# Who thinks *wh*-questions are exhaustive?

**Morgan Moyer (mcmoyer@stanford.edu)**

**Judith Degen (jdegen@stanford.edu)**
Department of Linguistics, Stanford University
Stanford, CA 94350 USA

## Abstract

Asking and answering questions is a staple of human communication. In order to answer a question effectively, a hearer must interpret the speaker's intention given the specific question asked. *Wh*-questions like *Where can I get coffee?* are underspecified for (non-)exhaustivity, i.e., how many answers the must be provided to resolve the speaker's goal. Intuitions from the semantic literature report that questions are generally exhaustive, and non-exhaustive only in the context of specific linguistic factors (e.g., the modal *can*, certain *wh*-words). To test these assumptions, we collected question paraphrase ratings for naturally occuring root questions in variable linguistic contexts. In contrast to previous claims, we find that questions are not biased for exhaustivity. However, other prior observations are supported by the data. We argue that a full account of the observed distribution of meanings must integrate discourse factors like the hearer's estimate of the speaker's goal, alongside (or subsuming the effect of) linguistic cues.

**Keywords:** psycholinguistics; experimental pragmatics; corpus linguistics; *wh*-questions; exhaustivity

## Introduction

Human communication proceeds remarkably fast and effortlessly despite the rampant underspecification of speakers' utterances with respect to the meaning they intend to convey. Resolving that underspecification requires that hearers integrate a wide range of possibly uncertain linguistic and extralinguistic cues.

This view of pragmatics, informed by psycholinguistic research on language processing that espouses a dynamic, non-modular view of comprehension, has provided a useful novel perspective on many phenomena at the semantics/pragmatics interface. For instance, one of the most-studied cases of underspecification in experimental pragmatics is the scalar inference from *some* to *not all* (e.g., *Scully ate some of the cookies* typically licenses the inference that she did not eat all of them). Recent research using a large dataset of naturally-occurring utterances has revealed a large amount of variability in whether hearers derive scalar inferences (Degen, 2015). Rather than being random, the observed variability was dependent on multiple features of the linguistic and discourse context. Data from controlled experimental tasks with artificially generated stimuli confirm these results: scalar inferences are systematically dependent on (the hearer's estimate of) the speaker's discourse goal (Zondervan, 2010), the speaker's epistemic state (Goodman & Stuhlmüller, 2013; Breheny, Ferguson, & Katsos, 2013), and which alternatives are contextually available (Huang & Snedeker, 2011), among other cues. These results were unexpected in light of the theoretical literature, which had predicted a higher prevalence of the inference and no systematic context-dependence (Levinson, 2000).

A good deal of experimental attention has been paid to pragmatic inferences, like scalar inferences, that are the result of reasoning about declarative utterances. In contrast, there is much less discussion about the cues guiding hearers' interpretation of non-declarative utterances like questions. Consider polar (*yes-no*) questions: these can be answered literally with a *yes* or *no*, but often the literal answer is neither the most appropriate, nor what the speaker intended. In Searle (1975)'s classic example, a dinner guest who asks *Can you reach the salt?* likely intends you to pass the salt, not say *yes*. Whether a hearer understands the speaker to want a literal or non-literal answer depends on what they infer about the speaker's goals. Clark (1979) surveyed liquor merchants to determine how they answered a polar question like *Does a fifth of Jim Beam cost more than $5?* when it was introduced by a brief sentence that made the speaker's goal explicit. If the speaker first stated *I want to buy some bourbon*, merchants were more likely to answer with the exact price of the whiskey. If the speaker instead stated *I've got $5 to spend*, merchants were more likely to provide the more literal *yes/no* answer. That is, merchants responded by addressing the inferred speaker goal. Research in computational cognitive science has followed suit by modeling question asking and answering as a species of rational, goal-directed behavior (Hawkins, Stuhlmüller, Degen, & Goodman, 2015; Rothe, Lake, & Gureckis, 2018).

*Wh*-questions are even more complex than polar questions in that even their literal interpretation is underspecified. To see this, consider that a *wh*-question can be answered in multiple ways:

(1) a. Where can I find coffee?
    b. Who came to the party?

