Contextual, Lexical and Structural Factors Licensing Mention-Some Embedded Questions

Morgan Moyer and Kristen Syrett
Rutgers, The State University of New Jersey - New Brunswick

1 Introduction

How to properly capture the truth conditions of sentences with embedded *wh*-questions, as in (1), has been a source of debate among semanticists since the mid-20th century.

(1) Mary knows where to find coffee

The crux of the issue is this: the truth conditions of such sentences depend on the interpretation of the embedded question they contain, which in turn, depend on the answers the subject provides to the corresponding root question *Where can one/I find coffee?* This question is claimed to license two sorts of answers, as in (2), which in turn correspond to two different readings of the question:

(2) a. **Mention-All** (MA)
Mary knows or names all the (contextually relevant/salient) places where one can find coffee.

b. **Mention-Some** (MS)
Mary knows or names (at least) one place where one can find coffee.

The acceptability of one interpretation over the other can be thought of as contextually determined by the goals of the speaker/questioner (Groenendijk & Stokhof 1982, 1984; Ginzburg 1995a,b; van Rooij 2004). For example, a caffeine-seeking tourist in town for a day who poses the root question would most likely have goals which are most compatible with Mary's providing an MS answer, where she names one or possibly more nearby and desirable places to find coffee. In contrast, a food blogger who is in town to review the local coffee scene would have goals which are more compatible with Mary providing an MA answer, where she names all of the local coffee spots. When Mary's (true) answer appropriately satisfies the Questioner's goals, then the embedded question report, as in (1), is judged true; when her answer does not, it is judged false or infelicitious.

Most semantic theories have focused on teasing apart different strengths of MA readings or answers–whether they are strongly (knowing all true answers and all that are false) or weakly (knowing all true, and not necessarily the false ones) exhaustive, and whether these are licensed in a given context or by a given predicate (Karttunen 1977; Groenendijk & Stokhof 1982, 1984; Heim 1994; Dayal 1996; Beck & Rullmann 1999; Lahiri 2002; Guerzoni & Sharvit 2007; Klinedinst & Rothschild 2011; George 2011; Fox 2014; Nicolae 2014; Uegaki 2014; Spector & Egré; Dayal 2016). Regardless of different theoretical stances, most researchers agree that at least one MA answer is available for every root or embedded *wh*-question.

Up for debate, however, is whether a MS answer is also available. There is no question that it is available for examples such as (1). However, a shared observation is that not all *wh*-questions allow an MS reading (George 2011, Fox 2013, Dayal 2016, Xiang 2016).

Consider, for example, (3). Putatively, no amount of sensitivity to a Questioner's goals makes the two readings equally available here as they are for (1). There seems to be consensus that in this case, Mary should be able to list all of the people who came to the party (and maybe also be able to identify those that didn't) in order for (3) to be considered true, and also for Mary's answer to the root question (*Who came to the party?*) to be permissible. However, it is not immediately apparent *why* this would be the case–why these two sorts of embedded ostensibly differ in their allowance of an MS answer or reading.

(3) Mary knows who came to the party.

The goal of this paper is to provide experimental support for a set of key factors licensing MS readings in any given embedded question. We show that a concert of linguistic and contextual factors conspire to give rise to MS readings, even in cases where previous theoretical work has claimed is not possible, namely when the embedded question is finite and there is no modal element involved. At the same time, we show that even in the best case scenario, a root question with a modal or embedded infinitival question, where both MA and MS readings should be viable options, the discourse context and the Questioner's (speaker's) goals in this context, may strongly favor one reading or the other. We argue, in line with van Rooij (2004) and Asher and Lascarides (1998) that both of these elements play a crucial role in determining the suitability of any given answer, and therefore to question-answer dynamics as a whole.

2 Background

To tease apart the differences between (1) and (3), one might start by comparing their surface forms. We make two observations. The first is that (1) and (3) differ in terms of the *wh*-word that heads the embedded clause (*where* or *who*).

While it is not typically assumed that *wh*-word should *a priori* differ semantically in terms of exhaustivity, there is a complicated body of observations suggesting that questions with different *wh*-words do yield differences in judgments related to exhaustivity. For one, the classic data supporting MS are typically *where*-questions, like (1), while the classic data supporting MA are typically *who*-questions like (3) (Groenendijk & Stokhof 1984, George 2011, Fox 2014). Asher & Lascarides (1998) desribe a notion of "defaultness": *who*-clauses seem to have an exhasutivity requirement by default. Given that Groenendijk & Stokhof's partition semantics focus *who*-questions, it should be no surprise that their semantics is MA.[1] Thus, when one considers non-*who*-questions, a purely MA semantics seems

[1] Asher & Lascarides also claim that *how*-questions as in (i), signal MS by default (following Hintikka (1975)), although we do not discuss *how*-questions in this paper:

(i) a. How do I get to the buried treasure?

less plausible and MS preferable (Asher & Lascarides 1998, Ginzburg 1995a,b, Hintikka 1983). At the same time though, there are classic exceptions where MS has been shown to be acceptable with a *who*-question, such as (4) from van Rooij (2004). Such a counterexample is consistent with the story told by Asher & Lascarides, who emphasize that apparent preferences may be overruled by context.

