Reasoning about the unsaid:
psycholinguistic investigations into pragmatic meaning

ESZTER RONAI (THE UNIVERSITY OF CHICAGO)

CITY UNIVERSITY OF HONG KONG, LT RESEARCH FORUM

Introduction

Integrated approach to study of language:

combine insights of **theory** with experimental **data**

Primary research:
experimental psycholinguistic investigation of **pragmatic meaning**

Secondary research:
syntax and sentence processing

Research questions

In language, meaning sometimes arises **without any explicit form.**
How does this happen? How are meanings with and without form distinguished in the mind?

Two domains I focus on:

Syntax: Ellipsis, long-distance dependencies

In the syntactic structure, words and phrases can be moved or **dropped entirely**, but their **meanings are still present**

Pragmatic meaning

In communication, **we reason about what is not said**, drawing inferences that were not stated explicitly

Syntax & syntactic processing

Cross linguistic investigation, with special focus on Hungarian

Ellipsis *(with Laura Stigliano)*

What's inside the ellipsis site

Experimentally testing interpretations

Interaction of syntax and morphology

2020, *Nat Language & Ling Theory*
2021, *Snippets*
2021, *Proc of WCCFL*
2021, *Proc of LSA*
in prep, *The Linguistic Review*

Relative Clauses

Working memory vs. frequency effects in processing difficulty

Ronai & Xiang in revision, *Cog Science*

Experimental pragmatics & semantics

Processing cost of pragmatic meaning

Ronai & Xiang 2021, *Journal of Linguistics*
Ronai & Xiang 2019, *Proc of NELS*

Integration of pragmatic cues with other sources of information

Ronai et al. 2019, Laboratory Phonology

Quantifier scope in heritage bilinguals

Ronai 2018, Proc of NELS

Introduction

Integrated approach to study of language:

combine insights of **theory** with experimental **data**

Primary research:
experimental psycholinguistic investigation of **pragmatic meaning**

Secondary research:
syntax and sentence processing

Introduction

Integrated approach to study of language:

combine insights of **theory** with experimental **data**

Primary research:
experimental psycholinguistic investigation of **pragmatic meaning**

Secondary research:
syntax and sentence processing

Pragmatic meaning

We use language based on our knowledge of sound patterns, word meanings, and grammar

We **reason about speakers' language use** and draw inferences:
Why did she choose *that* word instead of another?

In doing so, we recognize **meanings** that go **beyond the literal content** of what was said

Pragmatic meaning

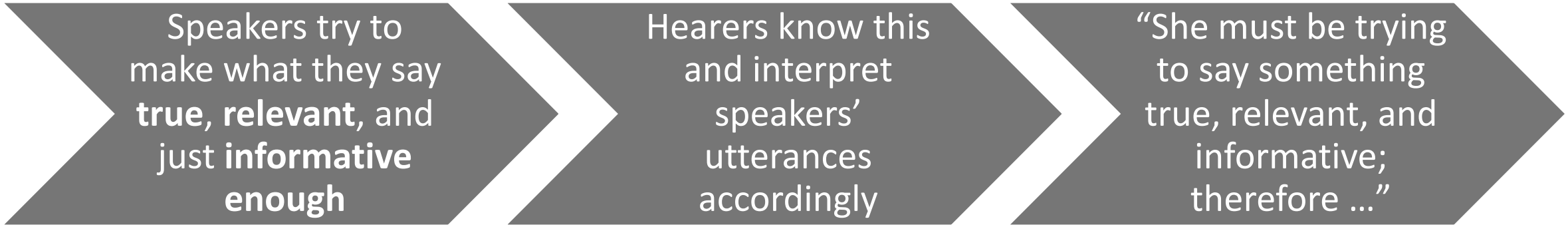
Can you pass me the salt?

Literal content: Do you have the ability to pass me the salt?

Why did the speaker say this?

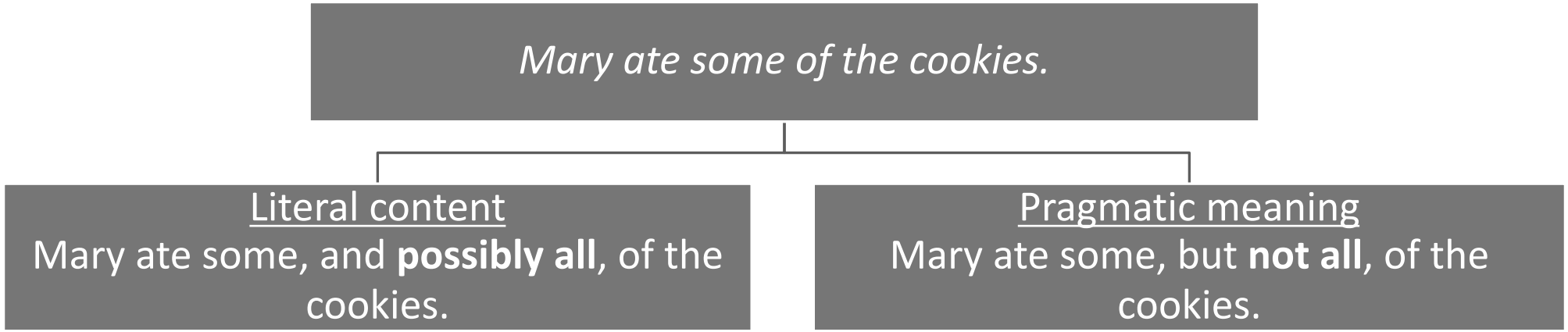
Pragmatic meaning: Please pass me the salt.

Pragmatic meaning



(Grice, 1967)

Scalar inference



Evidence: *Mary ate some of the cookies.* ***In fact, she ate all of them.***

If the literal content were *not all*, then this would be a contradiction!

How exactly does the pragmatic meaning arise here?

Scalar inference and alternatives

Inferential process:
Hearers reason about **alternative utterances** the speaker could have said...
....to recover the **intended meaning**

Scalar inference: *some*, *all* form a scale
all is logically **stronger** (**more informative**) than *some*

The speaker said *some* → If she thought *all* was true, she would have said it → But she chose not to say *all* → So, she must believe *all* is false

Scalar inference and context

Whether you make this inference depends partially on **the discourse context**:

A: *Did Mary eat any of the cookies?*
B: *She ate some of the cookies.*

In the context of A's question, all that matters is that she ate **at least some**

She ate all of the cookies is no longer a relevant alternative and **the scalar inference might not arise**

The role of alternatives and context

Alternatives and context are clearly relevant to how scalar inference works, but their respective **contributions are not well-understood**

I test their role in two open questions

Scalar inference: open questions

1. Inferential process: Is there empirical evidence for **alternative generation in real-time processing**?

   Theory gives us abstract characterization of the inferential process

   But what is the **underlying psycholinguistic mechanism**?

   It is possible that lexical alternatives (e.g., *all*) play no role

Scalar inference: open questions

2. **Scalar diversity**: variation in inference likelihood across different instances of scalar inference

   The movie is good.
   Scalar inference: *… but not excellent.*

   Theory assumes **uniformity** across scales

   What can explain the diversity?

This talk

crossroads of formal linguistics and psycholinguistics

Part I. Inferential process

Psycholinguistic evidence for lexical alternatives

Part II. Scalar diversity

Alternatives and context both contribute to variation

This talk

crossroads of formal linguistics and psycholinguistics

Part I. Inferential process

Psycholinguistic evidence for lexical alternatives

Part II. Scalar diversity

Alternatives and context both contribute to variation

Part I. Inferential process

Two theoretical approaches: **Neo-Gricean** vs. **Post-Gricean**

Neo-Gricean accounts (i.a. Horn, 1972; Katzir, 2007)

- Hearers infer the negation of unsaid, stronger alternatives
- Alternatives determined by the lexicon or grammar

Post-Gricean accounts (i.a. Sperber & Wilson, 1995)

- Scalar inference is a contextually driven, conceptual process
- Utterances "strengthened" by ad hoc concept construal
- Lexical scales play no special role

Semantic priming

Goal: Track the **retrieval and activation of alternatives** (like *all* and *excellent*)

Do we activate the meaning of *excellent* when we access *good?*

The movie is good.

Semantic priming

Goal: Track the **retrieval and activation of alternatives** (like *all* and *excellent*)

Do we activate the meaning of *excellent* when we access *good?*

Semantic priming with **lexical decision**

Sentential semantic priming

PRIME

TARGET

Item N=60
Participant N=46

- Task: decide whether *excellent* is a word or non-word
- Dependent measure: reaction time (RT)

Sentential semantic priming: conditions

Condition	Prime	Target
Related (scalar)	The movie is good.	excellent

Sentential semantic priming: conditions

Condition	Prime	Target
Related (scalar)	The movie is good.	excellent
Unrelated	The movie is foreign.	excellent

