

Objectives

Explore the role of Questions Under Discussion (QUDs) in explaining scalar diversity.

Background

Listeners reason about what is not said: the stronger alternative—*all* and *brilliant*.

(1) Mary ate some of the cookies. → SI: Mary ate some, but not all, of the cookies.

(2) The student is intelligent. → SI: The student is intelligent, but not brilliant.

→ Scalar inference (SI) calculation.

Scalar diversity

Considerable **variation across different scales in SI calculation rates**; e.g. the SI in (1) arises much more robustly than the one in (2) (van Tiel et al. 2016; see also Doran et al. 2012; Beltrama & Xiang 2013).

What properties of scales can explain this variation?

- Distinctness of the stronger scalar term (van Tiel et al. 2016).
- Local enrichability (Sun et al. 2018).
- Negative strengthening, polarity, extremeness (Gotzner et al. 2018).
- But: still a lot of variance unaccounted for in the empirical results.

The role of context

QUDs have an effect on rate of SI calculation: *Did Mary eat all of the cookies?* leads to higher SI rate than *Did Mary eat any/some of the cookies?* (i.a. Degen & Tanenhaus 2014; Ronai & Xiang 2020; Yang et al. 2018; Zondervan et al. 2008).

Previous work on scalar diversity: stimulus sentences presented without any context.

Open question: is there variation across scalar terms in what kind of QUD they most naturally bring to mind?

Hypothesis

Scalar diversity, in the absence of an explicit QUD, arises (in part) **due to the differential availability of a polar question containing the stronger scalar term** from the scale.

Intuition: the more likely a question such as *Is the student brilliant?* is, the higher the rate of SI calculation from the corresponding statement *She is intelligent*.

Experiments

Experiment 1: replication of van Tiel et al. (2016)

- 37 native speakers of American English; MTurk; IbexFarm.
- Inference task to investigate the likelihood of deriving an SI (from 43 scales).
- Participants saw “Mary: *The student is intelligent.*” + asked the question “Would you conclude from this that, according to Mary, the student is not brilliant?”
 - “Yes” response = SI was calculated.
 - “No” response = SI was not calculated.

Experiment 2: inference task with Question manipulation

- 40 native speakers of American English; MTurk; IbexFarm.
- Basic inference task identical to Experiment 1.
- Two-condition Question manipulation: Mary’s statement in a dialogue context.
 - Strong-scalar question: Sue: *Is the student brilliant?*
Mary: *She is intelligent.*
 - Weak-scalar question: Sue: *Is the student intelligent?*
Mary: *She is intelligent.*
- **Predictions:** 1) strong-scalar questions lead to higher SI rates
2) scalar diversity reduced