(4) Who has a light?

Further cementing the observation that *who* and *where* differ, Ginzburg (1995a,b) noted that these two *wh-* words diverge with respect to the granularity of their referential domains and what kinds of information allow a question to be resolved: *where* is vague and more flexibly allows for greater or finer levels of granularity, whereas *who* seems to typically only refer to individuals or a fine-grained level. Though Ginzburg does not directly relate this observation to MS/MA answers, he notes the contextually-provided goals of the Questioner determine the level of granularity and whether the answer resolves the question. Thus, previous observations motivating the role of contextual information, variability of answerhood, and the importance of addressing a Questioner's goals support a hypothesis that the *wh*-word plays a role in which answers are permitted, although there has as yet been no systematic investigation of this lexical cue.

The second difference is in the finiteness of the embedded clause: note that (1) contains a nonfinite (infinitival) embedded clause, while (3) contains a finite clause. As we mentioned in passing, the embedded question in (1) gives rise to a root question with an overt possibility modal *can*, whereas the corresponding question for (3) does not. It is possible that these additional surface-level differences in finiteness are linked with underlying semantic differences, explaining the acceptability patterns for MS answers.

Recently, some researchers in question semantics have argued that MS acceptability falls out naturally from the semantics of modal elements in the the *wh*-questions (George 2011, Dayal 2016, Xiang 2016). On the Kratzerian view of modality (Kratzer, 1981; 1991), a deontic modal like *can* ϕ contributes, in addition to an existential quantificational force, two elements: a modal base which picks out the set of worlds where the prejacent ϕ is satisfied, and an ordering, determined by the deontic flavor of *can*, over those worlds. What's more, the modal need not be overt. Bhatt (1999) has proposed that infinitival clauses contain covert modals, and contribute goal-oriented-ness. He enriches the Kratzerian picture to capture this goal-oriented-ness by contextually restricting the modal base to the worlds not only where ϕ is true, but where the agent's actions maximize likely satisfaction of her goals. Additionally, Bhatt further notes (footnote 12, p 140), that non-exhaustivity is linked to the absence of indicative tense. Following this logic, the fact that (1) contains an infinitival embedded clause (and is linked to a root question with a modal), and highlights a goal, while (3) has neither an infinitive nor an overt modal, could explain the perceived differences in MS acceptability.

While this explanation about the contribution of the modal appears at first glance to neatly explain the difference between the permissible answers for (1) and (3), it

b. You go to the hidden island.

leads to the generalization that MS readings should depend on a covert or overt modal, and should *never* be possible with an embedded finite clause (assuming that this clause does not capture a modal in the representation). However, we have seen that there are questions that lack a modal element and have no infinitival clause, which nonetheless naturally allow for MS answers/readings. For example, (4) does not require an exhaustive list of individuals, and yet this is a question without a modal element and with indicative tense. Thus, the claim about modal necessity appears to undergenerate instances where MS would be licensed. Furthermore, providing sound reasons for when an MS reading can be licensed does not tell us about what constrains it in such instances–why an MS reading appears might degraded or dispreferred when a modal is present. Thus, the first order of business is to test the hypothesis that a modal is necessary (either overtly or covertly), and the second is to deconstruct the linguistic and contextual aspects of the question-answer situation that modulate MS acceptability.

Finally, the two example sentences we have been discussing carry the same matrix verb (*know*), but this is, of course, not the only verb that takes a clausal complement with a question form. Other verbs include *predict*, *wonder*, *tell*, and so on (so-called "responsive" verbs in Lahiri (2002)'s typology). Matrix verbs typically encode selectional restrictions on the semantic and syntactic properties of their complements (Grimshaw, 1979). For example, the verb *know* can syntactically embed a question (e.g., *know* ***who*** *came to the party*) or a proposition (e.g., *know* ***that*** *Sam and Will came to the party*), while the verb *wonder* can only embed a question (*wonder* ***whether/if*** *X...* but **wonder* ***that*** *X...*). The verb *tell* reflects a distributional pattern similar to that of *know* in the syntactic complements it takes, but differs from *know* in factivity of the proposition it embeds (cf. Karttunen 1971, Kiparsky & Kiparsky 1968). However, *tell* still appears to impose a veridicality requirement on its question complement (Karttunen 1977). For example, Alice can tell Ben that Cynthia stayed in a particular city during her last trip to France, and it need not be true, but if Alice tells Ben *where Cynthia stayed*, then Alice has shared the actual location with Ben.