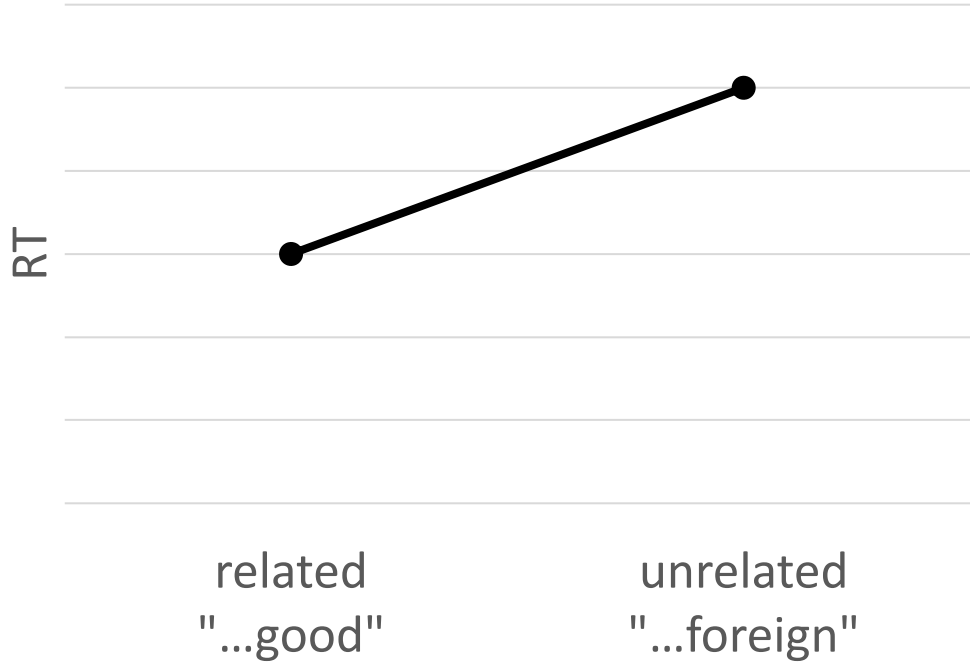
Sentential semantic priming: conditions

Condition	Prime	Target
Related (scalar)	The movie is good.	excellent
Unrelated	The movie is foreign.	excellent
Filler (non-word)	Susan decorated the cookies.	kleens

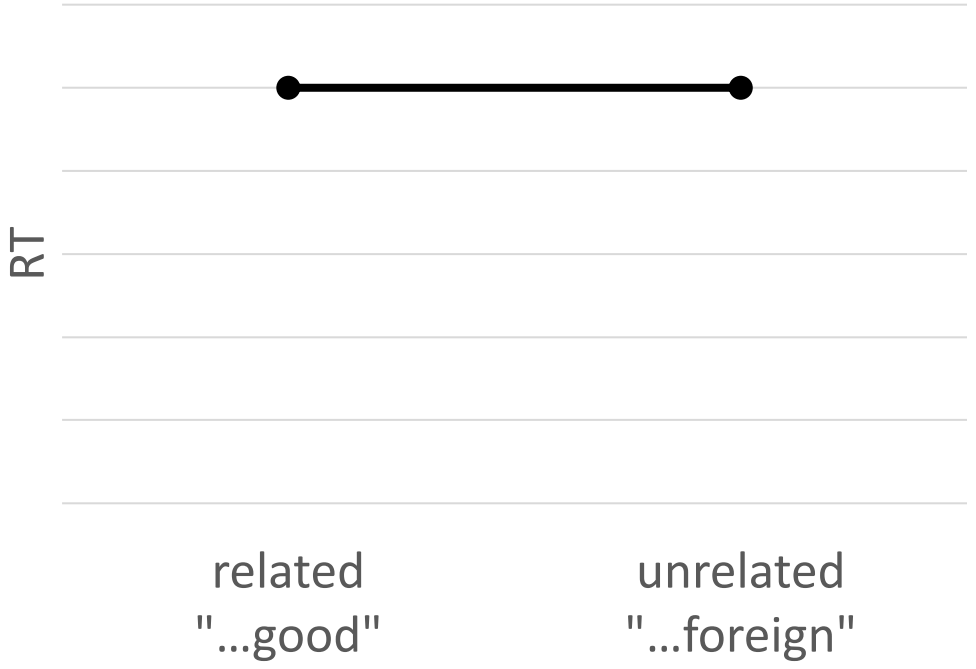
Predictions

target: “excellent”

(NEO-)GRICEANS:
LEXICAL SCALES PLAY A ROLE



POST-GRICEANS:
LEXICAL SCALES PLAY NO ROLE



Control: Lexical semantic priming

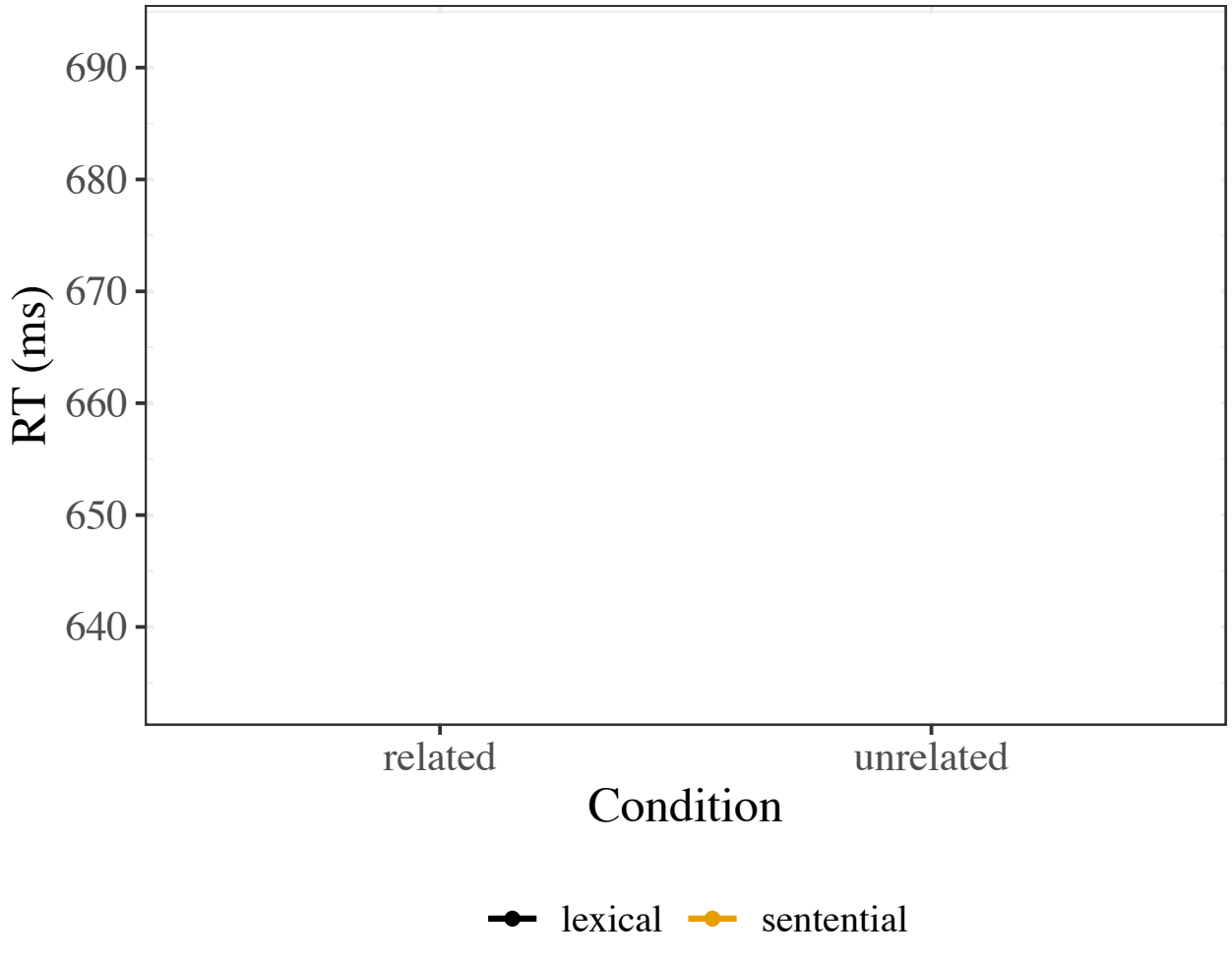
- What if the priming effect is not due to scalar inference?
- Non-sentential experiment to provide baseline for priming

