(Non-)exhaustivity in embedded questions: Contextual, lexical and structural factors[1]

Morgan MOYER — *Rutgers, The State University of New Jersey, New Brunswick*
Kristen SYRETT — *Rutgers, The State University of New Jersey, New Brunswick*

Abstract. *Wh*-questions often allow for either a Mention-Some (MS) or a Mention-All (MA) answer/interpretation, but what licenses MS answers remains unclear. In this paper, we systematically investigate a set of linguistic (*wh*-word, matrix verb, modality/non-finiteness) and discourse/contextual (questioner goals) factors that give rise to MS reading. We present evidence from two experiments showing that MS readings of embedded questions are not only available across various forms of embedded question, but also in environments that have been claimed not to license MS readings: specifically, finite clause embedded *who*-questions (e.g., *Mary knows who came to the party*). Moreover, acceptability of both MS and MA is influenced by contextual information. These results thus call into question a strictly semantic approach towards the resolution and interpretation of *wh*-questions, and an approach that claims obligatory licensing of MS by modals, and at the same time, provide empirical support for theories that incorporate the role of a speaker's context-dependent discourse goals.

Keywords: questions, embedded questions, mention some, non-exhaustivity.

1. Introduction

The truth-conditions of *wh*-questions, and more specifically, their embedded counterparts as in (1), have been the source of debate amongst semanticists over the past half century (Groenendijk and Stokhof, 1982, 1984; Hintikka, 1976).

(1) Mary knows **where to find an Italian newspaper**.

It is perhaps not surprising that the truth of (1) depends on the possible answers to the root question *Where can I find an Italian newspaper?* If there are four places where one can find an Italian newspaper, then the question is said to exhaustively specify that set of locations (see Groenendijk and Stokhof, 1984; Karttunen, 1977).[2] Thus, for (1) to be judged true, Mary must know (or demonstrate knowledge of) all four places. We could, however, also identify a second reading where Mary's knowledge need not be exhaustive. (1) may be true in a situation where Mary knows *at least one* true location. These two readings are known as **Mention-All** (MA) and **Mention-Some** (MS) answers, respectively (Hintikka, 1978).

This indeterminacy linked to how many answers Mary must know in order for (1) to be judged true can be resolved when we consider the larger discourse context and what kind of answer might satisfy the questioner's goals. If a tourist in town for a day asks Mary, *Where can I find an Italian newspaper?* it may suffice for Mary to name a single, nearby, shop (that is open). In contrast, if an economist interested in the local newspaper market or a realtor seeking to

[1] We are grateful to the research assistants in the Rutgers Laboratory for Developmental Language Studies for helping to design and administer these studies. We would like to thank audiences at Sinn und Bedeutung 23, Rutgers University, and Chicago Linguistic Society 54. Finally, thanks goes to Tim Dawson for his help in implementing the design for Experiment 2.

[2] We abstract away from issues of exhaustivity strength in order to focus on the distinction between non-exhaustive and exhaustive readings. For summary of debates on exhaustive readings, see (e.g., Klinedinst and Rothschild, 2011; Dayal 2016; George, 2011, Xiang, 2016).

determine whether Italian families will acclimate to the neighborhood asks the very same question, their goals likely require an exhaustive response from Mary. To wit, the truth-value judgment of (1) depends on the alignment between contextually-relevant discourse goals, and the answers that Mary gives to the question.

Now, while some questions, such as the embedded question in (1), or its root question counterpart, permit either MS or MA answers, other questions are more restrictive and seem to require an MS answer. Consider the sentence in (2).

(2) Mary knows **who came to the party**.

The presence of an MA reading for (2) is uncontroversial: every semantic theory (Beck and Rullmann, 1999; George, 2011; Groenendijk and Stokhof, 1982, 1984; Guerzoni and Sharvit, 2007; Hamblin, 1973; Heim, 1994; Karttunen, 1977; Klinedinst and Rothschild, 2011; Lahiri, 2002; Nicolae, 2013; Spector and Égre, 2015; Theiler, 2014; Uegaki, 2015; Xiang, 2016) predicts MA answers (with varying levels of exhaustivity, which we do not address here) to be permitted. By contrast, it been claimed to lack the MS reading (Fox, 2014; George, 2011). The question is *why*. In comparing (1) and (2), we can observe two surface-level differences: the *wh*-word and the finiteness of the embedded clause. There is also the issue of what is at stake: why are we asking about newspaper vendors or party attendees? What's more, embedded questions can be the complement of different embedding verbs, *know* among them. It is unclear what precisely the role of these factors is in MS answer licensing and MA answer requirement.

These factors would be assigned a different role of varying influence and centrality, depending on the theoretical approach to MA/MS readings in questions. Thus, it is imperative to carefully identify their contribution, whatever one's theoretical framework, in order to determine their contribution. We take space here to spell out in rough detail a number of major formal approaches to the study of MS/MA. We name the approaches based on the general features of how the theory accounts for MS.

First, general linguistic *ambiguity* approaches (Beck and Rullmann, 1999; George, 2011) posit at least two logical forms for each reading. Such a general ambiguity approach predicts that MS and MA readings should be equally available in any question regardless of form, *modulo* the context and the verb's selectional restrictions (George, 2011; Grimshaw, 1979). A more specific version of this approach holds that only *modal* questions are ambiguous (George, 2011; Fox, 2014; Xiang, 2016), and since non-finite clauses are said to be modal, then (1) permits both MS and MA in contrast to (2). A strong modal approach would predict that the MS reading will *only* be available in modal questions, as the ambiguity arises due to an interaction with a modal in the LF.

In contrast, *underspecification* approaches (Asher and Lascarides, 1998; Lahiri, 2002; van Rooij, 2003) posit a single representation whose logical form is semantically unspecified for exhaustivity. For example, Asher and Lascarides (1998) propose that information provided by semantics is weaker than that provided by pragmatics; pragmatics provides a *strengthening* of semantics. For them, the logical form allows both exhaustive and non-exhaustive answers and thus imposes no exhaustivity requirement *per se*. The prediction is that any given reading should be available whenever the relevant *contextual* factors license it, regardless of question

form.[3] Finally, *pragmatic* approaches (Groenendijk and Stokhof, 1982, 1984) hold that the semantic contribution of a question is always exhaustive, and a non-exhaustive meaning/answer is permitted when the context licenses *weakening*. Given these contrasting approaches to question/answer dynamics, our goal is to present empirical evidence for the individual contributions of the factors identified above in licensing MS answers. By so doing, we hope to adjudicate between theories and flesh out the balance between semantics and pragmatics in question/answer relations.