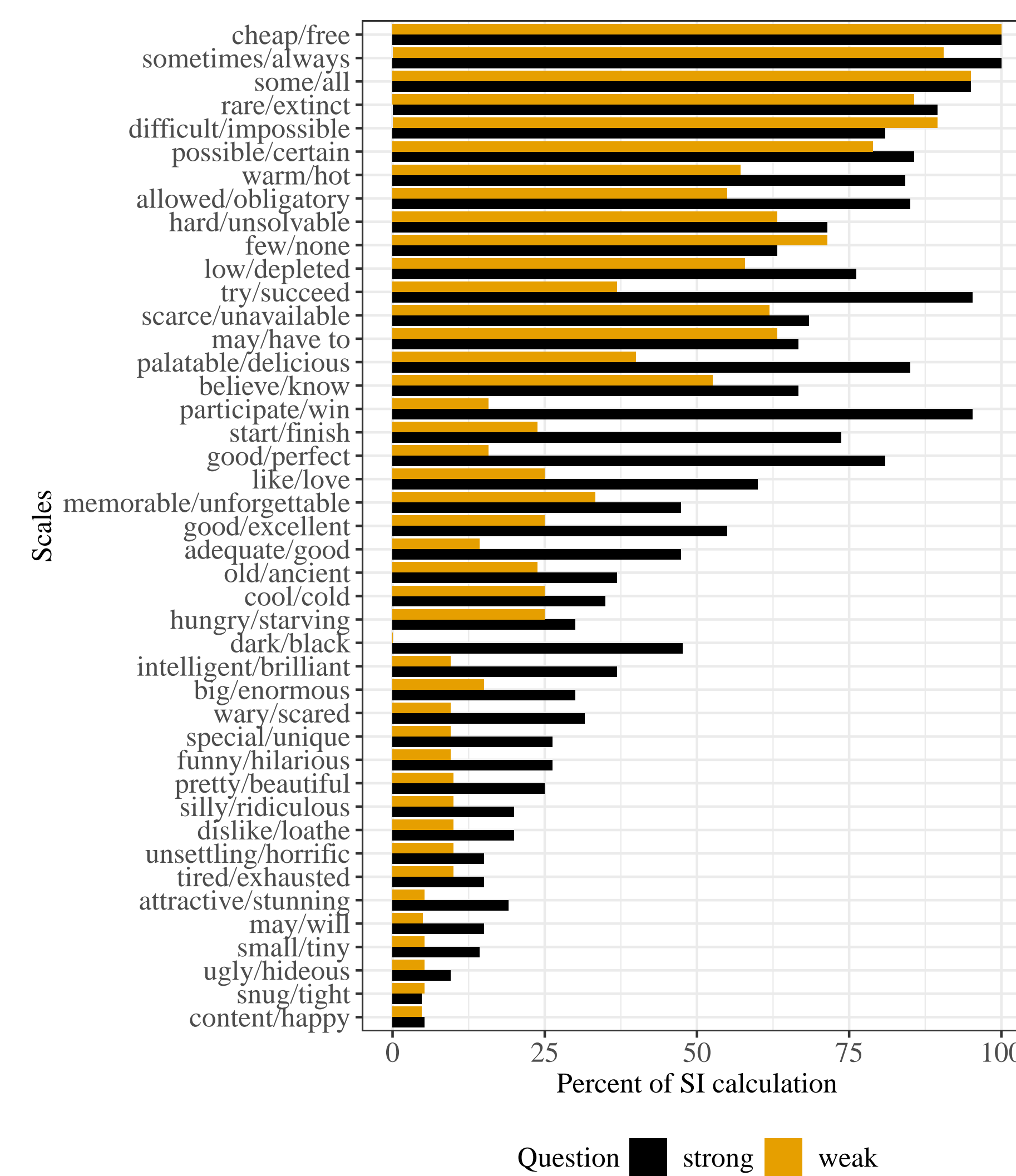
Experiment 3: question availability

- 35 native speakers of American English; MTurk; IbexFarm.
- Forced choice task: “Compare the following two questions about a student. Which one are you more likely to ask?”
 - choice between: *Is the student brilliant?* vs. *Is the student intelligent?*
- **Prediction:** forced choice results (henceforth **Question Availability**) should **predict scalar diversity**.
 - The more preferred the strong-scalar question in Exp. 3, the higher the SI rate for that scale in Exp. 1.

Results

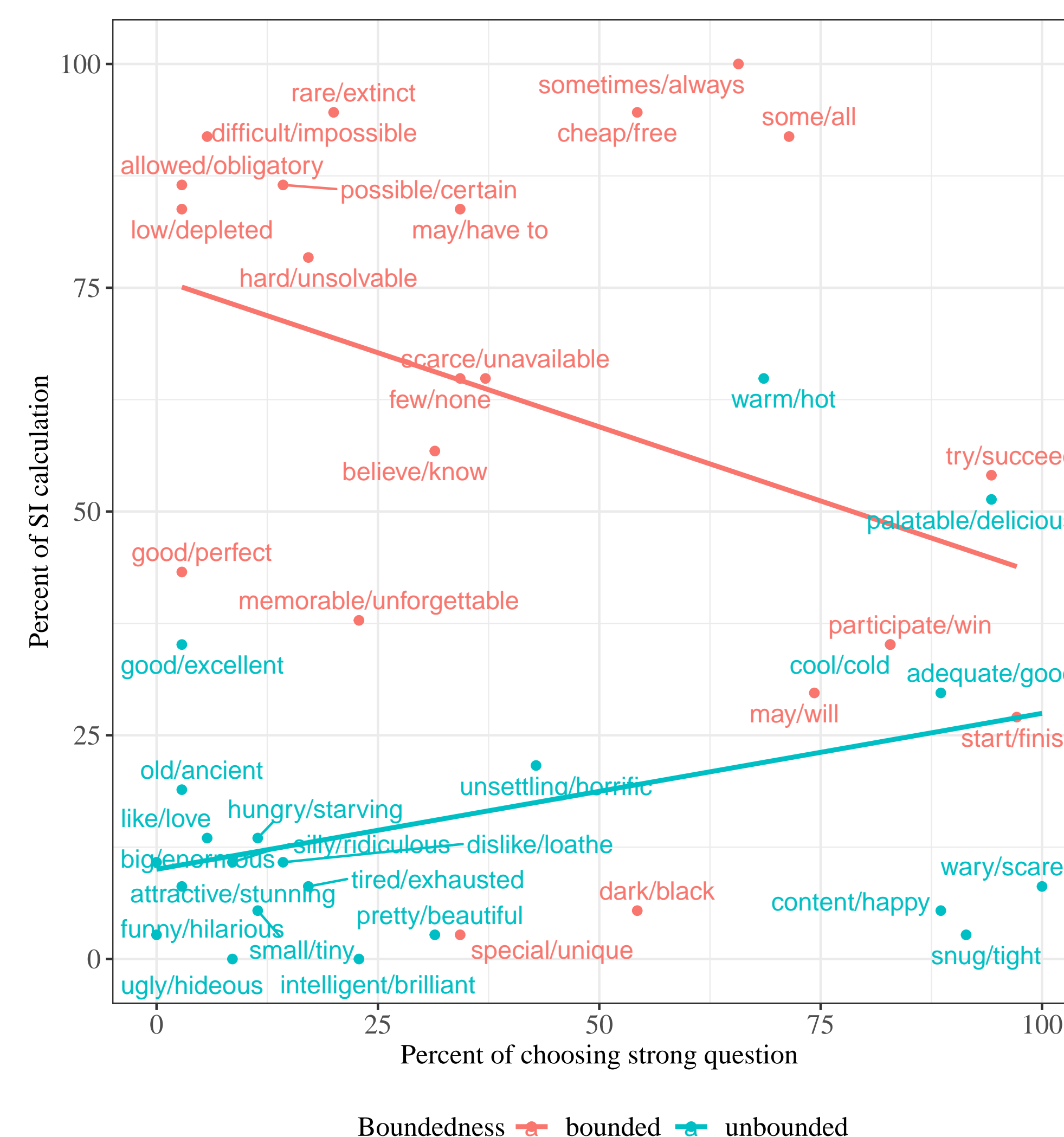
Exp. 2: more SIs derived with strong-scalar than weak-scalar question ($p < 0.01$).

