

# **Understanding the semantics and pragmatics of (non-)exhaustivity in *wh*-questions with corpus and experimental methods**

Morgan Moyer  
Linguae Seminar  
25 November 2021

# Where can I find coffee?



Mention-  
All (MA)

“The RR Diner,  
Starbucks,  
Peet’s,...”

“The RR Diner.”



Mention-  
Some (MS)

# Readings of *wh*-questions

## (1) Where can I find coffee?

**MS** What is *a* place that I can find coffee?

**MA** What is *every* place that I can find coffee?

## (2) Who came to the party?

**X—MS** Who is *a* person that came to the party?

**MA** Who is *every* person that came to the party?

# Roadmap

## A. Background

1. Insights from pragmatics/cognitive science
2. Insights from semantic theory
3. Experimental background
  - Moyer & Syrett 2019, Moyer 2020

## B. The current study

4. Step 1: Naturalistic stimuli from corpora
5. Step 2: Experiment
  - Root (1a)
  - Embedded (1b)
6. Quantifying Context (Experiments 2a/b)

## C. Discussion

# Literal vs. non-literal meaning

*Can you reach the salt? (Searle 1975)*



# Goal dependence

Clark (1979)

I want to buy \$5 worth of bourbon.

Does a 5th of Jim Beam  
cost more than \$5?

It costs \$5.95.



# Computational CogSci

Question asking and answering is a species of rational, goal-directed behavior

- Rational Speech Acts Models  
Hawkins, Stuhlmüller, Degen & Goodman 2015; Hawkins & Goodman, ms
- Questioning as program generation  
Rothe, Lake, & Gureckis 2018

# Semantic theories

- Questions denote the set of (possible?) answers to the question  
Hamblin 1973; Karttunen 1977; Groenendijk & Stokhoff 1982, 1984
- The focus has been on **strengths of MA** (weak/intermediate/strong exhaustivity)  
Theoretical: Karttunen 1977; Groenendijk & Stokhof 1982, 1984; Heim 1994; Dayal 1994, 1996; Lahiri 1991, 2002; Spector 2005, 2007, 2009; Guerzoni & Sharvit 2007; Sharvit 2007; Egré 2008; Kleindinst & Rothschild 2011; George 2011; Theiler 2014; Uegaki 2015; Romero 2015; Spector & Egré 2015; Mayr 2019, Uegaki & Sudo 2020....  
(not an exhaustive list)  
Experimental: Cremers & Chemla 2014, 2016; Chemla & George 2016; (Phillips & George 2018)
- Attempt to explain away MS as exceptional, presupposing it to be a marked case  
Groenendijk & Stokhoff 1982, 1984; George 2011, Ch. 6; Fox 2014, 2018; Nicolae 2014/5; Xiang 2016; 2020; Dayal 2016
- Although some theories argue for **ambiguity or contextual parameterization**  
Hintikka 1976; Beck & Rullmann 1999, Lahiri 2002; George 2011, Ch. 2; Ciardelli, Roelofsen & Groenendijk 2019—Ginzburg 1995; Asher & Lascarides 1998; Lahiri 2002; van Rooij 2003, 2004

# Some observations about linguistic factors affecting MS

(1) **Where** can I find coffee?

(2) **Who** came to the party?

(3) Scully **knows**  
predicted ...

... where we can find coffee

... who came to the party

Generally	<b>Questions are MA-biased</b> <b>MS is restricted...</b> ...see previous page for list
Wh-Word	<b>Who-Qs are MA-biased</b> <b>Non-Who-Qs are MS-biased</b> Ginzburg 1995; Asher & Lascarides 1998
Modal	<b>Modal-Qs are MS-biased</b> <b>Non-Modal-Qs are MA-biased</b> George 2011; Fox 2014; Nicolae 2014; Dayal 2016; Xiang 2016
Matrix Verb	<b>Know-wh is MA-biased?</b> Groenendijk & Stokhof 1982/1984; Berman 1991; Heim 1994, George 2011; Klindinst & Rothschild 2011; Schulz & Roeper 2011

# Experimental semantics

## Mention-All

- Matrix Verb restrictions on Weak/Intermediate/Strong exhaustivity, don't always track literature observations

Cremers & Chemla 2016, 2017; Chemla & George 2017

## Mention-Some

- Acceptability affected by proportion of false:true answers  
Phillips & George 2018
- Support for modality, *wh*?  
Xiang & Cremers, 2017

# Moyer & Syrett 2019/Moyer 2020

## Experiment 1

n=232 (14 excluded for non-native speaker status  
Binary T/F acceptability judgment task  
Run in-lab

Within subjects:

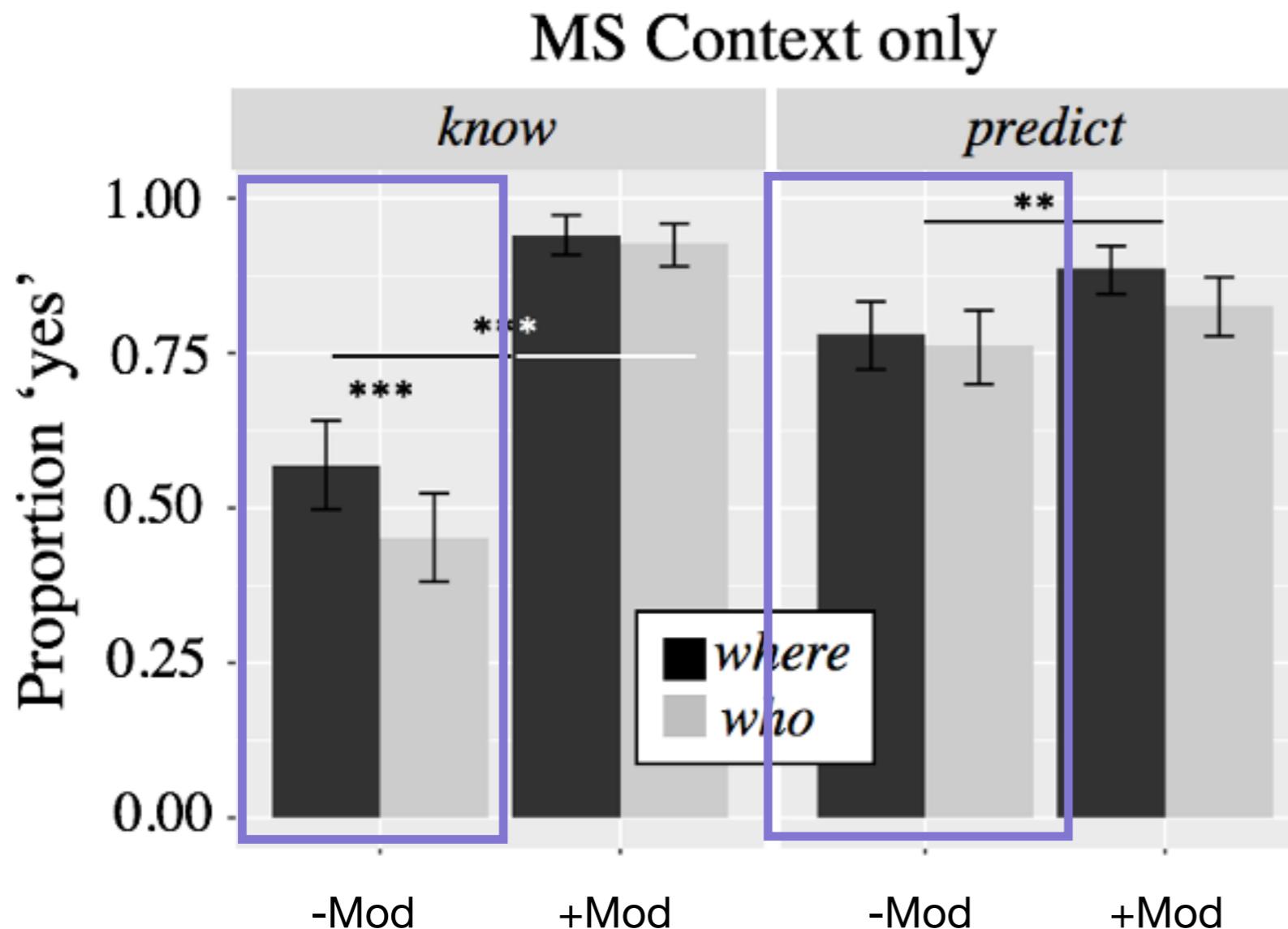
- **Answer** (MA, MS, MA+False, MF)
- **Matrix Verb** (*know*, *predict*)
- **Wh-word** (*who*, *where*)

Between subjects:

- **Finiteness** (+,-)

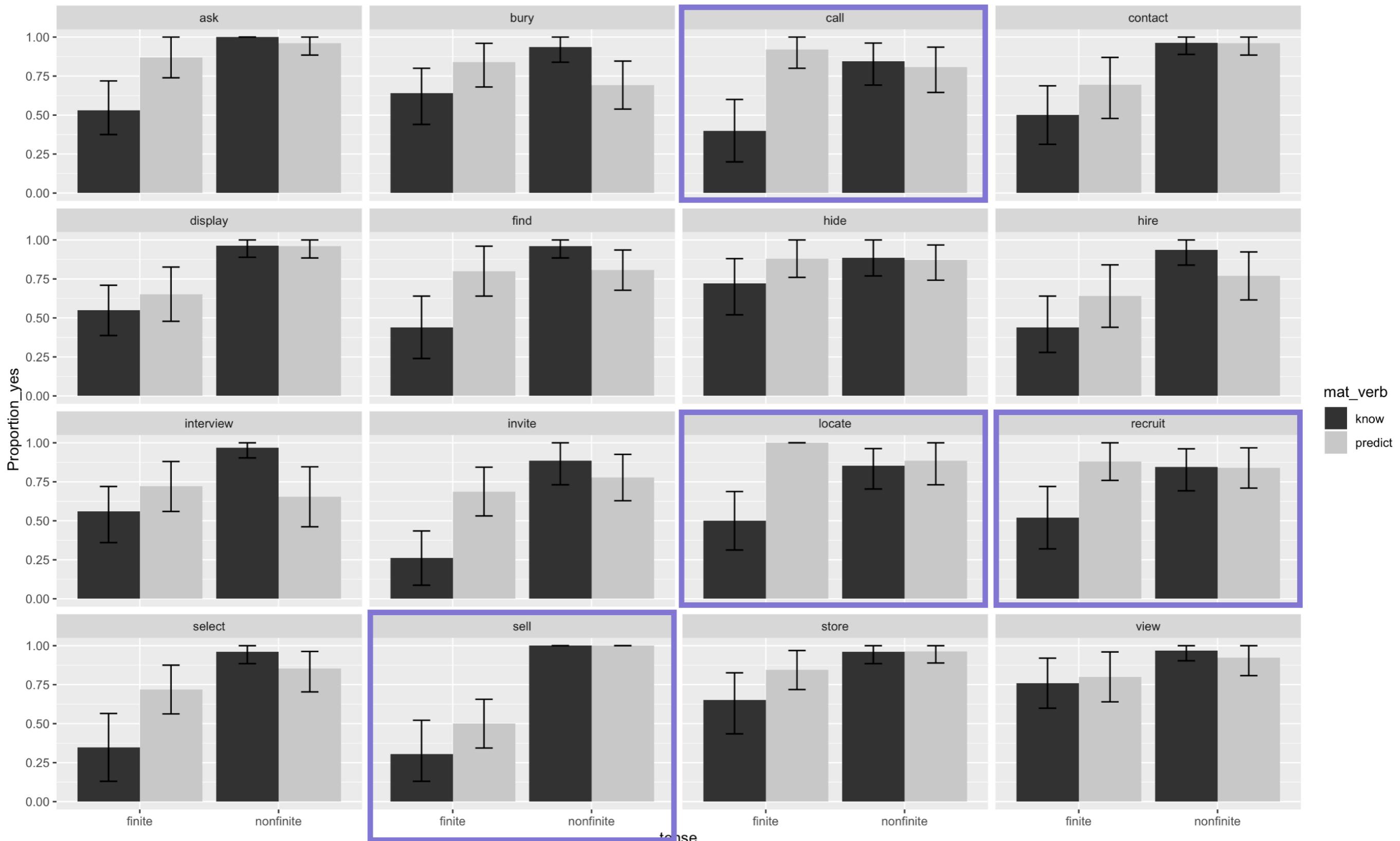
+ 16 different embedded predicates  
(8 *where*, 8 *who*)