At first blush, the most natural way to answer (1a) is to mention a nearby coffee shop, while the most natural way to answer (1b) is to provide an exhaustive list of party-goers. Thus, (1a) is interpreted non-exhaustively while (1b) is interpreted exhaustively. We refer to these interpretations as 'Mention-Some' (MS) and 'Mention-All' (MA), respectively, following Hintikka (1976). *Wh*-questions do not specify on the surface whether they are intended to be MS or MA—how hearers resolve this underspecification is the focus of this paper.

While linguistic theory has engaged with this question

for decades, the data to support the theoretical claims typically consists in received intuitions about a small number of researcher-generated examples. Our goal is to combine the forces of experimental psycholinguistics and formal semantics/pragmatics: we conduct a large-scale corpus study of naturally occurring *wh*-questions and experimentally test the interpretations these questions give rise to.

## Linguistic background on *wh*-questions

What determines whether a *wh*-question is interpreted exhaustively? Semantic theories have typically assumed that *wh*-questions are interpreted on MA readings (exhaustively) by default (Karttunen, 1977; Groenendijk & Stokhof, 1984; George, 2011; Fox, 2014; Nicolae, 2014). In contrast, non-exhaustive MS readings are taken to arise only in specific linguistic contexts (George, 2011; Fox, 2014; Nicolae, 2014; Dayal, 2016; Xiang, 2016). However, recent work has discussed both linguistic and (less so) discourse factors that modulate *wh*-question interpretation.

One linguistic form factor which has been suggested to modulate *wh*-question exhaustivity is the presence of a modal auxiliary like *can*, which appears to have a special licensing effect for MS readings (George, 2011; Fox, 2014; Nicolae, 2014; Dayal, 2016; Xiang, 2016): questions with such a modal felicitously allow MS, while questions without do not. This may be the reason that (1a), which includes the modal auxiliary *can* is naturally MS, while (1b), which does not, is not. Similarly, other modal questions like *How can I get to the buried treasure?* seem to be naturally MS.

A second form factor discussed as modulating question interpretation is the question's *wh*-word (e.g., (1a) is headed by *where*, while (1b) is headed by *who*). Ginzburg (1995) and Asher and Lascarides (1998) observe that different *wh*-words seem to exhibit different biases for MS and MA readings: *who*-questions seem particularly biased for MA; others, especially *how*- and *why*- questions seem biased for MS. Consider *How do I get to the buried treasure?*: while there might be many ways to get there, the speaker only needs one to achieve their goal of finding the treasure. Further, Asher and Lascarides (1998) point out that semantic theories that argue for MA meanings typically use data from *who*-questions, while theories that argue for an ambiguous or underspecified meaning typically use data from *how*- and *why*-questions. Thus, a second reason for the natural difference in interpretation between (1a) and (1b) may be the questions' varying *wh*-words.

To see the role of contextual speaker goals in *wh*-question interpretation, consider the following. If the speaker who asks (1a) is a tourist looking to have their morning coffee, it is likely that they want an MS answer. However, if the speaker instead is a food journalist writing about the local café culture, then it is likely that they want an MA answer. The best answer depends on what resolves the speaker's goal (Groenendijk & Stokhof, 1984; Ginzburg, 1995; Asher & Lascarides, 1998; van Rooij, 2003, 2004).

While the question of how discourse factors interact with linguistic form factors in guiding *wh*-question interpretation is very much under-explored, recent experimental research tested the above judgements in lab-controlled experiments with artificial stimuli. One study crossed the presence or absence of modality with *wh*-word, and another investigated the interaction of modality and exhaustive vs. non-exhaustive discourse goals (Moyer & Syrett, 2019; Moyer, 2020). In these studies, while the form factors had the effects theorized in the literature (albeit non-categorically), explicitly manipulating contextual goals could override the interpretational biases. In particular, participants judged non-modal questions as true on MS readings in contexts where the goals were non-exhaustive. For example, MS readings were rated higher than MA readings for a question like *Who knows where they went for coffee?* in a context where tourists were wandering around an unfamilar town. This invites a non-exhaustive answer; knowing one place the tourists went for coffee is sufficient to answer the question.

To date, there has been no systematic investigation of *naturally occurring questions* that tests the intuitions reported in the literature. We ask two questions: (Q1) How much does question interpretation vary in natural discourse contexts? Is there indeed an overall bias for MA interpretations? (Q2) Is the distribution of interpretations modulated by linguistic form, in particular by overt modality and *wh*-word? To address these questions, we conducted a two-part study combing a corpus analysis and a web-based experimental investigation, following Degen (2015). We first extracted naturally occurring root *wh*-questions from a corpus of spoken language. We then collected gradient (non-)exhaustivity ratings for each of those questions to determine the natural distribution of question readings.