PRIME

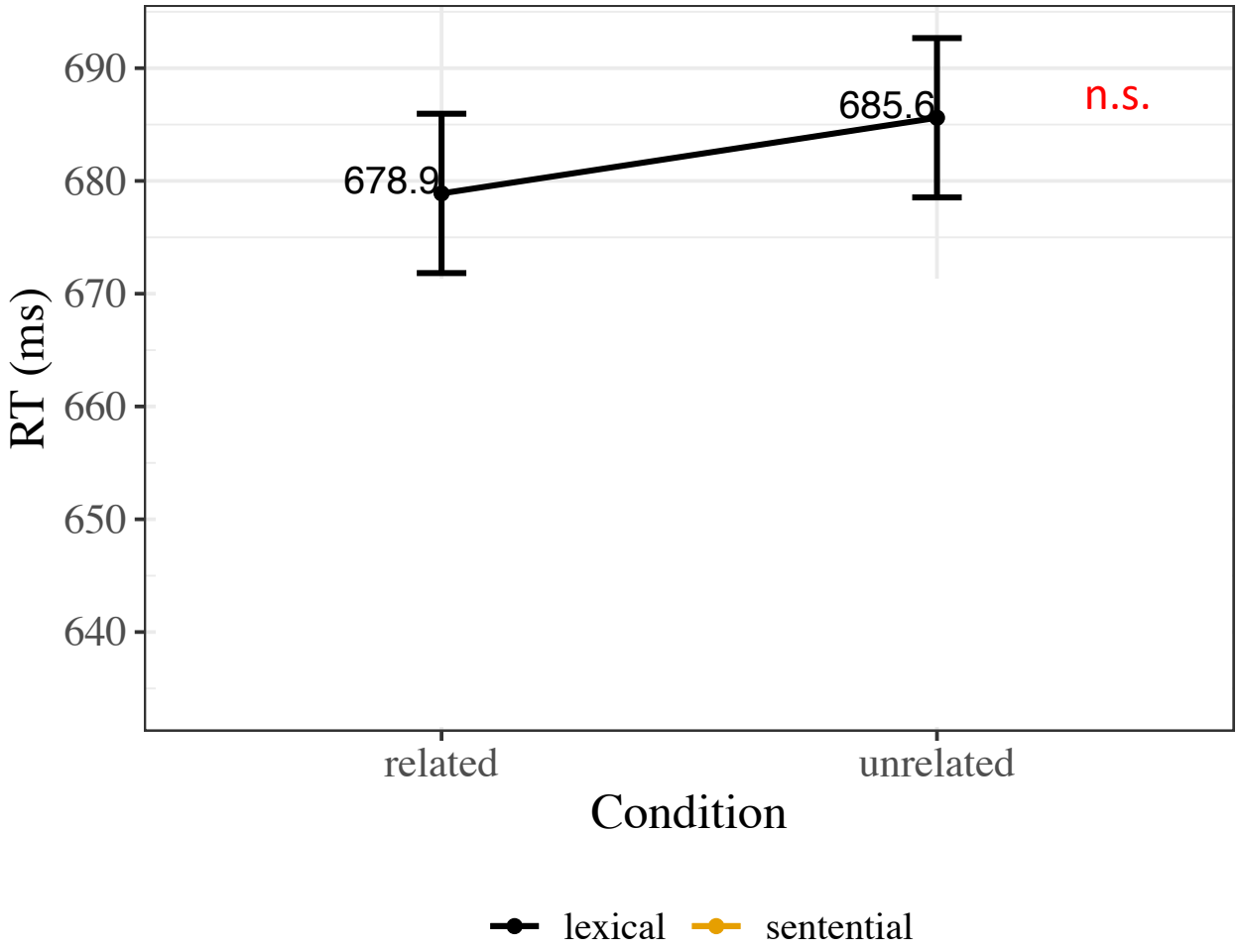
TARGET

Item N=60
Participant N=44

Results


690
680
670
660
650
640
RT (ms)
related
unrelated
Condition
lexical
sentential

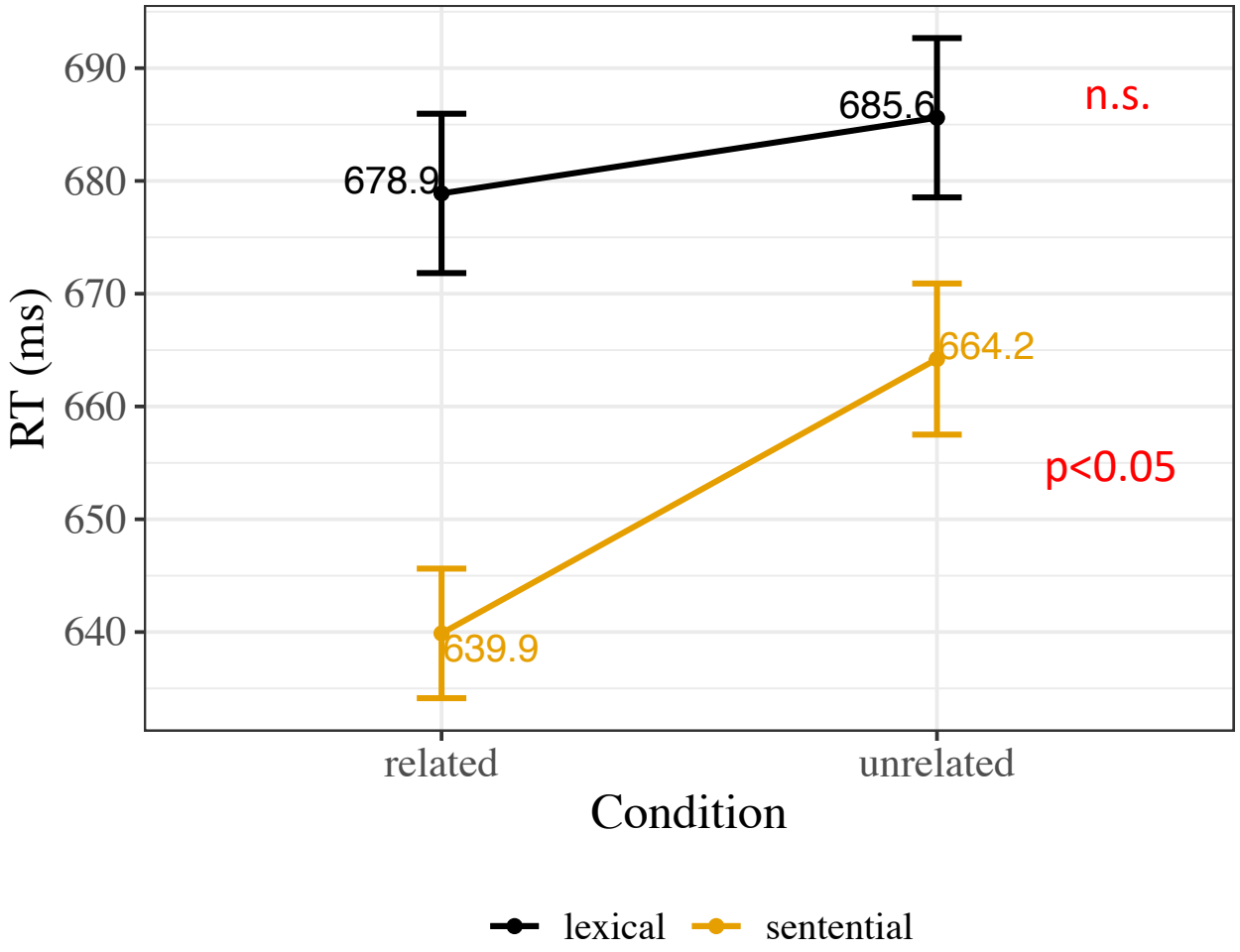
Results


690
680
670
660
650
640
RT (ms)
678.9
685.6
n.s.
related
unrelated
Condition
lexical
sentential

Results

- Facilitated RT to lexical alternatives in sentential experiment

- Predicted by Neo-Gricean, but not by Post-Gricean accounts



Inferential process: Conclusions

Lexical alternatives (*all*, *excellent*) **are** retrieved and activated in real-time processing of scalar inference-triggering sentences

Findings **support Neo-Gricean accounts** of scalar inference, in which hearers reason about particular lexical alternatives

This talk

crossroads of formal linguistics and psycholinguistics

Part I. Inferential process

Psycholinguistic evidence for lexical alternatives

Part II. Scalar diversity

Alternatives and context both contribute to variation

Part II. Scalar diversity

Formal accounts **implicitly assume uniformity** in the likelihood of scalar inference: the reasoning process should be the same no matter the scale

In reality: we find **robust variation** across lexical scales in calculation rates

(van Tiel et al., 2016)

My case study: 60 different lexical scales collected via corpus work (COCA)

Examples of different scales

(1) Mary ate **some** of the cookies. → Mary didn't eat **all** of the cookies.

(2) The movie is **good**. → The movie isn't **excellent**.

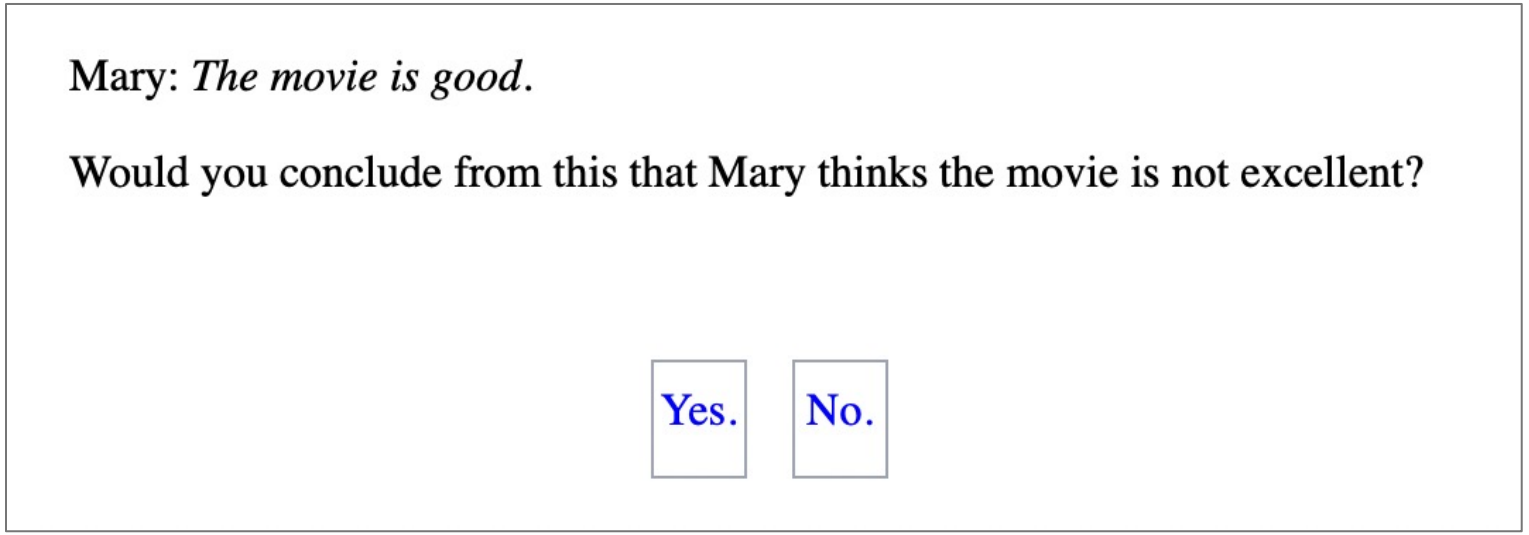
(3) The tank is **partially** full. → The tank isn't **completely** full.