- Explicit QUD influences SI calculation rates for a large number of scales (later replicated on 60 scales).
- Scalar diversity effect is still present (quantifiable using relative entropy).



Exp. 1 (y axis): replicated the scalar diversity effect.

Exp. 3 (x axis): Question Availability not an overall predictor of SI rates.



Boundedness (van Tiel et al. 2016)

Bounded scale: stronger scalar denotes an endpoint, e.g. *<some, all>* is bounded.

- Bounded scales led to higher SI rates than unbounded ones ($p < 0.001$; replicates van Tiel et al. 2016).

Interaction of Question Availability with Boundedness ($p < 0.05$):

- Unbounded scales: Question Availability showed a strong trend ($p < 0.08$) in predicting SI calculation.
 - The more likely participants were to choose the strong-scalar question (*Is the student brilliant?*), the higher the relevant SI rate (*intelligent*→*not brilliant*).
- Bounded scales: no effect of Question Availability ($p = 0.14$).

Discussion

Bounded scales: the stronger scalar term is not vague, but denotes a fixed point.

- Stronger scalar term is very salient as stronger alternative to vague, weaker term.
- → High rates of SI calculation; Question Availability makes no difference.

Unbounded scales: both scalar terms are vague; they denote intervals whose values vary according to context.

- Stronger scalar is less salient as an alternative → can get boost from context.
- The more available a QUD based on the stronger term is, the more likely hearers will be to reason about that term as the stronger alternative.
- → More likely to derive the SI.

Some of the stronger terms on unbounded scales are also extreme (Morzycki, 2012): “off the scale” scalar terms not generally salient alternatives.

Ongoing and future work

1) Exp. 2: QUDs that set up biasing contexts without mentioning the scalar terms.

- Right now, effects may be due to relevance implicature.

Question elicitation (Ronai & Xiang, 2020):

- Mary: _____?
 - Sue: *The student is intelligent.*
 - Mary: *Can I take that to mean she is not brilliant?* OR *Can I take that to mean she might be brilliant?* OR *Oh, I see.*
 - Sue: *Yes, that's right.* (not after *Oh, I see.*)
- Manipulation doesn't seem to have effect, completions mostly based on second line.

2) Different empirical measures of Question Availability?

- Question Availability may itself be context-dependent.

3) Is it boundedness or extremeness that matters? (Not all unbounded are extreme.)

4) Final puzzle: Exp. 2: effect for all scales vs. Exp. 3: effect only for unbounded scales.

Conclusions

- **Explicit questions robustly affect SI calculation rates for a large number of scales.**
- Yet there still remained substantial variation in SI rates across scales.
- **Likelihood of a question based on the stronger scalar contributes to scalar diversity, but only for unbounded scales.**

References

Beltrama & Xiang (2013). Is 'good' better than 'excellent'? An experimental investigation on scalar implicatures and gradable adjectives. *Proceedings of SuB 17*. | Degen & Tanenhaus (2014). Processing scalar implicatures: A constraint-based approach. *Cognitive Science*. | Doran et al. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*. | Gotzner et al. (2018). Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in Psychology*. | Morzycki (2012). Adjectival extremeness: Degree modification and contextually restricted scales. *NLLT*. | Ronai & Xiang (2020). Pragmatic inferences are QUD-sensitive: An experimental study. *Journal of Linguistics*. | 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*.