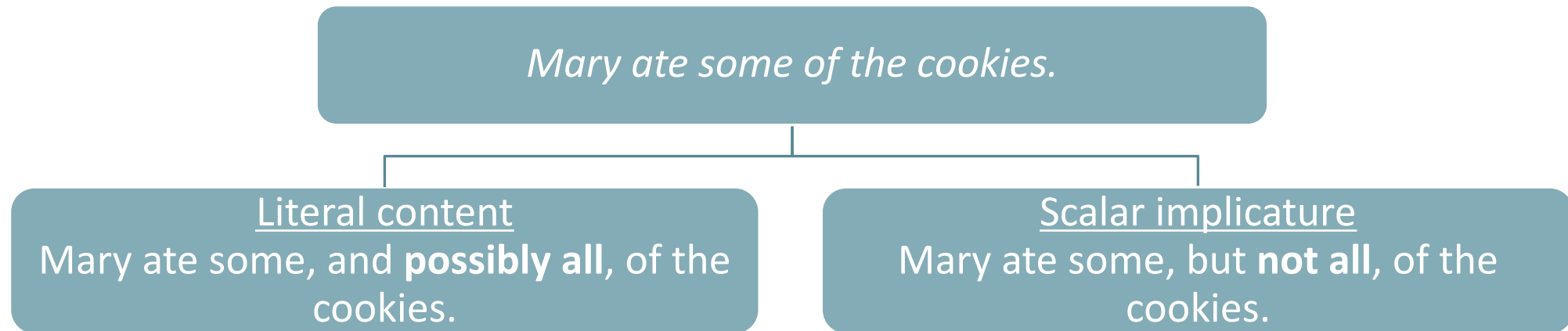


Degree estimates as a measure of inference calculation

Eszter Ronai (Northwestern University) & Ming Xiang (The University of Chicago)

LSA Annual Meeting 2023 (January 5-8)

Scalar implicature (SI)



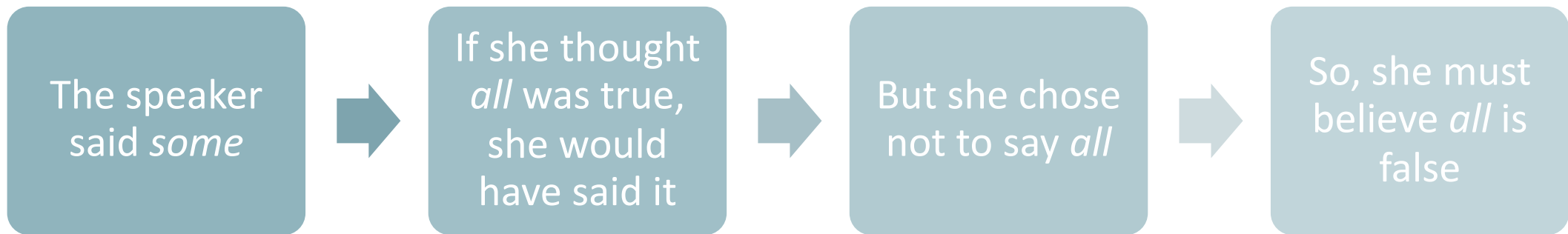
Comprehenders reason about **alternative utterances** the speaker could have said...
....to recover the **intended meaning**

(Grice, 1975; Horn, 1972)

Reasoning about alternatives

<*some*, *all*> form a scale

all is logically stronger (**more informative**) than *some*



Other lexical scales

The movie is **good**.

→

The movie isn't **excellent**.

The student is **intelligent**.

→

The student isn't **brilliant**.

Scalar diversity phenomenon (i.a. van Tiel, et al., 2016)

Inference task

Mary: *The movie is good.*

Would you conclude from this that Mary thinks the movie is not excellent?

Yes.

No.

- “**Yes**” = SI **was** calculated
- “**No**” = SI **was not** calculated

Geurts & Pouscoulous, 2009; van Tiel et al., 2016;
Gotzner et al., 2018; Sun et al., 2018;
van Tiel & Pankratz, 2021; Ronai & Xiang, 2022

Problem 1: bias

Mary: *The movie is good.*

Would you conclude from this that Mary thinks the movie is not excellent?