(4) The worker is **tired**. → The worker isn't **exhausted**.

(5) The plant **survived**. → The plant didn't **thrive**.

Frequency not found to be a relevant factor

Replicating scalar diversity



Item N=60
Participant N=40

Inference task:

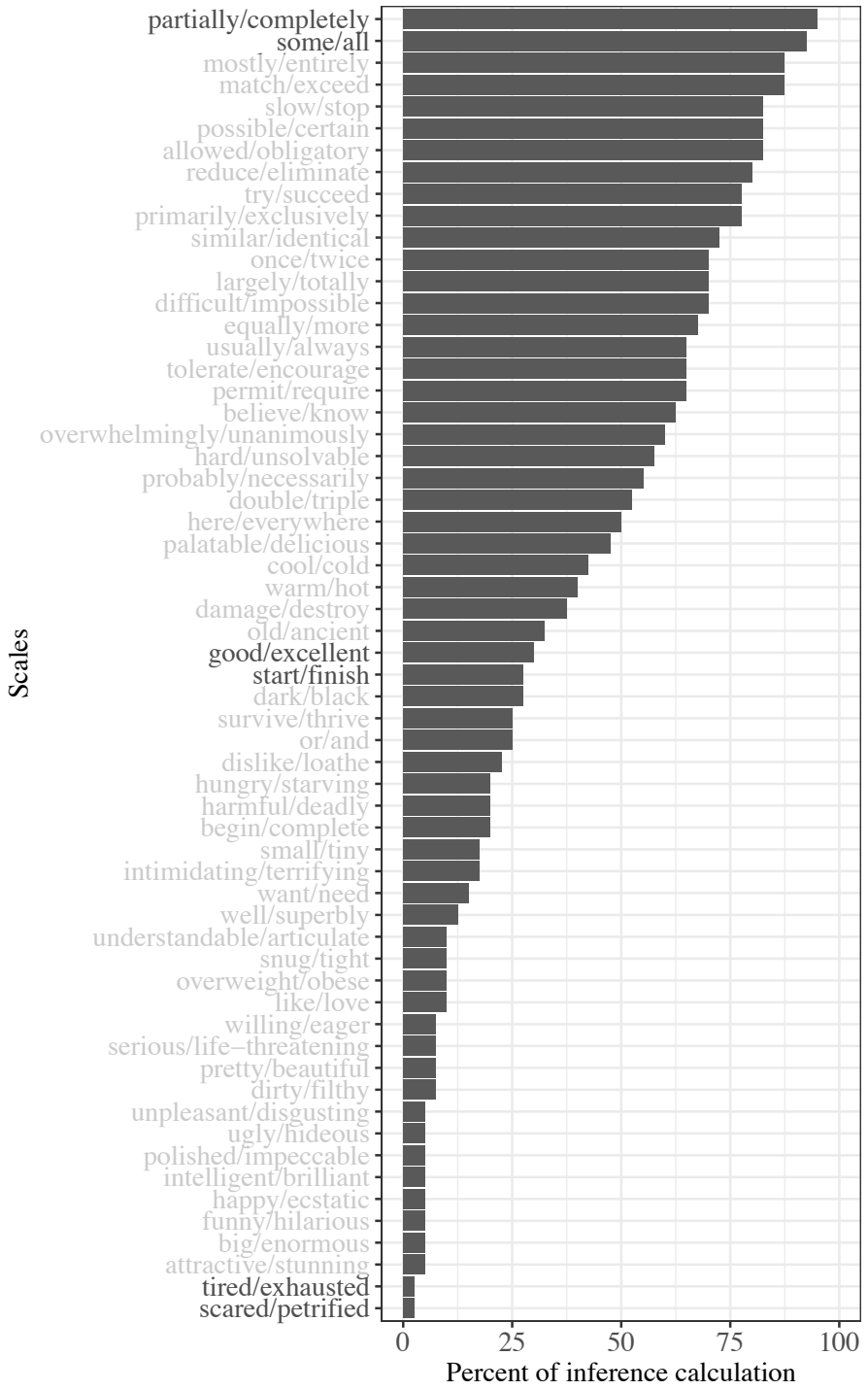
- "Yes" response = scalar inference was calculated
- "No" response = scalar inference was not calculated

Other tasks yield converging results

Scalar diversity

Scalar diversity

Robust variation



Explaining scalar diversity

Related to **alternative activation**?

Related to **context**?

Scalar diversity & alternatives

Hypothesis:
Scalar diversity can (in part) be explained by **how accessible a stronger alternative is**, given the weaker scalar

Causal mechanism

1) Scalar inference proceeds via reasoning about alternatives

2) Hearers **generate** a set of **alternatives**

3) The **more accessible** the alternative, the **more likely** hearers are to reason about it, and therefore **the more likely the scalar inference**

Ronai & Xiang 2021, Sinn und Bedeutung

Scalar diversity & alternatives

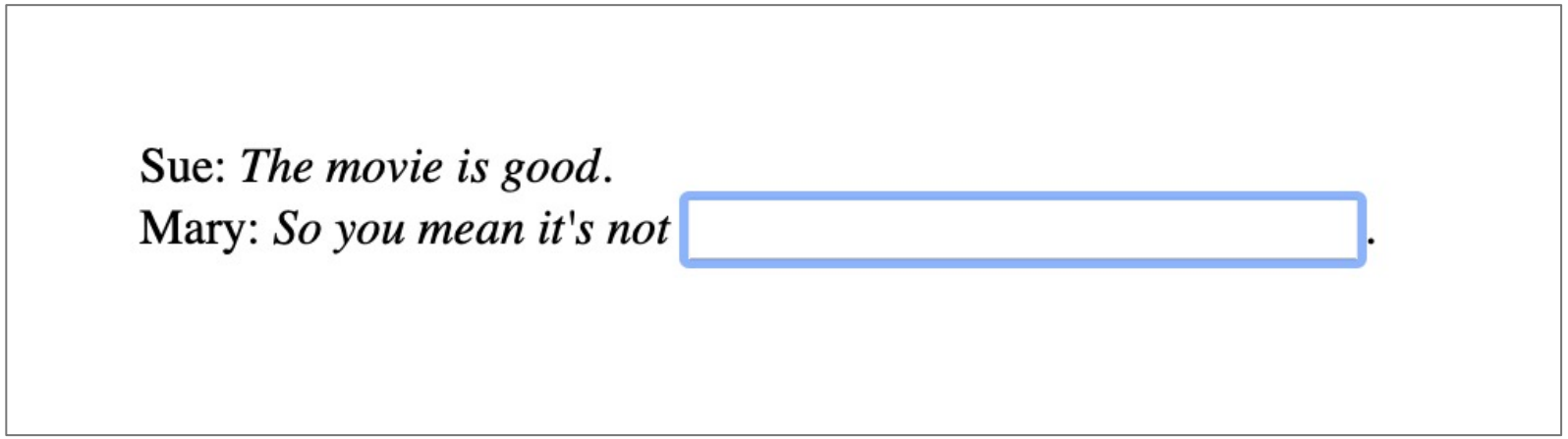
Intuition: Differences across scales in **how strongly the weaker scalar evokes** a stronger alternative

- *Some*: *All* always comes to mind

- *Good*: A number of competing alternatives may be activated:
  excellent, funny, thrilling, thought-provoking, …

Metric for alternative accessibility

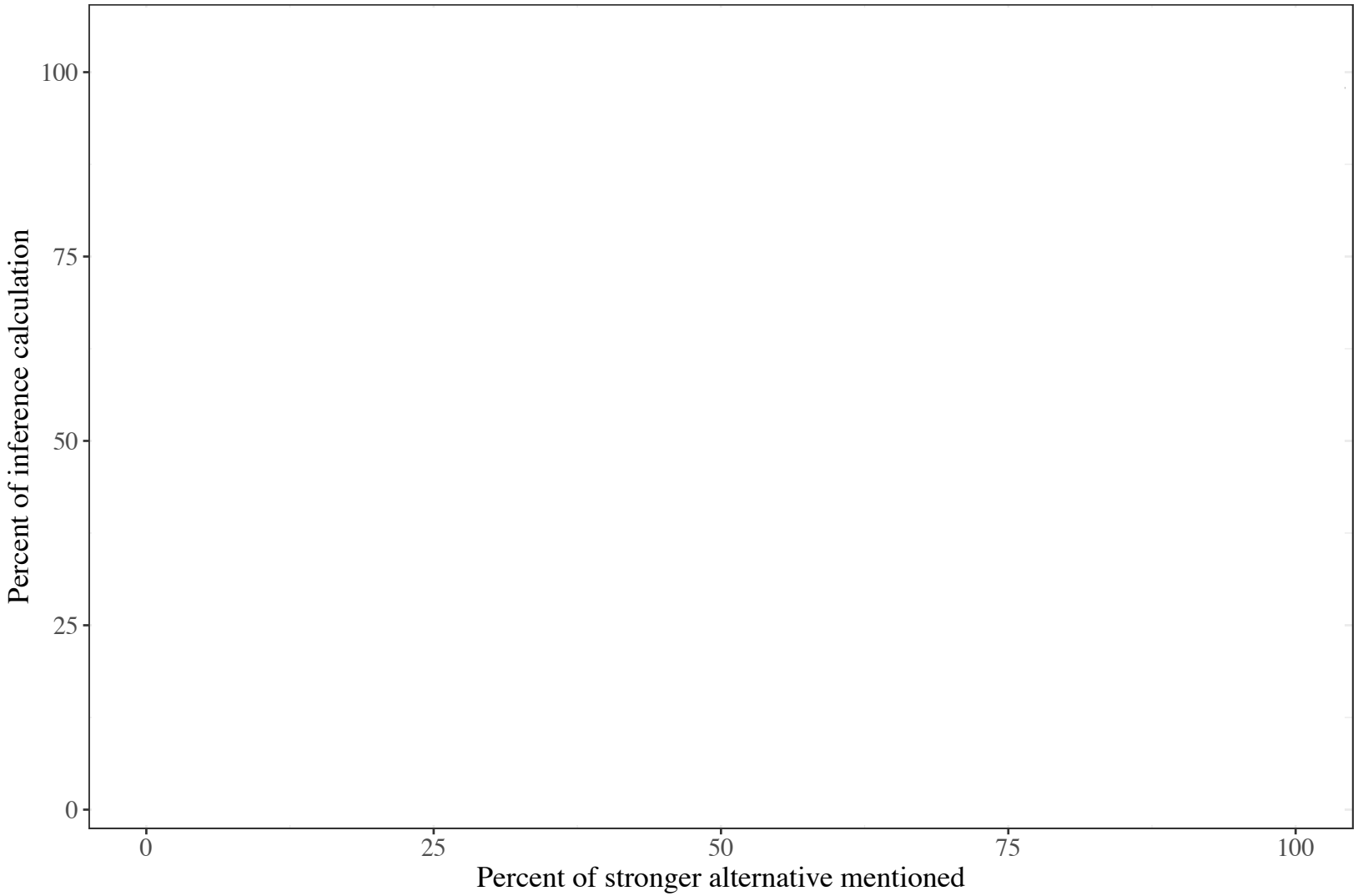
Cloze probability: measure of predictions hearers makes in language comprehension