## Step 1: database creation

We used TGrep2 (Rohde, 2005) and the TGrep2 Database Tools (Degen & Jaeger, 2011) to extract 10,192 occurrences of utterances containing a *wh*-phrase from the Switchboard corpus (Godfrey, Hilliman, & McDaniel, 1992). Each utterance was annotated automatically for features of interest, including presence/absence of modality, *wh*-word, and syntactic structure (e.g., embedded, root, adjunct).

Since our goal was to investigate the interpretation of ambiguous questions, we excluded those that are unambiguous with respect to exhaustivity. Moreover, while the literature on the MS/MA ambiguity in questions typically focuses on root and embedded questions, for simplicity we begin our experimental investigation with only root questions. Of the 10,192 *wh*-clauses, we thus retained only the 1,719 root questions (16.9%). We further removed degree questions (e.g., *How old are they?*), questions with complex *wh*-phrases (e.g., *Which group do you work in?*), and identity questions (e.g., *Who is their quarterback?*), because they have only one interpretation (in which MS and MA converge).

Additionally, we focused on just the *wh*-questions headed by *who*, *what*, *where*, *when*, *how*, and *why*. Table 1 presents

Table 1: Distribution of *wh*-words and modality in Switchboard root questions. Percentage of total (995).

.

| *Wh*-word | Modal present | Modal absent |
|---|---|---|
| *What* | 6.7% | 52.1% |
| *How* | 2.8% | 15.7% |
| *Where* | 0.4% | 9.3% |
| *Why* | 1.3% | 4.7% |
| *Who* | 0.8% | 4.7% |
| *When* | 0.1% | 1.4% |

the joint distribution of *wh*-words and the presence of modal auxiliary verbs in the database. *What*-questions comprise 58.8% of this constrained set, followed by *how*-questions at 18.5%. The final set included 995 unique questions.

## Step 2: experimental investigation of *wh*-question interpretation

Question interpretation for each case in the database was assessed via a web-based paraphrase rating task.

### Method

**Participants.** On Prolific, we recruited 660 self-reported native English speakers who were paid $2.5. Data from 4 participants were not recorded due to browser errors, and 35 participants were excluded for failing 2 of 6 control trials.

**Procedure and materials.** Participants were presented with each question and the 10 preceding lines of dialogue and rated the likely intended meanings (paraphrases) by adjusting a continuous slider for each paraphrase. Fig. 1 presents an example test trial. Question paraphrases were selected to reflect MS/MA readings: *a* indicates a (non-exhaustive) MS reading, *every* indicates an (exhaustive) MA reading, while in *the* paraphrases, the MS and MA readings converge due to the uniqueness presupposition introduced by the definite determiner—the single answer to *What is the place in which you have skied before?* is both MS (because it is a single answer) and MA (because it answers the question exhaustively). There was a fourth option (*something else*) in case none of the paraphrases was appropriate. Control items were polar questions containing either an indefinite, definite, or universally quantified noun phrase identical to the three paraphrases (e.g., *Did you grab all the cookies?*).

Participants were trained on four example dialogues: on two, the *a*/*the* paraphrases were best, and on the other two the *every* paraphrase was best. We included one modal and one non-modal question for each. Further, on each training trial, participants were instructed to interpret the ellipsis in each paraphrase relative to the content in the red question (to yield, e.g., "What is every place you have skied?" for the first paraphrase in Fig. 1).[1]

[1]Procedure, materials, analyses and exclusions were pre-registered at `https://bit.ly/3tp1FC1`.

