How a speaker signals, and a listener resolves, levels of exhaustivity

XPrag Workshop ‘Exhaustivity in Questions and Answers - Experimental and Theoretical approaches’

Morgan Moyer and Ranga Tirumala

June 13-14, 2019

Different questions warrant different answers

Who came to the party?

John, Bill, Sue, Chris,…came to the party.

How do you get to Central Park?

You take the subway to Columbus Circle.

Two main theories

The Form/Modal Theory

Modal/non-finite questions license MS (George 2011; Xiang 2015; Bhatt 1999 for non-finite clauses)

The "Pragmatic" Theory

Contextual goals license MS (Groenendijk & Stokhof 1982, 1984; Ginzburg 1995; Asher & Lascarides 1998;Lahiri 2002; van Rooij 2003, 2004; Schulz & van Rooij 2004)

Form is important.

Where **can** I find a coffee?
Dana knows where I can find a coffee.
Dana knows where to find a coffee.

Which numbers under 10 are prime?
Dana knows which numbers under 10 are prime.

What gas stations are open now?
Dana knows what gas stations are open now.

Experimental support (Xiang & Cremers 2017; Moyer & Syrett, in press)

Non-Modal MS Questions

Who’s got a light?

Dana knows who’s got a light

https://www.youtube.com/watch?v=Ri-0tEcmIDE

https://www.roberrera.com/modern-concerts-are-big-on-spectacle-short-on-feel/

Experimental support (Moyer & Syrett, in press)

Where are we now?

Though Form and Pragmatic theories both can't account for some data, they track a lot of other data.

These two theories identify at least two features relevant to non-exhaustivity.

The goal of this study is to take these features and a few more, and ask, "Can we design an experiment to test the role they play individually and jointly in determining exhaustivity?"

Roadmap of Today's Talk

I. What factors determine the mention-some reading of questions?

- Variation due to linguistic form of the question
- Variation due to contextual discourse goals

II. Study 1: Corpus study of cue distribution

III. Study 2: Judgements of acceptability and likelihood

IV. Discussion

Features important for MS

Linguistic Form:

1. WH-Word (Hintikka 1976; Ginzburg 1995; Asher & Lascarides 1998)
2. Modal/Non-finite (George 2011, Xiang 2015; Dayal 2016)
3. Matrix verb (Heim 1994; Sharvit 2002; Guerzoni & Sharvit 2007; George 2011; Klinedinst & Rothschild 2011; Uegaki 2014; Theiler 2014)

Contextual:

4. Speaker goals (Groenendijk & Stokhof 1982, 1984; Boër & Lycan 1985; Ginzburg 1995; Asher & Lascarides, 1998; van Rooij 2003, 2004, Schulz & van Rooij 2004)

Motivation for current studies

1. Understanding the input is useful for determining what language learners are exposed to (cf. Syrett 2007; Dudley 2017)

2. Cues are probabilistically linked to interpretation. The distribution of a cue, or cue co-occurrence, may effect the robustness of an interpretation
   (Degen & Tanenhaus 2015, 2016, 2018; Degen 2015; Elman, Hare, McRae 2004; MacDonald, Pearlmutter & Seidenberg 1994; Seigenberg & MacDonald 1999; Tanenhaus & Truswell 1995; McRae & Matsuki 2004, a.o.)

The Corpora

- British National Corpus (http://www.natcorp.ox.ac.uk/)

- Brown (NLTK)
- Reuters (NLTK)
- Penn Treebank (NLTK)
- Australian Broadcast Corpus (NLTK)

British National Corpus (BNC)

over 6 million sentences

- 90% Written (over 6 million sentences)
  - Extracts from newspapers, journals, periodicals all ages
  - Academic books and popular fiction
  - Letters and memoranda
  - School and university essays
- 10% Spoken, 49,292 sentences
  - Orthographic transcriptions of informal conversations (demographically balanced)
  - Other contexts too, like business or government meetings, radio shows

Coding in corpus study

Question Type: root, embedded, relative clause, fragment, ambiguous

Embedding/matrix verb

Clause Type: finite, modal, non-finite

Wh-word: *who*, *where*, *how*

How we searched for questions

Tagged for POS with NLTK POS tagger.

Pulled all sentences that had a *who*, *how*, *where.*

We applied a set of ***ordered heuristics*** to tag these sentences for Question Type and Clause Type.

questType Heuristics (Exclusionary)

For a given sequence of words or POS tags:

4. If VB_{RC}_WH → Relative Clause
5. If {RC}_WH → Relative Clause

{RC} = Part-of-Speech tags for all phrases that can be RC heads

1. If ¬∃.VB in sentence → Fragment
9. Else → Ambiguous

Code available on github.com/rangat/whAnalysis

questType Heuristics (Inclusionary)

2. If {RC}_VB_ ¬{RC}_WH — Embedded
3. If VB_WH and ¬S-Aux-Inv. — Embedded
8. If VB_WH — Embedded

6. If WH_{AUX}_{RC}_VB — Root
   Except: if WH_{RC}_{AUX}
   Except: if {AUX}_VB_{RC}
7. If '?' at end of sentence — Root