Item N=60
Participant N=20

Prediction: the **more frequently the stronger alternative is mentioned** in the cloze task, the **higher the scalar inference rate** for that scale

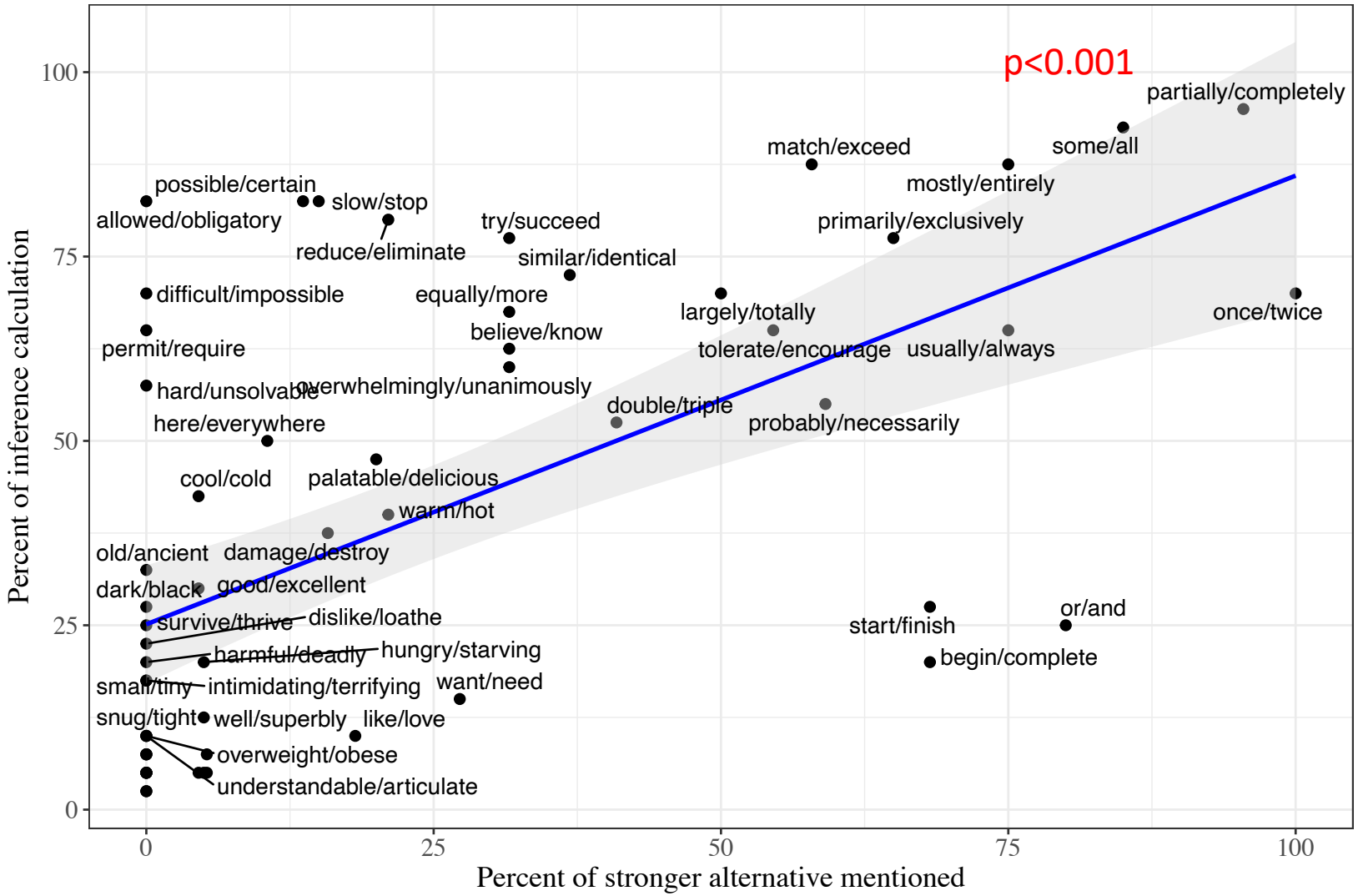
Results


100
75
50
25
0
Percent of inference calculation
0
25
50
75
100
Percent of stronger alternative mentioned

Results

Accessibility of alternatives predicts scalar inference rate

But much variance is still unexplained, R^2 =7.9%



Discourse context as QUD

Questions Under Discussion (**"QUDs"**; Roberts, 1996/2012):

- **Goal in discourse**: Give each other information
- Contributions modeled as attempts to answer **implicit or explicit questions**
- QUD determines what is relevant information (at a particular point)

Scalar inference is context-sensitive

QUDs affect the rate of scalar inference calculation:

(1) A: Did Mary eat all of the cookies?
B: Mary ate some of the cookies.

(2) A: Did Mary eat any of the cookies?
B: Mary ate some of the cookies.

Higher scalar inference rate in (1) than in (2)

(i.a. Cummins & Rohde 2015; Degen & Tanenhaus 2014; Ronai & Xiang 2021; Yang et al. 2018; Zondervan et al. 2008)

Scalar diversity & context

Ronai & Xiang 2021, Proc of LSA
Ronai & Xiang 2021, AMLaP

Hypothesis

Scalar diversity is a consequence of **differences in the implicit QUD that scales evoke**

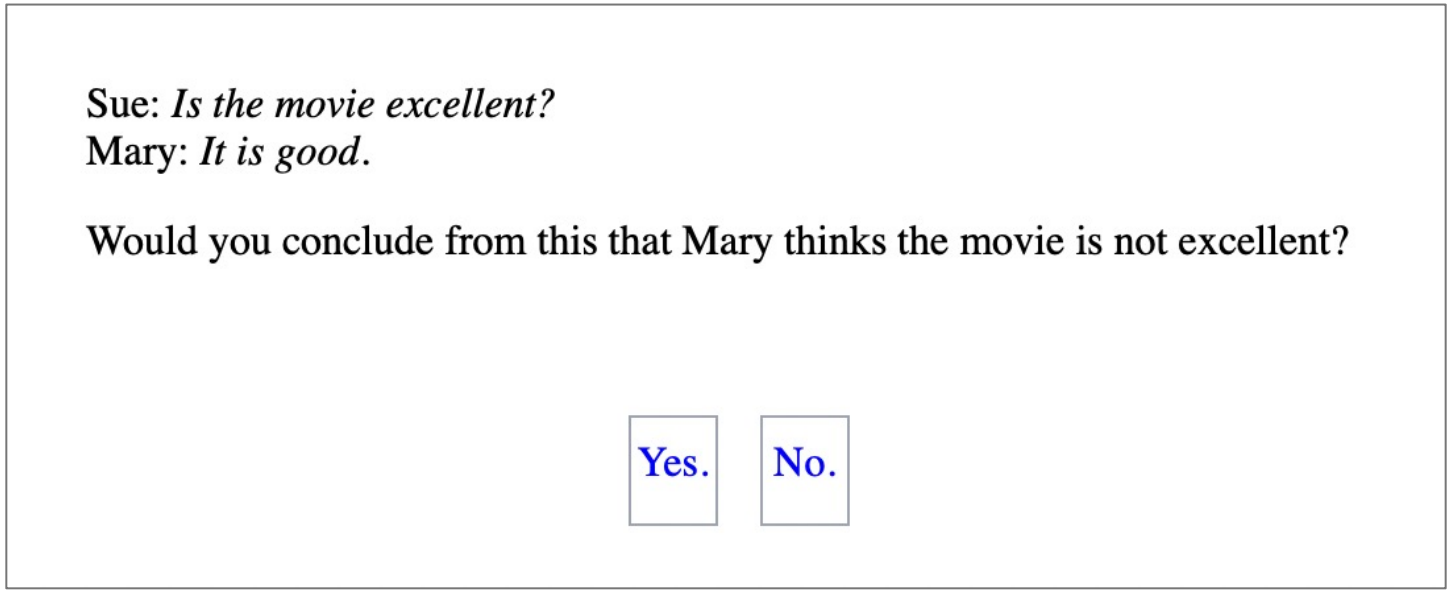
Intuition: *Did Mary eat all of the cookies?* is a **more likely question** than *Is the movie excellent?*

