

How many response options in a TVJT? It depends

Yuhan Zhang, Giuseppe Ricciardi, Kathryn Davidson



Department of Linguistics
Harvard University

Meaning & Modality (M&M) Lab

Experimental design features

Affect participant behavior in ways that we still need to understand better in order to link directly to syntactic, semantic, and pragmatic theory.

(Sprouse & Almeida 2017; Jasbi et al. 2019; Davidson 2020; Marty et al. 2020; Waldon & Degen 2020, among many others)

Today's question:

How do we choose the *right number of response options* for experimental tasks involving evaluating a sentence given a context?

- The most sensitive measures
- The most natural/intuitive for participants
- Whatever the lab is used to doing
- Depend on the semantic/pragmatic phenomenon ...if so, how?

Existing scalar implicature tasks

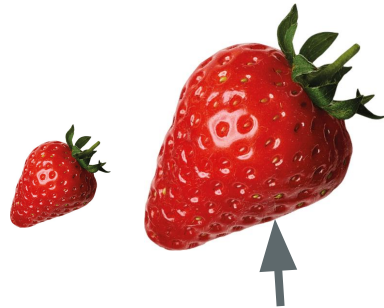


Picture 1. Starting point for scalar.

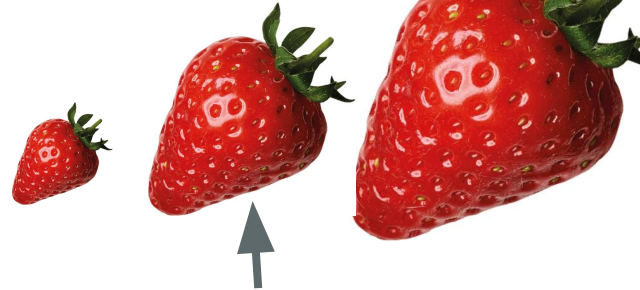


Picture 2. Outcome for scalar.

Caveman says: "The mouse picked up some of the carrots"



2 options: 70% pick large



3 options: 90% pick medium



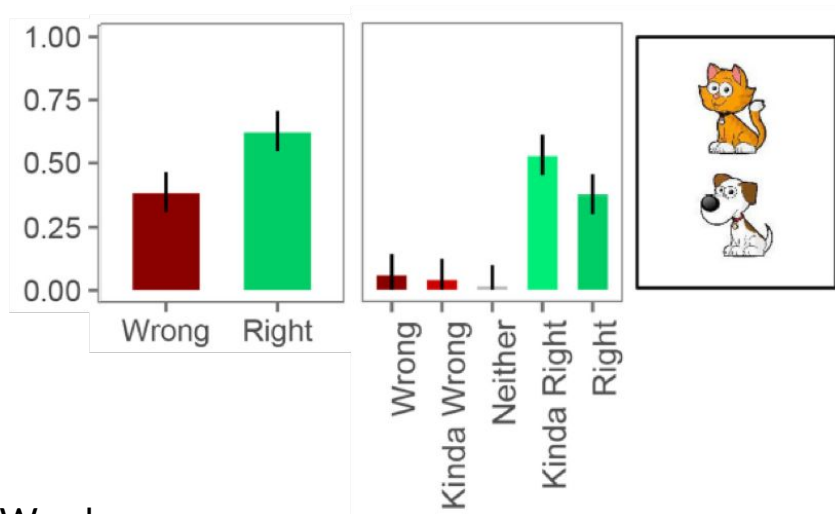
Our methodology

We compare adult responses given 2 vs. 5 options on sentence evaluation tasks given a context (roughly: TVJTs), across five different semantic phenomena:

- Three come from Jasbi, Waldon, & Degen 2019
 - *Scalar implicature of 'or'*
 - *Ad hoc scalar implicature*
 - *Boolean conjunction*
- Two from our experiments
 - *De re definite DPs*
 - *Novel de dicto type definite DPs*

Classic scalar implicature & Ad hoc scalar implicature (J19)