# Experiment 1 Results



- 3-way interaction between all forms!
- Degraded acceptability of MS driven by **non-modal *know-who***
- **Why?**

### Embedded Verbs in Mention-some context



# MS vs. MA is sensitive to...



**Goal/plan/decision problem/  
mental state**

Clark 1979; Groenendijk & Stokhoff 1982, 1984; Ginzberg 1995,  
Asher & Lascarides 1998, Lahiri 2002; van Rooij 2003, 2004

# Who has a light?



# Moyer & Syrett 2019/Moyer 2020

## Experiment 2

n=318 (6 excluded for non-native, 6 for browser incompatibility)

Multiple choice acceptability judgment task  
Ibex Farm / MTurk

Within subjects:

- **Stakes** (high, low)
- **Answer** (MA, Mention-some (MS), Mention-One (MO))
- **Informativity** (max, min)

Between subjects:

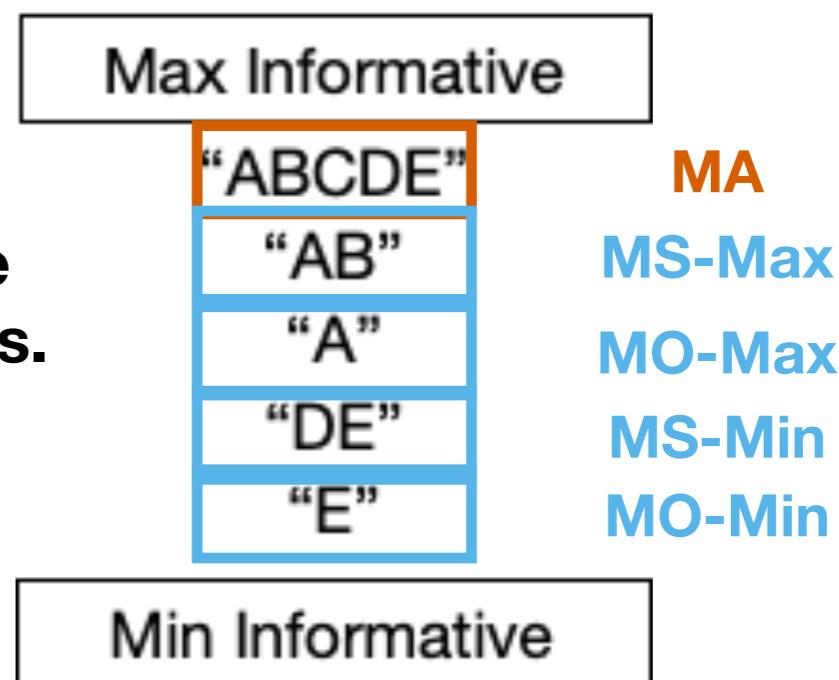
- **Finiteness** (+,-)

# Experiment 2

A local apartment building has caught on fire. The landlord's floor plan shows the following breakdown of tenants per floor:

Floor A has 10 tenants,  
Floor B has 8,  
Floor C has 5,  
Floor D has 2,  
Floor E has 1, and  
Floor F has 0.

Think: coffee  
shops close vs.  
far



The fire chief is deciding where to send his firefighters, because there are people trapped but they don't know on which floors.

He asks his firefighters, “Where should we go to find the trapped people?”

Firefighter A says, “Floor A.” MO-Max

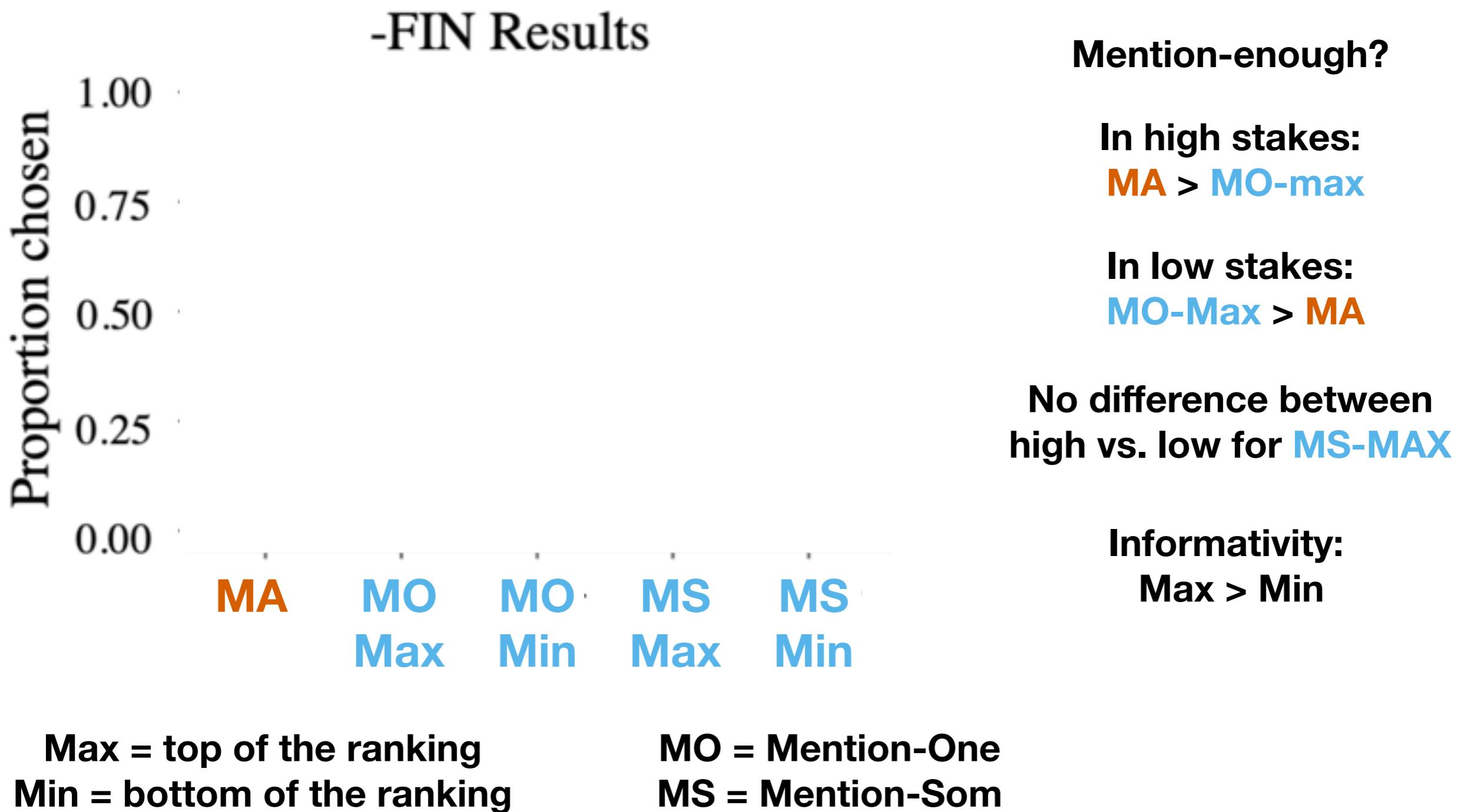
Firefighter B says, “Floors A, B, C, D, and E.” MA

Firefighter C says, “Floors D and E.” MS-Min

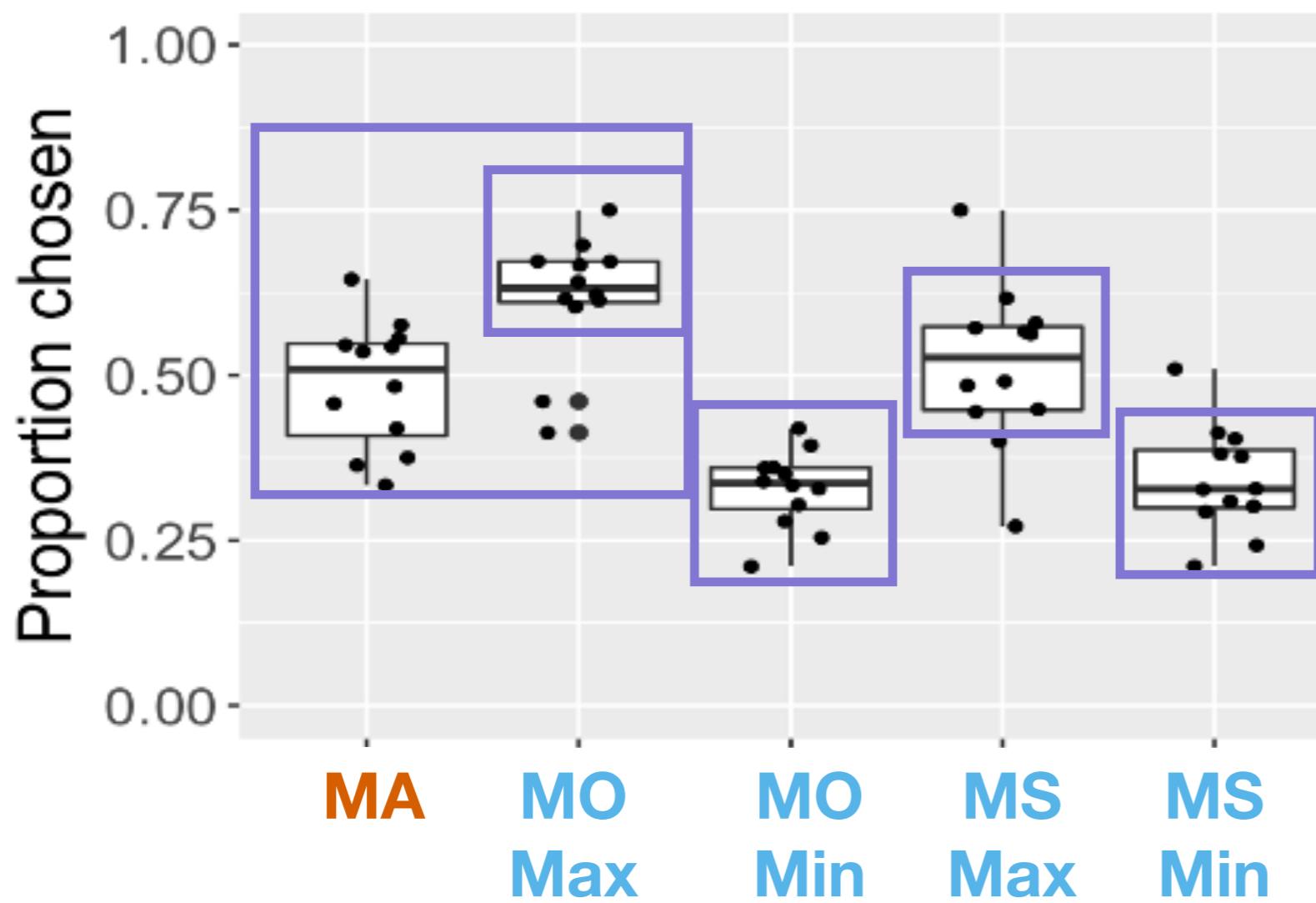
**Who knows where to go to find the trapped people?  
(Choose all that apply)**