In this paper, we present two sets of experiments that quantify both the linguistic and contextual discourse factors that modulate the availability of MS readings of embedded *wh*-questions. In Section 2, we carefully outline these factors, including discussion about each from the literature. Our focus will be on embedded questions. In Section 3, we present Experiment 1, which establishes a baseline to identify how generalizable the MS reading actually is (data that has, up to this point, been missing from discussions). To preview our findings, we show that the factors we have isolated do indeed modulate acceptability, and crucially, that while the presence of a modal boosts acceptability of MS readings, the absence of a modal does not yield categorical rejection of MS interpretations. In Section 4, we present Experiment 2, in which we manipulate the context and shows that exhaustivity in *wh*-questions is influenced by the speaker's goals in asking the question. We conclude in Section 5.

2. Semantics and Pragmatics of Questions: Linguistic and Discourse Licensors

Three factors that have been argued to modulate the availability of the MS reading in embedded questions include the matrix verb that embeds the question, the *wh*-word heading the question, and the presence of a modal element in the embedded question. We discuss each of these before turning to a fourth, contextual factor.

2.1. Linguistic Cues to interpretation

2.1.1. Matrix verbs

Clausal embedding verbs like *know* are known for their syntactic and semantic selectional restrictions (Grimshaw, 1979). *Know* can syntactically embed a question (e.g., *know who came to the party*) or a proposition (e.g., *know that Sam came to the party*). In contrast, a verb like *wonder* can only embed a question (*wonder whether/who …* vs. **wonder that…*). Some theoreticians have supposed that the distributional differences between readings of embedded questions arise from similar semantic restrictions. This discussion has typically revolved around three levels of exhaustivity (i.e., weak, intermediate, strong), and whether the distribution of those readings under different verbs is revealing about the semantic representation of *wh*-questions (see George, 2011; Guerzoni and Sharvit, 2007; Heim, 1994; Klinedinst and Rothschild, 2011; Lahiri, 2002; Spector and Egré, 2015). It is generally accepted by semanticists that some level of exhaustive reading is available for *know-wh*; it is often argued that a strong exhaustive reading is most available. However, Cremers and Chemla (2016) showed experimentally that *know-wh* actually gives rise to the range of exhaustive readings.

In contrast to *know*, some verbs, for example *predict*, are argued to prefer only weak exhaustive readings (Beck and Rullmann, 1999; Heim, 1994; Klinedinst and Rothschild, 2011; Sharvit,

[3] We discuss more specifically which contextual factors matter in Section 2.2.

2002). George (2011) argues that questions are ambiguous between a (strong) MA and an MS reading, and verbs will select for one (i.e., for the X-operator in the MA representation) or none. According to this account, weak exhaustivity is equivalent to the MS meaning. Thus, *predict* selects for MS, while *know* selects for MA. However, given that MS does easily arise in questions with *know* as in (1), we may question the power of these selectional restrictions on their own.

2.1.2. *Wh*-word

As we pointed out in the comparison between (1) and (2), one difference between these embedded questions lies in the type of *wh*-word. In fact, there are observations appearing throughout the literature suggesting that differences in the *wh*-word hearing embedded questions yield differences in judgments of (non-)exhaustivity. See, for example, Ginzburg (1995a, b) and Asher and Lascarides (1998). Constructed examples presented in support of MA readings usually feature *who*-questions, as in (2), while examples in support of MS readings feature *where*-questions, as in (1)—or more generally, non-*who*-questions. However, nothing about the lexical semantics of these *wh*-words predicts differences in exhaustivity, so these observations are curious. Consider, for example, (3), from Asher and Lascarides (1998):

(3) John knows **how to get to the treasure**.

It seems natural to interpret (3) as true just in case John knows at least one way (any way) to get to the treasure, and unreasonable that he should know all of the possible ways to get to the treasure. What matters is that John is able to find a way to the treasure.

Asher and Lascarides discuss the possibility that *wh*-words might encode preferences for resolving exhaustivity in certain ways (i.e., that *who*-questions are exhaustive), but that these preferences may be overridden by contexts that make explicit a questioner's goals and mental state. However, they argue given the variability of interpretations observed, a unified question semantics should provide a weak (monotonic) meaning (a non-exhaustive one). This may then be strengthened via pragmatics (to an exhaustive reading) given the questioner's plans and cognitive state, rather than the other way around (p. 262). This last aspect is crucial to their account, so we return to it in Section 2.2, which addresses discourse goals.

Further differences between *wh*-words may be observed in how the referential domain of a *wh*-word is determined. This may influence whether an MS reading is possible. For example, Ginzburg (1995a) notes that *who*- and *where*-phrases differ in the granularity of their referential domains. Consider the two contexts and in (4) and (5).

(4) a. Mary has just stepped off a plane in Helsinki.
b. Flight Attendant: Do you know **where** you are?
c. Mary: Helsinki.

(5) a. Mary has just gotten out of a taxi in front of her hotel.
b. Taxi Driver: Do you know **where** you are?
c. Mary: Helsinki.

(6) Mary knows **where** she is.

(6) seems true in (4), but false in (5). According to Ginzburg, *where* is vague with respect to specificity of location, while *who* typically only refers to individuals.[4] Though he does not relate this directly to MS/MA, he notes that with *where* questions, the questioner's contextually-provided goals determine the level of granularity appropriate.

However, we might create comparable scenarios with *who* questions, which also demonstrate granularity effects. Consider the following fictional scenarios.

(7) a. Luke Skywalker is talking to Han Solo about his dismay concerning the Galactic Empire's attempts to purge the galaxy of the Jedi. A menacing character dressed in all black with a breathing mask is suddenly revealed.
b. Han Solo: Do you know **who that is**?
c. Luke: Darth Vader.

(8) a. Luke Skywalker is expressing his despair to Obi-wan Kenobi about his lost opportunity to ever have one final moment to see his father. A menacing character dressed in all black with a breathing mask is suddenly revealed.
b. Obi-Wan: Do you know **who that is**?
c. Luke: Darth Vader.

(9) Luke knows **who that is**.

Just as with (6), we might argue that the truth of (9) depends on whether it is a response to (7) or (8): (9) appears to be true and felicitous in (7) but seems true but infelicitous in (8). As in (5), the response in (8) seems out of place, like the person who utters the embedded question is missing or not clued in on something. While we admit that this case is slightly different from Ginzburg's, it serves to demonstrate that for both *where* and *who*, the context may determine the level of specificity or granularity with which an embedded question is acceptable.

In a similar vein, we can also cite *who*-questions that seem to be naturally interpreted on an MS reading, as the examples in (10) and (11) show.