Subj-Aux Inversion

ClauseType Heuristics, ordered

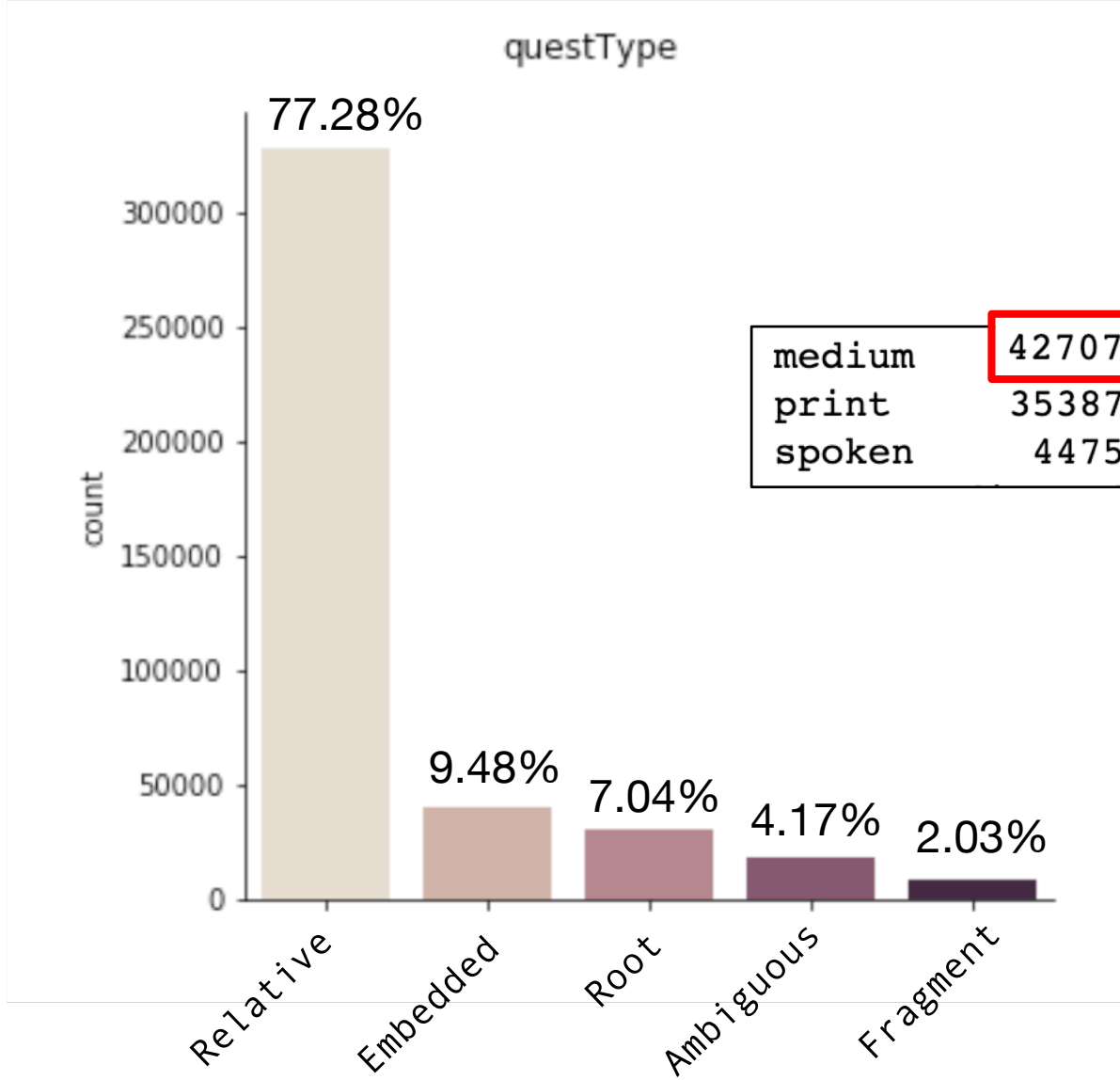
For Embedded Questions

1. Modal: WH_{Mod}
2. Non-Finite: WH_to_V
3. Finite: all else

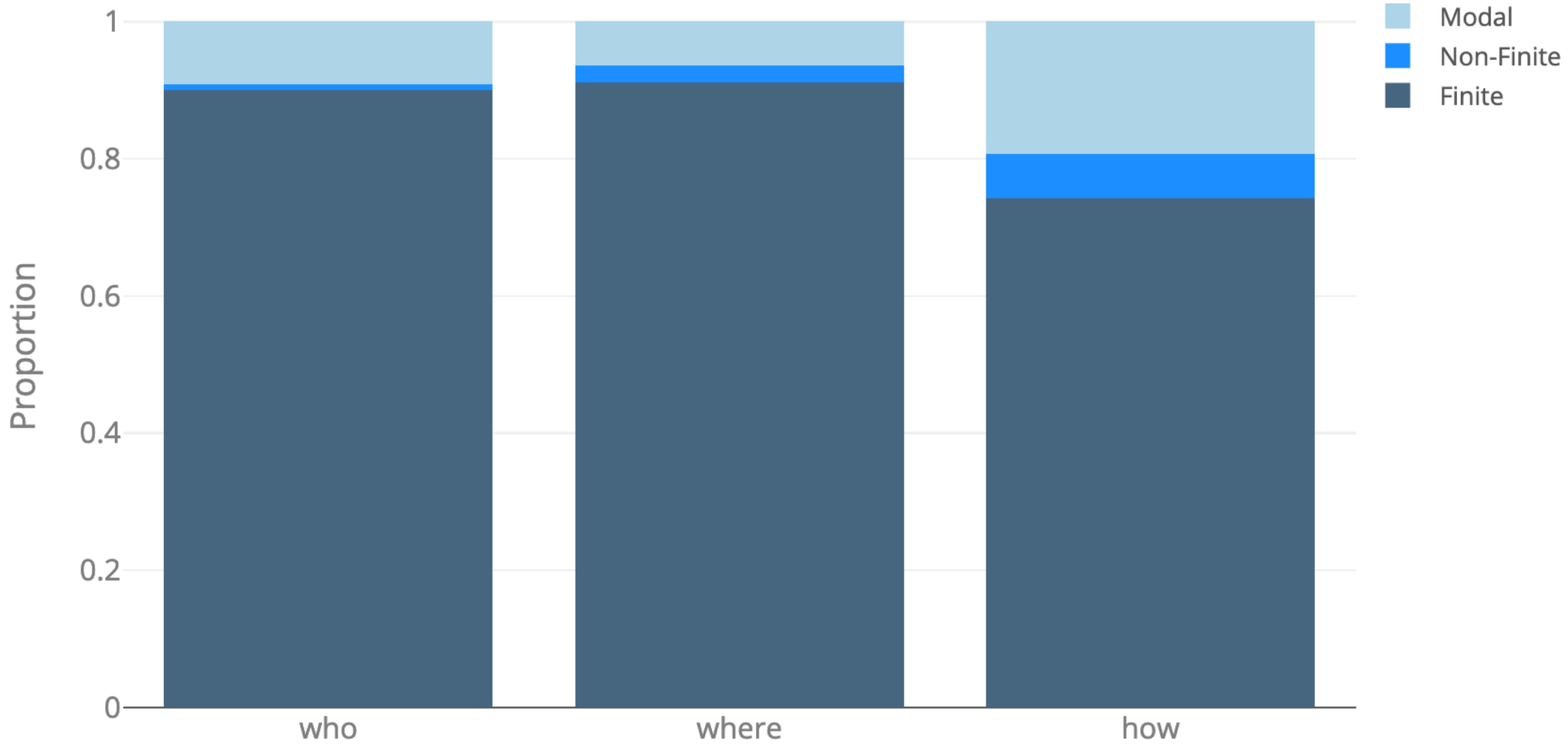