Prediction

Explicit question: such differences will be factored out, and **scalar diversity** will be **reduced** (or eliminated)

Scalar diversity & context

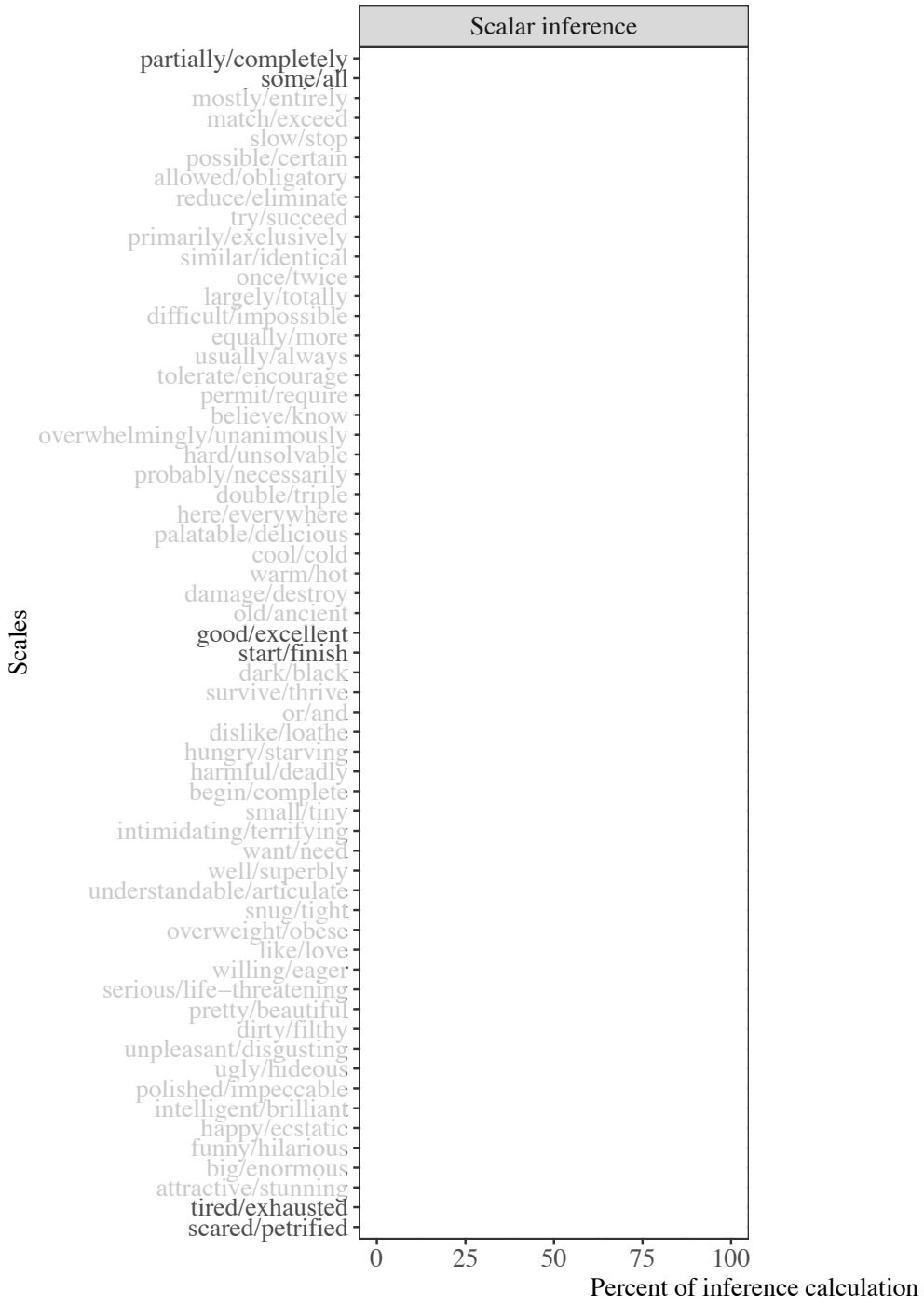
Item N=60
Participant N=40



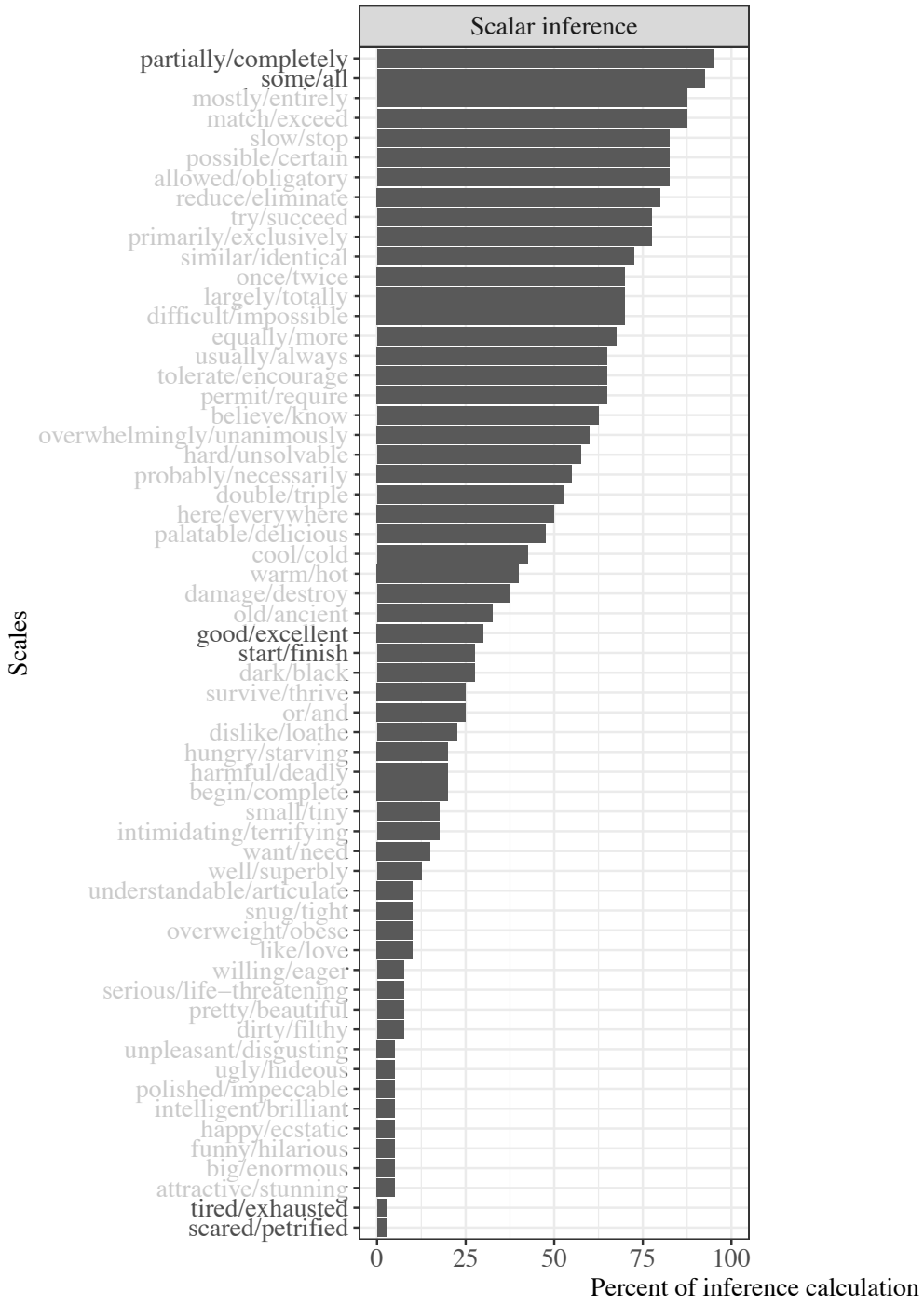
Inference task:

- "Yes" response = scalar inference was calculated
- "No" response = scalar inference was not calculated