Returning to the types of readings licensed in embedded questions, in addition to these structural features categorizing these verbs, many researchers have attempted to identify selectional restrictions related to exhaustivity (Lahiri 2002, Guerzoni & Sharvit 2007, George 2011, Klinedinst & Rothschild 2011, Spector & Egré 2015, a.o.). George (2011) and Phillips & George (2018) have argued that *know* favors a strongly-exhaustive MA reading by default. However, this is certainly not the only reading available, as the acceptability of (1) clearly shows. Guerzoni & Sharvit (2007), Klinedinst & Rothschild (2011), and Spector & Egré (2015) have argued, and Cremers & Chemla (2016) have shown experimentally, that with (*know*) and an embedded finite clause question, a variation of an exhaustive reading is available–one where the subject would not be required to know the false answers along with the true one, and may in fact have false beliefs ("weak exhaustivity"). A verb like (*predict*) is claimed *not* to favor a strongly-exhaustive MA reading (Heim 1994, Beck & Rullmann 1999, Klinedinst & Rothschild 2011, Sharvit 2002), and instead permits weak and (possibly) intermediate exhaustive readings (where one might not know about the false answers or have any false beliefs). An open question is whether non-exhaustive readings are also permitted in these cases, and if so,

whether these two verbs, which have been the focus of recent theoretical discussions, differ in the extent to which they permit non-exhaustive (i.e., MS) readings with a finite embedded clause. It is possible that (*predict*) is more permissive than *know* in this respect, given its less stringent requirements on exhaustivity in the first place.

Finally, there is an additional surface-level element that might play a role, which we do not explore here, for reasons of space: D-linking and plural marking of the *wh*-phrase (Pesetsky 1987). Comorovski (1996) argues, and Xiang & Cremers (2016) provide tentative experimental evidence for the claim, that plural-marked *which*-phrases (e.g., *which places*) may render an MS answer less available (than monomorphemic *who* questions) (e.g., *John remembered {which children/who} can lead the dance* vs. *John remembered {which children/who} have an accessory in common*), even when modalized.[2] They also found that the presence of a modal facilitated MS readings with *who*, but not with pluralized *which* phrases. However, we point out a confound in their design in that the target embedded questions with a finite clause without a modal aligned differently with the lead-in stimuli than the target modalized embedded question did. Specifically, the phrase *can lead the dance* was explicitly included in the lead-in and as part of the visual stimuli, requiring no memory load on the part of the participants and minimal recall of what the subject of the target sentence remembered (*Mary remembers wh- can lead the dance*), while the finite-clause embedded question *have an accessory in common* required an additional inference about Mary's memory, since her perspective on this was not stated explicitly, and evaluating the statement required additional work. Moreover, the predicates *have an accessory* and *lead the dance* were also not fully crossed for presence/absence of a modal. Given these design elements that leave too many questions open about the source of the results, we consider it empirically unresolved as to whether or not the factors they investigated exert an influence on MS reading.

Thus, given the host of observations and theoretical claims reviewed above, two questions arise: is an MS answer actually a permissible answer in the absence of an overt or covert modal, and (if so), what linguistic and contextual factors allow us to make a systematic prediction about when an MS answer will (or will not) be licensed? Answering these questions is the goal of the present research. Here, we zoom in on the three low-hanging fruits–the three surface-level characteristics outlined above: the type of matrix verb (and its corresponding selectional restrictions), the type of *wh*-word, and the finiteness of the embedded clause. We then turn to the role of the context, evaluating what's at stake relative to the Questioner's goals in a given context. Ultimately, the results we obtain here and in future research will allow us to pin down the source of the differences in answerhood patterns between embedded questions such as (1) and (3) and others like them.

In this paper, we present two experiments. The first experiment isolates the three surface-level linguistic aspects of the question (matrix verb, *wh*-word, finiteness), to tease apart each one's contribution to the interpretation of embedded questions. In this experiment, we seek to determine whether MS answers are available beyond instances where the question encodes a modal, and how surface-level cues recruited

[2]In the case of a singular-marked *wh*-pharse such as *which child*, the MS and MA answers are equivalent.

by the speaker may cue the listener as to the suitability of a particular answer. The second experiment addresses the role of the Questioner's goals in promoting with an MA or an MS answer in cases where both are semantically licensed. Here, we create contexts accompanied by a root question, which differ with respect to what's at stake relative to the speaker's goals, while holding non-finiteness and surface-level cues constant in the target embedded question.

Our experimental results offer three main findings about both root questions and embedded questions. First, they demonstrate that MS answers are (unsurprisingly) more likely with infinitival clauses than with finite embedded clauses, but not uniformly ruled out with the latter. Second, they indicate that the finiteness difference may be tied to particular verbs (and in particular, verb semantics): *know* displays a greater difference in finiteness than *predict*, such that embedded infinitival clauses are more likely to permit MS readings than embedded finite clauses. At the same time, the presence of a finite clause does not bar the MS reading with either verb. Third, they show that the difference between sentences like (1) and (3) does not appear to be due to the (*wh*-word. Finally, they reveal an interaction between the type of answer licensed (whether an MA answer or a specific version of an MS answer where the highest ranked singleton is mentioned) and what goals are at stake in the context at hand. We take these results to indicate that the suitability of a given answer is not determined purely semantically in the absence of contextual information, but rather is contingent upon establishing the discourse-relevant goals. Moreover, the Speaker (i.e., the Questioner) may explicitly signal to the Hearer that they are seeking to elicit an exhaustive MA or an MS answer by manipulating the surface-level form of the question (e.g,. a modal, infinitive, or a finite clause) to facilitate the conversational exchange and contribute to the Question/Answer dynamics.