For Root Questions

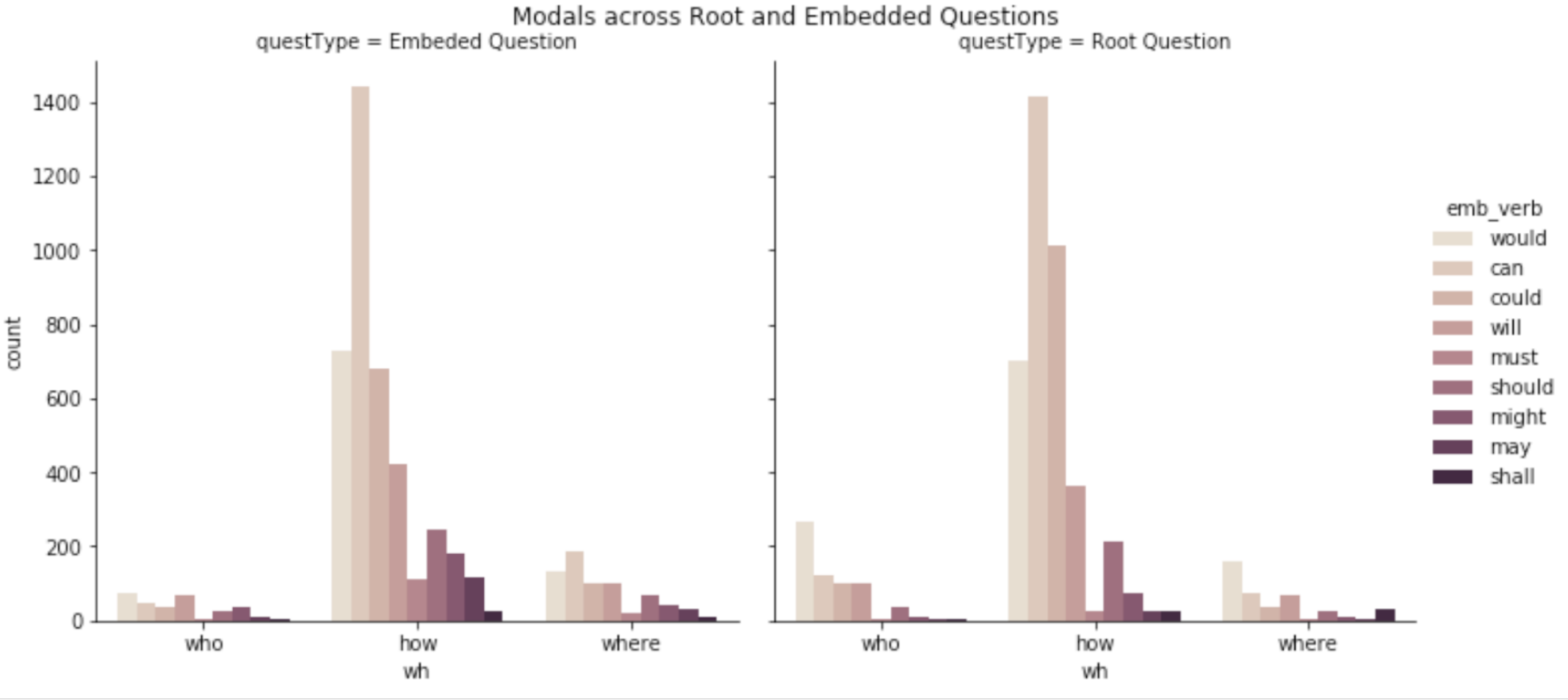
1. Modal: WH_{Mod}
2. Finite: all else

Overall Results



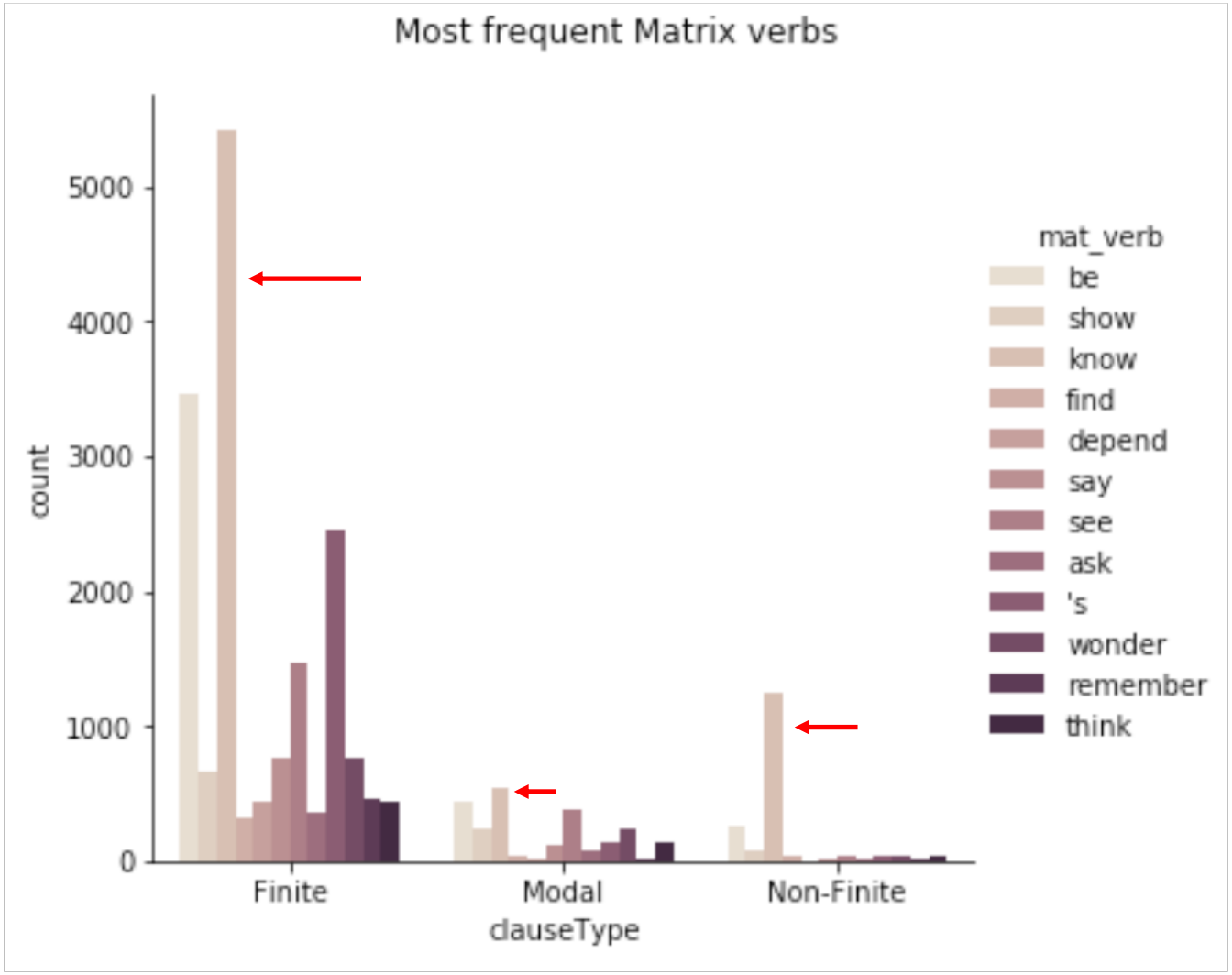
- Highest proportion clauseType is Finite