(10) Who's got a light? (Groenedijk and Stokhof, 1984; van Rooij, 2003)
(11) I need a ride. Who's going to the party? (Dayal, 2016; p.c.)

Both questions are headed by *who*, and yet both permit an MS answer. If one person steps forward and truthfully offers, "Me," the speaker should be satisfied.

Likewise, Asher and Lascarides discuss another example where we naturally have a non-exhaustive *know-who*. Imagine that Jill is a gossip columnist, writing on the celebrities who attended Elton John's party. (12) can be true even if Jill doesn't know any cameramen who were at the party.

(12) Jill knows who attended the post-Oscar party at Elton John's house.

Some might argue that (12) reveals an exhaustive answer when the domain is restricted to the set of celebrities (e.g., *who [of the celebrities/relevant party-goers] attended the post-Oscar*

[4] Ginzburg does not present a comparable example with *who* to show that a similar scenario with *who* cannot give rise to the same effects.

party?) (see discussion in George, 2011, Section 6.2). However, if this is the case, it is still unclear how domain restriction alone could be the determining factor and why with rampant domain restriction in natural language (e.g., with quantifiers and definite descriptions), MS readings still seem to be blocked in same cases but not others.

Domain restriction of the set picked out by *who* also does not seem to readily explain other examples we might create, where an MS answer seems felicitous. Consider the following scenario.[5] Imagine that our friend Mark is incredibly cliquish, and typically only invites philosophers to his parties. I am trying to prove that he's biased, while you are defending him. In fact, Mark had a party just last night, so we have the following dialogue in (13).

(13) a. Me: Who came to the party last night? / Who was invited to the party last night?
b. Jill, a linguist.

Note that your response in (13b) is both felicitous and non-exhaustive. When I ask either question in (13a), I may intend a restriction to the set of philosophers (i.e., *who, of the philosophers*), because of my beliefs about Mark.[6] Indeed, this set would be the natural restriction available from the common ground (cf. von Fintel, 1994). Alternatively, one might argue that the restriction is to the set of party goers, thus if a linguist went, then a linguist is included in the restriction. On the first restriction, (13b) should be odd. On the second, (13b) should no longer be odd, but this posited restriction now misaligns with the intent behind my question, namely, to show that Mark only invited philosophers.[7] No obvious semantic mechanism is present to trigger the restriction.

What exactly is changing between these examples such that the truth conditions of embedded question reports seem starkly different, and the felicity of MS answers varies? Let us suggest that for any question, we can construct a context where discourse goals license non-exhaustivity. It is the contribution of these discourse goals that matters and give rise to interesting interactions with the linguistic features of the (embedded) question.

While levels of granularity and exhaustivity may be different sorts of beasts, they share one salient commonality: in order to establish truth conditional content, precisification of the speaker's referential intent is necessary. A hearer must recruit whatever information is available to them in order to resolve the intended level of granularity/specificity. This could include information conveyed in the linguistic form of her utterance as well as any contextual information that may elucidate the speaker's goals.

Finally, it has also been observed that the availability of MS readings can be conditioned by the use of a D-linked *wh*-phrase (Pesetsky, 1987). While we do not manipulate this factor in our studies, as we focus on *who* and *where*, we take some space here to remark on the contrast. While a singular-marked *which* phrase (e.g., *which child*) can give rise to both MA and MS readings (although it is unclear whether they have equivalent availability, see Dayal (2016) and Groenendijk and Stokhof (1984)), Comorovski (1996) observes, as does Spector (2007), that

[5] Thanks to Caley Howland for bringing this scenario to our attention.
[6] To boot, I may even plausibly intend you to give me an exhaustive answer. However, again, not only is your answer felicitous but it's non-exhaustive as well.
[7] A third candidate explanation might be possible. You may think that a restriction is to the *complement* of the contextually available set, (i.e., any non-philosophers). There may be independent reasons to prevent this kind of move. Further, the question remains nonetheless of how this kind of restriction is licensed.

plural-marked *which*-phrases (e.g., *which places*) block an MS answer (as opposed to monomorphemic *who*-phrases). Xiang and Cremers (2016) provide tentative evidence for this claim. They also found that the presence of a modal facilitated MS readings only with *who*, but not with plural *which N* phrases (e.g., *Mary remembered {which children/who} can lead the dance* vs. *Mary remembered {which children/who} have an accessory in common*).

However, aspects of the experimental design may have led to, or at least influenced, this pattern. Notably, the modal predicate *can lead the dance* was explicitly included in the lead-in and as part of the visual stimuli prior to appearing in the target statement, while the non-modal predicate *have an accessory in common* never did. As a result, in the non-modal condition, participants may have had to undergo additional inferences to calculate both Mary's perspective and what she remembers, given that this information was not explicitly stated. It is possible that the additional task demands involved in this condition incur processing costs, resulting in the observed response patterns. Moreover, the two predicates *have an accessory* and *lead the dance* were not fully crossed for presence/absence of a modal, so a tight comparison between the two conditions cannot be made. Given that these design points leave open questions open about the source of the results, we consider it empirically unresolved as to whether and to what extent these factors give rise to an MS reading.

2.1.3. Modals and non-finite clauses

Perhaps the most robust explanation for cuing non-exhaustivity is the presence of a modal (Fox, 2014; George, 2011; Nicolae, 2013; Xiang, 2016). The aspects of the modal relevant for non-exhaustivity are both its existential force and the contextually-determined conversational background, which provides a goal-oriented interpretation. The conversational background is comprised of a modal base, which picks out the set of worlds where the prejacent φ in *can* φ is satisfied, and an ordering over those worlds, determined by the deontic flavor of *can* (Kratzer, 1981; 1991). In Portner's (2009) classification, these are the existential priority modals.

While the example in (1) does not seem to have a modal in it, it can easily be paraphrased with the modal *can* (*Mary knows where she can find an Italian newspaper*); both forms allow MS. This comparison highlights the natural relationship between infinitival clauses and modality, and supports Bhatt's (1999) proposal that infinitival clauses contain *covert* modals. Bhatt enriched the Kratzerian picture to capture this goal-orientedness in infinitival clauses by contextually restricting the modal base to the worlds where not only φ is true, but where the agent's actions maximize the likely satisfaction of their goals. Bhatt further notes (fn. 12, pg. 140), that non-exhaustivity is linked to the absence of indicative tense. Following this logic, the fact that (1) contains an embedded infinitival, while (2) has neither a covert nor an overt modal, could explain the perceived difference in MS acceptability between the two.