3 Experiment 1

In this experiment, we manipulated three surface-level cues to explore their effect on the types of answers licensed in questions, and consequently, their role leading to the truth conditions of sentences with related embedded questions. These cues were the matrix verb (*know*, *predict*), the *wh*-word heading the embedded question (*who*, *where*), and the finiteness of the embedded question (+FIN or -FIN).

3.1 Design

3.1.1 Participants

232 undergraduates participated. These participants were recruited from the Rutgers University Linguistics and Cognitive Science subject pool, and were students in introductory-level courses. 14 participants were removed from final analysis for non-native speaker status.

3.1.2 Stimuli

The experiment had a 2x2x2x4 design: Matrix Verb (*know*, *predict*), *wh*-Word (*who*, *where*), Finiteness of embedded clause (+FIN, -FIN), and Answers: Mention-All (MA), Mention-Some (MS), Mention-All+False Report (MA+F), and False Re-

port(FR)). Three of the factors were within subjects, while one factor (finiteness) was between subjects. We manipulated this factor between subjects in order to be sure there was no influence of this factor within a participant's experience. We chose to feature four types of answers within a given context to see which answers were or were not licensed, given the surface-level factors we manipulated. In this way, we could determine whether exhaustive and non-exhaustive readings of various degrees ? include those wtih a false report (FR) (so-called "weak exhaustive" readings) were available in each case.

3.1.3 Procedure

The experiment was designed and administered using Qualtrics survey software (Qualtrics, Provo, UT). Each participant was run in a laboratory setting at an individual console, 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. A sample context is given in (5). Each context was comprised of 3-4 sentences, ending with a description of a question that was posed to someone based on the material at hand. The person then delivered an answer, corresponding to one of the Answer types manipulated, as shown in (a-d). This was then followed by a report provided by another character. Participants chose either 'Yes' or 'No' in response to the prompt (e.g., *Is Jane right?*).

(5) 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 combination of the three factors apart from Answer. The target sentences featured eight different embedded verbs for each *wh*-word to allow for generalization across predicates within the embedded clause (e.g., *who*: *find*, *view*; *where*: *ask*, *invite*). 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 filler sentences, which were root questions. These fillers were *wh*-questions of the form *Which of the following X is not Y?*, with four possible answers listed. Filler questions addressed common world-knowledge based category membership, for example *Which of the creatures is not a mammal?* In this way, the fillers were a check for comprehension and attention.

3.1.4 Predictions

Recall that our main motivation for conducting this research is to investigate the conditions under which MS readings are licensed. We therefore couch our predictions in terms of when MS readings are expected to be available or not. However, we also make further predictions about baseline items based on previous research. For example, we expect that MA answers should be uniformly acceptable, and blatant

FR answers uniformly unacceptable. Moreover, we expect that MA+FR answers will be degraded. If *know* resists MS readings and *predict* permits them more readily, we might expect for the target MS items with *know* to display more rejections than those with *predict*. It is also quite possible that *predict* is more permissive of MA+FR answers than *know*, because the latter is strongly exhaustive by default, and prefers an MA answer. However, this pattern might not hold across the board; it might be linked to the other factors we manipulated: the *wh*-word and finiteness.

Although the semantics of *wh*-words do not yield ready-made predictions for why they should be different in terms of MS answers, based on the observations we summarized earlier, we might expect *where*-headed questions to give rise to more affirmations relative to *who*-headed questions in reaction to an MS answer. Finally, and most importantly, if the (covert) modal in the non-finite embedded clause drives MS acceptability, we should see significant differences between +FIN and -FIN clauses, where affirmations of MS answers are more likely with -FIN targets than with +FIN ones. We might also expect to see flat-out rejections connected to MS answers in response to +FIN targets.

3.1.5 Results

The results of Experiment 1 are presented in Figure 1, where each graph corresponds to a factor tested. All analyses conducted were non-parametric Kruskal-Wallis tests. We found overall main effects of Verb ($\chi^2(1) = 53.714$, p<0.0001), *Wh*-word ($\chi^2(1) = 9.71$, $p < 0.005$), Finiteness ($\chi^2(1) = 43.567$, p<0.0001), and Answer ($\chi^2(3) = 823.42$, $p < 0.0001$). Breaking down each factor per Answer, all effects are significant for both MS and MA+FR: Verb (MS: $\chi^2(1)=18.892$, $p < .0001$, and MA+FR: $\chi^2(1)=51.731$, $p < .0001$); *Wh*-word (MS: $\chi^2(1)=6.61$, $p < .05$; MA+FR: $\chi^2(1)=30.219$, $p < .0001$); and finally Tense (MS: $\chi^2(1)=156.7$, $p < .0001$; and MA+FR: $\chi^2(1)=9.513$, $p < .005$). We then focused in on the critical MS answers to see the full breakdown of all factors, as shown in Figure 2. Here as well, all factors related to MS availability were significant: Verb ($\chi^2(1)=18.892$, $p < .0001$), *Wh*-Word ($\chi^2(1)=6.61$, $p < .05$), and Finiteness ($\chi^2(1)=156.7$, $p < 0.0001$).