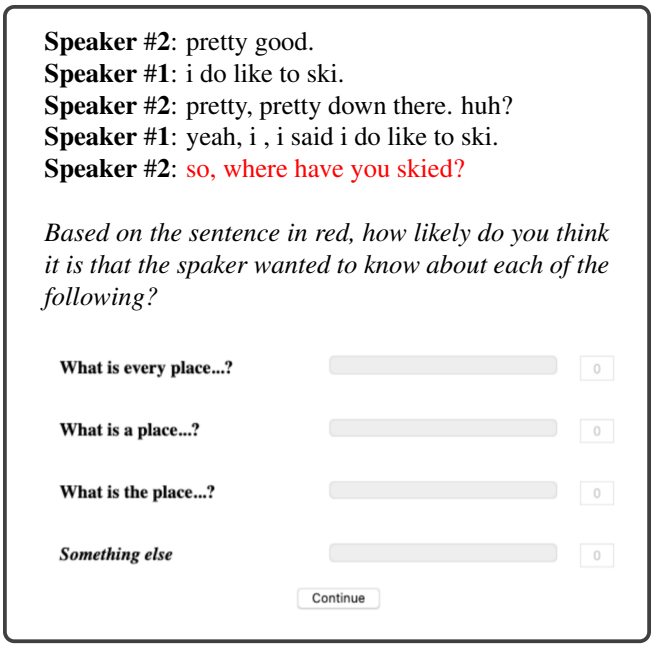


Figure 1: Example trial. Slider values had to sum to 100, which we rescaled to interpret ratings as a probability distribution reflecting subjective beliefs about intended meaning.

**Predictions** The previous semantic literature predicts an overall MA bias (Q1) (Karttunen, 1977; Groenendijk & Stokhof, 1984; George, 2011; Fox, 2014; Nicolae, 2014; Xiang, 2016). It further makes two main predictions concerning the effect of overt modality and *wh*-word on question interpretation (Q2): first, that the MS paraphrase should receive higher ratings in the presence of existential priority modality, (e.g., *can*, George, 2011; Fox, 2014; Nicolae, 2014; Dayal, 2016; Xiang, 2016). Second, the MS paraphrase should receive higher ratings in *how* and *why* questions than in *who*-questions (Asher & Lascarides, 1998; Ginzburg, 1995). The predictions are far less clear for the understudied *when*, *what*, and *where* questions.

### Results and discussion

**Exclusions and pre-processing** Questions that received higher ratings for *something else* than any other option were removed (15%). These tended to be rhetorical questions (e.g., *Who knows?*, *Who has the time?*, *What are we becoming?*), whose interpretation is orthogonal to the question of whether *wh*-questions are interpreted exhaustively. After exclusion, ratings were normalized such that for each participant and item, the three remaining slider values summed to 1.

**Qualitative analysis** Because this is a novel task for testing *wh*-question interpretation, we begin by qualitatively assessing whether the ratings given for particular items accord with intuitions about the best paraphrase.

Questions like *Where do you live?* (mean=1,sd=0), *Where*

*do you work?* (.99,.03), *What do you drive now?* (.99,.06), *How do you spell that?* (.99,.04), *When did you first take your first piano lesson?* (.92,.24), *Why are you cutting off the phone?* (.81,.36) all received mean ratings for *the* at or near 1. For these questions, it is indeed possible but unlikely that there is more than one answer.

Questions that received high *every* ratings included *What does it have in it?* (.86, .31) and *Where have you skied?* (.73, .37). The first occurred in a context about cooking a casserole; the second in a conversation about the hearer's love for skiing. The exhaustive interpretation—wanting to know all the ingredients in the casserole, wanting to know all the places the hearer has skied—is sensible in both cases.

Questions that received high *a* ratings often involved recommendations, e.g., *What is a good brand, a inexpensive?* (.63, .39), which occurred in a discussion about computers where the hearer was an expert. Many questions involved discussions about books or movies, e.g., *What have you seen lately?* (.66, .3). Other interesting cases included *How can I tell you?* (.66, .42), where the speaker struggled to articulate (tell) why they like a certain movie, and *What else can we talk about?* (.99, .21) where the speaker is struggling to find a conversation topic. In both cases, there are presumably multiple answers (ways to tell, things to talk about), but a single one is sufficient to achieve the speaker's goal.

Overall, the qualitative assessment of individual items suggests that participants understood the task and that the ratings are interpretable.

**Data analysis** Analyses were conducted to assess overall question interpretation bias and the effect of modality and *wh*-word on question interpretation. To this end, we conducted a mixed effects linear regression predicting rating from fixed effects of PARAPHRASE (reference level: *every*), WH-WORD (reference level: *when*), a dummy-coded and mean-centered measure of whether a modal auxiliary verb was present (MODALPRESENT), all 2-way interactions between fixed effects, and the 3-way interaction. We included the maximal random effects structure justified by the design: random by-item and by-subject intercepts, as well as by-item and by-subject slopes for PARAPHRASE, and by-subject slopes for WH and MODALPRESENT.