Semantic theories that support the modal account attribute the MS/MA ambiguity to a scope interaction between an exhaustivity operator and the modal (George, 2011; Nicolae, 2013; Xiang, 2016). Essentially, the MS reading is derived when the modal takes wide scope. A natural assumption here is that non-modal questions do not give rise to these scope interactions, and thus cannot give rise to MS readings. While our experimental data do not definitively adjudicate between these particular accounts, we briefly highlight their main points, because our data bear on their viability insofar as we illuminate the contribution of the modal.

George (2011) proposes that the modal (or an existential quantifier) syntactically scopes over an exhaustivity operator X by undergoing Quantifier Raising (May, 1985). Nicolae (2013) proposes that the modal quantifies into a subset of the *wh*-domain such that, for each possible subset, it constitutes a maximally informative answer.[8] For her, these readings are licensed via presupposition failure when the context fails to provide a maximally informative answer. In contrast to these two syntactic approaches, Xiang's proposal achieves a *semantic* scope effect without QR, via a type-shifting operator LIFT, modified from the type-lifting operator employed in continuation-based grammars (Barker, 2002; Barker and Shan, 2014; Charlow, 2014; Shan 2004; Shan and Barker, 2006), which is in turn modified from Partee's (1986) LIFT. Here, the competing operator is the trace of the *wh*-word: a type-lifted *wh*-trace out-scopes the modal to derive the MA reading; while a non-type-lifted trace is out-scoped by the modal to derive the MS reading.

2.2. Discourse Cues to Interpretation

We have suggested that a common theme across the cases we have considered is that any perceived preference for (non-)exhaustivity may be the result (at least in part) of the context highlighting non-exhaustive discourse goals. Some have even argued that semantic interpretation of a question relies crucially on contextual aspects of discourse—namely, the speaker/questioner's goals that drive the search for information. For example, Ginzburg (1995) argues that question interpretation is governed by what *resolves* the question, which depends on the questioner's goals and her mental state. His notion of *resolution* is dependent on contextually-determined factors. Asher and Lascarides (1998) similarly argue that the questioner's cognitive state and plan (what she intends to do with the answer she receives to her question) are crucial. On both accounts, these factors enter into determining both the appropriateness of answers to root questions, and the interpretation of embedded questions. Finally, van Rooij (2003) formalizes this issue using the game-theoretic notion of decision problems, and Bayesian decision theory (Savage, 1954), where interpretation is governed by its utility in resolving the questioner's goal/problem. These notions allow what counts as a suitable answer to vary with context, and further highlight the importance of mental states in evaluation (cf. Boër and Lycan, 1985).

Our review of the linguistic and discourse factors in this section has revealed significant variability in the availability of MS readings, and the extent to which this availability is influenced by (and may even depend on) linguistic factors, such as the embedding matrix verb, the *wh*-word heading the question, and the presence/absence of a modal element, and on higher-level features of the discourse context, including the questioner's goal—which we argue is central. We crucially observe that for any question which seems to semantically block MS, we can nonetheless construct a context where explicit discourse goals make MS natural to a degree.

Thus, we are left asking how systematic and robust are the linguistic constraints on MS, and to what extent context exerts an influence? That is, to what extent is (non-)exhaustivity derived from or licensed by the linguistic form of the question, and to what extent can context override the influence of these linguistic cues? We conducted two sets of experiments to answer these questions. In Experiment 1, we establish the baseline contribution of the three surface-level

[8] Several problems have been identified with QR approaches. The main objection is that while existential quantifiers are found to undergo QR, modals are typically not. See e.g., Fox (2014), or Dayal (2016) for discussion.

linguistic cues. In Experiment 2, we provide evidence that discourse goals play a role and can in fact override constraints associated with the linguistic form of the question.

3. Experiment 1: Surface-level linguistic cues to interpretation

Experiment 1 focuses on linguistic cues, by manipulating the matrix verb (*know*, *predict*), the *wh*-word (*who*, *where*), and the finiteness of the embedded clause (+FIN, -FIN), in embedded question reports to determine the effect of each factor on the availability of MS.

3.1. Methods

3.1.1. Design & Materials

The experiment had a 2x2x2x4 design with three Question form factors Matrix Verb (*know*, *predict*), *Wh*-Word (*who*, *where*), and Finiteness of embedded clause (+FIN, –FIN), and Answer type: Mention-All (MA), Mention-Some (MS), Mention-All+False Report (MA+F), and False Report (FR). These four Answer types differ in how many answers a character provides in the specific trial. FRs were included as a control. Finiteness was the only factor manipulated between subjects, in order to make sure that there was no influence of this factor within a participant's experimental session. Contexts were minimally changed across this factor to satisfy the felicity conditions of finite/non-finite clauses. An example of a trial with *know*, *where*, –FIN and an MS Answer appears in (14).

(14) The places that serve cappuccinos around the neighborhood are A, B, C, and D. E and F do not. Mary usually gets her cappuccino at D. Jane is going to be in the neighborhood tomorrow. She loves cappuccinos, and texts Mary to ask where to get a cappuccino. Mary responds, "D."

Jane reports, "Mary knows where to find cappuccinos."
Is Jane right?

There were eight total sentence frames, for every possible combination of the three factors pertaining to question form. The target sentences featured eight different embedded verbs following the *wh*-word to allow for generalization across predicates within the embedded clause (e.g., *where*: *find*, *view*; *who*: *ask*, *invite*, etc.). This manipulation yielded a total of 64 unique sentence tokens. Stimuli were assigned to four lists in a pseudo-randomized Latin square fashion. In addition to the 64 unique test items, there were 10 root question filler sentences of the form *Which of the following X is not Y?*, with four possible answers listed. Filler questions served as comprehension and attention checks and addressed common world-knowledge based category membership, for example, *Which of the creatures is not a mammal?*

3.1.2. Participants and Procedure

232 undergraduates enrolled in introductory-level courses were recruited from the Rutgers University Linguistics and Cognitive Science subject pool. 14 participants were removed from final analysis for non-native speaker status. The experiment was designed and administered using Qualtrics survey software. Each participant was run in a quiet laboratory setting, seated

at an iMac. Participants were asked to read a series of brief contexts, and after each one, respond to a question corresponding to a preceding statement. Each context was comprised of 3-4 sentences, and ended with a question. A person then delivered an answer to the question, corresponding to one of the Answer types manipulated. Participants chose either 'Yes' or 'No' in response to the prompt (e.g., *Is Jane right?*). (See (14) above.)