3.1.6 Discussion

Our predictions were largely borne out, with some interesting twists. There was, as predicted, a significant effect of Finiteness (with +FIN embedded clauses resisting MS readings), which played out more with the matrix verb *know* than with *predict*. However, +FIN clauses did not render MS answers completely unacceptable; rather, they appear to have been dispreferred, relative to MA answers. Targets with a +FIN embedded clause elicited an affirmative response in the context of an MS answer approximately 65% of the time. Interestingly, this effect was driven by questions embedded under *know*, as shown in Figure 2. Thus, MS answers do not depend upon the presence of a (c)overt modal element. We further observed a significant effect for *Wh*-Word, especially in MA+FR answers but also in MS answers, where *where*-questions allowed such answers more so than *who* questions did.

If +FIN embedded clauses *do* permit MS answers, why is the % of "yes" answers in response to +FIN target sentences not as high as the % observed with -FIN

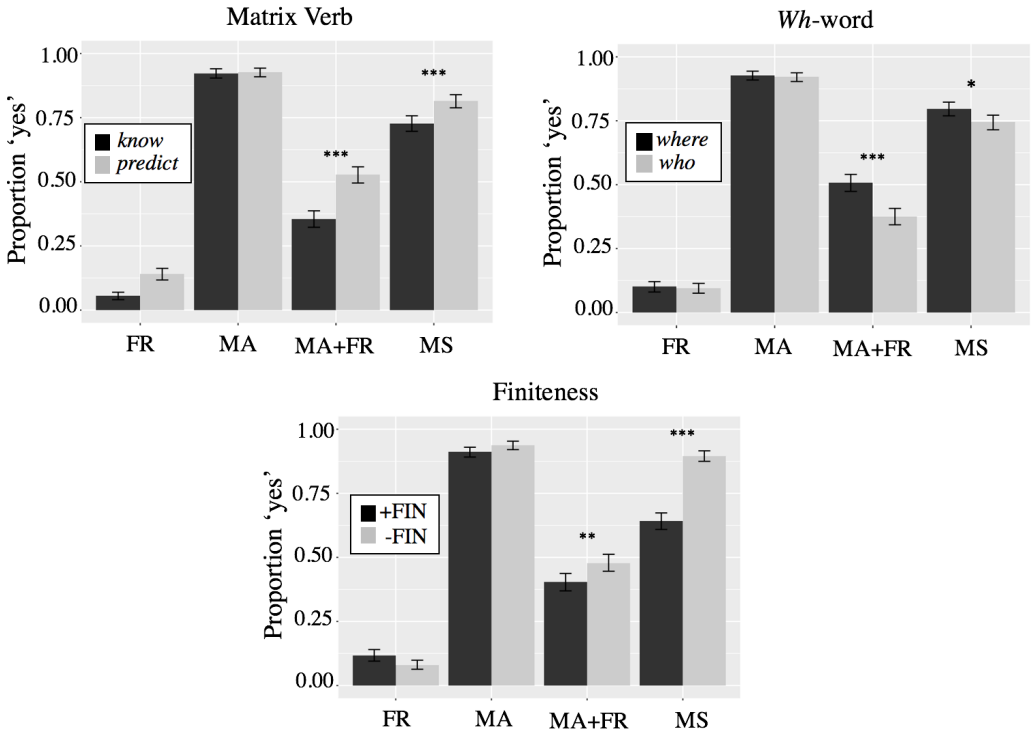

Figure 1: Experiment 1 results for each factor manipulated.

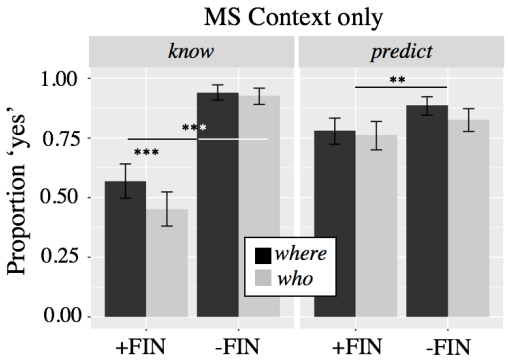

Figure 2: Experiment 1 results for MS Answers only.

target sentences? We would like to argue that in the absence of explicit surface-level cues indicating what the speaker's goals are, and that an MS answer addressing such goals is licensed, the listener is left to make the inference that an MA answer was preferred with a +FIN clause. More specifically, we argue that an MS answer should be possible in principle, but that it is contingent not only upon characteristics of the discourse context in which a root question is asked or an embedded question is presented, but also upon the speaker cuing the listener in to what their goals are, and what kind of an answer is suitable and would resolve the question at hand. This kind of breadcrumb dropping might be done by using an overt or covert modal (i.e., an infinitival clause), or saying something like, "for example" or using a superla-

tive or other modifier (e.g., *Mary knows where to find [a quick, the best, a nearby] cappuccino.*). In the absence of such explicit cues, an MS reading is unlikely.