Firefighter A    Firefighter B    Firefighter C    None of the Above

# Experiment 2 Results



# Finite (non-modal) Clauses

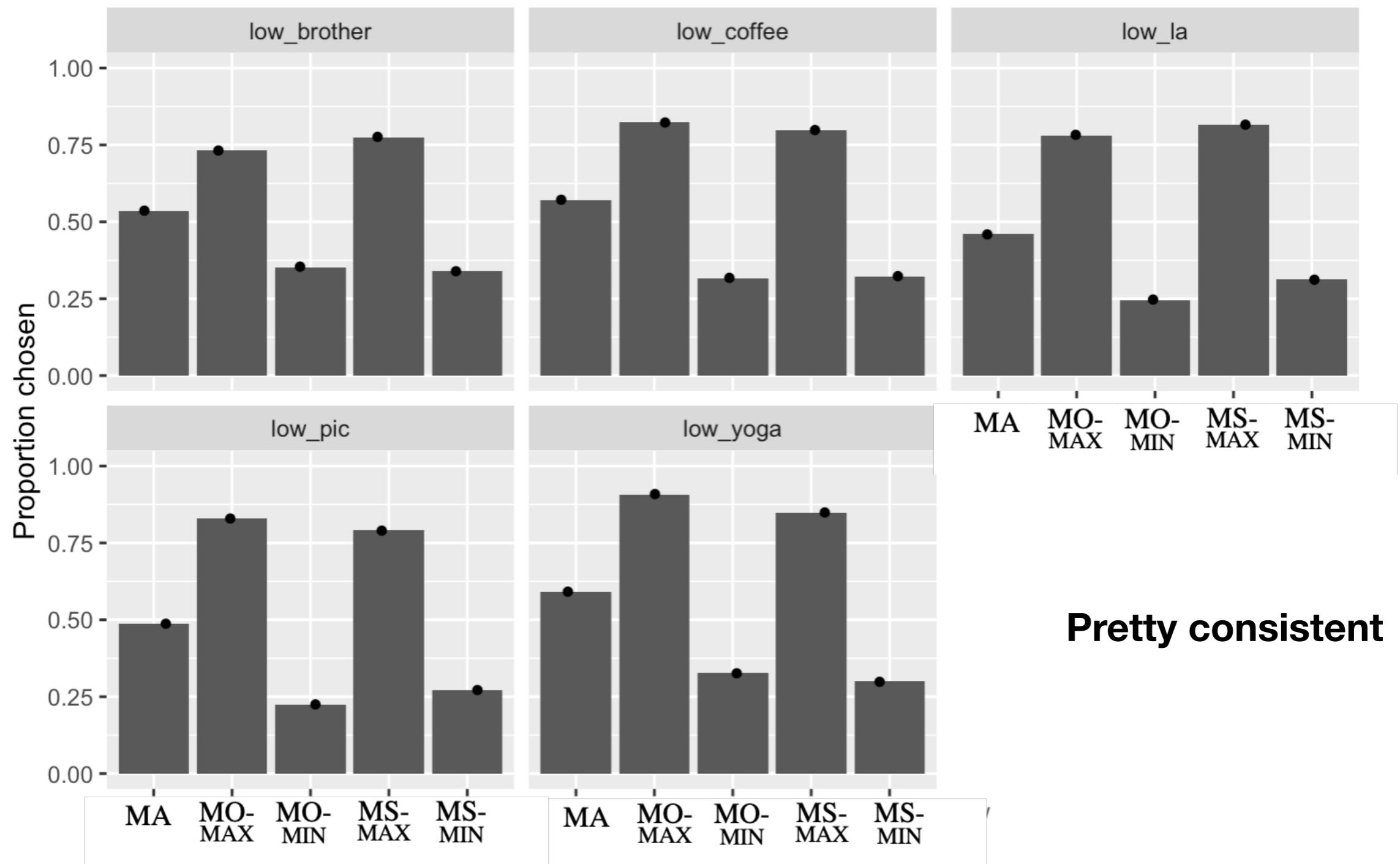


**Low stakes only**

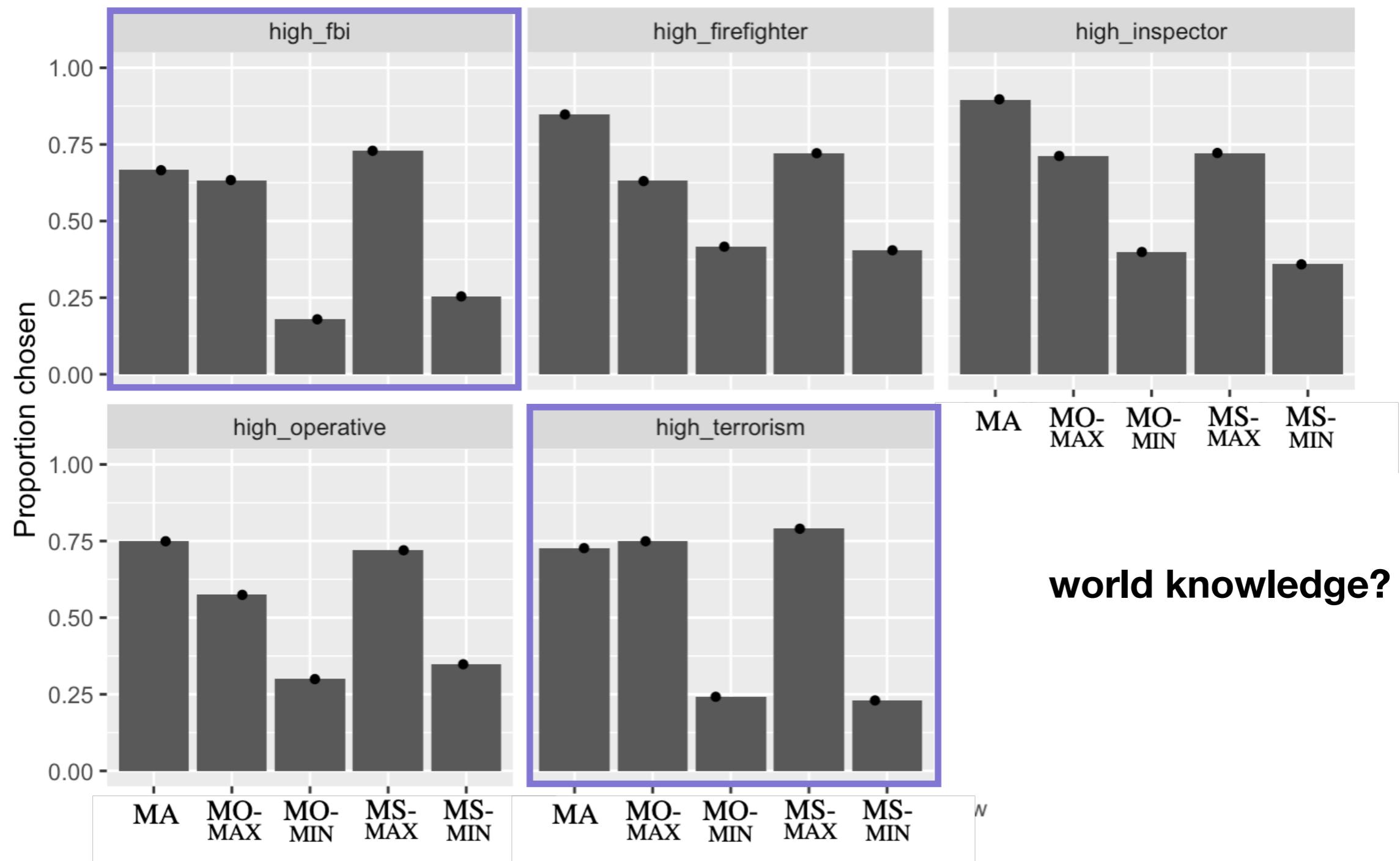
**MO-Max > MA**

**Informativity:**  
**Max > Min**

# Low stakes



# High stakes



# Summary

**Evidence for role of linguistic form (“semantics”):**  
Modal and verb (and even *wh*)

**Evidence for goal sensitivity (“pragmatics”):**  
“Mention-enough” determined by contextual stakes  
Tracked by Item effects?

# The current investigation

A systematic investigation of **naturally-occurring** questions

- (Q1)** Is there an MA bias?
- (Q2)** Is interpretation modulated by linguistic form?
- (Q3)** How much does discourse context contribute to interpretation?