Yes.

No.

Task question **explicitly provides** the alternative (*excellent*)

Bias to reason about it

Bias towards calculating the **SI**

Problem 2: other inferences

Mary: *The movie is good.*

Would you conclude from this that Mary thinks the movie is not excellent?

 Yes. No.

Negative strengthening: *not excellent* \approx *mediocre*

(Horn, 1989; Gotzner et al., 2018)

Response doesn't just reflect SI

Effect of task question

Sun & Breheny (2022)

stronger alternative under negation (*not... all*) vs. possibility modal (*could be... all*)

Mary says: Some of the questions are easy.

*Would you conclude from this that, according to Mary, **not all** of the questions are easy?*

*Would you conclude that, it **could be** that Mary thinks, **all** of the questions are easy?*

→ <*some, all*> and <*possible, certain*> : more SI with “not”

→ numerals : more SI with “could”

Effect of response options

Jasbi et al. (2019) (also Katsos & Bishop, 2011; Sikos et al., 2019)

Sentence-picture rating:

binary: wrong, right

ternary: wrong, neither, right

quaternary: wrong, kinda wrong, kinda right, right

quinary: wrong, kinda wrong, neither, kinda right, right

Number of response options makes a difference

What do we take to correspond to SI? (“wrong” or not “right”)

Degree estimate task

What **world states** comprehenders come to have in mind, given an utterance

- *The movie is good.*

Degree estimates on the underlying degree scales

- What degree of goodness?

More fine-grained measure than the binary inference task (“Yes” vs. “No”)

Avoids the bias of directly presenting stronger alternative

Experiment 1

Validate methodology:

- weaker scalar (*good*)
- stronger alternative (*excellent*)
- negated stronger alternative (*not excellent*)


91 participants

60 lexical scales

Experiment 1: weak

The movie is good.

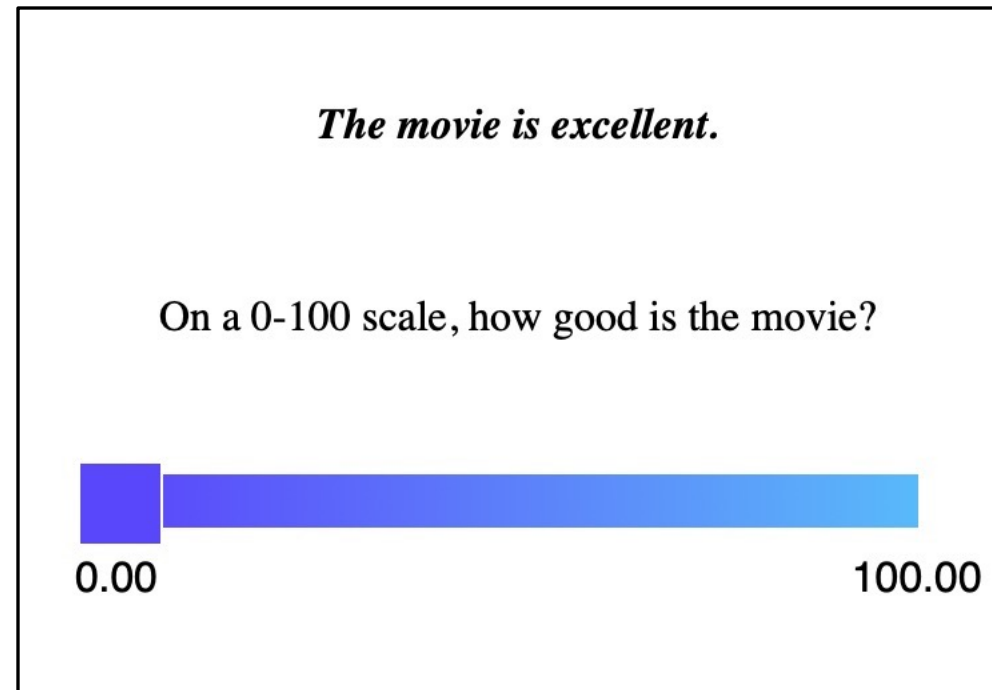
On a 0-100 scale, how good is the movie?