Having shown that an MS reading can surface under certain circumstances, including where it was not expected to surface, given previous theoretical claims, and that certain linguistic and contextual factors modulate this availability, we might now ask if an MS reading could be suppressed in cases where it is expected to be acceptable. We turn to this question now, in order to demonstrate the importance of the discourse context and the Speaker's/Questioner's goals.

Several theorists argue that certain aspects of context are essential for determining the semantic contribution of a question; that exhaustivity is parasitic on the goals provided by a context. Ginzburg (1995) argued for different notions of resolvedness, set by the Questioner's goals and mental state. The appropriateness of an answer is thus determined by how well it resolves the question. Asher & Lascarides (1998) similarly argue for the importance of the Questioner's cognitive state and their plans to execute an action, which underlie the question asked. Finally, Van Rooij (2004) formalizes this notion in a game-theoretic framework of decision problems, appealing to Bayesian decision theory (Savage 1954), where interpretation is governed by the utility in resolving the Questioner's goal/problem. Here again, what counts as a suitable answer varies with context and the speaker's mental state (cf. Boër & Lycan 1975). Importantly for these accounts, these factors determine the question's truth conditions by virtue of determining the appropriateness of answers, and enter into the compositional semantics of an embedded question. Thus, one cannot speak of an exhaustive or non-exhaustive meaning disconnected from a goal behind the question.

In Experiment 2, we investigate the role of Questioner's goals by creating contexts that highlight different contextually-determined goals. We refer to these as "high-" and "low-stakes" contexts, because what's at stake influences the goal(s) at hand, and consequently, what kind of answer is suitable. A "high stakes" context is one where e.g., people's health or lives are at risk, while a "low stakes" context is one where what matters is something much more trivial (e.g., choosing a good diner or hair salon). We assume that in a "high stakes" context, the goal might be to ascertain as much complete information as possible, and therefore that an MA answer is preferable, while in a "low stakes" context, an MS answer not only suffices, but may be preferred–and further, that not just any MS answer suffices. We explore this hypothesis in Experiment 2.[3] We also manipulate one further variable based on a claim by Xiang (2016) that an answer that is not a singleton MS answer but is not exhaustive is "unacceptable" and "ignorance-marked."

[3]We do recognize that "stakes" may not track exhaustivity isomorphically. For example, it may be possible 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 pressure for time. However, we present these two factors as correlated in this experiment, assuming that in most cases, when the stakes are high, one values above all an answer that is not only true, but thorough. We further note that what counts as "high" or "low" stakes is also context-dependent.

4 Experiment 2

4.1 Participants

114 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. 3 participants were removed for browser incompatibility issues, and 2 participants were removed for non-native speaker status.

4.1.1 Stimuli and Procedure

Experiment 2 had a 4x2x2 design. There were three within-subject factors: Answer Type (Mention-All (MA), Mention-Some (MS), Mention-One (MO) or False Reports (FR)), Informativity (MAX, MIN for MO and MS answers), and Stakes (High, Low). Target sentences (root questions) were always in -FIN form in order to maximize the availability of MS answers. We included two kinds of MS answer in this experiment: Mention-Some (MS) where two answers were given, and Mention-One (MO) where a singleton answer was given. This manipulation was included in response to Xiang (2016)'s proposal that "intermediate" answers where some number of answers between one and the exhaustive list are given, are unacceptable as answers to root questions and as interpretations in embedded questions (Xiang 2016, p.45-46).

As in Experiment 1, participants were presented with a brief 3- to 5-sentence context, and were asked to render an assessment based on this information and a key question. Each context featured a main topic, a main character conducting a search for some sort of contextually-relevant information, and a set of ranked entities relevant to the topic. The main character in search of the information asks a *wh*-question of a group of individuals, who then each provide an answer connected to the ranking. The participant's task was to evaluate the knowledge of these individuals, based on their answers and the context at hand. These questions all used non-finite embedded *where*-clauses, which have been previously shown to be maximally compatible with MS answers. Thus, we purposely favor MS conditions in this experiment.

The answer types were MA, MS, MO, or FR. There was also an additional answer type manipulation. MO and MS answers were further distinguished based on their level of informativity, whereby answers at the top of the list were Maximally Informative (MAX), and answers at the lower end of the ranking were Minimally Informative (MIN) (i.e., MS-MAX, MS-MIN, MO-MAX, MO-MIN). Thus, there were 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 below.