Step 1: stimuli from corpus database  
Step 2: crowd-source annotation for meaning



# A Corpus Database

- Extracted 10,199 occurrences of *wh*-phrases from Switchboard corpus (Godfrey et al 1992) using TGrep2 (Rohde 2005) and TGrep2 Database tools (Degen & Jaeger 2011)
- Coded for syntax (e.g., root, embedded,...), *wh*-word, presence/absence of modality, matrix verb

QuestionType	Example	% of Total
Root	<i>What are you studying?</i>	16.9%
Embedded	<i>I don't know how the Peace Corps works.</i>	16.6%
Relative	<i>I get a charge out of Texas highway, which is uh, very colorful.</i>	13.5%
Clause		
Cleft	<i>That's what they say.</i>	6.4%
Other (Pseudo- cleft, Adjunct)	<i>Where I work is predominantly male. I'm not even sure if we had one when I was little.</i>	45%
Fragment/ Exclamation	<i>I think when babies</i>	1.6%

# Naturalistic stimuli for a paraphrase rating task

- Only Root and Embedded Qs
- Excluded Qs where MS/MA collapse:
  - Degree Qs: (*How much sugar do you need?*)
  - Identity Qs: (*Who is Mohammad Ali?*)

995 root questions

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

1075 embedded questions

Wh-word	Modal present	Modal absent
<i>What</i>	8%	38.2%
<i>How</i>	13%	14.9%
<i>Where</i>	1.7%	6.9%
<i>Why</i>	1.8%	7%
<i>Who</i>	0.5%	5.5%
<i>When</i>	0.48%	1.9%

# Paraphrase Task: Root Questions

- Native English speakers on Prolific (n=660)
- 19 removed for performance on controls
- Paid ~\$14/hr for ~15 mins

10 preceding lines of discourse

Target question

MA

MS

MS=MA

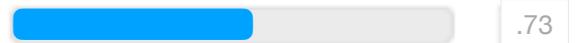
In theory these could have multiple answers, but they seem to **presuppose** a unique answer

Dayal, 1991; Fox 2014, 2018

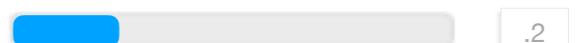
Speaker #2: pretty good.  
Speaker #1: i do like to ski.  
Speaker #2: pretty, pretty down there. huh?  
Speaker #1: yeah, i , i said i do like to ski.  
Speaker #2: so, where have you skied?

*Based on the sentence in red, how likely do you think it is that the speaker wanted to know about each of the following?*

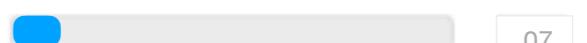
What is every place...?



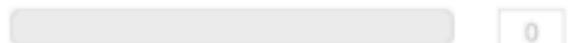
What is a place...?



What is the place...?



Something else



Continue

# Paraphrase Task: Embedded Questions

- English speakers on Prolific (n=1080)
- 25 removed for reporting non-native status
- 36 removed for performance on controls
- Paid ~\$14/hr for ~15 mins

**Speaker #2:**

**Speaker #1:** um.

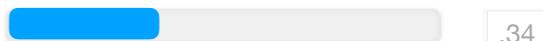
**Speaker #2:** so. well, uh, did you hear about that killeen massacre or whatever?

**Speaker #1:** yeah, the, did it happen at a cafeteria or something?

**Speaker #2:** yeah, right. that kind of i mean it just makes **you wonder how people get guns**

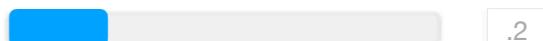
*Based on the sentence in red, how likely do you think it is that the speaker means each of the following?*

...what is the way...



.34

...what is every way...



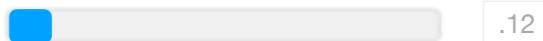
.2

...what is a way...



.32

Something else



.12

Continue

**every  
paraphrases  
(MA)**

- Nobody can predict what's going to happen in 20 years. (.69)
- He knew who worked there (.79)
- I have got to be so much more careful with what I do (.77)
- [i read the newspaper]...to see what's going on in our little community (.73)

- What does it have in it? (.86)
- Where have you skied? (.73)
- Who do you play with? (.54)

- When would people do it? (.74)
- What have you seen lately? (.66)
- Where do you like to eat? (.53)

**the  
paraphrases  
(MS = MA)**

- I'm not real sure why anybody would need a full automatic weapon (.61)
- I don't know how to make it better for them (.57)
- I don't know when I'm going to get them all (.49)

- Where do you live? (1)
- What school you going to? (1)
- How do you spell that? (.99)
- Who do you work for? (.95)
- When did it happen? (.97)
- Why are you cutting off the phone? (.82)

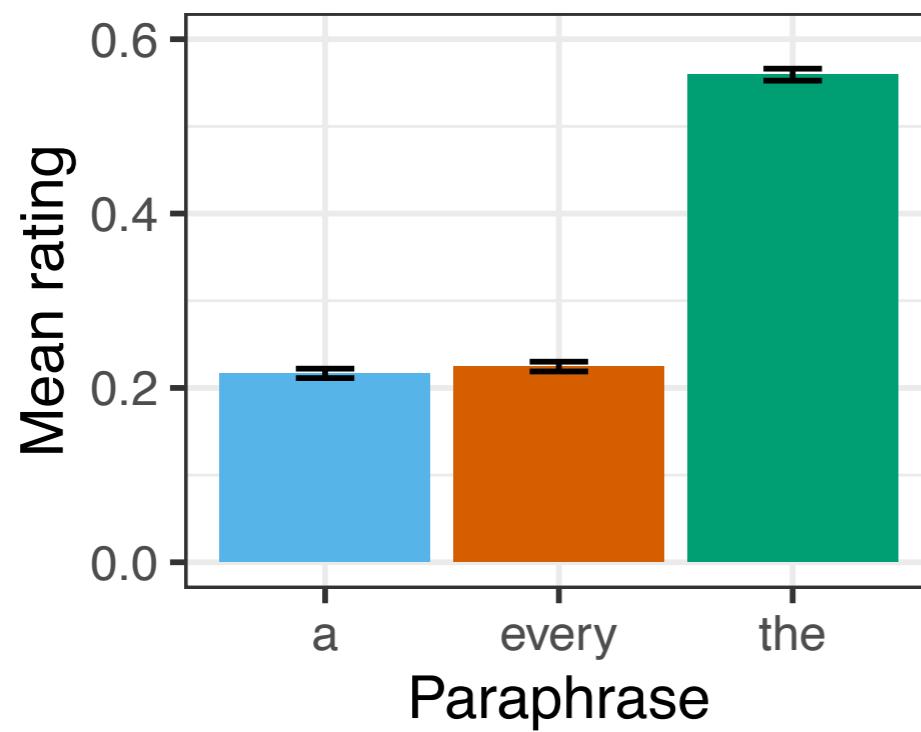
- I know who you're talking about (.94)
- Do you know who the guy was that was playing the wagon driver? (.95)
- My mother taught me how to make it (.93)
- I forget how she got it (.86)

**a  
paraphrases  
(MS)**

# Root vs. Embedded

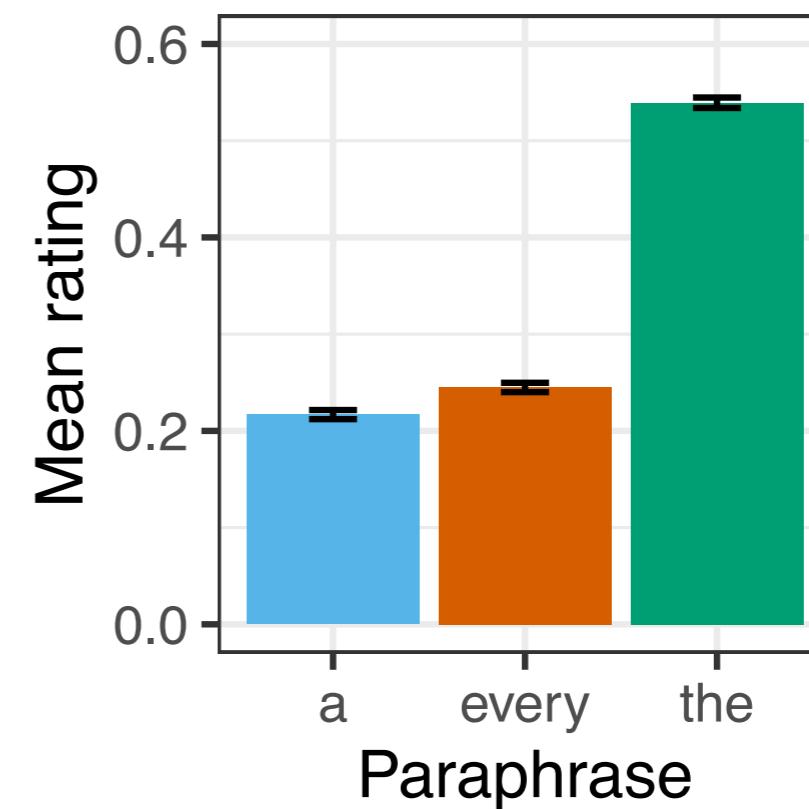
**Root Questions Task**

No significant difference  
between 'a' and 'every'