- *How*-q occurring more than other WH-q with Modal & Non-Finite Clause Types





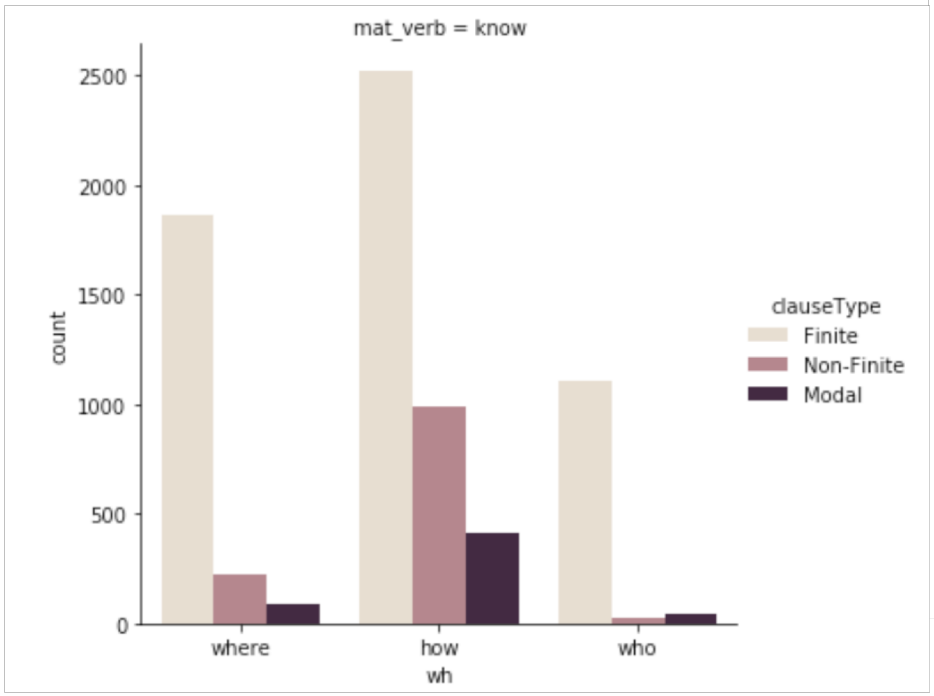
- *Can* is the most frequently occurring modal
- Modals are co-occurring with *how*-qs more than any other q's