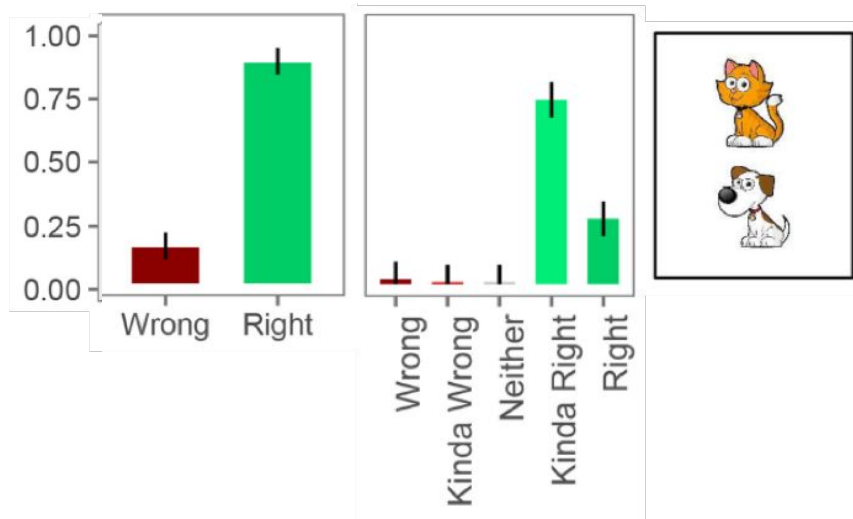
Guess: "There is a cat or a dog."



We observe:

- ✓ Participants are aware of the pragmatic oddity.
- ✓ Choosing binary "wrong" was merely a resistance to choosing "right" in the binary condition.

Guess: "There is a cat."



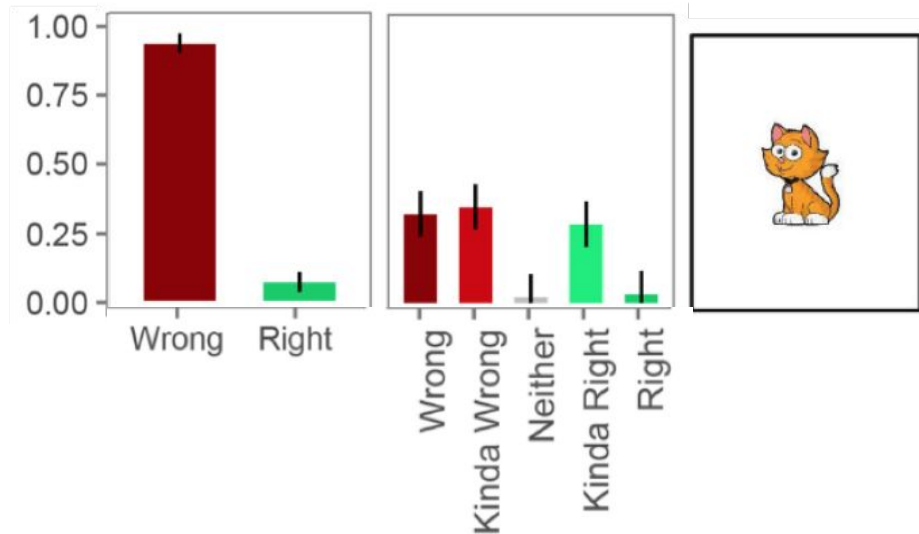
We observe:

- ✓ Binary scale suggests no implicature, but quinary has same distribution as the classic scalar implicature ("kinda right"!).

➔ **Quinary is more sensitive**

Conjunctive statements (J19)

Guess: "There is a cat and a dog."