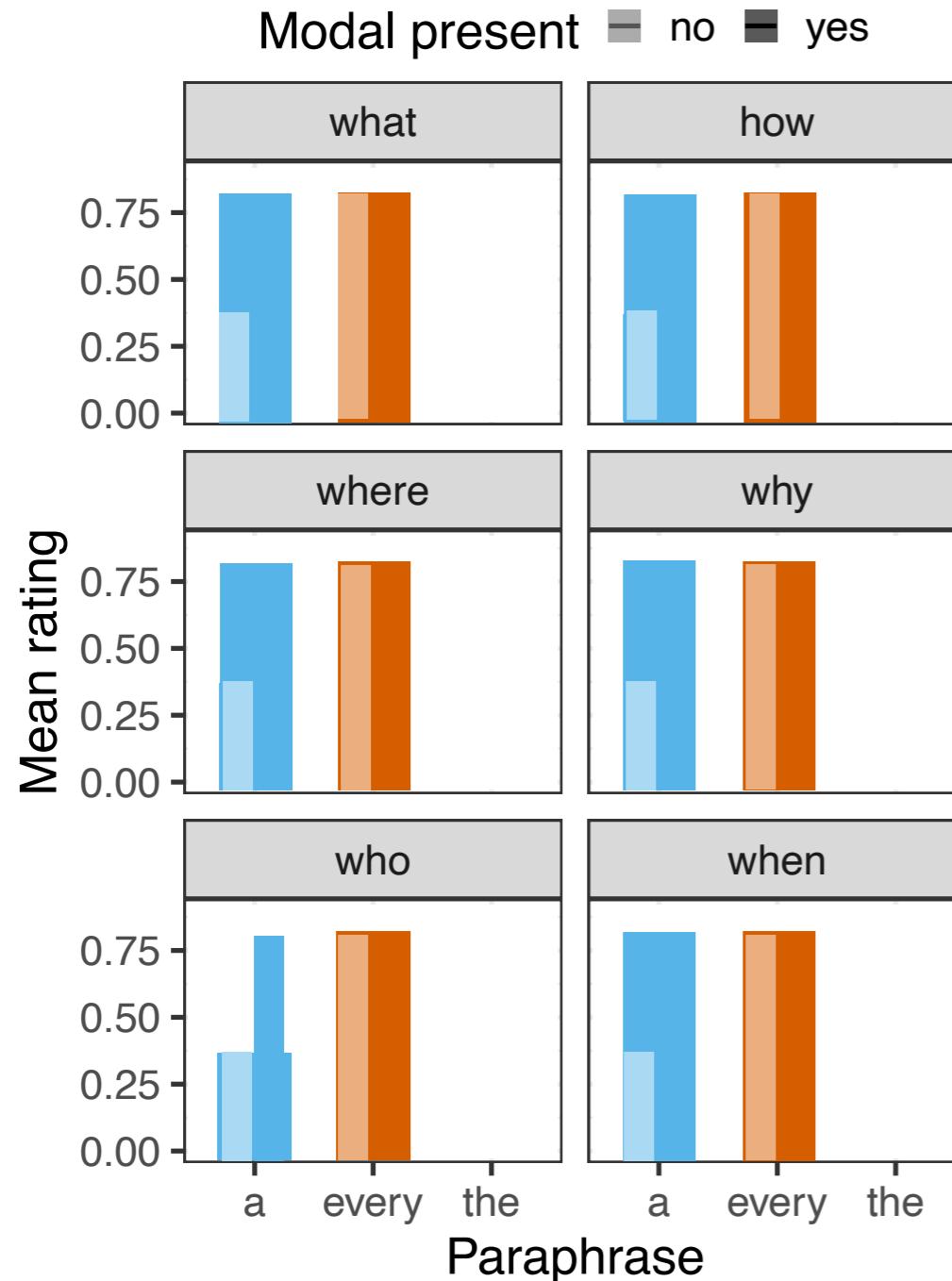


**Embedded Questions Task**

Significant initial bias for  
'every' over 'a'



# Predictions Modals x WH



Generally

Questions are MA-biased

MS is restricted...

Karttunen 1977, Groenendijk & Stokhof 1984, George 2011...

Wh-Word

Who-Qs are MA-biased  
Non-Who-Qs are MS biased

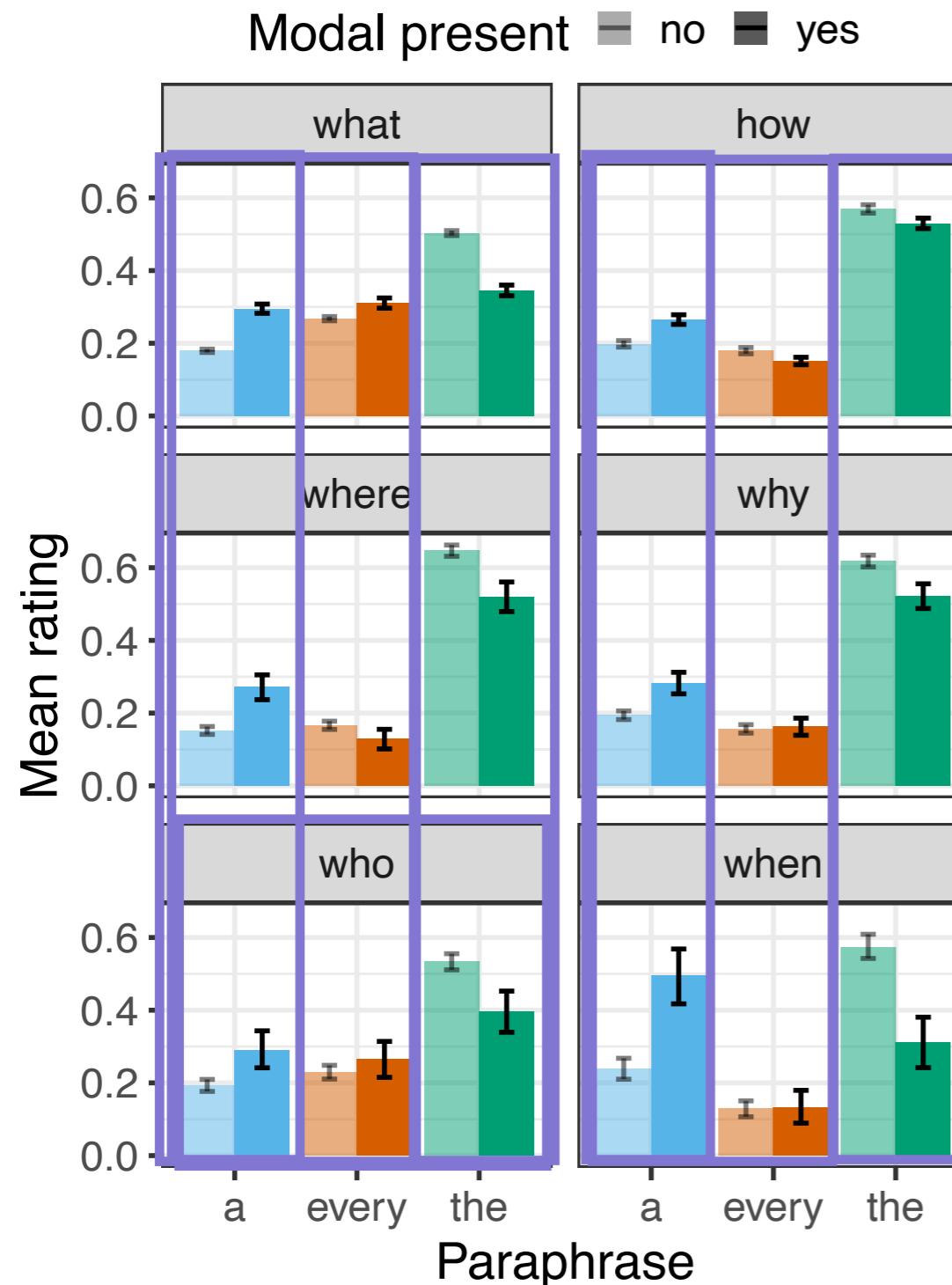
Ginzburg 1995; Asher & Lascarides 1998

Modal

Modal-Qs are MS-biased  
Non-Modal-Qs are MA-biased

George 2011; Fox 2014; Nicolae 2014; Dayal 2016; Xiang 2016

# Modals x Wh



**Generally**

No overall MA bias

**Wh-Word**

Who-Qs not MA-biased

**Modal**

Modal-Qs are MS-biased

# Embedded Questions: Matrix Verb Predictions

## Know-Wh is (strong) MA

Groenendijk & Stokhof 1982/1984;  
Berman 1991; Heim 1994, George 2011;  
Klindinst & Rothschild 2011; Schulz &  
Roeper 2011

**BUT!** Interactions between  
modality, wh-word (*who*, *where*),  
and matrix verb (*know* vs. *predict*)

Bhatt 1999, Moyer & Syrett (2019), Moyer (2020)