0.00 100.00

Continue


Experiment 1: strong



Experiment 1: negated strong

The movie is not excellent.

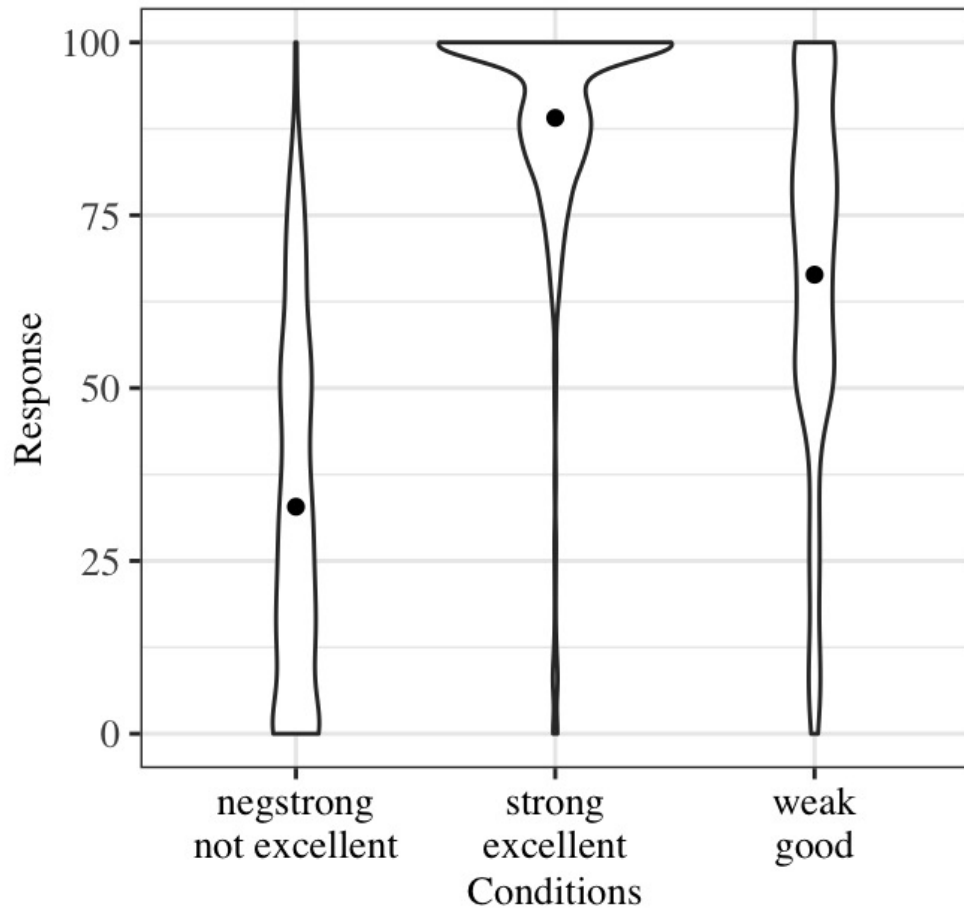
On a 0-100 scale, how good is the movie?



0.00 100.00

Continue

Experiment 1: results



strong higher than weak ($p < 0.001$)

→ reality check

→ SI (?)

negated strong lower than weak ($p < 0.001$)

→ negative strengthening

(Horn, 1989; Gotzner et al., 2018)

Experiment 2

Reassessing prior findings:

- Question Under Discussion (QUD; Roberts, 1996/2012)
- *only*

92 participants

60 scales

Experiment 2

Ronai & Xiang (2022)

SI rates higher in a **supportive discourse context**

(1) A: Is the movie excellent?
B: It is good.

(2) A: Is the movie good?
B: It is good.

same as no context

Focus particle *only*: inference rates even higher

(3) The movie is only good.

Inference task: **(3) > (1) > (2)**

Experiment 2: strong QUD

Sue: "Is the movie excellent?"

Mary: "It is good."

On a 0-100 scale, how good is the movie?



Continue

Experiment 2: weak QUD

Sue: "Is the movie good?"

Mary: "It is good."

On a 0-100 scale, how good is the movie?