We observe:

- ✓ Intermediate options create a task demand not fitting the immediate theoretical goal

➔ **Quinary is misleading**

New data: *De re* interpretations

Sentence-context evaluation task

Context: Julie is one of several judges of an ongoing poetry competition. The best poem that she's read so far is an extremely intriguing poem about the ocean. She believes that this poem will win the competition. Julie remembers being told that Nicole, one of the best-known contemporary poets, submitted a poem about the ocean to the competition. Therefore, Julie concludes that the first prize will be going to Nicole. However, this poem was actually written by Elizabeth, a younger and lesser-known poet. It is just a coincidence that the two poets wrote on the same topic.

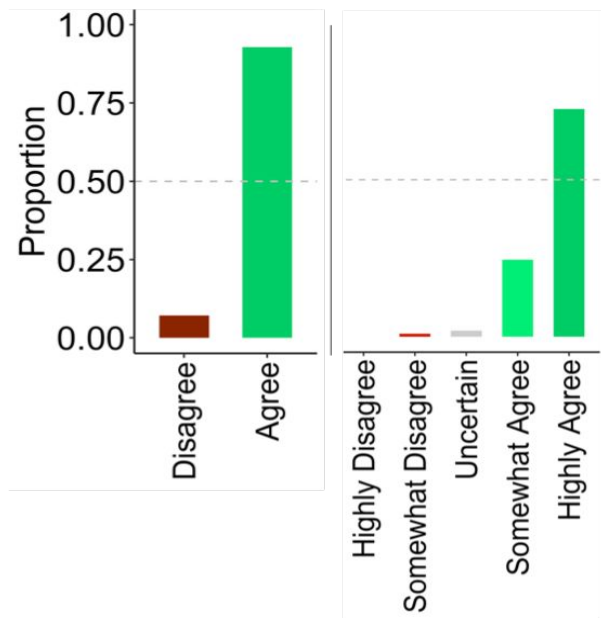
Please indicate whether/to what extent you agree or disagree with the following statement.

S_{Target} : Julie believes that [Nicole's poem] is going to win the competition. (*de dicto*-type)

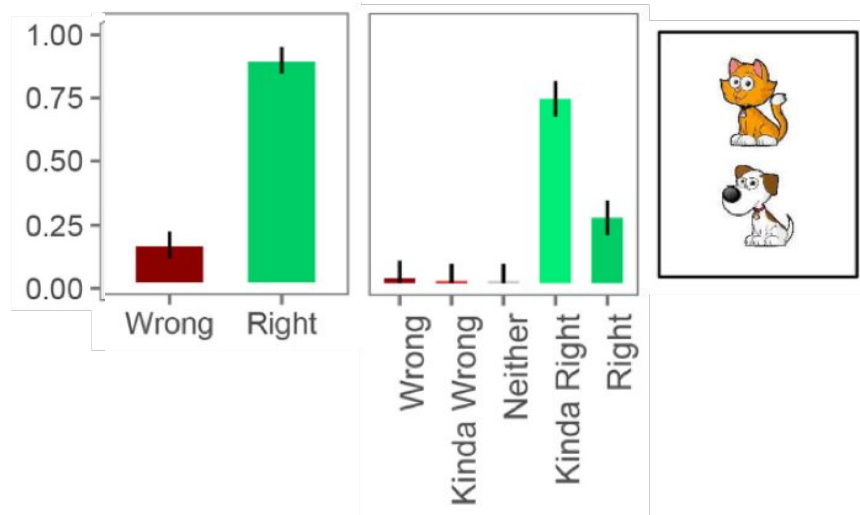
S_{Target} : Julie believes that [Elizabeth's poem] is going to win the competition. (*de re*)

Similar responses to binary, different responses to quinary

De dicto-type: Julie believes that Nicole's poem is going to win the competition.



Ad hoc scalar implicature: "There is a cat."