## Everything else??? Is MS/(weak) MA

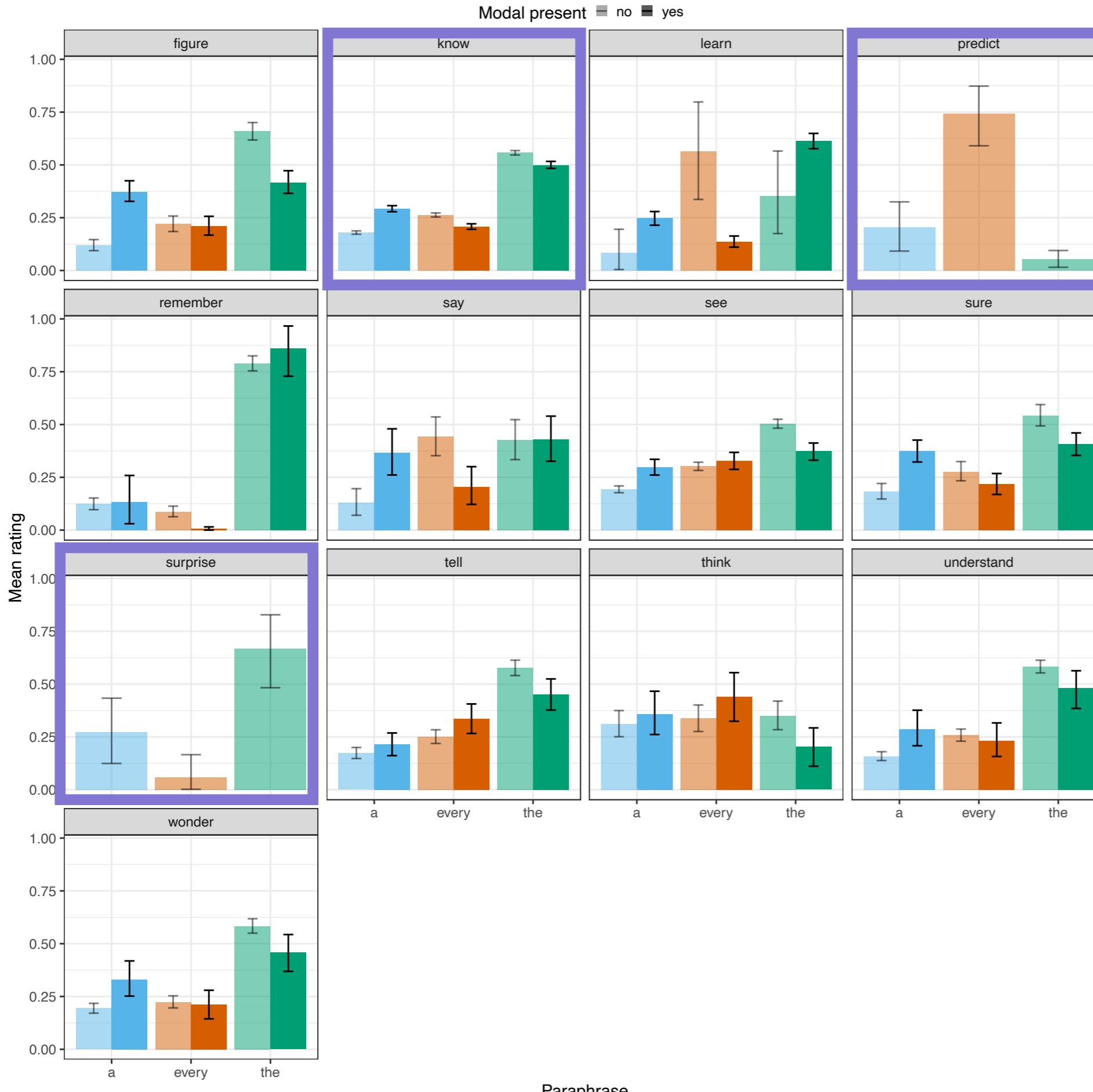
George 2011, Ch. 2; Heim 1994

*Scully knows where we can find coffee*  
*Scully knows who came to the party*

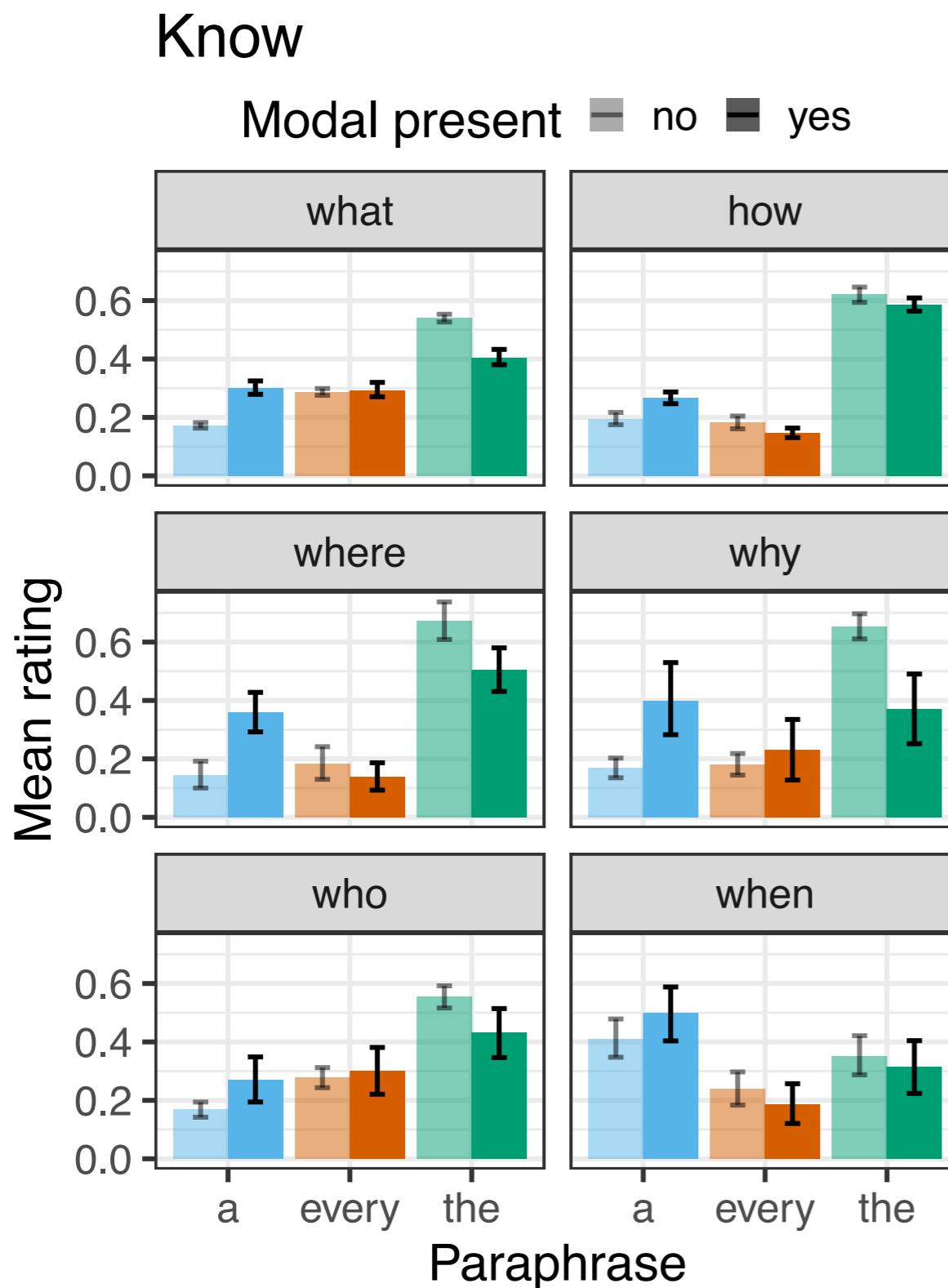
# matrix verbs

## Sparse data problem

- Lots of variability
  - role of lexical semantics?
- *Know-wh* isn't overwhelmingly MA-biased
- In contrast, *predict-wh* was? (only one item)
- And *surprise-wh* not? (1 or 2 items)



# know-wh



- No overwhelming preference for **MA** over **MS**
- Interactions with modal and *wh*-word

# Understanding the effect of context (in progress results)

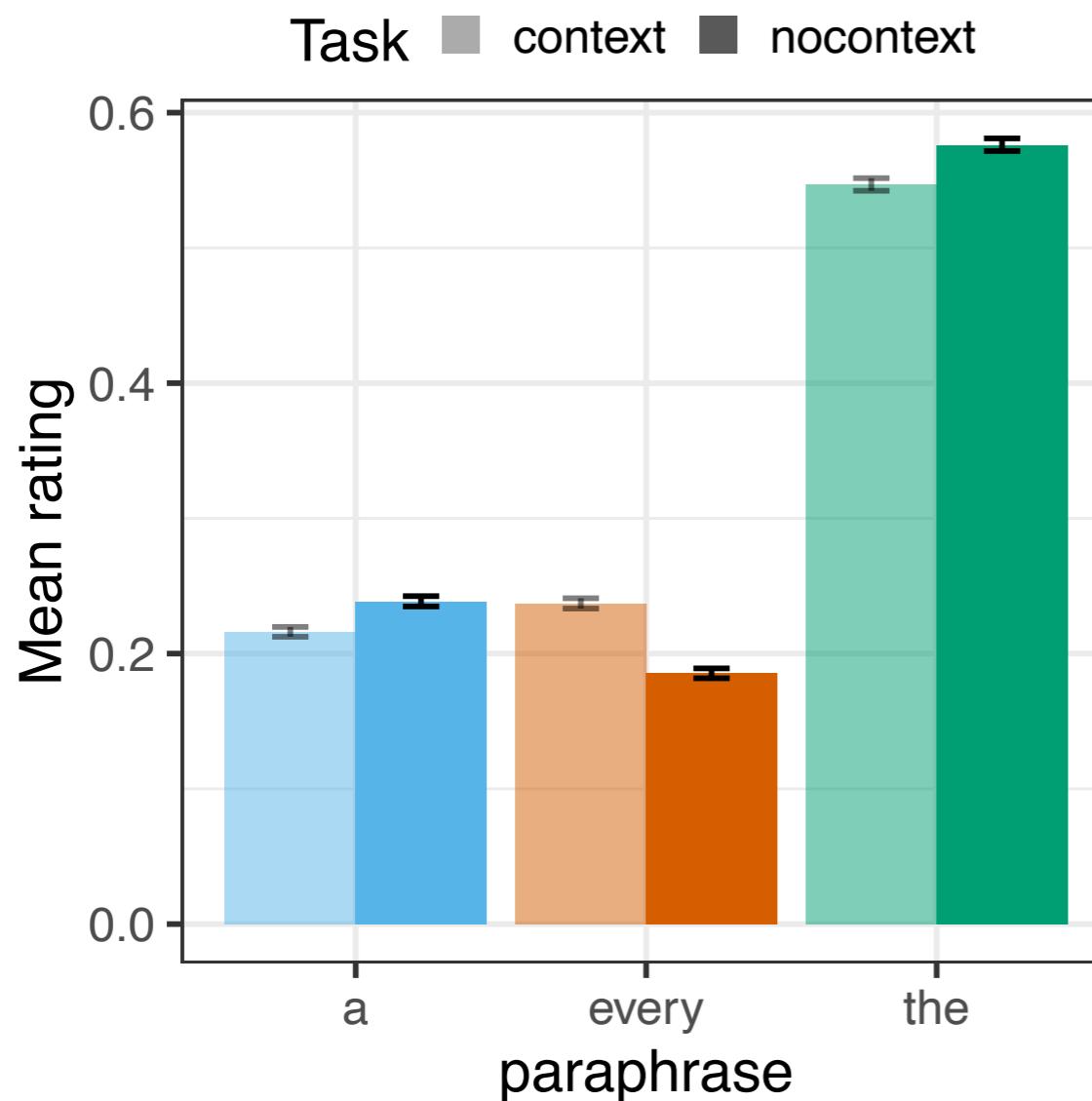
**Same study without 10 preceding lines of discourse.**

1. If MS needs crucial contextual licensing, then it should be less acceptable without that information.
2. If hearers are uncertain about the contextual goals, they may rely on the speaker's utterance more as a cue to the intended meaning.

# Understanding the effect of context (in progress results)

- 1. Does Context predict paraphrase rating?**  
—> Yes, but not in the expected way!
- 2. Is crucial information lost in the NoContext Task?**  
—> Yes, but not as much as you might expect.
- 3. Are linguistic factor effects stronger in NoContext?**  
—> No, potentially weaker
- 4. Does the context make a plausible domain restriction?**  
—> No real explicit linguistic information available, except exceptionally

# Task x Paraphrase



Significant interaction between Task and Paraphrase

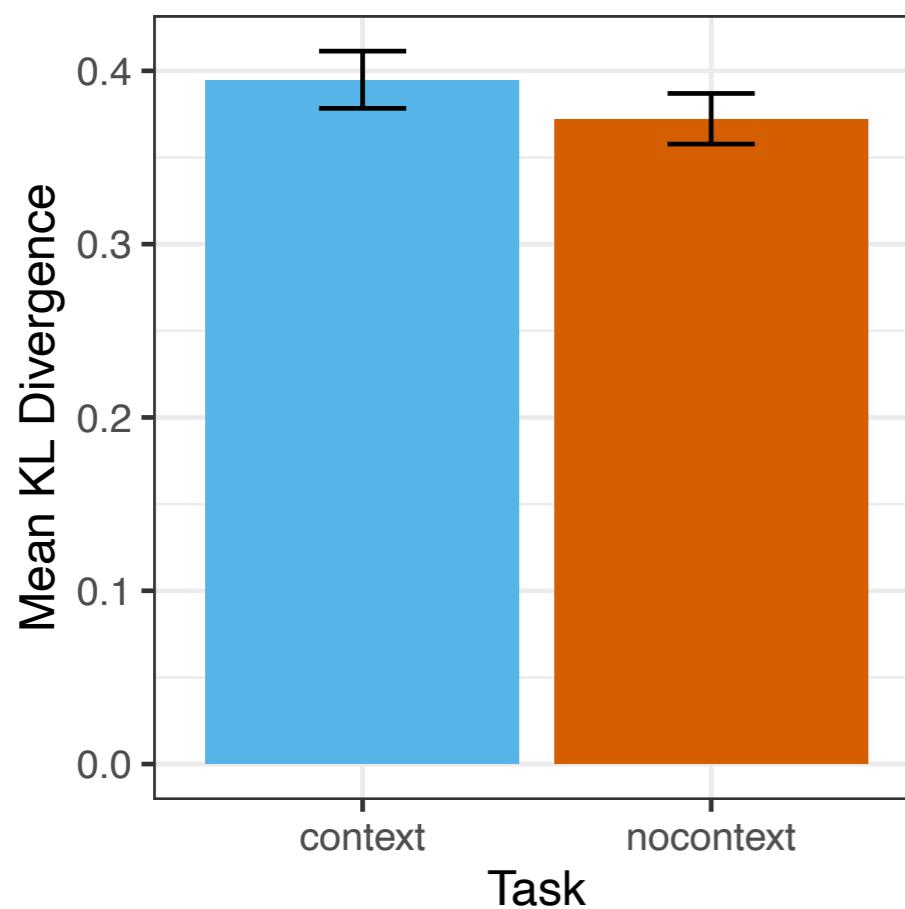
NoContext task lead to a significant increase in 'a' ratings compared to 'every' ratings

'A' ratings not more acceptable with context!

