set that the coordinated decreasing GQs range over.

For example, the GQ-coordination *every article but no book* is excluded because it ensures existence with respect to the set of articles but not to the set of books. Intuitively, when

[17] 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, the equivalence in (i) shows that $A \cup B$ is a live-on set of π: replacing X with $\underline{X \cap (A \cup B)}$ in the set description does not change the set.

(i) $$\left\{X \,\middle|\, \begin{matrix}[A \subseteq (\underline{X \cap (A \cup B)})] \wedge \\ [B \cap (\underline{X \cap (A \cup B)}) = \emptyset]\end{matrix}\right\} \Leftrightarrow \left\{X \,\middle|\, \begin{matrix}[A \subseteq X \wedge A \subseteq (A \cup B)] \wedge \\ [(B \cap (A \cup B)) \cap X = \emptyset]\end{matrix}\right\} \Leftrightarrow \pi$$

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

(ii) $$\left\{X \,\middle|\, \begin{matrix}[A \subseteq (\underline{X \cap ((A \cup B) - \{a\})})] \wedge \\ [B \cap (\underline{X \cap ((A \cup B) - \{a\})}) = \emptyset]\end{matrix}\right\} \Leftrightarrow \left\{X \,\middle|\, \begin{matrix}[A \subseteq X \wedge A \subseteq (A \cup B - \{a\})] \wedge \\ [(B \cap ((A \cup B) - \{a\})) \cap X = \emptyset]\end{matrix}\right\}$$
$$\Leftrightarrow \left\{X \,\middle|\, \begin{matrix}[A \subseteq X \wedge A \subseteq (A \cup B) \wedge a \notin A] \wedge \\ [(B - \{a\}) \cap X = \emptyset]\end{matrix}\right\}$$
$$\Leftrightarrow \pi \text{ if and only if } a \notin A \text{ and } a \notin B$$

existence is ensured with respect to set (A) but not to set (B), the exclusions expressed by the coordinated decreasing GQs would be irrelevant to the inclusions expressed by the coordinated increasing GQs. For example, in the sentence *John has to read every article but no book*, how many books John has to read is irrelevant to how many articles John has to read. Hence, I strengthen the Positive-GQ Constraint in (36) to (38). The definition of 'homogeneously positive' will be given in (40) below.

(38) *Homo-Positive-GQ Constraint* (final)
GQs in the Q-domain of a *wh*-question must be homogeneously (homo-)positive.

In most GQ-coordinations, the GQs being coordinated cannot be semantically retrieved from their coordination. For one thing, some GQs cannot be decomposed into a simple coordination of monotonic GQs. For example, the decomposition in (39a) has to involve at least disjunctions over conjunctions of monotonic GQs. For another, even for a GQ that can be decomposed into a simple coordination, there can be multiple ways to decompose it, as seen in (39b,c).

(39) a. $[\![$*exactly 2 A or exactly 4 B*$]\!]$
$\Leftrightarrow [\![$*exactly 2 A*$]\!] \cup [\![$*exactly 4 B*$]\!]$
$\Leftrightarrow ([\![$*at least 2 A*$]\!] \cap [\![$*at most 2 A*$]\!]) \cup ([\![$*at least 4 B*$]\!] \cap [\![$*no more than 4 B*$]\!])$
b. $[\![$*every A and no B*$]\!]$
$\Leftrightarrow [\![$*every A*$]\!] \cap [\![$*no B*$]\!]$
c. $[\![$*every A and no B*$]\!]$
$\Leftrightarrow [\![$*every A or some B*$]\!] \cap [\![$*no B*$]\!]$
$\Leftrightarrow ([\![$*every A*$]\!] \cup [\![$*some B*$]\!]) \cap [\![$*no B*$]\!]$

Despite this challenge, in order to determine whether a GQ-coordination is homo-positive we just need to find out the strongest increasing and decreasing GQs involved which determine the lower and upper bounds of this GQ-coordination. I define homo-positiveness as in (40). π^+ is the logically strongest increasing GQ entailed by π which determines the lower bound of π, and π^- is the logically strongest decreasing GQ entailed by π which determines the upper bound of π.[18]

(40) For any π of type $\langle et, t\rangle$, π is *homogeneously positive* if and only if
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 — monotonicity, positiveness, and homo-positiveness — for a broader range of GQs. The GQs are divided into three groups depending on whether they are simplex or have to be expressed as a GQ-disjunction or a GQ-conjunction. A,B,C are three sets of entities; C is a superset of A, and B is not. (The conjunction 'every A but no C' is ignored since it denotes an empty set.) Observe that whether a GQ π is homo-positive is independent from whether π is simplex/complex, disjoined/conjoined, (non-)increasing, (non-)monotonic, or upper (un-)bound.[19]

[18]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 homogeneous 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 *two or four cards*.

[19]If a GQ π is unbound, then one or both of the strongest GQs retrieved from π are trivial (viz., equivalent to

Generalized quantifier π	π^+	π^-	Mon?	Pos?	Homo-Pos?
at least 2 A	at least 2 A	$D_{\langle e,t\rangle}$	$\uparrow_{\text{MON}}$	Yes	Yes
an even number of A	at least 2 A	$D_{\langle e,t\rangle}$	N.M.	Yes	Yes
exactly 2 A	at least 2 A	no more than 2 A	N.M.	Yes	Yes
exactly 2 to 4 A	at least 2 A	no more than 4 A	N.M.	Yes	Yes
exactly 2 or 4 A	at least 2 A	no more than 4 A	N.M.	Yes	Yes
no more than 4 A	$D_{\langle e,t\rangle}$	no more than 4 A	$\downarrow_{\text{MON}}$	No	No
less than 2 or more than 5 A	$D_{\langle e,t\rangle}$	$D_{\langle e,t\rangle}$	N.M.	No	No
every or no A	every A	no A	N.M.	No	No
every A or no B	every A	no B	N.M.	No	No
every A or no C	every A	no C	N.M.	No	No
every A but no B	every A	no B	N.M.	Yes	No
~~every A but no C~~					
at least 2 A but no more than 4 B	at least 2 A	no more than 4 B	N.M.	Yes	No
at least 2 A but no more than 4 C	at least 2 A	no more than 4 C	N.M.	Yes	Yes

Table 2: Monotonicity (Mon) versus positiveness (Pos) versus homo-positiveness

The generalizations from Table 1 still hold here. First, every GQ that is not positive is also not homo-positive. Second, for any increasing π, the retrieved π^- is trivial (viz., $\pi^- = D_{\langle e,t\rangle}$; see footnote 19), and thus this increasing π being positive ensures π being homo-positive.