(6) *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.)

(7) *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, Cafe F has 0.

Johanna asks three of her classmates originally from the city, "Where should I go to eat out?"
Classmate A says, “Cafés A and B.” **MS-MAX**
Classmate B says, “Cafés E and F.” **MS-MIN**
Classmate C says, “Café F.” **FA**
Who knows where to eat out in LA? (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.

4.1.2 Predictions

We generated the following predictions, based on the previous theoretical accounts reviewed above. First, we predicted that MA answers would generate a high rate of knowledge ascription, since MA answers are unequivocally available for the questions presented in this experiment. We also predicted that MS and MO answers would also be acceptable and lead to relatively high rates of knowledge ascription, perhaps even just as high as with MA. However, we further predicted that the type of context and the informativity of the MS or MO answer would influence participant responses. Specifically, we predicted that with "high stakes" contexts, MA answers would be preferred to any type of MS or MO answer, while in "low stakes"

contexts, MS and MO answers might be preferred to MA answers. We further predicted that MS or MO MAX-ranked answers would be consistently favored over MIN-ranked answers. FR answers should display floor-level performance.

4.1.3 Results

The results for Experiment 2 are presented in Figure 3. Our predictions were largely borne out: MA answer types were selected more often in "high stakes" contexts ($\chi^2(1) = 6.988$, $p < 0.01$), while MO answers were selected more often in "low stakes contexts ($\chi^2(1) = 19.996$, $p < 0.0001$, for MO-MAX). Level of informativity (MAX v. MIN) had a significant overall effect for the two answer types for which this was manipulated: MO-MAX answers were significantly more likely to be chosen than MO-MIN answers ($\chi^2(1)=92.986$, $p < .0001$). We found no significant difference between either MO-MAX and MS-MAX ($\chi^2(1)=1.593$, $p=.21$) or MO-MIN and MS-MIN ($\chi^2(1)=2.4$, $p=.12$).There was also an interaction between MA and MO-MAX and stakes ($\chi^2(3)=56.726$,$p > .0001$). The two contexts did not differ for any other answer types.

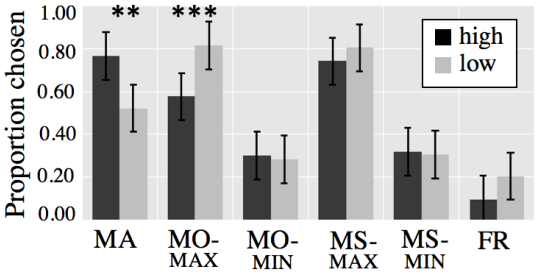

Figure 3: Experiment 2 results.

4.1.4 Discussion

In this experiment, we tested the hypothesis that even when MA and MS(/MO) answers are made available via the form of the question (specifically, because it encodes modal semantics), the context and the contextually-relevant Questioner goals modulate the acceptability of MA and MS answers. As predicted, MA answers generated higher acceptance in contexts where the stakes were high, while MS answers generated higher acceptance when the stakes were low. Moreover, not just any MS/MO answer did the trick: participants weighted maximally-informative answers over those that were minimally informative (i.e., had reduced information entropy). Finally, we found that both MS and MO answers were judged as acceptable answers to the root question with the modal, and yielded affirmations of speaker knowledge, with non-singleton MS answers being more acceptable in "high stakes" contexts, in line with MA answers. We therefore present evidence in favor of a pragmatic account of MS/MO licensing.

The dual roles of answer type (MA v. MS/MO) and informativity in question answers is perhaps not so surprising, pragmatically speaking. Both Gricean and Relevance Theoretic frameworks would favor a conversational participant who provides maximally informative, relevant information in response to an interlocu-

tor's query. However these frameworks do not make clear predictions about how such information feeds in to evaluations of an embedded question. This experiment demonstrates that answers to a salient question in a given context associated with certain goals tied to the Questioner lead to different assessments of a speaker's knowledge, and therefore that the discourse context and Questioner goals play a key role in question-answer dynamics.

5 Discussion and Conclusion

We began this paper by sharing a long-standing observation about the answers licensed by two different sorts of embedded *wh*-questions: those with a modal (overt or covert) and those without. The former are claimed to licensed Mention-Some (MS) answers, while the latter are said not to. However, we raised the possibility that certain surface-level linguistic cues (i.e., the type of *wh*-word heading question, finiteness, and the matrix verb), and the context at hand make a contribution to MS availability. To date, no research has systematically investigated whether and how answer type is influenced by these factors.

We presented two experiments investigating the factors that license exhaustivity in embedded questions. we confirmed that questions with modal elements contribute a strong baseline of non-exhaustivity, but that a modal element is not required for a MS answer, and certain surface level cues contribute to promoting or demoting the availability of MS answers. We further showed that the answer type licensed in response to questions with modals depends on contextually-relevant goals, such that MA answers may be favored in certain contexts, and MS in others, depending on what is at stake. We therefore show that context plays a crucial role in question-answer dynamics.