3.2. Predictions

Accounts that assume ambiguity of possible answer predict no difference in responses based on question form manipulations; nor do they in principle predict any difference between the Answer types. However, given the robust sense from the literature that MA answers are generally available by default, we might expect to see a higher percentage of 'yes' answers, given an MA answer than an MS answer. By contrast, accounts that hold that MS answer availability is tied to the presence of a modal would predict that non-modal (+FIN) questions should *not* permit MS answers, and participants should categorically answer 'no' with these questions, but 'yes' with –FIN questions, *modulo* other factors. Accounts that attribute stronger exhaustivity requirements to *know* than e.g., to *predict* (George, 2011; Cremers and Chemla, 2016) might predict fewer 'yes' answers in response to MS answers for the former relative to the latter. Accounts such as those espoused by Asher and Lascarides (1998) and Ginzburg (1995) would predict that 'yes' responses to MS answers would surface more often with *where* questions than with *who* questions. All accounts would predict that the presence of false information —especially in an entirely false report—should generate a significant percentage of 'no' answers. It is an open question whether or not a combination of accurate and false reports (a weak exhaustive answer) would be permitted, allowing an affirmative response to the question. Experimental results from Cremers and Chemla (2016, 2017) and Phillips and George (2018) indeed revealed high rejection rates, though not as high as false controls. Klinedinst and Rothschild (2011) also conducted a small survey and found that participants accepted weak, intermediate, and strong exhaustive readings with *predict*. We predict that these mixed reports will also receive degraded acceptability. Accounts requiring *know* to be strongly exhaustive would predict uniform rejection of all FR-type answers.

3.3. Results

Experiment 1 results are presented in Figure 1. Each graph corresponds to a factor tested.

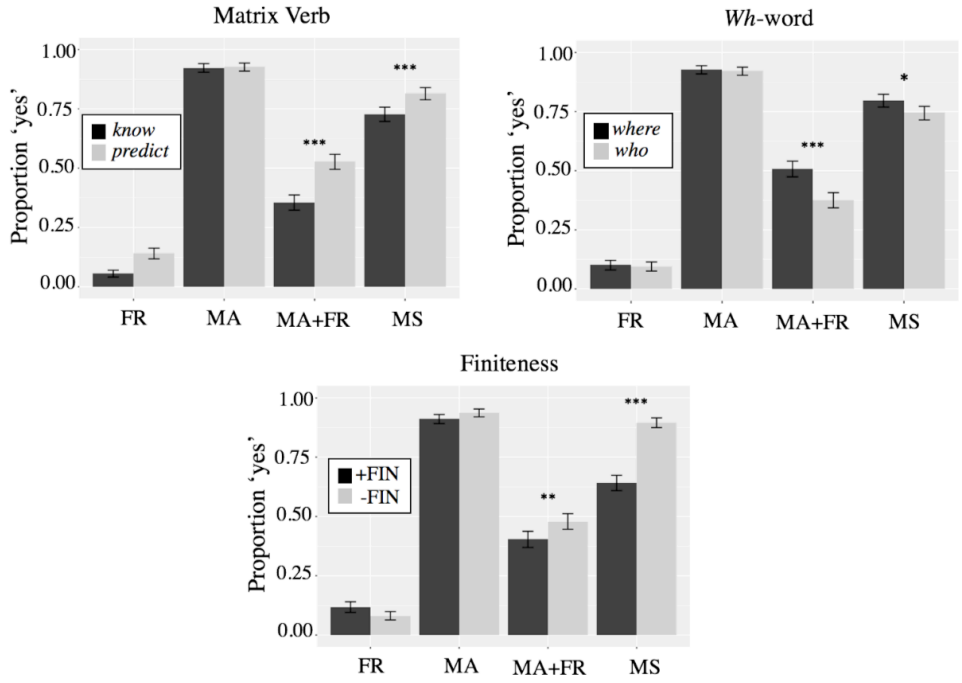

Figure 1: Experiment 1 results given three question form factors and 4 answer types.

All analyses conducted were non-parametric Kruskal-Wallis tests. We found overall main effects of Verb ($X^2$1) = 53.714, p<0.0001), *Wh*-word X^2(1) = 9.71, p<0.005), ±FIN ($X^2$1) = 43.567, p<0.0001), and Answer (X^2(3) = 823.42, p<0.0001). Breaking down each factor per Answer, all effects are significant for both MS and MA+FR: Verb (MS: X^2 (1) = 18.892, p<0.0001; MA+FR: X^2(1) = 51.731, p<0.0001); *Wh*-word (MS: X^2(1) = 6.61, p<0.05; MA+FR: X^2(1) = 30.219, p<0.0001); and ±FIN (MS: X^2(1) = 156.7, p<0.0001; MA+FR: X^2(1) = 9.513, p<0.005). We then zoomed in on the critical MS Answer condition, as shown in Figure 2. Here too, all factors were significant: Verb (X^2(1) = 18.892, p<0.0001), *Wh*-Word (X^2(1) = 6.61, p<0.05), and ±FIN (X^2(1) = 156.7, p<0.0001).

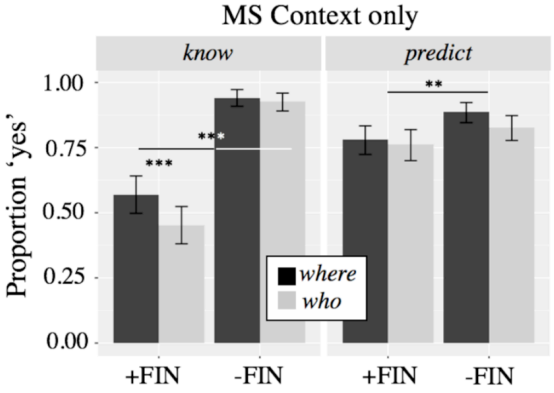

Figure 2: Experiment 1 results for MS Answer Type given three question form factors.

3.4. Discussion

The results of Experiment 1 demonstrated that the form of the speaker's question matters for the types of answer provided by the addressee. There were significant effects of all factors. Whereas MA answers were consistently supported across factor types and conditions, and FRs led answers to be degraded, the locus of the action was in MS answers. In particular, the presence of an infinitival (–FIN) embedded clause boosted MS acceptability in comparison to +FIN clauses. At the same time, however, the acceptance rate in +FIN clauses was far from 0%—indeed, it was over 50%. Thus, while the difference in acceptability lends partial support to a modal account, insofar as infinitival clauses are modal, it is only partial, since MS answers are not categorically ruled out with finite embedded questions.

While there was an effect of Finiteness for both *know* and *predict,* the effect was more pronounced for *know-wh*: for MS answers, there was a depression in acceptance (but not complete rejection) with +FIN *know-wh*. These results could be taken as providing initial evidence in support of accounts based on ambiguity and/or underspecification, which both predict MS to be (nearly) as widely available as MA, given appropriate contextual licensing.