For non-monotonic GQs, however, the Homo-Positive-GQ Constraint yields a different prediction. The simplex non-monotonic GQ *exactly two books* is positive as well as homo-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 of books. In contrast, the complex non-monotonic GQ-coordination *every article and no book* is positive but not homo-positive: the retrieved π^- *no book* ranges over the set of books, but *every article and no book* does not entail *some book*.[20]

$D_{\langle e,t\rangle}$), ranging over the discourse domain D_e. In particular, increasing GQs are upper-unbound, and decreasing GQs are lower-unbound.

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

b. If π is lower-unbound, i.e., $\forall P[P \in \pi \to \exists P' \in \pi[P' \subseteq P]]$, then $\pi^+ = D_{\langle e,t\rangle}$.
Examples: *less than two or more than four books*, *at most four books*

[20] A puzzle arises with non-monotonic GQ-coordinations such as *some book but no leisure book*, where the set that the coordinated decreasing GQ ranges over (viz., the set of leisure books) is a proper subset of the set that the coordinated increasing GQ ranges over (viz., the set of books). 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 his reading requirement. It is appealing to say that the question-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 homo-positive: it does not entail existence with respect 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-Positive-GQ Constraint still holds here. In the given scenario, the completeness condition of the embedding sentence (i) should be (ic), which is stronger than (ia) and weaker than the conjunction of (ia) and (ib). The GQ that serves as the complete true short answer to *what John has to read* is *some non-leisure book*, which is homo-positive.

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

3.4. Interim summary

To sum up, not all GQs can be used as semantic answers to *wh*-questions. In general, increasing GQs are qualified answers while decreasing GQs are not; however, there is a prominent higher-order reading of *wh*-questions in which some of the non-monotonic GQs are also qualified answers. Homo-positiveness captures the subtle difference among the non-monotonic GQs: the GQs that are homo-positive are more readily available to be used as semantic answers than GQs that are non-decreasing but not homo-positive. In this reading, the Q-domain yielded by the *wh*-phrase '*wh*-A' is the set consisting of the homo-positive GQs ranging over a subset of A. I write this set as '$^{\text{H}}A$', defined as follows:

(41) For a set of entities A, we have:
$^{\text{H}}A = \{\pi_{\langle et,t\rangle} \mid \textsc{smlo}(\pi) \subseteq A \wedge \pi \text{ is homogeneously positive}\}$, where
π is homogeneously positive if and only if $\pi \subseteq \mathbb{E}(\textsc{smlo}(\pi^+))$ and $\pi \subseteq \mathbb{E}(\textsc{smlo}(\pi^-))$.

It is yet unclear where the Homo-Positive-GQ Constraint comes from. It could be in the lexicon of a type-shifting operator, presupposed by the higher-order *wh*-trace, a constraint on semantic reconstruction, or even just a matter of pragmatics. For now, I treat the H-shifter in '$^{\text{H}}A$', which turns a set of entities A into a set of GQs, as a type-lifting operator syntactically presented within the *wh*-phrase. I also assume that this shifter asserts homo-positiveness; however, if compelling evidence suggests that homo-positiveness does not come from the *wh*-phrase, the meaning of the H-shifter can be altered accordingly.

With the above assumptions, I attribute the first-order/higher-order ambiguity of a *wh*-question to the absence/presence of the H-shifter within the *wh*-phrase. (For distributional constraints on the H-shifter and where this operator is placed within a noun phrase, see Sect. 4.) As exemplified in (42), 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 homo-positive GQs, and then the *wh*-phrase binds a higher-order trace π across the modal.

(42) Which books does John have to read?

a. First-order reading
[$_{\text{CP}}$ which-books λx_e [$_{\text{IP}}$ have-to [$_{\text{VP}}$ John read x]]]
$[\![wh\text{-Q}]\!] = \lambda x_e\colon x \in \textit{books}_@.\Box[\lambda w.\textit{read}_w(j,x)]$

b. Higher-order reading ($\Box \gg \pi$) (revised from (11))
[$_{\text{CP}}$ which-$^{\text{H}}$books $\lambda \pi_{\langle et,t\rangle}$ [$_{\text{IP}}$ have-to [π λx_e [$_{\text{VP}}$ John read x]]]]
$[\![wh\text{-Q}]\!] = \lambda \pi_{\langle et,t\rangle}\colon \pi \in {}^{\text{H}}\textit{books}_@.\Box[\lambda w.\pi(\lambda x_e.\textit{read}_w(j,x))]$

4. Distributing the 'conjunction-admitting' higher-order reading

As discussed in Sect. 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:

(43) a. Which children came?
$\not\rightsquigarrow$ *Only one of the children came.*

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

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

(44) a. Which children formed a team?
$\not\rightsquigarrow$ *Only one group of children formed any team.*

b. Which two children formed a team?
$\rightsquigarrow$ *Only one pair/group of children formed any team.*

According to Dayal (1996), the singular-marked *wh*-question (43a) presupposes uniqueness because its strongest true answer exists only when it has exactly one true answer. This analysis also extends to the numeral-modified *wh*-questions (43b,c), as argued in Sect. 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 a higher-order reading. Since the uniqueness effect is only related to the diagnostic for Boolean conjunctions, I call the higher-order reading in which a *wh*-question admits higher-order conjunctive answers the 'conjunction-admitting' higher-order reading. In the above examples, this reading is available in (43a) and (44a), but not in the other singular-marked and numeral-modified *wh*-questions. In contrast, Section 5 will discuss another reading of questions that rejects conjunctive answers but admits disjunctive answers, which I will call the 'conjunction-rejecting' higher-order reading.

Strikingly, in contrast to a numeral-modifier, a PP-modifier does not block higher-order readings. Compare (46) and (47), for example. Although *students (who are) in a group of two* is semantically similar to *two students*, the embedded question in (47) 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.

(45) I know { *who* / which *students* } presented a paper together, ...

a. ... the two boys.

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

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

a. ... the two boys.

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

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

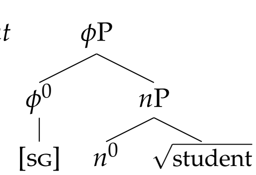
a. ... the two boys.

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

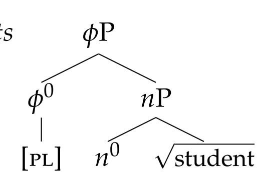
To account for the above distributional constraints, I assume that the H-shifter, which turns a set of entities into a set of GQs, must be applied locally to the *n*P within the *wh*-complement. In what follows, I argue that the application of the H-shifter is blocked in singular nouns and numeral-modified nouns due to conflicts in meaning and type.