Continue

Experiment 2: *only*

The movie is only good.

On a 0-100 scale, how good is the movie?



0.00 100.00

Continue

Recap: inference task results

strong QUD > weak QUD (or no context)

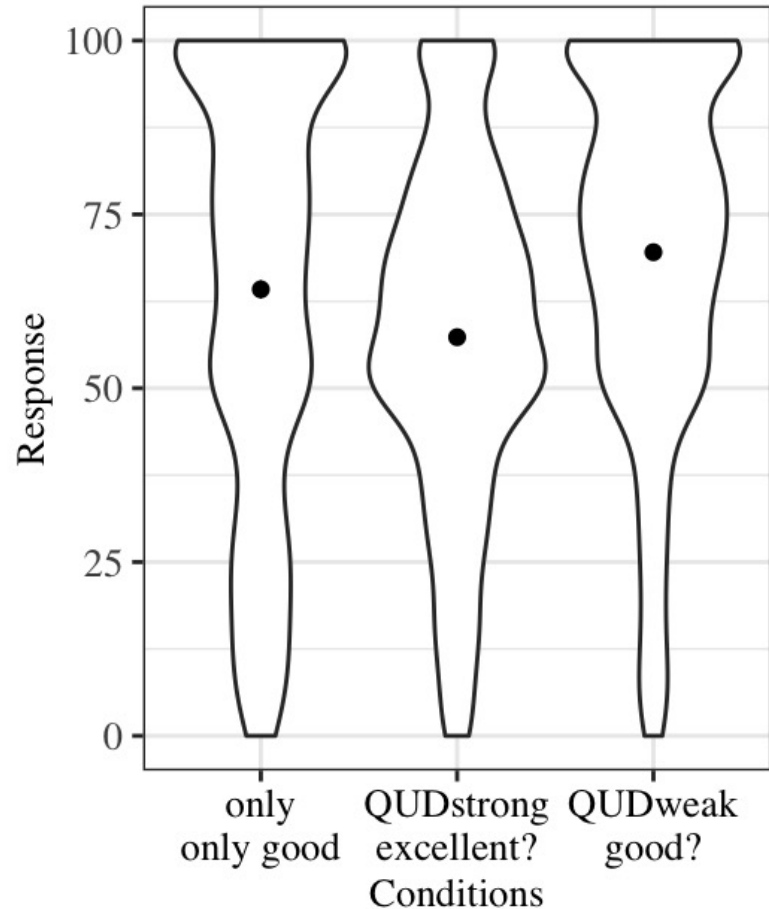
only > strong QUD

only **semantically** encodes **exclusion of alternatives** (Rooth, 1985, 1992)

biasing **question encourages SI** calculation only **pragmatically**

(i.a., Hulseley et al., 2004; Degen, 2013; Zondervan et al., 2008)

Experiment 2: Results



baseline weak QUD higher than *only* ($p < 0.05$)

strong QUD lower than *only* ($p < 0.01$)
→ reverse of previous inference task results!

Experiment 2: Discussion

Inference task:

more *not excellent* inferences with *only* than with strong QUD

Degree estimate task:

lower degree of goodness with strong QUD than with *only*

Reason 1:

***only* doesn't specify what **alternative** gets excluded (*only good* → not funny)**

inference task specifies scalar alternative (→ not excellent)

→ **inflated rates** of “Yes” responses

Experiment 2: Discussion

Reason 2:

A: Is the movie excellent?

B: It is good.

B intends to give a negative answer but avoids “No” out of politeness

by *good*, **B intends to communicate *not excellent***

negative strengthening of *not excellent*

→ less than good

→ **lower degree estimate**

inference task: *good but not excellent* and *less than good* → both “Yes”

Conclusion

Inference task: a common measure of SI (especially scalar diversity)

Bias: **explicit** stronger **alternative**

Obscures **other** non-SI **inferences**

Degree estimate task: more fine-grained measure

Test the role of QUDs and *only* in modulating inference calculation

Results not in line with prior work

Open question

What corresponds to the **SI-enriched meaning**?