Results

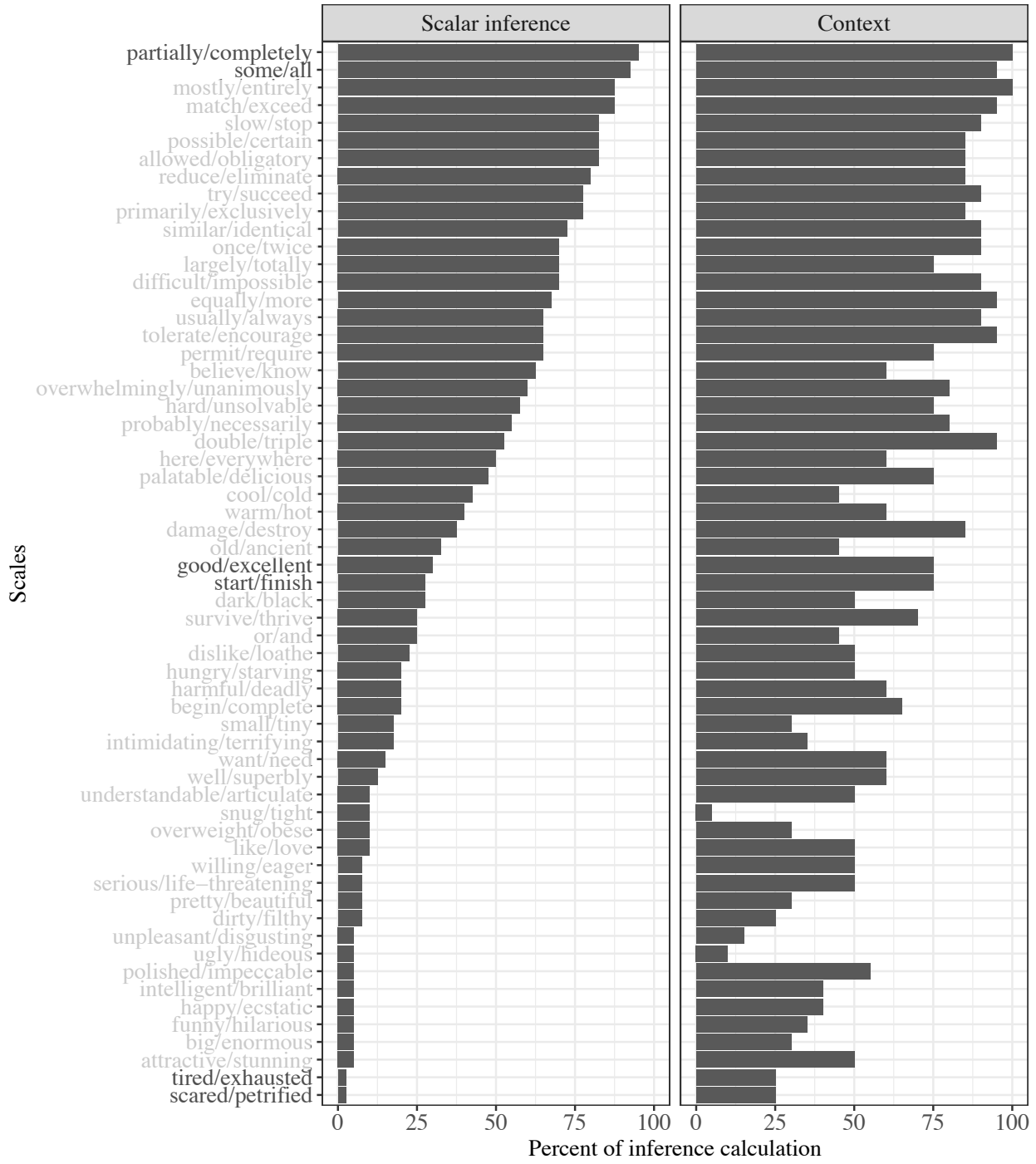

Scalar inference
partially/completely
some/all
mostly/entirely
match/exceed
slow/stop
possible/certain
allowed/obligatory
reduce/eliminate
try/succeed
primarily/exclusively
similar/identical
once/twice
largely/totally
difficult/impossible
equally/more
usually/always
tolerate/encourage
permit/require
believe/know
overwhelmingly/unanimously
hard/unsolvable
probably/necessarily
double/triple
here/everywhere
palatable/delicious
cool/cold
warm/hot
damage/destroy
old/ancient
good/excellent
start/finish
dark/black
survive/thrive
or/and
dislike/loathe
hungry/starving
harmful/deadly
begin/complete
small/tiny
intimidating/terrifying
want/need
well/superbly
understandable/articulate
snug/tight
overweight/obese
like/love
willing/eager
serious/life−threatening
pretty/beautiful
dirty/filthy
unpleasant/disgusting
ugly/hideous
polished/impeccable
intelligent/brilliant
happy/ecstatic
funny/hilarious
big/enormous
attractive/stunning
tired/exhausted
scared/petrified
Scales
0
25
50
75
100
Percent of inference calculation

Results


Scalar inference
partially/completely
some/all
mostly/entirely
match/exceed
slow/stop
possible/certain
allowed/obligatory
reduce/eliminate
try/succeed
primarily/exclusively
similar/identical
once/twice
largely/totally
difficult/impossible
equally/more
usually/always
tolerate/encourage
permit/require
believe/know
overwhelmingly/unanimously
hard/unsolvable
probably/necessarily
double/triple
here/everywhere
palatable/delicious
cool/cold
warm/hot
damage/destroy
old/ancient
good/excellent
start/finish
dark/black
survive/thrive
or/and
dislike/loathe
hungry/starving
harmful/deadly
begin/complete
small/tiny
intimidating/terrifying
want/need
well/superbly
understandable/articulate
snug/tight
overweight/obese
like/love
willing/eager
serious/life–threatening
pretty/beautiful
dirty/filthy
unpleasant/disgusting
ugly/hideous
polished/impeccable
intelligent/brilliant
happy/ecstatic
funny/hilarious
big/enormous
attractive/stunning
tired/exhausted
scared/petrified
Scales
0
25
50
75
100
Percent of inference calculation

Results

Overt question reduced scalar diversity somewhat



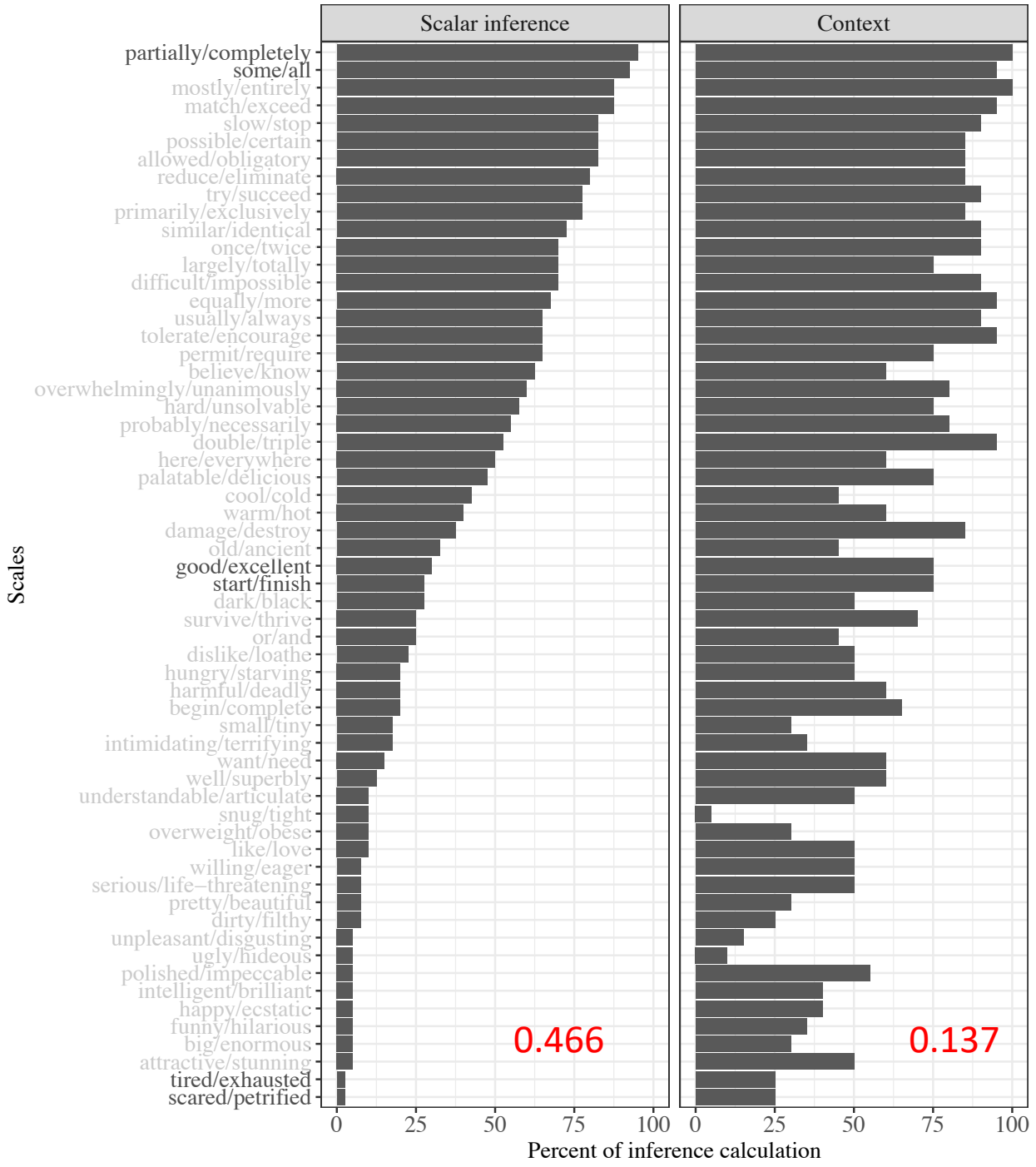
Results

Overt question reduced scalar diversity somewhat but did not eliminate it

Relative entropy to quantify this observation:

- SI rates = probability distribution
- Compare to uniform distribution

$$D\left(p||q\right) = \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right)$$



Scalar diversity: Conclusions

- Production-based measure of alternative (*excellent*) accessibility captures scalar diversity
  - But much variance still remains
- Scalar diversity is reduced by supporting context
  - But context did not eliminate diversity

Both context and alternatives contribute to scalar diversity

Other factors in scalar diversity

- Semantic distance between scalars (van Tiel et al. 2016)
- Boundedness of the scale (van Tiel et al. 2016)
- Local enrichability (Sun et al. 2018)
- Extremeness (Gotzner et al. 2018; Beltrama & Xiang 2013)
- Polarity (Gotzner et al. 2018)
- Negative strengthening (Gotzner et al. 2018)
- The relevance of the scalar inference (Pankratz & van Tiel 2021)

Ronai & Xiang (2021):
- Availability of the relevant QUD (*Proc of LSA*)
- Distinctness of the two scalar terms (*Sinn und Bedeutung*)
- Meaning of the the negated strong scalar term (*Sinn und Bedeutung*)

Complex problem, full empirical range of scalar diversity unexplained

Conclusions

Part I. Inferential process: lexical alternatives are activated in real-time processing

Part II. **Alternatives and context** both contribute to likelihood of inference calculation across different lexical scales

Pragmatic meaning arises both as a function of **global properties of context,** and as a function of **local properties of the scalar terms** themselves

Zooming out

Communication is not just about what is said, but also about what is **not said**

Speakers and hearers **coordinate**:

Speakers reason about **hearer expectations**;
Hearers reason about **speaker intentions**

Understanding this process helps us **understand how communication works** in context

Scalar inference: window into the precise roles of linguistic and extra-linguistic factors

discourse-syntax interface

multilingual population

brain mechanisms

Future work

Thank you!