First, I assume the following structure for a singular/plural bare noun:

(48) a. *student*



b. *students*



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

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

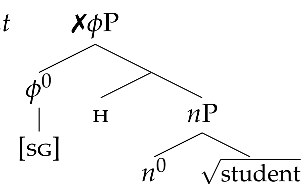
b. $[\![[\text{SG}]]\!] = \lambda P_{\langle e,t\rangle} \lambda x_e . \text{ATOM}(x) \wedge P(x)$
(or: $[\![[\text{SG}]]\!] = \lambda P_{\langle e,t\rangle} \lambda x_e : \text{ATOM}(x) . P(x)$)

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

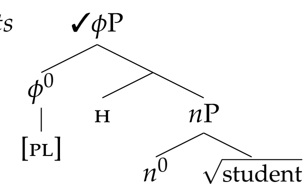
d. $[\![[\text{SG}](n\text{P})]\!] = \{a, b, c\}$

The above assumptions straightforwardly explain why the H-shifter cannot be used in a singular noun. In (50a), 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 amendment 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.[21]

(50) a. *student*



b. *students*



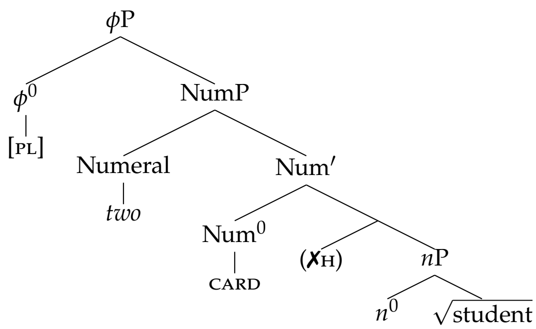
Next, let us consider numeral-modified nouns. Following Scontras (2014), I place cardinal numeral-modifiers at [Spec, NumP] and assume that Num^0 is located between n^0 and ϕ^0 and is occupied by a cardinality predicate CARD. As defined in (51a), CARD combines with a predicate P and a numeral N and returns 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 noun: 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.)

[21]This claim holds regardless of whether plurals are treated inclusively or exclusively. I also assume that the bare *wh*-words *who* and *what* have a structure similar to *which people/things*. Alternatively, one can treat [PL] as a predicate restrictor that asserts/presupposes/anti-presupposes non-atomicity (see also footnote 6). The latter option is more suited for languages in which plural-marking is not vacuous. For example, Spanish bare *wh*-words can be singular-marked or plural-marked. Interestingly, as seen in (i), while the singular form admits both atomic and non-atomic answers (see also footnote 7), the plural form admits only non-atomic answers (Maldonado 2020). For languages with non-vacuous plural-marking, it is plausible to assume that the the singular morpheme is semantically neutral, while the plural morpheme asserts/presupposes non-atomicity (Alonso-Ovalle and Rouillard 2019; cf. Elliott et al. 2020).

(i) a. Quién se fue?
who.SG REFL left
'Who.SG left?'
'$^{\text{OK}}$John.'/ '$^{\text{OK}}$John and Billy.'

b. Quiénes se fue?
who.PL REFL left
'Who.PL left?'
'#John.'/ '$^{\text{OK}}$John and Billy.'

(51) *two students*



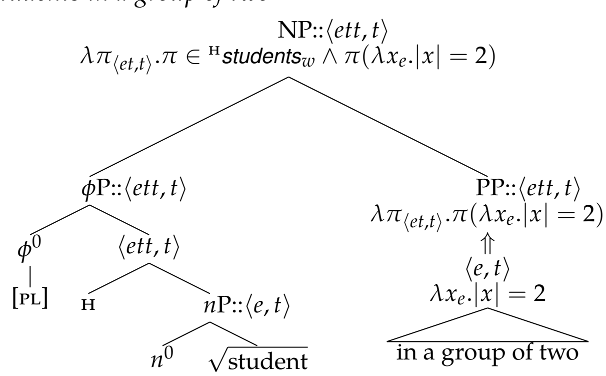
a. $[\![\textsc{card}]\!] = \lambda P \lambda N \lambda x.P(x) \wedge |x| = N$

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

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

In contrast to numeral-modifiers, PP-modifiers are adjoined to the entire NP/ϕP. Hence, the H-shifter can be used within the modified NP without causing type mismatch. As illustrated in (52), all we need is to apply argument-lifting to the PP-modifier (indicated by '⇑') and shift it into a set of GQs. The lifted PP composes with the higher-order ϕP standardly via Predicate Modification. This analysis also extends to NPs modified by a relative clause.

(52) *students in a group of two*



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

5.1. The puzzles

In Sect. 2.2, in the context of presenting stubborn collectivity and explaining uniqueness effects, I showed that singular-marked and numeral-modified *wh*-questions do not admit answers naming Boolean conjunctions. I further concluded in Sect. 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 the fact that applying the H-shifter yields semantic consequences that conflict with the atomicity requirement of singular nouns and the cardinality requirement of numerals.

Surprisingly, I find that narrow-scope disjunctions are not as bad as conjunctions when used as direct answers to a $\Box$-question that has a singular-marked or numeral-modified *wh*-phrase. This contrast is seen clearly in (53) and marginally in (54).[22]

(53) 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*)

(54) I know which two books John has to read, ...
  a. ?? ... the two French novels and the two Russian novels.
  b. ? ... the two French novels or the two Russian novels. (#*or* $\gg \Box$, ?$\Box \gg$ *or*)

Narrow scope readings of disjunctive answers are even more readily available in discourse. In (55), 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 given in Sect. 2.1, the fact that the disjunctive answer has a narrow-scope reading suggests that here the $\Box$-question admits higher-order answers, which conflicts with the aforementioned generalization that singular-marked *wh*-questions do not have a higher-order reading.

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

Moreover, I observe the same contrast between conjunctive and disjunctive answers in questions with a possibility modal (called '$\Diamond$-questions' henceforth). In what follows, I will first introduce the general facts about $\Diamond$-questions and their answers, and then discuss specifically the answers to singular-marked and numeral-modified $\Diamond$-questions.

It is well-known that $\Diamond$-questions are ambiguous between mention-some (MS-)readings and mention-all (MA-)readings (Groenendijk and Stokhof 1984; for a recent discussion on the characteristics and distribution of MS-readings, see Xiang 2016b: Chap. 2). As exemplified in (56), if interpreted with a MS-reading, the $\Diamond$-question can be naturally addressed by an answer that specifies only one feasible option; while in MA-readings, the $\Diamond$-question requires the addressee to exhaustively list all the feasible options. Crucially, MA-answers of $\Diamond$-questions can have either an elided conjunctive form, as in (56b), or an elided disjunctive form read as free-choice, as in (56c). Despite having different forms, both of the MA-answers convey the same conjunctive inference that we can use *Heim & Kratzer* for this class and we can use *Meaning & Grammar* for this class.

