Exploring the connection between Question Under Discussion and scalar diversity

Eszter Ronai & Ming Xiang

The University of Chicago

95th Annual Meeting of the Linguistic Society of America (LSA 95)
7-10 January 2021
Introduction

In conversation, comprehenders draw inferences beyond literal meaning: scalar inference, e.g. some but not all.

Scalar diversity: likelihood of drawing such an inference varies across scales.

We explore the role of Question Under Discussion (QUD; Roberts 1996/2012) in explaining this variation.

Upshot of findings:

- QUDs affect rate of inference calculation for all scales.
- Question availability predicts variation in inference rates, but only for unbounded scales.
Roadmap

1. Background
   1. Scalar inference.
   2. Scalar diversity.

2. Experiment 1: replication of scalar diversity.

3. Experiment 2: QUD manipulation.

4. Experiment 3: question availability as a predictor of scalar diversity.

5. Conclusions.
Scalar inference

Scalar inference (SI) calculation:

(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.

Comprehenders reason about what is not said: the stronger alternative

▶ all in (1)
▶ brilliant in (2)

(Grice 1967)
Considerable **variation across different scales in SI calculation rates.**

E.g. *some but not all* SI arises much more robustly than *intelligent but not brilliant* —finding about 43 scales (van Tiel et al. 2016; see also Doran et al. 2012; Beltrama & Xiang 2013).
Explaining scalar diversity

What properties of scales can explain this variation?

Existing work has identified:

- Distinctness of the stronger scalar term (van Tiel et al. 2016).
  - Semantic distance: the more distant a weak and a strong scalar terms are, the stronger the SI from the weak term.

  (3) a. Many of the senators voted against the bill.
   b. Most of the senators voted against the bill.
   c. All of the senators voted against the bill.

  SI from (3a): more likely the negation of (3c) than of (3b) (Horn 1972).

  - Boundedness: bounded scales, where the stronger scalar term refers to an end point, lead to higher SI rates.
Explaining scalar diversity

Other known factors:

- Local enrichability (Sun et al. 2018).
- Extremeness (Gotzner et al. 2018; Beltrama & Xiang 2013).
- Polarity (Gotzner et al. 2018).

But: still a lot of variance unaccounted for in the empirical results. That is, a lot of scalar diversity is unexplained.
The role of context

**QUDs have an effect on the rate of SI calculation:**

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

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

Higher SI rate in (4) than in (5) (i.a. Cummins & Rohde 2015; Degen & Tanenhaus 2014; Ronai & Xiang 2020; Yang et al. 2018; Zondervan et al. 2008).
The role of context

In previous work on scalar diversity: stimulus sentences presented in the absence of context.

Open question: is there variation across scalar terms in what kind of QUD they most naturally bring to mind?
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?*...
- ...the higher the rate of SI calculation from *She is intelligent.*
Overview of experiments

- Experiment 1: replication.
- Experiment 2: QUD manipulation.
- Experiment 3: question availability.
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 different scales.

Mary: *Success is possible.*

Would you conclude from this that, according to Mary, success is not certain?

- “Yes” response = SI was calculated.
- “No” response = SI was not calculated.
Replicated the scalar diversity effect.
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 embedded in a dialogue context.
  - Question containing stronger scalar: *Is the solution perfect?*
  - Question containing weaker scalar: *Is the solution good?*

```
Sue: Is the solution perfect?
Mary: It is good.

Would you conclude from this that, according to Mary, the solution is not perfect?

Yes.  No.
```
Experiment 2 results

Across the board:
- More SIs when the preceding question contains the stronger scalar term.
- Significant effect of Question ($p < 0.001$).

Explicit QUD influences SI rates for a large number of scales.
Experiment 3: question availability

- 35 native speakers of American English; MTurk; IbexFarm.

- Forced choice task: participants had to choose which of two polar questions (containing the stronger vs. the weaker scalar term) they would be more likely to ask.

Compare the following two questions about a girl. Which one are you more likely to ask?

1. Is the girl beautiful?
2. Is the girl pretty?

**Prediction**: forced choice results (Question Availability) should predict scalar diversity.
- The more preferred the stronger question (Exp. 3), the higher the SI rate (Exp. 1).
Question Availability not an overall predictor of SI rates, contra our predictions.
Bounded scale: the stronger scalar denotes an endpoint, e.g. \( all \rightarrow <\text{some}, \ all> \) is bounded.

Unbounded scale: both scalar terms denote an interval, e.g. \( \text{brilliant} \rightarrow <\text{intelligent}, \ \text{brilliant}> \) is unbounded.
Bounded scales: higher SI rates than unbounded ($p < 0.001$; replicates van Tiel et al. 2016).

Question Availability-Boundedness interaction ($p < 0.05$).
Interaction of Question Availability with Boundedness

Significant interaction of Question Availability and 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 question (Is the student brilliant?), the higher the rate of calculating the relevant SI (intelligent $\rightarrow$ not brilliant).

- Bounded scales: no effect of Question Availability ($p = 0.14$).
Sketch of an account

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

- The stronger term is **very salient** as an alternative to the vague, weaker term (see van Tiel et al.’s “distinctness”).
- → High rates of SI calculation; the **QUD makes no difference**.

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

- Salience of stronger alternative is subject to **contextual support**.
- 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.
Open questions

Other types of QUDs for Exp. 2, e.g. setting up biasing contexts without explicitly mentioning the scalar terms.

▶ Distinguishing relevance implicature from scalar implicature.

Better empirical measure of question availability than Experiment 3?

▶ Which question a speaker is more likely to choose may itself be context-dependent.
Conclusion

- **QUDs robustly affect SI calculation rates for a large number of scales:** questions based on the stronger of two scalar terms lead to higher SI rates.

- **Likelihood of a question based on the stronger scalar contributes to scalar diversity**, but only for **unbounded scales**.
Thank you!
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.

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.

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.

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

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.