“good” vs. “good but not excellent”

“good”:

we **don't** know whether **SI** was calculated


“good but not excellent”:

definitely **“SI”**

but **in the asserted content**

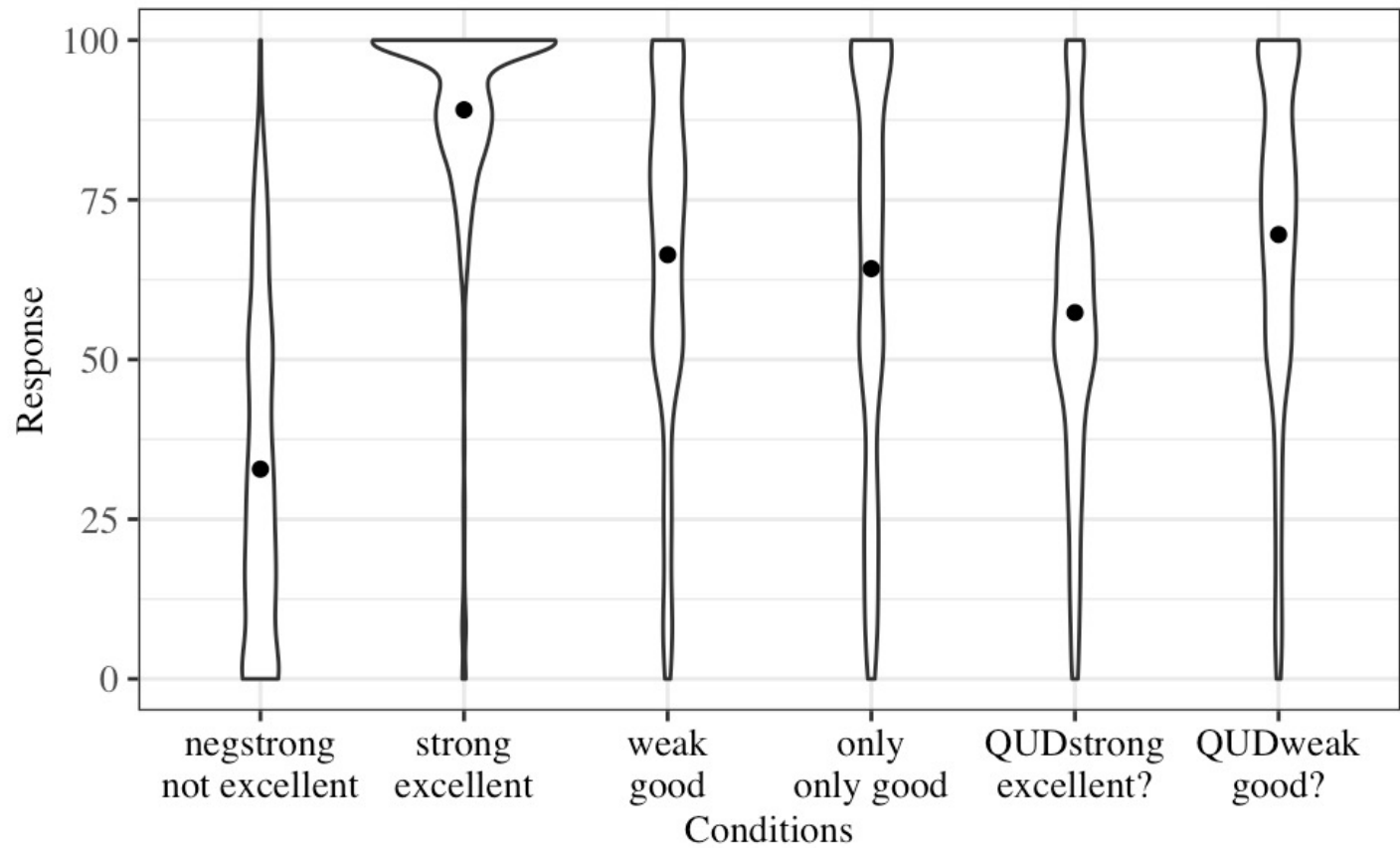
The movie is good but not excellent.

On a 0-100 scale, how good is the movie?



0.00 100.00

Continue



Thank you!



National Science Foundation
 DDRIG Grant #BCS-2041312

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>