(56) 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: Chap. 2) proposes that MS-readings are higher-order readings: in the LF of a $\Diamond$-question with a MS-reading, the *wh*-phrase binds a higher-order trace across the possibility modal. MA-readings of $\Diamond$-questions arise as long as one of the following two

[22]The conjunctive continuation in (54a) is intuitively more acceptable than the conjunctive continuation in (53a), as pointed out by Gennaro Chierchia (pers. comm.). A reviewer of *Natural Language Semantics* also reported that they found no clear contrast between (54a) and (54b). One possible explanation of the improvement in (54a) is that the numeral modifier *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 has to read?'.

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 Sect. 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: Chap. 3). The infelicity of the continuations in (57) supports this view: the continuations name multiple choices of textbooks, whereas the preceding question-embedding sentence implies that there is only one feasible choice.

(57) 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 (58) does have a multi-choice reading.[23] They argue that the uniqueness presupposition triggered by the singular-marked *wh*-phrase can be accommodated under the scope of the modal verb *could*. The question can be read as 'Which x is such that x could be the unique letter missing in *fo__m*?'.

(58) Which letter could be missing in *fo__m*? (Hirsch and Schwarz 2019)
  a. (The missing letter could be) *a*.
  b. The missing letter could be *a* and the missing letter could be *r*.

Note that in Hirsch and Schwarz's original example (58), the multi-choice answer (58b) is not a direct answer. As seen in (59a,b), both the conjunctive response congruent with the question and the respective short answer are greatly degraded. In contrast, the multi-choice inference can be felicitously expressed via an elided free-choice disjunction, as in (59c). The same pattern is seen with numeral-modified ◇-questions, as shown in (60). (For reasons why the conjunctive answer (60a) is marginally acceptable, see footnote 22.)

(59) 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.)

(60) [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 (56), I re-illustrate Hirsch and Schwarz's idea in (61). According to Hirsch and Schwarz, the uniqueness presupposition triggered by the singular-marked *wh*-phrase *which textbook* can be interpreted either globally or locally. The global uniqueness reading says: 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: we will only use a single textbook for this class, and the questioner asks the addressee to specify one feasible option, as in a MS-reading, or all feasible options, as in a MA-reading. In

[23] This example is cited from the original SALT 29 poster presentation of Hirsch and Schwarz (2019). In a recently published proceedings paper, Hirsch and Schwarz 2020, they instead use a short answer "*a* or *r*", not a full sentence.

contrast to the number-neutral ◇-question (56) 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 (61a,b). The disjunction–conjunction contrast is also seen with the universal free-choice item *any book*, which is argued to be existential in the lexicon (Chierchia 2006, 2013), and the basic universal quantifier *every book*, as in (61c,d).

(61) 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 by a free-choice disjunction, not a conjunction.

Outside the realm of *wh*-questions, Gentile and Schwarz (2018) observe similar local uniqueness readings with *how many*-questions. Like singular-marked and numeral-modified *wh*-questions, *how many*-questions presuppose uniqueness. For example, the question in (62) cannot be felicitously responded to 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.[24]

(62) 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 (2018) observe that possibility modals can obviate violations of uniqueness in *how many*-questions: the question in (63) admits multi-choice answers like (63a,b). In analogy to the examples in (59)–(61), I add that the multi-choice answer cannot be expressed by an elided conjunction, as seen in (63c).

(63) How many students are allowed to solve this problem together?

[24] Fox (2020) disagrees with Gentile and Schwarz's conjecture. He argues that *how many*-phrases can range over higher-order conjunctions of degrees/intervals, in light of the following data:

(i) How high are we not allowed to fly?
  Below 50 meters (too low) and above 2000 meters (too high). ($and \gg \neg \gg \Diamond$)

I don't think the above example can knock down Gentile and Schwarz's conjecture. The information conveyed by the above conjunctive answer can even more preferably be expressed as a narrow-scope disjunction, namely, *below 50 meters or above 2000 meters* ($\neg \gg \Diamond \gg or$). To argue against Gentile and Schwarz (2018), one would have to find a case where the strongest true answer can and must be expressed as a higher-order conjunction over degrees/intervals. For example, in a context where the only group work constraint is that we cannot have two students work together while simultaneously another three students work together, the strongest true answer to the following negative ◇-question, if available, would have to be a narrow-scope conjunction. However, as seen in (ii), this conjunctive answer does not appear felicitous.

(ii) How many students are not allowed solve this problem together?
  # Two and three.
  (Intended: 'It is not allowed that [two students solve the problem together and simultaneously (another) three students solve this problem together].')

a. Two are OK and three are OK.

b. Two or three.

c. # Two and three.

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

The following subsections provide two approaches to deriving the 'conjunction-rejecting' reading and explaining its distributional constraints. One approach treats the 'conjunction-rejecting' reading and the 'conjunction-admitting' reading as the same reading and 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 two puzzles.

5.2. A uniform approach

The uniform approach treats the 'conjunction-rejecting' reading and the 'conjunction-admitting' reading as a single 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 a singleton.[25] For example, the Boolean conjunction $a^{\Uparrow} \cap b^{\Uparrow}$ is ruled out because its only minimal witness set $\{a, b\}$ is not a singleton. The (b)-condition applies the 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 (64b) because their smallest live-on sets (viz., $\{a \oplus b\}$ and $\{a \oplus b, c\}$, respectively) include a non-atomic element $a \oplus b$.

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

(65) A GQ π has the cardinality N if and only if
   a. the minimal witness sets of π are all singleton sets;
   b. every member in the smallest live-on set of π has the cardinality N.

Incorporating the above assumptions into the semantics of the singular feature [sg] and the cardinality predicate CARD yields the revised definitions in (67). 'MWS(A, x)' is read as 'A

[25] 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 if and only if $A \subseteq B$ and $\pi(A)$. For example, given a discourse domain including three students *a*,*b*,*c*, 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.

is a minimal witness set (MWS) 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.

(66) Old definitions

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

b. $[\![\text{CARD}]\!] = \lambda P_{\langle e,t\rangle} \lambda N \lambda x_e . P(x) \wedge |x| = N$

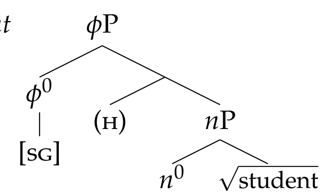
(67) New definitions

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

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

With the revised definitions, the H-shifter can be used regularly in singular nouns and numeral-modified nouns. In a discourse with three students a,b,c, 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: $[\![n\text{P}]\!] = \{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$.