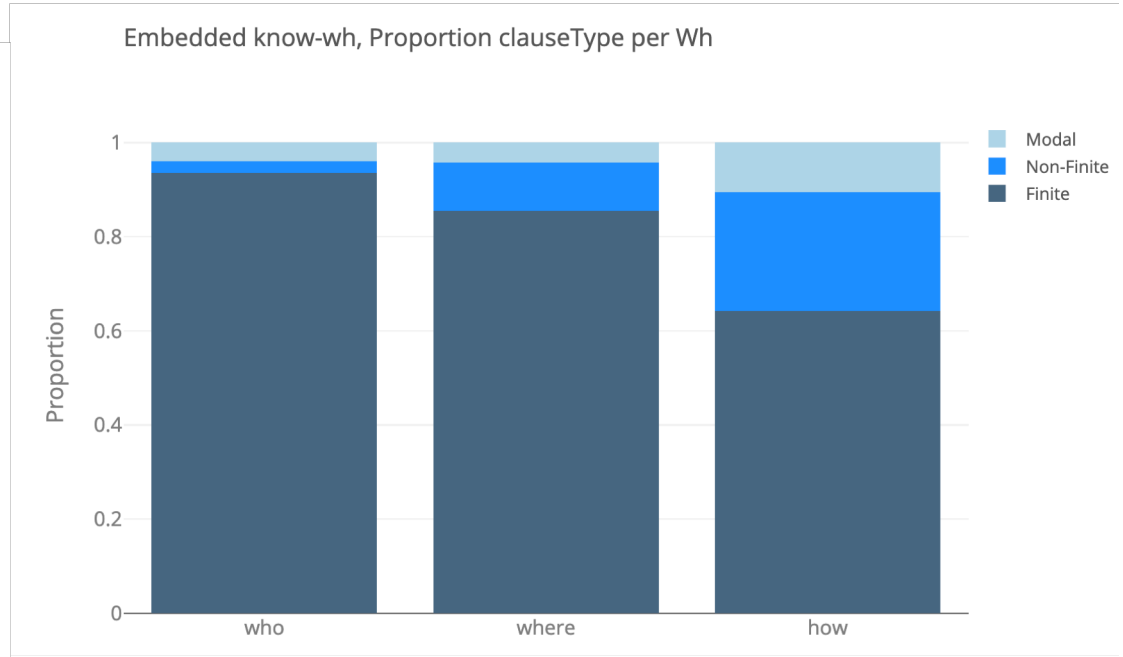
Matrix verbs



- Know is the most frequent across clauseType
- Occurs overwhelmingly with finite clauseTypes

clauseType and wh across *know-q*





- *Know* co-occurs overwhelmingly with Finite clauses
- *Know-how* questions have the highest proportion of Non-Finite and Modal clauses, relative to other Whs

Discussion of Corpus

The co-occurrence data validate some intuitions from the literature on the basis of one big assumption

Big Assumption: The cues we've given are indeed cues for exhaustivity/non-exhaustivity

We want to link the distributional data to the likelihood of interpretation

Experiment (n=238)

Two Tasks:

"How ***likely*** is someone to give an answer, given the cues in the question form"

"How ***acceptable*** is it for someone to give an answer, given the cues in the question's form"

5-point Likert scale (unacceptable/unlikely to acceptable/likely)

Experiment 2

- Designed using Qualtrics (Provo, UT)
- Run on Amazon Mechanical Turk

- Workers restricted to US location
- HIT accept rate greater than or equal to 99%
- HIT completion at 1,000 or above

2x3x2(x3) Factorial Design

12 Test trials, with embedded *know*-questions

Between-subjects factors:

- Task (Likely, Accept)
- Answer (MS, MO, MA)
- Modal (modal, nomodal)

Within-subjects:

- WH (who, where, how)

32 Fillers

Pseudo-randomized in a Latin-square, 12 lists

Likely, who-nomodal-MO

The Paris Hotel only accommodates seven people in a hotel room. The Vellucis have decided to go on vacation and stay in a fancy hotel. Mom and Dad allowed Amanda to invite three friends to come on vacation with them.

Amanda invited Jenny, Buster, and Janice, but not Robbie or Sally.

Likely, who-nomodal-MO

Mom asks Dad, "**Who did Amanda invite?**"

WH + NoModal

Based on Dad's answer, Mom concludes,

"**Dad knows who Amanda invited.**"

How likely is it that Dad gave an answer like, "Jenny"?