Both ambiguity and underspecification accounts admit the role of the context. It is possible that the non-categorical judgments we observed in Experiment 1 indicate that the underlying representations are underspecified for exhaustivity, and pragmatics fills that value.[9] To tease apart these theoretical approaches and better pin down the division of labor between semantics and pragmatics, we require a better understanding of the extent to which context licenses certain answers regardless of question form. We now turn to Experiment 2.

In Experiment 2, we investigate the role of the questioner's goals by manipulating contextual information. We refer to "high-stakes" and "low-stakes" contexts, because what's at stake influences the goal(s) at hand, and consequently, what kind of answer can resolve the question. We define a High Stakes context as one where human health or lives are at risk, while we define a Low Stakes context as one without any such life-threatening issue (for instance, choosing a good diner or hair salon). By design, in our High Stakes contexts, the goal is to ascertain as much information as possible, therefore an MA answer is most informative. In contrast, our Low Stakes contexts present goals where multiple answers are possible, thus an MS answer not only suffices, but may be preferred. In addition, we manipulated the level of informativity of MS answers in order to tease apart the types of answers that could resolve the question, since we assumed that not just any MS answer would suffice in any context.

We note two things about our notion of Stakes. First, it is not isomorphic to exhaustivity. It is possible in principle to have a High Stakes context where the Questioner's goals are non-exhaustive, and a Low Stakes context where the goals are exhaustive. For example, this difference might arise under constraint of time pressure, or where a High Stakes goal may only be satisfied by a single person, etc. Nonetheless, one might assume that in most cases, when the stakes are High, one values above all an answer that is not only true, but thorough. Second, we recognize that what counts as High or Low Stakes is also context-dependent. However, this approach at least gives us a first look at the contribution of contextually-relevant goals and their influence on answer reports.

[9] See discussion in Sprouse (2007), Sprouse et al. (2012) for arguments about gradient vs. categorical grammaticality judgments in syntax and what they might indicate about the underlying representations.

In addition to the manipulation of context, we also manipulated – as we did in Experiment 1— the form of the answer. However, here, we introduced a further manipulation in MS type. Xiang (2016) has claimed that an answer that is not a singleton MS answer (a Mention-One answer, MO) but is not exhaustive is "unacceptable" and must be "ignorance-marked." We therefore controlled for whether non-exhaustive answers were MO or MS (mention-a-few). While we recognize that there could be contexts in which a non-MO answer could signal ignorance on the part of the speaker, here the contrast between MO/MS and MA could diffuse the MO/MS contrast.

4. Experiment 2: Discourse-level cues to interpretation

4.1. Methods

4.1.1. Participants and Procedure

318 native speakers of English participated. The study was constructed and administered using Alex Drummond's Ibex Farm platform. Participants were recruited online through Amazon Mechanical Turk. IP addresses were restricted to US only, and further questions were included to ascertain native speaker status. 6 participants were removed for browser incompatibility issues, and 6 participants were removed for non-native English speaker status.

4.1.2. Materials and Procedure

As in Experiment 1, in Experiment 2 we manipulated Finiteness in target sentences as a between-subjects factor. The –FIN target condition contained three within-subjects factors: Answer Type (Mention-All (MA), Mention-Some (MS), Mention-One (MO), and False Report (FR)), Informativity (MAX, MIN, for MO/MS Answers), and Stakes (High, Low). The +FIN target condition had two within-subjects factors (Answer Type and Informativity). We did not manipulate Stakes in the +FIN condition, and only targeted Low Stakes for the following reason: we predicted that in a High Stakes condition, an MA answer would be favored. Given that +FIN embedded questions already favor MA answers, we would predict to see little to no change. The question is whether a Low Stakes context can pull answers away from MA towards MO/MS. For both conditions, we included both Mention-Some Answer Types, where two answers were given, and Mention-One Answer Types, where a singleton answer was given.

Each context featured a main topic, a main character conducting a search form some contextually-relevant information, and a set of ranked entities relevant to the topic. The main character in search of the information posed a *wh*-question to a group of individuals, who then each provided an answer connected to the ranking. The participant's task was to evaluate the knowledge of these individuals, based on their answers and the given context.

The Answer Type and Informativity manipulations yielded 6 possible answers in the set of answers, which were randomized so that the same answers did not always appear together, and so that participants would see different answers for each story. Thus, there were six answer type permutations. At any given time, only three Answer Types were randomly displayed by an algorithm, in order to reduce the cognitive load on the experimental participants, and to ensure that it was not the case that the same types were always pitted against each other (thereby

forcing certain comparisons and reducing the probability of a response bias from surfacing on every trial). An example of a High Stakes and a Low Stakes trial type follow.

(15) *High Stakes Trial Type*
Scientists have discovered a new strain of a dangerous virus that has contaminated oysters in the Mid-Atlantic. The Center for Disease Control is trying to prevent as much contamination as possible by tracking down all the oysters. In this area, luckily only 6 restaurants usually buy oysters from the contaminated area: Restaurant A ordered 10 crates, Restaurant B ordered 8, Restaurant C ordered 5, Restaurant D ordered 2, Restaurant E ordered 1, Restaurant F ordered 0.

The supervisor for this county asks his inspectors, "Where should we check for contaminated oysters?"

Inspector A says, "Restaurant A, B, C, D and E."	MA
Inspector B says, "Restaurant A."	MO-MAX
Inspector C says, "Restaurants D and E."	MS-MIN

Who knows where to look for oysters? (Choose all that apply.)

(16) *Low Stakes trial type*
Johanna is new to Minneapolis and wants to try local coffee shops. The Ultimate Coffee Guide 2018 ranks cafes on a ten-point scale, where ten is the highest number of points. Minneapolis has the following ranking for coffee roasteries: Café A has 10 stars, Café B has 8, Café C has 5, Café D has 2, Café E has 1, Café F has 0.

Johanna asks three of her classmates originally from the city, "Where should I go for coffee?"

Classmate A says, "Cafés A and B."	MS-MAX
Classmate B says, "Cafés E."	MO-MIN
Classmate C says, "Café F."	FR

Who knows where to go for coffee? (Choose all that apply.)

At the end of each trial, participants were instructed to answer the question about the individuals' knowledge by *choosing all that apply*. There was also a "None of the above" option. This multiple-choice question allowed participants to choose more than one answer, allowing us to determine if multiple answer types were permitted in a given scenario.