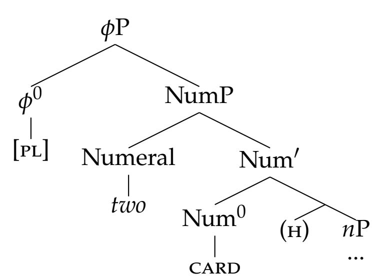
(68) *student*



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

b. With H:
$[\![\phi\text{P}]\!] = \{a^{\Uparrow}, b^{\Uparrow}, c^{\Uparrow}, a^{\Uparrow} \cup b^{\Uparrow}, a^{\Uparrow} \cup c^{\Uparrow}, b^{\Uparrow} \cup c^{\Uparrow}, a^{\Uparrow} \cup b^{\Uparrow} \cup c^{\Uparrow}\}$
$= \{\bigcup A \mid A \subseteq \{x^{\Uparrow} \mid x \in \{a, b, c\}\}$

(69) *two students*



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

b. With H:
$[\![\phi\text{P}]\!] = [\![\text{NumP}]\!]$

$$= \left\{ \begin{array}{ccc} (a \oplus b)^{\Uparrow}, & (b \oplus c)^{\Uparrow}, & (a \oplus c)^{\Uparrow}, \\ (a \oplus b)^{\Uparrow} \cup (b \oplus c)^{\Uparrow}, & (a \oplus b)^{\Uparrow} \cup (a \oplus c)^{\Uparrow}, & (b \oplus c)^{\Uparrow} \cup (a \oplus c)^{\Uparrow}, \\ & (a \oplus b)^{\Uparrow} \cup (b \oplus c)^{\Uparrow} \cup (a \oplus c)^{\Uparrow} & \end{array} \right\}$$
$$= \{\bigcup A \mid A \subseteq \{x^{\Uparrow} \mid x \in \{a \oplus b, b \oplus c, a \oplus c\}\}$$

In sum, in the uniform approach, higher-order readings of *wh*-questions are uniformly derived in two steps as follows.[26] First, applying an H-shifter to the *n*P within the *wh*-complement yields a higher-order domain. (In the following examples, BOOK/CHILD abbreviates the *n*P within the noun *book(s)/child(ren)*.) In particular, if the *wh*-complement is a bare plural as in (70) or the *wh*-phrase is number-neutral, the yielded domain is the set of homo-positive GQs ranging over a subset of $[\![nP]\!]$. If the *wh*-complement is singular-marked as in (71), the yielded domain is a set of disjunctions of Montagovian individuals $x^{\Uparrow}$ where x is an atomic entity in $[\![nP]\!]$. If the *wh*-complement is modified by a numeral N as in (73), the yielded domain is a set of disjunctions of $x^{\Uparrow}$ where x is an entity in $[\![nP]\!]$ 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. This uniform analysis easily explains the availability of narrow-scope disjunctive answers in (70) and (71), while accounting for the unavailability of conjunctive answers in (71) and (73) (cf. (70) and (72)).

(70) Which books does John have to read?
[$_{\text{CP}}$ [$_{\text{DP}}$ which [PL]-$^{\text{H}}$BOOK] $\lambda\pi_{\langle et,t\rangle}$ [$_{\text{IP}}$ have-to [π λx_e [$_{\text{VP}}$ John read x]]]]
  a. The French novels.
  b. The French novels or the Russian novels. The choice is up to him.
  c. The French novels and the Russian novels.

(71) Which book does John have to read?
[$_{\text{CP}}$ [$_{\text{DP}}$ which [SG]-$^{\text{H}}$BOOK] $\lambda\pi_{\langle et,t\rangle}$ [$_{\text{IP}}$ have-to [π λx_e [$_{\text{VP}}$ John read x]]]]
  a. Book A.
  b. Book A or Book B. The choice is up to him.
  c. # Book A and Book B.

(72) Which children formed a team?
[$_{\text{CP}}$ [$_{\text{DP}}$ which [PL]-$^{\text{H}}$CHILD] $\lambda\pi_{\langle et,t\rangle}$ [$_{\text{IP}}$ π λx_e [$_{\text{VP}}$ x formed a team]]]
  a. The two girls.
  b. The two girls and the two boys.

(73) Which two children formed a team?
[$_{\text{CP}}$ [$_{\text{DP}}$ which [PL]-two-$^{\text{H}}$CHILD] $\lambda\pi_{\langle et,t\rangle}$ [$_{\text{IP}}$ π λx_e [$_{\text{VP}}$ x formed a team]]]
  a. The two girls.
  b. # The two girls and the two boys.

[26] 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 propose that *wh*-questions do not have first-order readings at all. At this point I do not see any direct evidence for first-order readings of questions. However, denying their existence outright would lead to the 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.

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 it accounts for the contrast between disjunctive and conjunctive MA-answers (Sect. 5.3.2).

5.3.1. Questions with a necessity modal

Let me start with a singular-marked $\Box$-question. (74) 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 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 (74b). Second, uniqueness is evaluated at whichever scopal position the reconstructed noun adjoins to. In both (74a) and (74b), uniqueness takes scope below the necessity modal.[27]

(74) Which book does John have to read?

a. First-order reading ($\Box \gg \iota$)
'Which x_e is such that the book that John reads has to be x?'

i. $[_{\text{CP}}$ which$_D$ λx_e $[_{\text{IP}}$ have-to $[_{\text{VP}}$ x is the book John read $]]]$

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

b. Higher-order reading ($\Box \gg \pi \gg \iota$)
'Which (homo-positive) $\pi_{\langle et,t\rangle}$ is such that the book that John reads has to be π?'

i. $[_{\text{CP}}$ which$_{^{\text{H}}D}$ $\lambda\pi_{\langle et,t\rangle}$ $[_{\text{IP}}$ have-to $[$ π λx_e $[_{\text{VP}}$ x is the book John read $]]]]$

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

The following trees illustrate the two LF structures in more detail. 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 reads* denotes a one-place predicate. Second and third, THE-insertion introduces uniqueness, and variable insertion introduces a variable bound by the *wh*-phrase.[28] In particular, in the LF (76) for

[27] 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 pair of matching cards, either red aces or black aces, but this alone does not guarantee a win.