We observed significant 3-way interactions. However, interpreting the interaction terms in this full model is very complex because two of our predictors include $> 3$ levels. We thus take the significant three-way interactions as evidence that effects varied by *wh*-word and report the outcome of separate specific models on each *wh*-word subset of the data: each model included fixed effects of PARAPHRASE, MODALPRESENT, and their interaction, coded as in the full model.

An exhaustive MA bias is evidenced as a significantly negative coefficient of the 'a vs. every' PARAPHRASE contrast; a non-exhaustive MS bias as a significantly positive coefficient. An introduction or strengthening of an MS bias in the presence of a modal is evidenced in a significantly positive interaction of MODALPRESENT with the 'a vs. every' PARAPHRASE contrast. The results of each model are shown in Table 2, with the two relevant contrasts highlighted in gray.

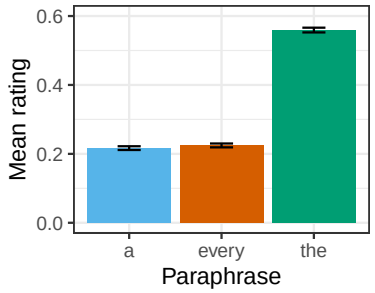


Figure 2: Mean ratings by paraphrase. Here and below, error bars indicate 95% bootstrapped confidence intervals.

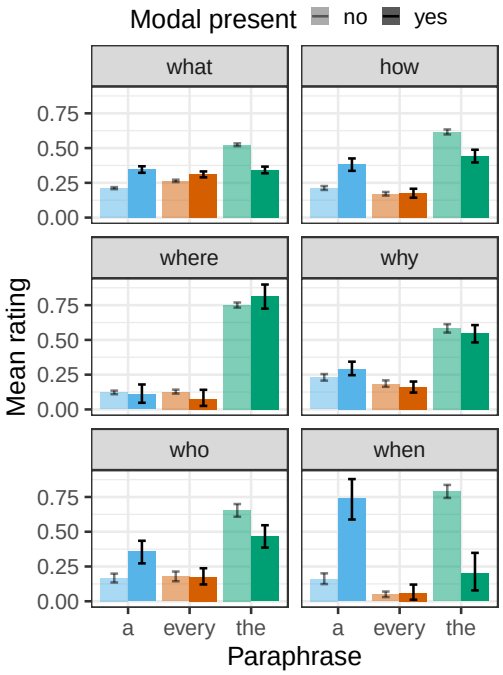


Figure 3: Mean ratings by paraphrase, *wh*-word and modality.

**Q1: Is there an overall bias for MA?** Overall mean ratings are shown in Fig. 2. Rather than a preference for MA over MS readings, we observed a clear preference for the *the*-paraphrase. An MA bias was observed only for *what*-questions, while an MS bias was observed for *how*, *why* and *when* questions. *Who* and *where* questions did not demonstrate evidence of a bias one way or another.

**Q2: Do modality and *wh*-word modulate question interpretation?** Simply, yes. Fig. 3 presents mean ratings by paraphrase and the presence of a modal, separately for each *wh*-word. We focus the following discussion on the coefficients and *p*-values in the Table 2 cells highlighted in gray.

For *what* questions, the overall MA bias is inverted to an

Table 2: Coefficient table (predicted β coefficient, standard error *SE*, *t* value, and *p* value) for *wh*-word-specific models. 'MP' stands for MODALPRESENT, and 'Para' for PARAPHRASE. The reference level for PARAPHRASE is *every*.

| WHAT | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | .26 | .01 | 24.61 | <.0001 |
| MP | .05 | .03 | 1.7 | .08 |
| Para.a | -.04 | .01 | -3.11 | <.002 |
| Para.the | .23 | .02 | 10.3 | <.0001 |
| MP:Para.a | .08 | .03 | 2.8 | <.006 |
| MP:Para.the | -.22 | .06 | -3.55 | <.0005 |

| HOW | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | .16 | .01 | 14.4 | <.0001 |
| MP | .002 | .03 | .07 | .95 |
| Para.a | .06 | .02 | 3.83 | <.0002 |
| Para.the | .41 | .03 | 15.29 | <.0001 |
| MP:Para.a | .16 | .04 | 4.23 | <.0001 |
| MP:Para.the | -.17 | .07 | -2.39 | <.02 |