4.2. Predictions

Given the theoretical approaches reviewed earlier, we might make the following predictions regarding the influence of context and informativity on acceptability of answer type. We first spell out predictions regarding contextual Stakes. An Ambiguity theory might predict that either reading will be licensed, and that context simply pulls out one reading over the other. Thus, these accounts might expect no difference between different question forms (–FIN and +FIN conditions), but rather have all differences housed between High and Low Stakes. Underspecification theories also predict that a reading will be as available as made appropriate by context (contextual goals), thus differences should be seen between High and Low Stakes,

perhaps with modulation by question form. The difference will be apparent in the –FIN condition, where Stakes was manipulated. However, given the long-standing assumption that MA, but not MS, is available to every question, extant theories might predict that MA is trivially available. If so, MA Answer Types should receive high acceptance across-the-board, and the influence of Stakes will only be seen with MS Answer Types. Modal theories would predict that non-exhaustive (MS/MO) answers would receive high acceptance only in the –FIN condition, perhaps regardless of the context.

Regarding the manipulation of answer Informativity, we expect MAX-informative MS/MO answers to be more acceptable than MIN-Informative ones, because by design maximally-informative answers are better question resolvers than minimally-informative ones. This prediction is in line with underspecification theories, which explicitly argue that question interpretation is driven by issues surrounding the resolution of the questioner's goals (see Ginzburg, 1995a; Asher and Lascarides, 1998; van Rooij, 2003). While this result could also be consistent with other theories, they offer no explicit account of how such effects arise.

4.3. Results

Experiment 2 results are in Figure 3. MO-MAX answers are more acceptable than MA answers in Low-Stakes scenarios (-FIN: $X^2(1) = 21.56$, $p<0.0001$; +FIN: $X^2(1) = 17.56$, $p<0.0001$). In the -FIN condition, there was a further Stakes x Answer Type (MA/MO-MAX) Interaction ($X^2(3) = 60.35$, $p<0.0001$) reflecting the fact that MO-MAX is more acceptable in Low- vs. High-Stakes, and MA responses are more acceptable than MO in High-Stakes. The effect of Stakes disappears in MS Answer Types (MIN v. MAX). Finally, Informativity significantly affected acceptability of an MS/MO answer: MIN answers were chosen significantly less than MAX (+FIN: $X^2(1) = 55.98$, $p<0.0001$; -FIN: $X^2(1) = 93.93$, $p<0.0001$).

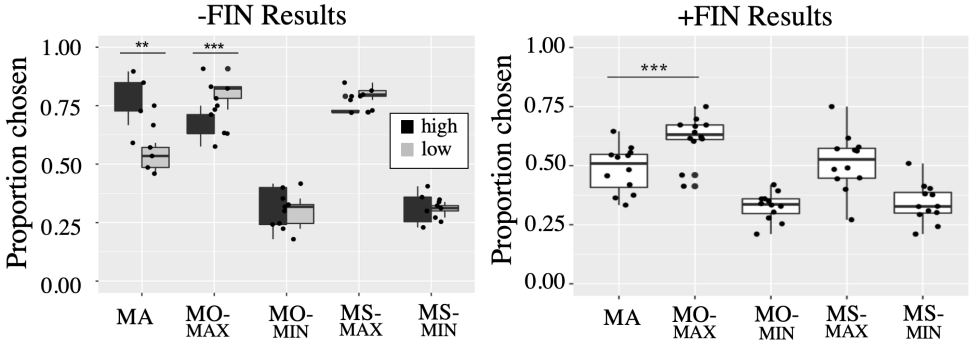

Figure 3: Experiment 2 Results.

4.4. Discussion

The results reveal the following. In the –FIN condition (left graph), MA answers were preferred over MS and MO answers in High Stakes contexts, with the exception of MS-MAX answers. Participants accepted both MA and MS-MAX answers at comparable levels in High Stakes contexts. By contrast, in the Low Stakes –FIN condition, MS-MAX and MO-MAX answers were preferred over MA answers. In the +FIN condition, MA answers did not consistently win out: here, too, participants were willing to accept a maximally informative MS or MO answer.

Thus, accounts that incorporate a role of pragmatics and contextually-relevant goals are best able to explain these patterns of results, while accounts that rigidly appeal to particular semantic or structural features as licensing answerhood type fall short. Still, the proportion of accepting non-exhaustive and maximally informative MO and MS answers was higher in the –FIN condition than in the +FIN condition, suggesting that a weaker modal account has explanatory power. Such a theory could say, consistent with our data, that the presence of a modal boosts the acceptability of non-exhaustive readings, while the absence of one does not block them.

That participants were sensitive to the Informativity of MS and MO answers (with maximally informative answers being more acceptable than minimally informative ones) is consistent with underspecification and ambiguity theories. In both types of accounts, context plays the crucial role of disambiguating, in the case of ambiguity, or of precisifying the relevant value(s) of the underspecified logical form. Whatever mechanism(s) are involved in this process, more information about the behavioral signatures of precisification and disambiguation would be incredibly illuminating.

These findings therefore demonstrate that the discourse context provides central information that is relevant to resolving a given question, including which answers are most informative and how much information is needed to resolve exhaustivity and the questioner's goal. This effect of context was observed *regardless* of finiteness in the embedded clause, and therefore did not depend on the presence of a modal element to license either an MO or MS answer. Finally, our results show no signs that a partially exhaustive non-singleton answer that is maximally informative is significantly degraded. This result held even when participants could perform an explicit comparison between MS and MO answers. If such MS answers can signal ignorance, our empirical results show that they do not consistently and obligatorily do so.

5. Conclusion

Across two experiments, we have presented empirical data that reveal the following. First, MS answers are not as constrained as has been previously assumed. They are acceptable when associated with either infinitival embedded questions or with finite embedded questions. The presence of a modal element facilitates licensing of a non-exhaustive answer, but is not a necessary precondition for it. Second, the type of embedding verb plays a role in answer acceptability, highlighting the role of lexical semantics. Third, pragmatics in the form of contextual information and a questioner's goals in the discourse context play a key role in the resolution of exhaustivity in a question. Finally, the type of mention-some answer matters: those that are more informative are valued more than those that are not. Thus, in this work, we have identified a set of surface-level and contextual cues that the speaker can manipulate and that the listener can recruit to arrive at an intended interpretation.

While these findings may make sense intuitively, our experiments are the first, we believe, to systematically probe the surface-level linguistic and discourse-level contextual cues that influence question-answer dynamics, and that they therefore are valuable to any theory of question semantics and pragmatics that seeks to identify what licenses or constrains answers to a given question. While both ambiguity and underspecification theories allow for wide availability of MS, ambiguity theories give no explicit account of contextual modulation, nor how certain questions encode stronger preferences for particular interpretations. At the same time, current modal-based theories and pragmatic theories that assume exhaustivity by default with optional weakening undergenerate MS. Our results highlight specific factors that license

and constrain non-exhaustive answers, and thus prompt current theories to take seriously the form and context of the question-answer exchange, and the central role of the speaker and hearer.