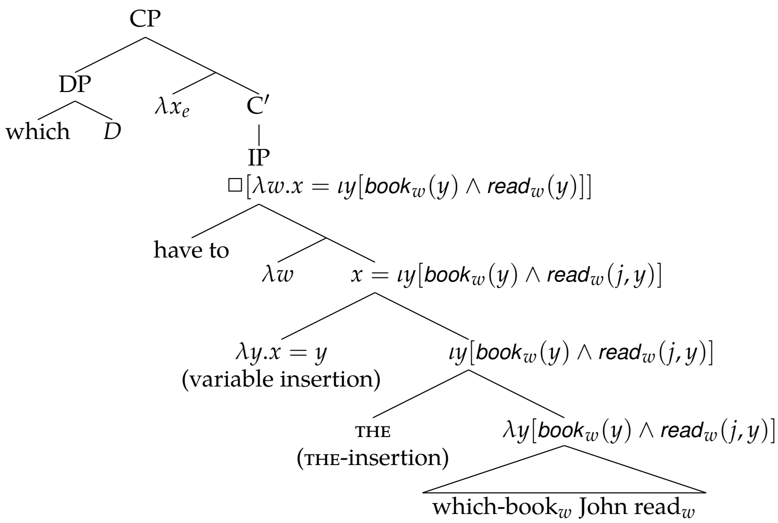
(i) Which two cards do you need to win the game?
The two red aces or the two black aces.

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.

[28] One might have concerns with the assumed syntax for reconstruction. On the one hand, the assumed THE-insertion and variable insertion 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

the higher-order reading, 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.[29] This derivation is similar to what I assumed for the conjunction-admitting higher-order reading.

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



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

---

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 structure used for such derivation is unconventional, but this is not necessarily a problem for considering (76) as the structure that derives the 'conjunction-rejecting' reading. As seen in Sect. 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 (53), (54), and (57)). 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 question under discussion.

[29] I assume a locality constraint to the effect that the variable introduced by variable insertion has to be 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. By contrast, the structure in (i), where variable insertion (underlined) introduces an individual variable x bound by the higher-order *wh*-trace, is ruled out.

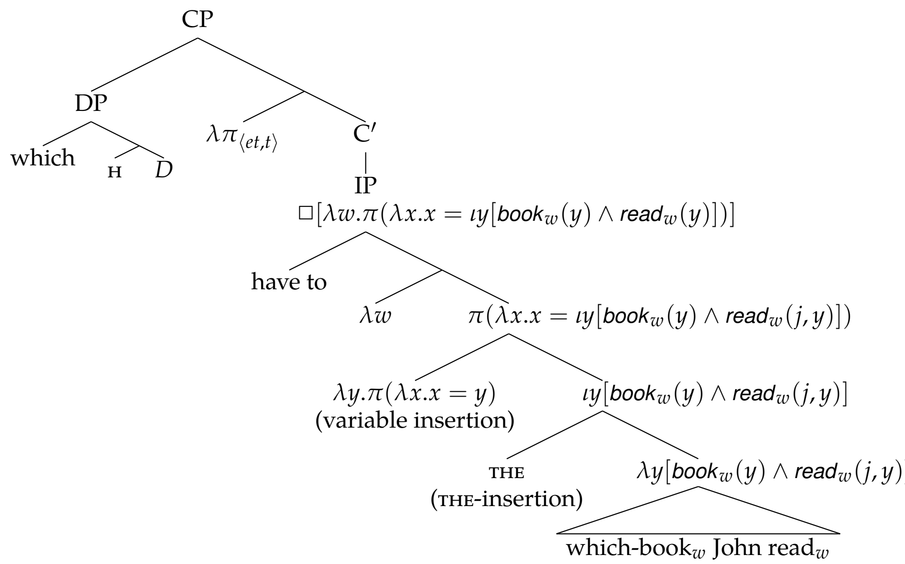
(i) * [*wh*P $\lambda\pi_{\langle et,t\rangle}$ [have-to [π λx_e [$\underline{\lambda y.x = y}$ [THE [which book John read]]]]]

This constraint avoids unattested split-scope readings of conjunctive answers to questions with an existential quantifier. Observe that the singular-marked question in (ii) cannot be felicitously responded to by a conjunction. The infelicity of the conjunctive answer suggests that this answer cannot be interpreted with a split-scope reading such as 'For a math problem x_1, Andy is the unique student who solved x_1, and for a math problem x_2, Billy is the unique student who solved x_2' (*and* $\gg \exists \gg \iota$). The unavailability of this reading suggests ruling 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]]]]]]



The above derivation predicts that the higher-order *wh*-trace π immediately scopes over uniqueness. This prediction explains why a *wh*-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 (77b), unless Book A and B are the same book, applying the Q-function to the conjunction $a^{\Uparrow} \cap b^{\Uparrow}$ yields a contradiction.

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

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

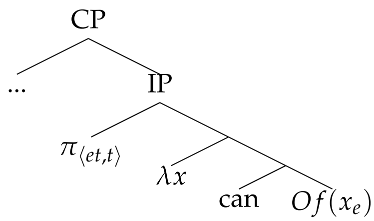
5.3.2. Questions with a possibility modal

The MA-answer to a question is the strongest true answer which entails all the true answers to this question. As seen in Sect. 5.1, in responding to a $\Diamond$-question, the MA-answer can be expressed in the form of an elided conjunction or a free-choice disjunction. However, if the *wh*-phrase in the $\Diamond$-question is singular-marked or numeral-modified, the MA-answer can only be expressed in the form of a free-choice disjunction. In Xiang 2016b (Chap. 2), I argue that the MA-reading calling for a conjunctive answer and the MA-reading calling for a disjunctive answer are derived via different LF structures. With the assumed reconstruction, the contrast in LF between these two MA-readings naturally predicts the conjunction–disjunction asymmetry in singular-marked and numeral-modified $\Diamond$-questions.

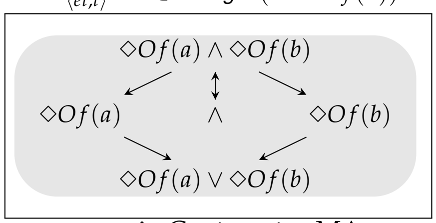
In the conjunctive MA-reading, the *wh*-phrase binds a higher-order trace which takes scope above the possibility modal. The following considers the interpretations of a number-neutral $\Diamond$-question in cases where the higher-order *wh*-trace π takes scope below and above the possibility modal *can*. For these two cases, the structure of the question nucleus and the yielded Q-function and answer space are illustrated in (78a,b). '$f(x)$' abbreviates the proposition *we use x for this class*. For example, $\Diamond Of(a)$ is read as 'It can be the case that we use only Book A [as a textbook] for this class.'[30] The illustrations of the answer space consider only the propositions derived by applying the Q-function to the following four GQs: 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 entailments among the propositions, and shading marks the propositions that are true in the considered world.