TASK

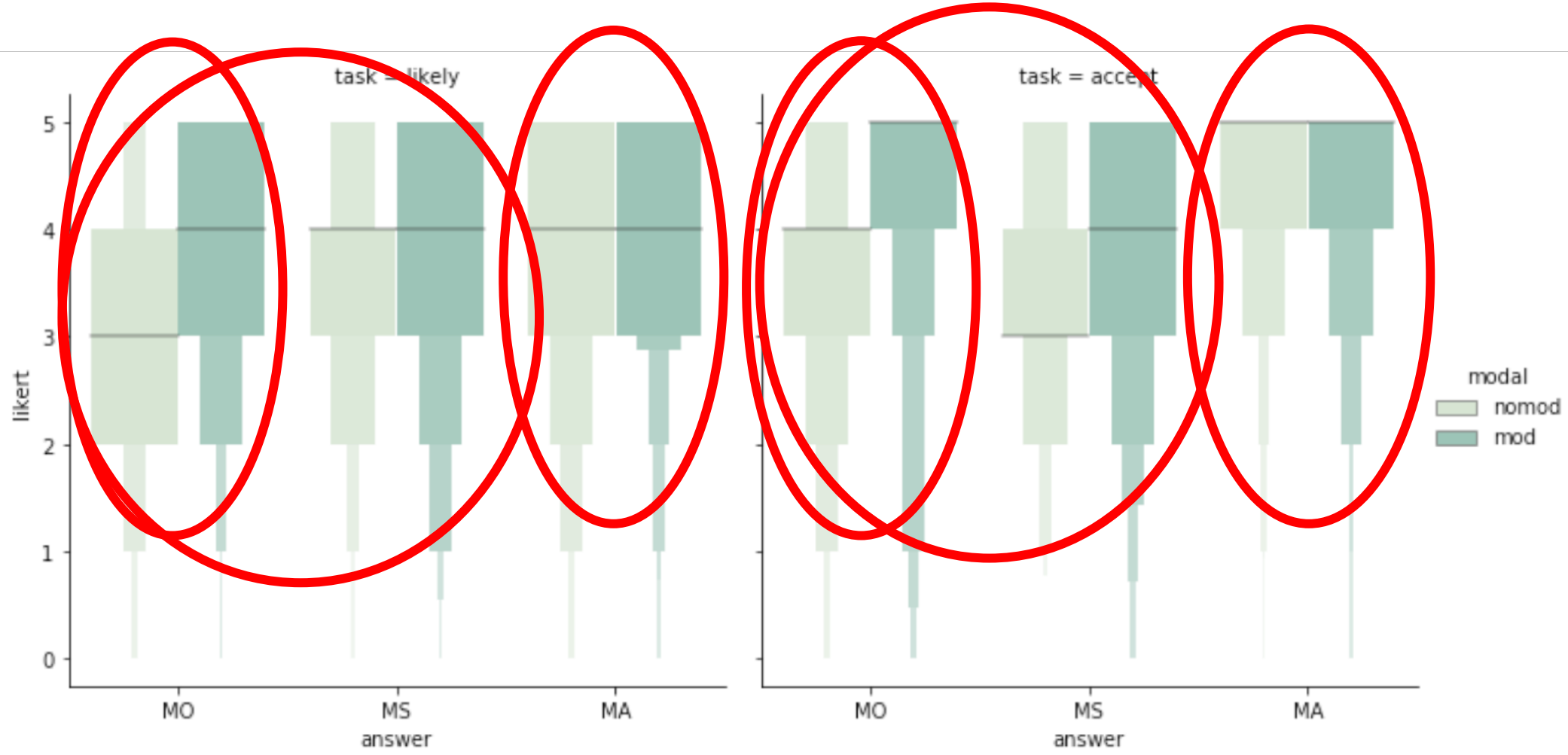
Answer

Predictions

- MO/MS answer conditions should receive higher ratings for Modal questions (cf. Modal Hypothesis)
- MO/MS answer conditions should receive lower ratings for NonModal questions
- MA answer conditions should receive high ratings across the board