References

Asher, N. and A. Lascarides. (1998). Questions in dialogue. *Linguistics and Philosophy*, *21*(3), 237-309.

Barker, C. (2002). Continuations and the nature of quantification. *Natural Language Semantics, 10*(3), 211–242.

Barker, C. and C-C. Shan. (2014). *Continuations and natural language*. Oxford: Oxford University Press.

Beck, S. and H. Rullmann. (1999). A flexible approach to exhaustivity in questions. *Natural Language Semantics*, *7*(3), 249–298.

Bhatt, R. (1999). *Covert modality in non-finite contexts*. PhD thesis, University of Pennsylvania.

Boër, S. E. and W. G Lycan. (1985). Knowing who. *Philosophical Studies*, *28*(5), 299–344.

Charlow, S. (2014). *On the semantics of exceptional scope*. PhD thesis, New York University.

Comorovski, I. (1996). *The Interpretation of Interrogative Phrases*. Dordrecht: Springer Netherlands.

Cremers, A. and E. Chemla. (2016). A psycholinguistic study of the exhaustive readings of embedded questions. *Journal of Semantics*, *33*(1), 49–85.

Cremers, A. and E. Chemla. (2017). Experiments on the acceptability and possible readings of questions embedded under emotive–factives. *Natural Language Semantics 25*(3), 223–261.

Dayal, V. (2016). *Questions*. In C. Barker and C. Kennedy (Eds.), *Oxford Surveys in Semantics and*
Pragmatics. Oxford: Oxford University Press.

von Fintel, K. 1994. *Restrictions on quantifier domains*. PhD thesis, University of Massachusetts
Amherst.

Fox, D. (2014). The semantics of questions. Class notes, MIT.

George, B.R. (2011). *Question embedding and the semantics of answers.* PhD thesis, UCLA.

Ginzburg, J. (1995a). Resolving questions, I. *Linguistics and Philosophy*, *18*(6), 459–527.

Ginzburg, J. (1995b). Resolving questions, II. *Linguistics and Philosophy*, *18*(6), 567–609.

Grimshaw, J. (1979). Complement selection and the lexicon. *Linguistic Inquiry*, *10*(2), 279–326.

Groenendijk, J. and M. Stokhof. (1982). Semantic analysis of wh-complements. *Linguistics and Philosophy*, 5(2), 117–233.

Groenendijk, J. and M. Stokhof. (1984). *Studies in the semantics of questions and the pragmatics of answers*. PhD thesis, University of Amsterdam.

Guerzoni, E. and Y. Sharvit. (2007). A question of strength: on NPIs in interrogative clauses. *Linguistics and Philosophy*, *30*(3), 361–391.

Hamblin, C.L. (1973). Questions in Montague English. *Foundations of Language*, *10*, 41–53.

Heim, I. (1994). Interrogative semantics and Karttunen's semantics for *know*. In R. Buchalla and A. Mittwoch (Eds.), *Proceedings of The Israel Association for Theoretical Linguistics*, *1*, pp 128–144. Hebrew University of Jerusalem.

Hintikka, J. (1974). Questions about questions. In M. K. Munitz and P. Unger (Eds.), *Semantics and Philosophy*. New York, NY: NY University Press.

Hintikka, J. (1976). *The Semantics of questions and the questions of semantics*. Ithaca, NY: Cornell University Press.
Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*(1), 3–44.
Klinedinst, N., and D. Rothschild. (2011). Exhaustivity in questions with non-factives. *Semantics & Pragmatics*, *4*(2), 1–23.
Kratzer, A. (1981). The Notional Category of Modality. In H. J. Eikmeyer and H. Rieser (Eds.), *Words, Worlds, and Contexts*, 38–74. Berlin: de Gruyter.
Kratzer, A. (1991). Modality. In A. von Stechow and D. Wunderlich (Eds.), *Semantics: An International Handbook of Contemporary Research*, 639–50. Berlin: de Gruyter.
Lahiri, U. (2002). *Questions and Answers in Embedded Contexts.* Oxford: Oxford University Press.
May, R. (1985). *Logical form: Its structure and derivation*, Linguistic Inquiry Monograph 12. Cambridge, MA: MIT press.
Nicolae, A. (2013). *Any questions? Polarity as a window into the structure of questions*. PhD thesis, Harvard University.
Partee, B. H. (1986). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, and M. Stokhof (Eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, 357–381. Dordrecht: Foris Publications.
Pesetsky, D. (1987). Wh-in-situ: Movement and unselective binding. In Eric J. Reuland and Alice ter Meulen (Eds.), The representation of (In)definiteness (pp. 98–129). Cambridge, MA: MIT Press.
Phillips, J. and B.R. George. (2018). Knowledge *wh* and False Beliefs: Experimental Investigations. *Journal of Semantics*, *35*(3), 467–494.
Savage, L.J. (1954). *The Foundations of Statistics*. New York: John Wiley and Sons.
Shan, C. (2004). Binding alongside Hamblin alternatives calls for variable-free semantics. *Proceedings of Semantics and Linguistic Theory*, *14*, 289–304.
Shan, C. and C. Barker. (2006). Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy*, *29*(1), 91–134.
Spector, B. and P. Egré. (2015). A Uniform semantic for embedded interrogatives: An answer, not necessarily the answer. *Synthese*, *192*(6), 1729–1784.
Sprouse, J. (2007). *A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge*. PhD Thesis, University of Maryland.
Sprouse, J., M. Wagers, and C. Phillips. (2012). A test of the relation between working memory capacity and syntactic island effects. *Language* 88(1), 82–132.
Theiler, N. (2014). *A multitude of answers: Embedded questions in typed inquisitive semantics.* MSc thesis, ILLC, University of Amsterdam.
Uegaki, W. (2015). *Interpreting questions under attitudes*, PhD thesis, MIT.
van Rooij, R. (2003). Questioning to resolve decision problems. *Linguistics and Philosophy*, *26*(6), 727–763.
Xiang, Y. (2016). *Interpreting questions with non-exhaustive answers*. PhD thesis, Harvard University.
Xiang, Y., and A. Cremers. (2017). Mention-some readings of plural-marked questions: Experimental evidence. *Proceedings of the North East Linguistics Society 47*, pp 261–274. Amherst, Mass: GLSA.