(78) (Context: 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.

a. $\pi \gg \Diamond$

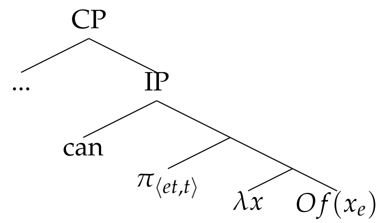


$\lambda\pi_{\langle et,t\rangle} : \pi \in {}^{\text{H}}\textit{thing}.\pi(\lambda x.\Diamond Of(x))$

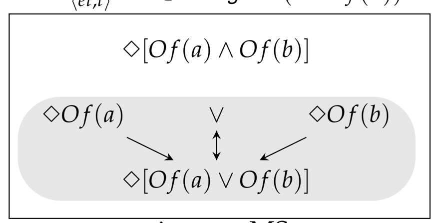


$\pi \gg \Diamond$: Conjunctive MA

b. $\Diamond \gg \pi$



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



$\Diamond \gg \pi$: MS

In (78a) 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 can be the case that we only use Book A for this class, and it can be the case that we only use Book B for this class.' In contrast, in (78b), 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 can be the case that we only use Book A for this class and only use Book B 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 ($\pi \gg \Diamond$).

[30] In both structures, an exhaustivity operator 'O' ($\approx$ *only*) is inserted under the possibility modal and is associated with the individual *wh*-trace x.

(i) $O_C = \lambda p \lambda w.p(w) = 1 \wedge \forall q \in C[p \not\subseteq q \rightarrow q(w) = 0]$ (Chierchia et al. 2012)
(For any proposition p and world w, $O_C(p)$ is true if and only if p is true in w, and any proposition in C that is not entailed by p is false in w.)

In Xiang 2016b, I propose that the MS-reading arises if the higher-order *wh*-trace π scopes below the possibility modal. The local O-operator is assumed to account for 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.

Next, consider the corresponding singular-marked $\Diamond$-question in (79). Again, the puzzle is that multi-choice answers to this question cannot have an elided conjunctive form. As assumed in Sect. 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 yields the following scopal pattern: $\pi \gg \iota \gg \Diamond$. As shown in (79b), unless A and B are the same book, applying the derived Q-function to the Boolean conjunction $a^{\Uparrow} \cap b^{\Uparrow}$ yields a contradiction.

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

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

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

(# *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 a free-choice-triggering operator, regardless of whether this trace takes scope below or above the possibility modal. In what follows, drawing on observations about the uses of the Mandarin particle *dou* in questions and declaratives, I assume that disjunctive MA and free-choice in English are obtained via the application of the covert operator DOU, whose meaning is akin to the Mandarin particle *dou*.

The Mandarin particle *dou* has many uses, but the most important ones in the context of interpreting $\Diamond$-questions are the following: in $\Diamond$-questions, associating *dou* with a *wh*-phrase blocks the MS-reading, as seen in (80a); in $\Diamond$-declaratives, associating *dou* with a pre-verbal disjunction yields a free-choice inference, as shown in (80b). (For other uses of *dou* and a unified analysis, see 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.[31]

(80) 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 semantics of sub-alternatives varies by the item that *dou* is associated with and the prosodic pattern of the sentence that *dou* appears in, causing alternations in function of *dou* (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.

[31] There is a rich literature on the semantics of *dou*, but very few analyses can account for the widely discussed distributor use and the *even*-like use of *dou* while also explaining its free-choice-triggering effect in declarative sentences. Besides the account of Xiang (2016b, 2020) adopted here, another possible candidate is Liu 2016. Although Liu (2016) does not discuss free-choice disjunctions in particular, his analysis predicts the mandatory use of pre-/recursive exhaustifications in the presence of *dou*. See Xiang 2020 for a review.

With this semantics, applying *dou*/DOU to a disjunctive sentence yields a universal free-choice inference.

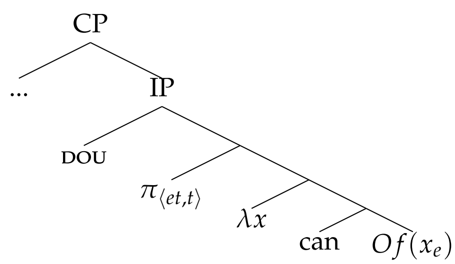
(81) $[\![dou_C]\!] = \lambda p \lambda w : \exists q \in \textsc{Sub}(p, C).p(w) = 1 \wedge \forall q \in \textsc{Sub}(p, C)[O_C(q)(w) = 0]$
(For any proposition p and world w, $[\![dou_C]\!](p)(w)$ is defined only if C contains a sub-alternative of p. When defined, $[\![dou_C]\!](p)(w)$ asserts that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

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

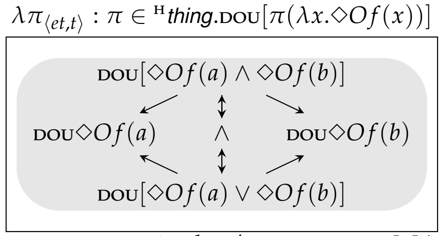
The following illustrates two possible structures of the question nucleus for the disjunctive MA-reading, as well as the Q-function and the answer space yielded by each structure. In both structures, a covert DOU-operator is presented at the left edge of the question nucleus and is associated with the 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 (84), 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.

(83) (Context: 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 $\gg \pi \gg \Diamond$

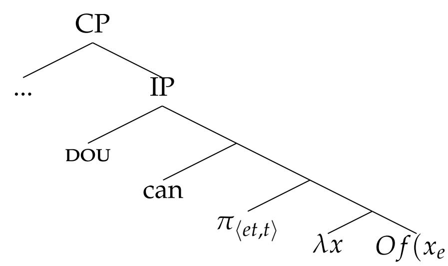


$\lambda \pi_{\langle et,t \rangle} : \pi \in {}^{\text{H}}\textit{thing}.\textsc{dou}[\pi(\lambda x.\Diamond Of(x))]$

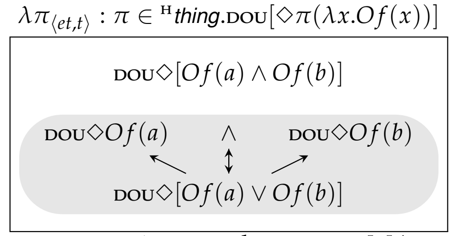


DOU $\gg \pi \gg \Diamond$: disj/conj-unctive MA

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



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



DOU $\gg \Diamond \gg \pi$: disjunctive MA

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

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

$$= \Diamond[Of(a) \vee Of(b)] \wedge [\Diamond Of(a) \leftrightarrow \Diamond Of(b)]$$
$$= \Diamond Of(a) \wedge \Diamond Of(b)$$

Next, let us 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 *wh*-trace. For the number-neutral $\Diamond$-question in (83), 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 above or below the possibility modal yields a crucial difference with free-choice disjunctive answers. If π takes wide scope, as seen in (85a), the derived free-choice inference is a contradiction, just like the case of the wide-scope conjunctive answer in (79). In contrast, as seen in (85b), if π takes narrow scope relative to the possibility modal, the derived free-choice inference is not contradictory and is a desired MA-answer.