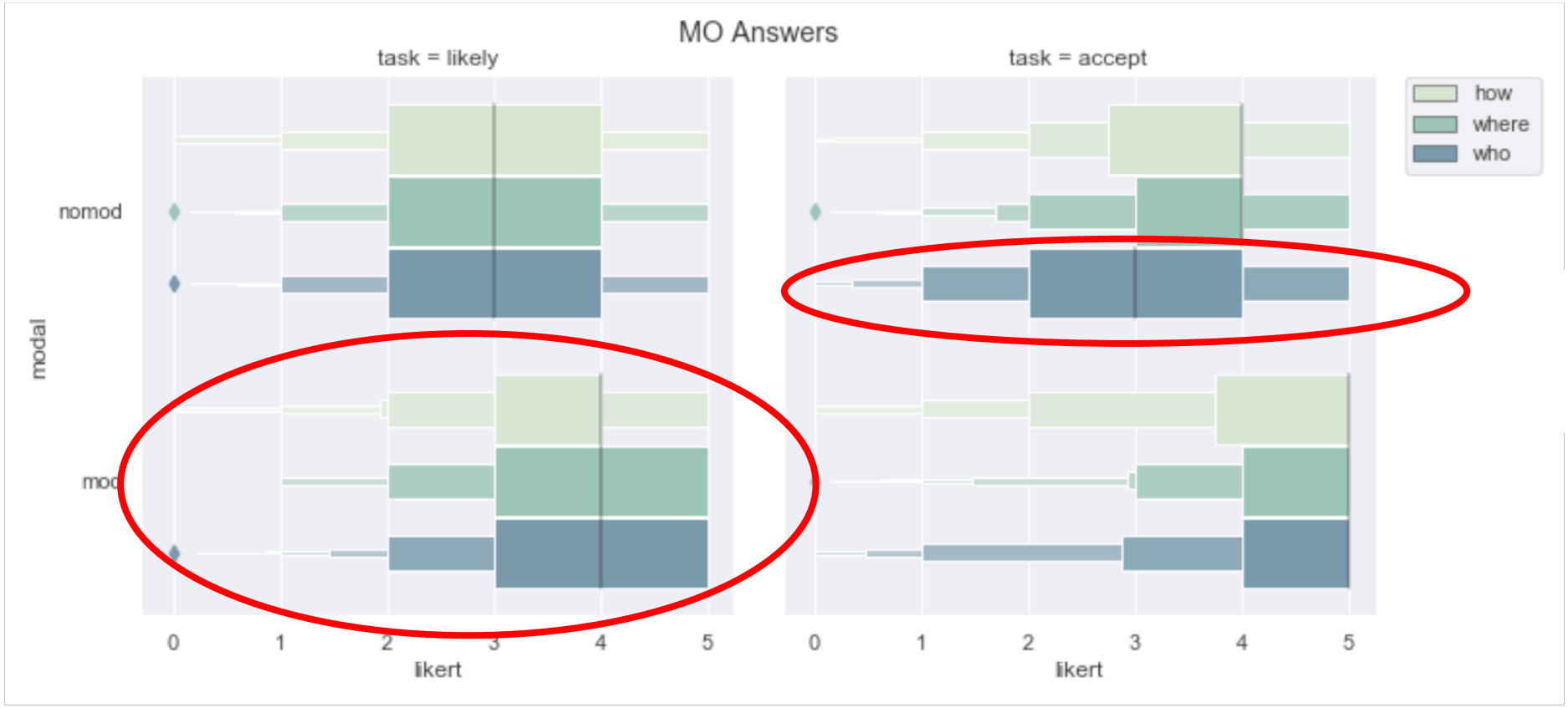
Overall

NonMod MS/MO medians degraded wrt MA
MS/MO Medians still 3 or higher!
In MO: NoMod degraded wrt Mod



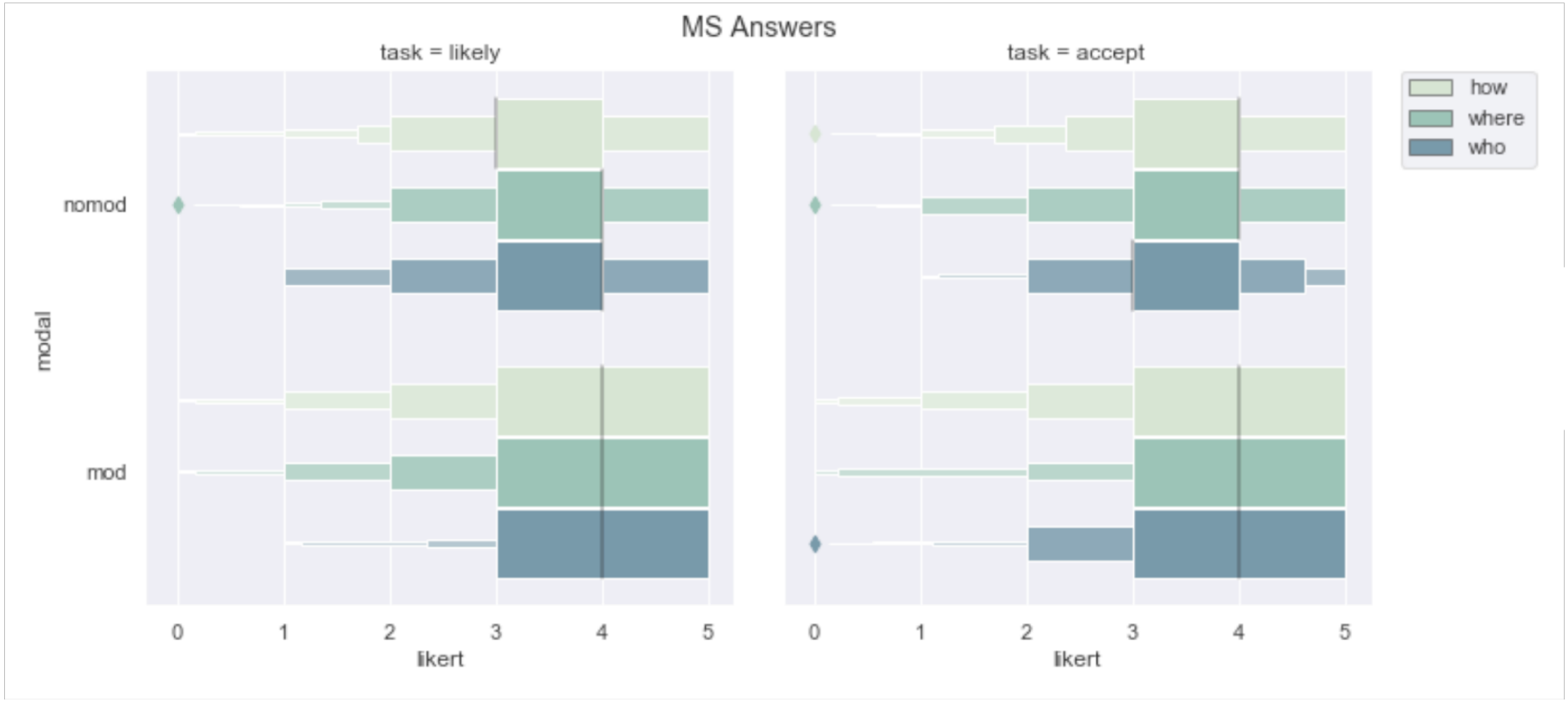
MO Answers

Consistently, NoMod degraded wrt Mod
NonMod *who* degraded in Accept Task
Lower medians for Modals in Likely Task