| WHERE | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | .12 | .02 | 6.88 | <.0001 |
| MP | -.05 | .1 | -.46 | .65 |
| Para.a | -.004 | .02 | -.26 | .8 |
| Para.the | .61 | .04 | 14.01 | <.0001 |
| MP:Para.a | .04 | .09 | .47 | .64 |
| MP:Para.the | .07 | .24 | .3 | .8 |

| WHY | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | .18 | .02 | 11.59 | <.0001 |
| MP | -.02 | .04 | -.65 | .52 |
| Para.a | .06 | .02 | 3.02 | <.004 |
| Para.the | .38 | .03 | 11.74 | <.0001 |
| MP:Para.a | .09 | .04 | 2.05 | <.05 |
| MP:Para.the | -.02 | .08 | -.29 | .77 |

| WHO | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | .17 | .04 | 4.93 | <.0001 |
| MP | -.001 | .08 | -.01 | .99 |
| Para.a | .03 | .05 | .74 | .47 |
| Para.the | .41 | .08 | 5.16 | <.0001 |
| MP:Para.a | .2 | .11 | 1.78 | .09 |
| MP:Para.the | -.18 | .19 | -.96 | .35 |

| WHEN | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | .05 | .02 | 2.38 | .03 |
| MP | .01 | .08 | .16 | .88 |
| Para.a | .15 | .03 | 5.06 | <.0001 |
| Para.the | .68 | .05 | 12.55 | <.0001 |
| MP:Para.a | .55 | .12 | 4.65 | <.0002 |
| MP:Para.the | -.57 | .22 | -2.65 | .02 |

MS bias when a modal is present (note the significant positive shift in coefficient value when a modal is present). *How*, *why*, and *when* pattern together in showing an intitial MS bias, confirming the prediction from Ginzburg (1995) and Asher and Lascarides (1998). *Who* and *where* show no initial bias, only the overall preference for *the*-paraphrases. Thus, the prediction that *who* is biased for MA is not confirmed. Only *what* displayed an MA bias. Qualitative inspection of those questions which received the highest *every* ratings revealed that some of these were questions with a plural-marked complex *wh*-phrase (e.g., *What cities are they looking at*, .92, .15) that slipped through our initial filters. These were intended to be excluded precisely because they are expected to not be ambiguous between MS and MA. However, future work should explicitly include and test complex *wh*-questions. If they really are unambiguous, then plural-marked complex questions should show a clear preference for *every*, and singular-marked ones for *a* or *the*.

All *wh*-questions—except for *where*—showed a significant increase in MS ratings in the presence of a modal. In general, the presence of a modal auxiliary shifts probability away from *the*-paraphrases, and redistributes it to *a*-paraphrases: in Fig. 3 this can be seen in the higher red than blue *a* bars accompanied by the lower red than blue *the* bars, while the changes to *every* are negligible (indeed, the simple effect of MODALPRESENT never reached significance). Altogether, these results confirm the observation that modal auxiliaries facilitate MS readings.

## Conclusion

Our investigation of *wh*-question interpretation, contrary to theoretical predictions, yielded no evidence of an overall bias for exhaustive MA question readings in naturalistic dialogue; rather, for most cases we observed the opposite bias. Many questions seem to presuppose a unique answer, in virtue of general facts about the world. Qualitative analysis of cases where questions were interpreted as MS or MA tended to occur in contexts with a clear discourse goal. Future work should further extend the database with independently quantified discourse goals to investigate the interaction between discourse goals and linguistic form factors.

We did find support for some, but not all, observations regarding the effect of linguistic form on question interpretation reported in the literature. In particular, modal questions significanly increased the likelihood of an MS reading. However, this result should not be interpreted as revealing a grammatical constraint *against* MS in non-modal questions. Indeed, we have already discussed two non-modal questions that were interpreted non-exhaustively: *What have you seen lately?* and *What is a good brand?* One might be tempted to explain this result by appealing to either the temporal adverb *lately*, which could be argued to restrict the domain, or the indefinite description *a good brand*, which introduces an ordering of brands that picks out one on the top. However, such linguistic explanations are not available across the board: nei-

ther *What do you find boring?* nor *What do you like to read?* have any linguistic element which could be argued to give rise to semantic MS. Rather, it seems that in all cases, the speaker's goal is to keep the conversation active. Thus, an MS answer will suffice, even if an MA answer is available.