# Kullback-Leibler Divergence

KLD is a way to determine the difference between a probability distribution and a reference distribution.

Reference = uniform distribution (the uncertain state)



- Task predicts KLD
- Going from NoContext to Context significantly increases the KLD
- Participants in the NoContext task were on average less certain about the intended meanings

# Annotating contexts (In progress)

Some people have suggested that MS is just MA with domain restriction. (George 2011, Fox 2018, many pcs with different people)

If that's the case, it's not really clear there's anything explicit.

Often, restriction comes from world knowledge, e.g., "What do you want for lunch?"

# Informative examples

**speaker A:** uh-huh.

**speaker B:** we wore shorts.we didn't know 0 that was a big no, no.

**speaker A:** oh no.

**speaker B:** and the supervisor from another area came up to us and, you know, this area where record retention is \*t\*-1, there's only a few people that \*t\*-2 work there, and he, sho-, he knew who \*t\*-1 worked there,

**Context:** [a: .09, every: .83, the: .08]

**NoContext:** [a: .35, every: .35, the: .29]

# Discussion

- **The naturalistic distribution of meanings is different than expected**—most intuitions reported in the literature not supported
- There are a host of fine-grained linguistic factors that affect MS/MA
  - Many others we didn't discuss...definite/indefinite descriptions
- Keep the semantics simple (Asher & Lascarides 1998) because you can't build everything in to the semantics

# Conclusion

- Rather than there being clear default interpretations for wh-questions, these results point towards accounts of meaning that assign an important role to **hearers reasoning about the speaker's goal in the context** (Degen & Tanenhaus 2019, Moyer & Syrett 2019, Moyer 2020)
- The results suggest that building theories of meaning based on the same small dataset of intuitions handed down through generations may be misleading



# Thank you!

Alex Lascarides, Paul Pietroski, Veneeta Dayal, Yimei Xiang

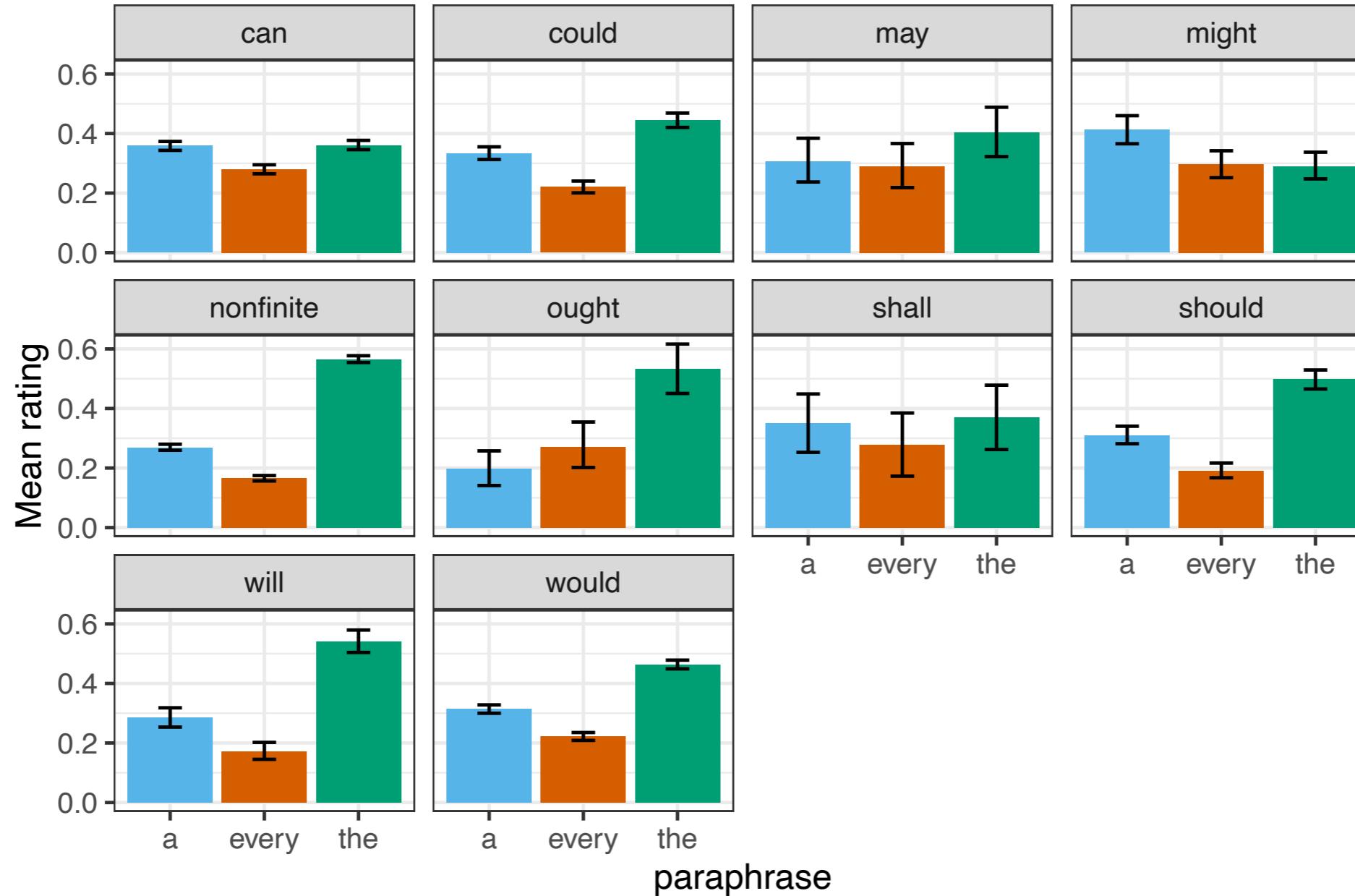
Leyla Kursat, Vera Gor, Taylor Martinez, Rangaraj Tirumala, Divya Appasamy, Knyck Sutherland

Stanford ALPS lab, Stanford Cog Sci Seminar, Audiences at ELM 1, CUNY 2021, CogSci 2021, CAMP 2021, LLF, Rutgers Lab for Developmental Language Studies

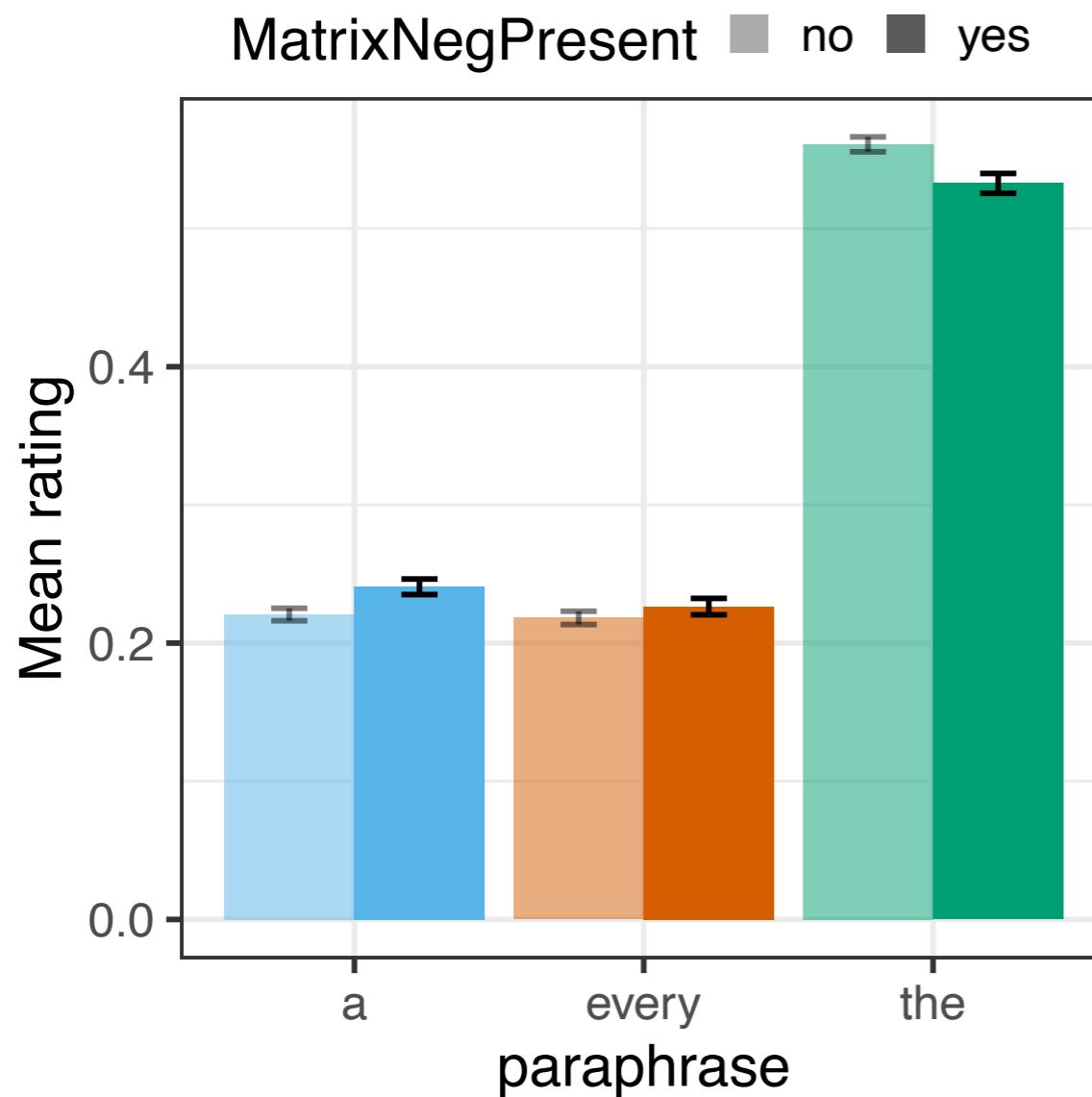
Funding from NSF-SBE-SPRF Grant #2005042 to Moyer/Degen, Stanford HAI Hoffman-Yee Grant “Toward Grounded, Adaptive Communicative Agents”, NSF BCS-DDRIG #1918068 to Syrett/Moyer, the Rutgers Center for Cognitive Science