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

a. If $\textsc{dou} \gg \pi \gg \iota \gg \Diamond$:

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

(#*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 $\textsc{dou} \gg \Diamond \gg \pi \gg \iota$:

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

(*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, number-neutral $\Diamond$-questions admit three types of MA-answers, including wide-scope conjunctions, wide-scope free-choice disjunctions, and narrow-scope free-choice disjunctions. As for a singular-marked $\Diamond$-question, however, the MA-answer can only be a narrow-scope disjunction: due to the uniqueness inference associated with the individual variable bound by the higher-order *wh*-trace, wide-scope conjunctions and wide-scope free-choice disjunctions yield contradictory inferences. This analysis also applies to numeral-modified $\Diamond$-questions.

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 *wh*-questions admit higher-order disjunctive answers. 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 *wh*-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 the individual variable immediately bound by the higher-order *wh*-trace triggers uniqueness, and conjoining two uniqueness inferences yields a contradiction.

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

These two approaches, however, are not notational equivalents. 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 (85a) shows an exception: if disjunctions are interpreted as wide-scope free-choice, they would, just like conjunctions, result in contradictions. In contrast, the uniform approach does not predict disjunctions to be deviant in any case. Unfortunately, it is hard to check these 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, but on the other hand, only 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 semantic 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 GQs 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, namely, that singular-marked and numeral-modified questions admit disjunctive answers but not conjunctive answers. I provided two potential explanations for 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 syntactically reconstructed, which gives rise to local uniqueness and yields contradictions for conjunctive answers.

Acknowledgement This paper significantly expands on Xiang 2019, published in *Proceedings of the 22nd Amsterdam Colloquium*. For helpful discussions, I thank Luis Alonso-Ovalle, Lucas Champollion, Gennaro Chierchia, Danny Fox, Michael Glanzberg, Manuel Križ, Floris Roelofsen, Vincent Rouillard, Benjamin Spector, Bernhard Schwarz, and the audiences at the Spring 2019 Rutgers Seminar, Georg-August-Universität Göttingen, Ecole Normale Supérieure, and the 22nd Amsterdam Colloquium. I am grateful to two anonymous reviewers of *Natural Language Semantics*, whose comments led to notable improvements, and to Christine Bartels for her masterful editorial work. All errors are mine.

References

Alonso-Ovalle, Luis, and Vincent Rouillard. 2019. Number inflection, Spanish bare interrogatives, and higher-order quantification. In *Proceedings of the North East Linguistic Society 49*, ed. Maggie Baird and Jonathan Pesetsky, 25–38. Amherst, MA: GLSA. http://people.linguistics.mcgill.ca/~luis.alonso-ovalle/papers/AlonsoOvalle_Rouillard.pdf.

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

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(3):561–622.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry* 37(4): 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. Berlin: Mouton de Gruyter.

Cremers, Alexandre, and Emmanuel Chemla. 2016. A psycholinguistic study of the exhaustive readings of embedded questions. *Journal of Semantics* 33(1): 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. Manuscript, ZAS Berlin (Leibniz-Center of General Linguistics). https://ling.auf.net/lingbuzz/003959

Erlewine, Michael Yoshitaka. 2014. Movement out of focus. PhD dissertation, MIT.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33(1): 63–96.

Fox, Danny. 2013. Mention-some readings. MIT seminar notes. http://lingphil.mit.edu/papers/fox/class1-3.pdf.

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

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

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

George, B. 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(1):185–229.

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

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.

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. Jerusalem: IATL. 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 the 29th Semantics and Linguistic Theory, 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 29*, ed. Katherine Blake, Forrest Davis, Kaelyn Lamp, and Joseph Rhyne, 748–767. Washington, D.C.: LSA. 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(1): 19–40.

Jacobson, Pauline. 2016. The short answer: implications for direct compositionality (and vice versa). *Language* 92(2):331–375.

Keenan, Edward L, and L.M. Faltz. 1985. *Boolean Semantics for Natural Language*. Dordrecht: Reidel.

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 University.

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

Liu, Mingming. 2016. Varieties of alternatives: Mandarin focus particles. *Linguistics and Philosophy* 40:61–95.

Maldonado, Mora. 2020. Plural marking and d-linking in Spanish interrogatives. *Journal of Semantics* 37(1):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. Berlin: De Gruyter.

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

Rullmann, Hotze, and Sigrid Beck. 1998. Presupposition projection and the interpretation of *which*-questions. In *Proceedings of Semantics and Linguistic Theory 8*, ed. Devon Strolovitch and Aaron Lawson, 215–232. Ithaca, NY: CLC Publications. https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/2811.

Sauerland, Uli. 2003. A new semantics for number. In *Proceedings of Semantics and Linguistic Theory 13*, ed. Robert B. Young and Yuping Zhou, 258–275. Ithaca, NY: CLC Publications. 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. PhD dissertation, Harvard University.

Sharvy, Richard. 1980. A more general theory of definite descriptions. *The Philosophical Review* 89(4): 607–624.

Spector, Benjamin. 2007. Modalized questions and exhaustivity. In *Proceedings of Semantics and Linguistic Theory 17*, ed. M. Gibson and T. Friedman, 282–299. Ithaca, NY: CLC Publications. 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(4): 677–686.

Srivastav, Veneeta. 1991. WH dependencies in Hindi and the theory of grammar. PhD 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: Kluwer.

Theiler, Nadine, Floris Roelofsen, and Maria Aloni. 2018. A uniform semantics for declarative and interrogative complements. *Journal of Semantics* 35(3): 409–466.

Uegaki, Wataru. 2015. Interpreting questions under attitudes. PhD dissertation, MIT.

Uegaki, Wataru. 2018. On the projection of the presupposition of embedded questions. In *Proceedings of Semantics and Linguistic Theory 28*, ed. Sireemas Maspong, Brynhildur Stefánsdóttir, Katherine Blake, and Forrest Davis, 789–808. Washington, D.C.: LSA. 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. Accepted by *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 20*, ed. Polina Berezovskaya Nadine Bade and Anthea Schöller, 815–830. University of Tüebingen. https://semanticsarchive.net/sub2015/SeparateArticles/Xiang-SuB20.pdf.

Xiang, Yimei. 2016b. Interpreting questions with non-exhaustive answers. PhD 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. Amsterdam: ILLC. 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* 37(2): 171-217.