Finally, our findings have methodological implications: data hand-selected during theory-building may be biased and not reflect a realistic distribution of meanings (Degen, 2015; Gibson, Piantadosi, & Fedorenko, 2011). As Asher and Lascarides (1998) note, given the limited data sets which have been used to construct semantic theories of *wh*-questions, important generalizations about the role of context-dependent speaker goals have been ignored.

The recipe for pragmatics is this: resolving interpretation requires hearers to reason about the speaker's goal in light of the speaker's utterance. This reasoning process requires integrating linguistic and non-linguistic cues to meaning, as spelled out in constraint-based accounts of pragmatics (Degen & Tanenhaus, 2015, 2019). Formally characterizing the interaction of linguistic and discourse factors in *wh*-question interpretation is an exciting avenue for future work.

## References


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

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive psychology*, *11*(4), 430–477.

Dayal, V. (2016). *Questions* (C. Barker & C. Kennedy, Eds.). Oxford, United Kingdom: Oxford University Press.

Degen, J. (2015). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*, 11–1.

Degen, J., & Jaeger, T. F. (2011). The TGrep2 database tools. *Unpublished manuscript*.

Degen, J., & Tanenhaus, M. (2015). Processing scala implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667–710.

Degen, J., & Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. In C. Cummins & N. Katsos (Eds.), *Handbook of experimental semantics and pragmatics* (chap. 3). Oxford University Press.

Fox, D. (2014). *Mention-some readings.* MIT seminar notes.

George, B. R. (2011). *Question embedding and the semantics of answers*. Doctoral dissertation, Linguistics Department, University of California, Los Angeles.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using mechanical turk to obtain and analyze english acceptability judgements. *Language and Linguistics Compass*, *5*(8), 509–524.

Ginzburg, J. (1995). Resolving questions, I. *Linguistics and Philosophy*, *18*(5), 459–527.

Godfrey, J., Hilliman, E., & McDaniel, J. (1992). SWITCHBOARD: A telephone speech corpus for research and development. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 517–520.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. , *5*(1), 173–184.

Groenendijk, J., & Stokhof, M. (1984). *Studies on the semantics of questions and the pragmatics of answers*. Doctoral dissertation, University of Amsterdam.

Hawkins, R., Stuhlmüller, A., Degen, J., & Goodman, N. (2015). Why do you ask? good questions provoke informative answers. In *Proceedings of the 37th annual conference of the cognitive science society.*

Hintikka, J. (1976). *The semantics of questions and the questions of semantics: Case studies in the interrelations of logic, semantics, and syntax*. North-Holland Publishing Company.

Huang, Y. T., & Snedeker, J. (2011). *Logic and conversation* revisited: Evidence for a division between semantic and pragmatic content in real- time language comprehension. *Language and Cognitive Processes*, *26*(8), 1161–1172.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*(1), 3–44.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press, Cambridge, MA.

Moyer, M. (2020). *The question of questions: resolving (non-) exhaustivity in* wh*-questions*. Doctoral dissertation, Linguistics Department, Rutgers University.

Moyer, M., & Syrett, K. (2019). (Non-)exhaustivity in embedded questions: Contextual, lexical and structural factors. In M. T. Espinal, E. Castroviejo, M. Leonetti, L. McNally, & C. Real-Puigdollers (Eds.), *Proceedings of Sinn und Bedeutung 23* (pp. 207–224).

Nicolae, A. C. (2014). *Any questions? polarity as a window into the structure of questions*. Doctoral dissertation, Linguistics Department, Harvard University.

Rohde, D. L. (2005). Tgrep2 User Manual. *Unpublished manuscript*.

Rothe, A., Lake, B. M., & Gureckis, T. M. (2018). Do people ask good questions? *Computational Brain & Behavior*, *1*(1), 69–89.

Searle, J. R. (1975). Indirect speech acts. In *Speech acts* (pp. 59–82). Brill.

van Rooij, R. (2003). Questioning to resolve decision problems. *Linguistics and Philosophy*, *26*(6), 727–763.

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*. Doctoral dissertation, Linguistics Department, Harvard University.

Zondervan, A. (2010). *Scalar implicatures or focus: An experimental approach*. Doctoral dissertation, University of Utrecht.