# Only existential modals?

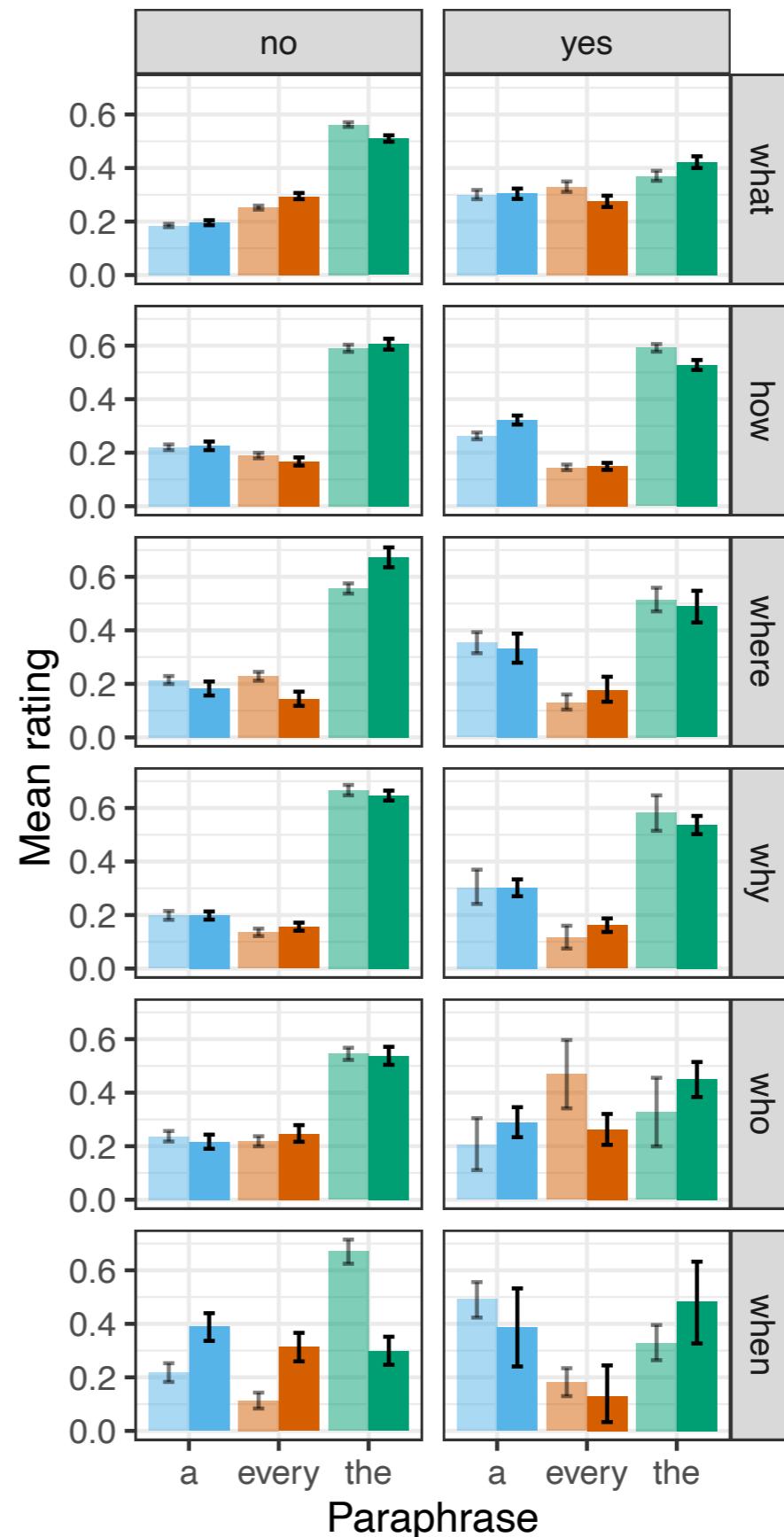


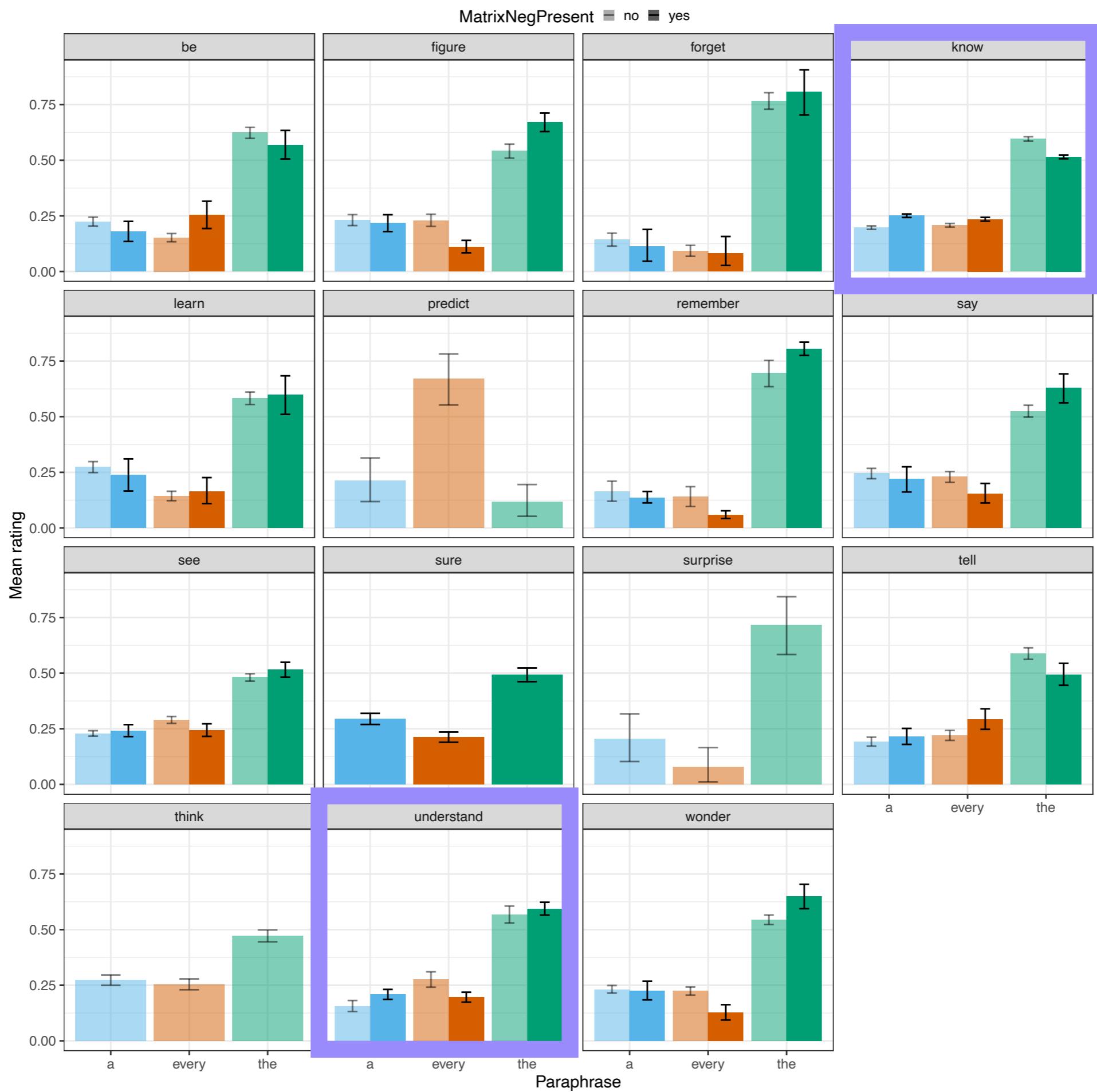
# negation?



# negation?

MatrixNegPresent no yes





# Methodological discussion

## A: What is the meaning of *wh*-questions?

linguistic theory,  
theoretical semantics

- Independent of speakers, listeners, maybe even context?
- “semantics” vs. “pragmatics”

- Introspective judgement experiments on expert participants
- Small number of constrained examples, passed down

- generalizations that don’t scale
- generalizations that are overlooked

Degen & Goodman 2014, Degen 2015; Tonhauser, Bever, Degen 2018; Degen & Tonhauser (submitted); Tonhauser & Degen (under review), amongst others.

## B: How do people interpret *wh*-questions?

x prag, psycholinguistics,  
cognitive science

- Acknowledges/embraces speakers, listeners, contexts
- S vs. P: not really an interesting question



# Example questions

Wh		Paraphrase	Example	Mean Rating
When	<b>the</b>	<i>Oh, when did it happen?</i>		.97
		<i>When do you get out?</i>		.9
	<b>a</b>	<i>When would people do it?</i>		.74
		<i>When was the last time you took your truck out?</i>		.39
	<b>every</b>	<i>When does it bloom?</i>		.16
		(Actually highest for <i>the</i> )		
How	<b>the</b>	<i>How do you spell that?</i>		.99
		<i>How'd you get into it?</i>		.93
	<b>a</b>	<i>How can I tell you?</i>		.66
		<i>How can it be bad?</i>		.54
	<b>every</b>	<i>How do you think we've changed in the last 10-20 years?</i>		.54
Where	<b>the</b>	<i>Where do you live?</i>		1
		<i>Where'd you get the car?</i>		.99
	<b>a</b>	<i>Where do you like to eat?</i>		.53
		<i>Where do you go for steaks?</i>		.42
	<b>every</b>	<i>Where have you skied?</i>		.73
		<i>Where have you been?</i>		.62

Wh		Paraphrase	Example	Mean Rating
Who	<b>the</b>	<i>Who's the last one?</i>		.99
		<i>Who do you work for?</i>		.95
	<b>a</b>	<i>Who would steal a newspaper?</i>		.73
		<i>Who wants to clean their junk before they throw it away?</i>		.64
	<b>every</b>	<i>Who do you play with?</i>		.54
		<i>Who's not very well educated?</i>		.39
Why	<b>the</b>	<i>Why are you cutting off the phone?</i>		.82
		<i>Why am I waiting, you know?</i>		.81
	<b>a</b>	<i>Why bother working?</i>		.39
		<i>Why get rid of him now?</i>		.39
	<b>every</b>	<i>Why has crime increased?</i>		.54
		<i>Why did you choose a private school?</i>		.37
What	<b>the</b>	<i>What school you going to?</i>		1
		<i>What do you drive now?</i>		.98
	<b>a</b>	<i>What else can we talk about?</i>		.91
		<i>What have you seen lately?</i>		.66
	<b>every</b>	<i>What cities are they looking at?</i>		.92
		<i>What does it have in it?</i>		.86