We take our experimental findings as presenting evidence that exhaustivity (i.e., MA vs. MS/MO answer acceptability) falls out naturally from the semantics of the question combined with information tied to the context. A speaker/Questioner has certain goals related to this context, and can make these goals more or less explicit for the listener by recruiting surface-level linguistic cues to signal different levels of answer exhaustivity. The listener is called upon to make use of these cues, paired with the context and the semantic aspects of the question, to provide an answer in respond to the speaker's question. The question may or may not fully resolve the question. Thus, the picture of exhaustivity in question-answer dynamics that we paint is one necessarily situated in a discourse context.

References

Asher, N., & A. Lascarides. 1998. Questions in Dialogue. *Linguistics and Philosophy* 21.237–309.

Beck, S., & H. Rullmann. 1999. A flexible approach to exhaustivity in questions. *Natural Language Semantics* 7.249–298.

Boër, S. E., & W. G. Lycan. 1975. Knowing who. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 28.299–344.

Comorovski, I. 1996. *The Interpretation of Interrogative Phrases*. Dordrecht, Netherlands: Springer.
Cremers, A., & E. Chemla. 2016. A psycholinguistic study of the exhaustive readings of embedded questions. *Journal of Semantics* 33.49–85.
Cremers, A., & E. Chemla. 2017. Experiments on the acceptability and possible readings of questions embedded under emotive-factives. *Semantics Archive* .
Dayal, V. 1996. *Locality in WH Quantification: Questions and Relative Clauses in HIndi*. Dordrecht: Kluwer.
Fox, D., 2013. Mention-some readings. Class notes.
George, B. R. 2011. *Question embedding and the semantics of answers*. University of California, Los Angeles dissertation.
Ginzburg, J. 1995a. Resolving questions, I. *Linguistics and Philosophy* 18.459–527.
Ginzburg, J. 1995b. Resolving questions, II. *Linguistics and Philosophy* 18.567–609.
Grimshaw, J. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10.279–326.
Groenendijk, J., & M. Stokhof. 1982. Semantic analysis of wh-complements. *Linguistics and Philosophy* 5.175–233.
Groenendijk, J., & M. Stokhof. 1984. *Studies on the semantics of questions and the pragmatics of answers*. University of Amsterdam dissertation.
Guerzoni, E., & Y. Sharvit. 2007. A question of strength: on NPIs in interrogative clauses. *Linguistics and Philosophy* 30.361–391.
Heim, I. 1994. Interrogative semantics and Karttunen's semantics for *know*. In *Proceedings of IATL*, volume 1, 128–144.
HIntikka, J. 1975. Answers to questions. In *Questions*, ed. by H. Hiz. Reidel: Dordrecht.
HIntikka, J. 1983. New foundations for a theory of questions and answers. In *Questions and Answers*, ed. by F. Kiefer. Reidel: Dordrecht.
Karttunen, L. 1971. Implicative verbs. *Language* 47.340–358.
Karttunen, L. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1.3–44.
Kiparsky, C., & P. Kiparsky. 1971. Fact. In *Semantics: an Interdisciplinary Reader*, ed. by Steinberg & Jacobwits. Cambridge University Press: Cambridge.
Klinedinst, N., & D. Rothschild. 2011. Exhaustivity in questions with non-factives. *Semantics and Pragmatics* 4.1–23.
Kratzer, A. 1981. The notional category of modality. In *Words, Worlds, and Contexts: New Approaches in Word Semantics*, volume Research in Text Theory, 6, 38–74. de Gruyter: Berlin.
Kratzer, A. 1991. Modality. In *Semantics, An International Handbook of Contemporary Research*, ed. by A. von Stechow & D. Winderlich. de Gruyter: Berlin.
Lahiri, U. 2002. *Questions and answers in embedded contexts*. Oxford University Press on Demand.
Nicolae, A. C. 2013. *Any questions? Polarity as a window into the structure of questions*. Harvard University dissertation.
Pesetsky, D. 1987. Wh-in-situ: Movement and unselective binding. *The representation of (in) definiteness* 98.98–129.
Phillips, J., & B. George. (forthcoming). Knowledge wh and false beliefs, experimental investigations. *Journal of Semantics* .
Savage, L. 1954. *The Foundations of Statistics*. New York: Wiley.
Sharvit, Y. 2002. Embedded questions and 'de dicto' readings. *Natural Language Semantics* 10.97–123.
Spector, B., & P. Egre. 2015. A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese* 192.1729–1784.
Uegaki, W. 2014. Predicting the variation in exhaustivity of embedded interrogatives. *Sinn und Bedeutung* 19.
van Rooij, R. 2004. The utility of mention-some questions. *Research on Language and Computation* 2.401–416.
Xiang, Y. 2016. *Interpreting questions with non-exhaustive answers*. Harvard University dissertation.
Xiang, Y., & A. Cremers. 2017. Mention-some readings of plural-marked questions: Experimental evidence. In *Proceedings of North East Linguistics Society 47*.