NATIONAL SCIENCE FOUNDATION
DDRIG GRANT #BCS-2041312 (TO ESZTER RONAI)

References

Beltrama & Xiang (2013). Is 'good' better than 'excellent'? An experimental investigation on scalar implicatures and gradable adjectives. Proceedings of SuB 17.

Cummins & Rohde (2015). Evoking context with contrastive stress: Effects on pragmatic enrichment. Frontiers in Psychology.

Degen & Tanenhaus (2014). Processing scalar implicature: A constraint-based approach. Cognitive Science.

Gotzner et al. (2018). Scalar Diversity, Negative Strengthening, and Adjectival Semantics. Frontiers in Psychology.

Grice (1967). Logic and conversation. Studies in the way of words.

Horn (1972). On the Semantic Properties of Logical Operators in English. Ph.D. thesis.

Katzir (2007). Structurally-defined alternatives. Linguistics & Philosophy.

Pankratz & van Tiel (2021). The role of relevance for scalar diversity: a usage-based approach. Language and Cognition.

Roberts (1996/2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. Semantics and Pragmatics.

Sperber & Wilson (1995). Relevance: Communication and Cognition. Wiley-Blackwell.

Sun et al. (2018). A link between local enrichment and scalar diversity. Frontiers in Psychology.

van Tiel et al. (2016). Scalar diversity. Journal of Semantics.

Zondervan et al. (2008). Experiments on the role of the question under discussion for ambiguity resolution and implicature computation in adults. Proceedings of SALT 18.

Yang et al. (2018). Context-sensitivity and individual differences in the derivation of scalar implicature. Frontiers in Psychology.

References

Ronai, E. (2018). Quantifier scope in heritage bilinguals: a comparative experimental study. In NELS 48: Proceedings of the 48th Annual Meeting of the North East Linguistic Society (Vol. 3, pp. 29–38). University of Massachusetts, Amherst: GLSA.

Ronai, E., & Stigliano, L. (2020). Adjectival sluices in Hungarian: An argument for isomorphic sources. Natural Language & Linguistic Theory 38(4), pp. 1197–1210. doi: 10.1007/s11049-020-09468-w

Ronai, E., & Stigliano, L. (2020). NP ellipsis bleeds allomorphy in Hungarian. Snippets 39, pp. 10-12.

Ronai, E., & Stigliano, L. (2021). Is multiple fronting necessary for multiple sluicing? The view from Hungarian. In Proceedings of the 38th West Coast Conference on Formal Linguistics, pp. 377–385. Cascadilla Proceedings Project.

Ronai, E., & Stigliano, L. (2021). Licensing of nominal ellipsis in Hungarian possessives. In Proceedings of the Linguistic Society of America 6(1), pp. 600–607. DOI: 10.3765/plsa.v6i1.4994.

Ronai, E., & Stigliano, L. (in prep). Experimentally testing the interpretation of multiple sluicing and multiple questions. Invited contribution to the The Linguistic Review special issue: Empirical approaches to elliptical constructions.

Ronai, E., Sun, Y., Yu, A. C. L., & Xiang, M. (2019). Integration of contextual-pragmatic and phonetic information in speech perception: An eye-tracking study. Laboratory Phonology: Journal of the Association for Laboratory Phonology 10(1): 20, pp. 1–15. doi: 10.5334/labphon.186

Ronai, E., & Xiang, M. (2019). Calculating scalar inference under QUDs. In NELS 49: Proceedings of the 49th Annual Meeting of the North East Linguistic Society (Vol. 3, pp.73–86). UMass, Amherst: GLSA.

Ronai, E., & Xiang, M. (2021). Exploring the connection between Question Under Discussion and scalar diversity. In Proceedings of the Linguistic Society of America 6(1), pp. 649–662. DOI: 10.3765/plsa.v6i1.5001

Ronai, E., & Xiang, M. (2021). Pragmatic inferences are QUD-sensitive: an experimental study. Journal of Linguistics 57(4), pp. 841-870. doi: 10.1017/S0022226720000389

Ronai, E., & Xiang, M. (2021). Three factors in explaining scalar diversity. Talk presented at Sinn und Bedeutung 26.

Ronai, E., & Xiang, M. (2021). Quantifying scalar diversity: a first look. Talk presented at the 27th Architectures and Mechanisms for Language Processing Conference.

Ronai, E., & Xiang, M. (ms.). Memory vs. expectation: processing relative clauses in a flexible word order language. In revision at Cognitive Science.

Relative entropy (information theory)

- Normalized % of "Yes" responses (i.e., the SI rates) across different scales = probability distribution
  - Test: does a given SI rate provide enough information to identify the scale that it came from?
- Compared to the uniform distribution: each scale leads to the same SI rate
  - The % of "Yes" responses gives 0 information about the identity of the scale it came from
  - Scales cannot be identified by their associated SI rates
- Resulting measure: relative entropy (entropy of the uniform distribution minus the entropy of the given SI rates) → quantify how "diverse" the SI rates are
- Results: Scalar inference 0.466, Context manipulation 0.137
- Benchmarks:
  - Uniform distribution: relative entropy is 0
  - If we evenly distribute 60 items (=lexical scales) over a 0-100 scale (=SI rates): relative entropy is 0.2912
  - Including "only" in the experiment (*The movie is only excellent*): relative entropy is 0.046

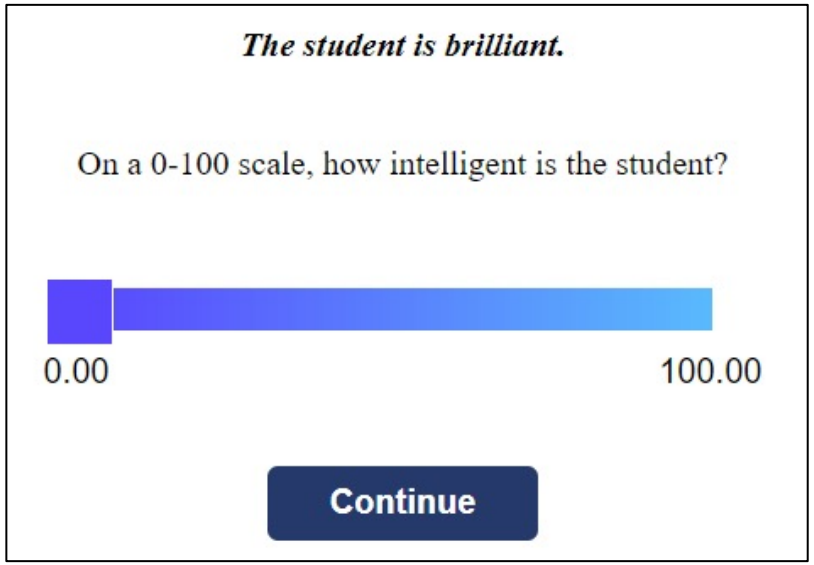
$$D\left(p||q\right) = \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

p(x): observed % of "Yes" responses across scales; X: items
q(x) = 1/60: uniform probability mass function over the 60 scales

Other tasks for scalar inference

Problem with inference task: metalinguistic; biasing question ("*Would you conclude... not excellent*"); triggers other kinds of pragmatic reasoning

New task: posterior **degree estimates**:



- Inspired by Bayesian pragmatics, which models recursive reasoning between speaker and hearer (Goodman & Frank 2016; Lassiter & Goodman 2015; Xiang et al. under review)

- What world states do hearers think the speaker had in mind when she uttered "*The student is* {*intelligent/brilliant/not brilliant/intelligent bot not brilliant*}"?

List of scales

Adjective	<allowed, obligatory>; <attractive, stunning>; <big, enormous>; <cool, cold>; <dark, black>; <difficult, impossible>; <dirty, filthy>; <funny, hilarious>; <good, excellent>; <happy, ecstatic>; <hard, unsolvable>; <harmful, deadly>; <hungry, starving>; <intelligent, brilliant>; <intimidating, terrifying>; <old, ancient>; <overweight, obese>; <palatable, delicious>; <polished, impeccable>; <possible, certain>; <pretty, beautiful>; <scared, petrified>; <serious, life-threatening>; <similar, identical>; <small, tiny>; <snug, tight>; <tired, exhausted>; <ugly, hideous>; <understandable, articulate>; <unpleasant, disgusting>; <warm, hot>; <willing, eager>
Verb	<begin, complete>; <believe, know>; <damage, destroy>; <dislike, loathe>; <double, triple>; <like, love>; <match, exceed>; <permit, require>; <reduce, eliminate>; <slow, stop>; <start, finish>; <survive, thrive>; <tolerate, encourage>; <try, succeed>; <want, need>
Adverb	<equally, more>; <here, everywhere>; <largely, totally>; <mostly, entirely>; <once, twice>; <overwhelmingly, unanimously>; <partially, completely>; <primarily, exclusively>; <probably, necessarily>; <usually, always>; <well, superbly>
Quantifier	<or, and>
Connective	<some, all>