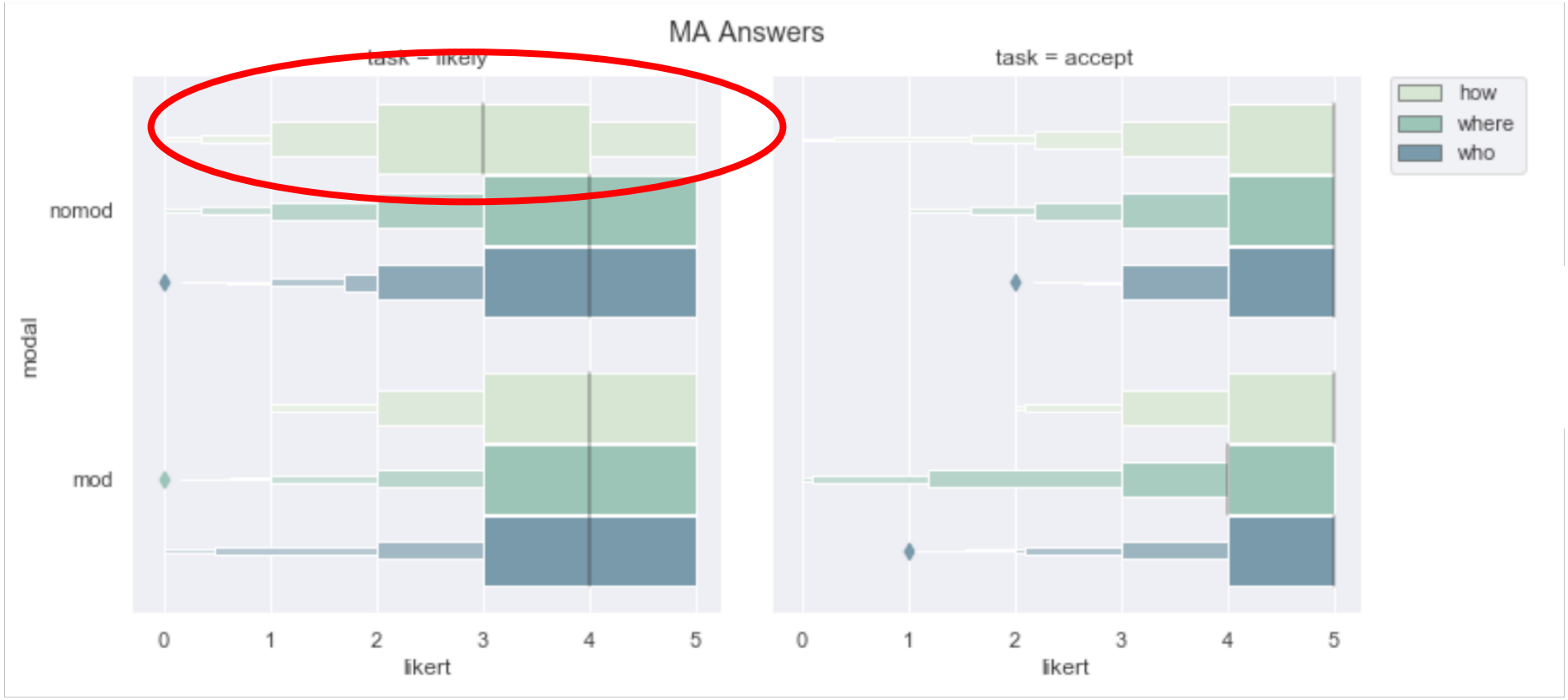
MS Answers

Not many differences across Factors!



MA Answers

Lower medians in Likely Task
NoMod *How* degraded wrt other *WH* in Likely Task



Discussion of results

The Likert rating of an answer is sensitive to form factors

BUT ratings for all answers were generally high – medians were all above ‘3’

Do the Task differences track a particular kind of contextual evaluation sensitivity (cf. Roberts, 2017)?

General Discussion

Co-occurrence data support some intuitions about form differences

But experimental data suggest that those form considerations don’t make as big a difference as you might think

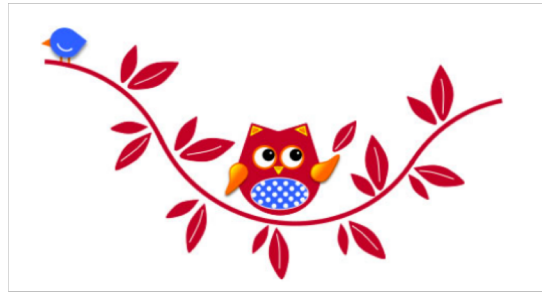
The structure in the input isn’t robustly affecting interpretation

General Discussion

What do we make of the corpus data and the link to interpretation?

Acceptability of an answer is neither seemingly purely a form or purely a contextual matter. It’s a combination of the two.

Thank you!



Kristen Syrett
Yimei Xiang
Veneeta Dayal
Paul Pietroski
Liz Camp

Divya Appasamy
Knyckolas Sutherland
Caley Howland
Tim Dawson

Rutgers Lab for Developmental Language Studies
Semantics-U-Name-it-at-Rutgers (SURGE)

Funding from Rutgers Off-Campus Dissertation Fellowship, Rutgers Aresty Research Center, and the Rutgers Center for Cognitive Science