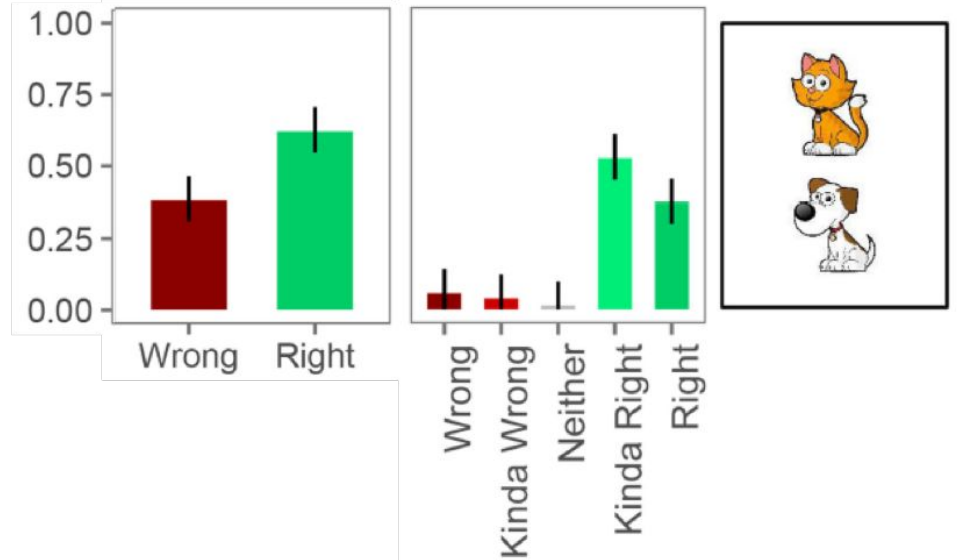
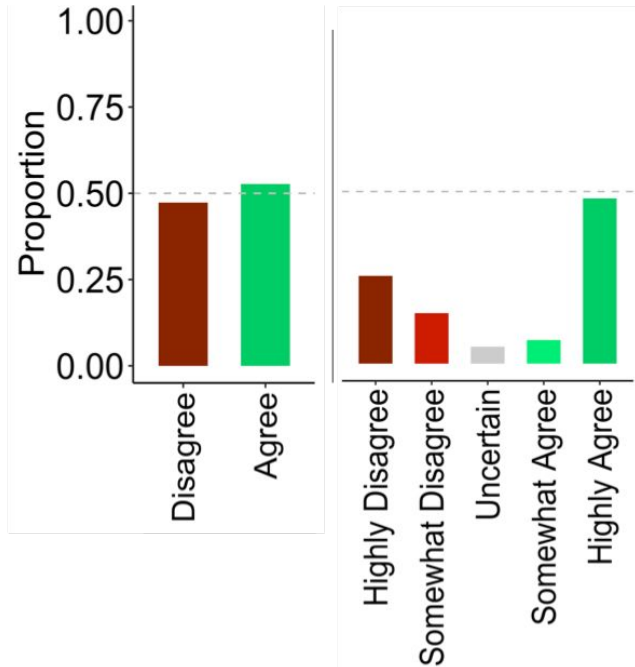
→ Binary and quinary are equally informative.

→ Quinary is more sensitive

Similar responses to binary, different responses to quinary

De re: Julie believes that Elizabeth's poem is going to win the competition.

Classic scalar implicature: There is a cat or a dog.



Our summary

	Phenomenon		
<i>Jasbi et al.</i>	Classic scalar implicature	Binary ?	Quinary ✓
	Ad hoc scalar implicature	Binary ?	Quinary ✓
	Conjunctive statement	Binary ✓	Quinary ✗
<i>New data</i>	De Dicto-type	Binary ✓	Quinary ✓
	De Re	Binary ?	Quinary ✓

✓ informative, accurate. ? underinformative. ✗ misleading

Take-away:

- (1) It is **NOT** the case that quinary scale is always better; it depends on the phenomenon under investigation and the research question.
- (2) To really ensure the understanding of a phenomenon, **we'd better manipulate the number of response options in a sentence evaluation task.**

Acknowledgements & References

Special thanks to members at the Meaning & Modality Lab in the Department of Linguistics at Harvard University. This project was funded by the Institute for Quantitative Social Science at Harvard.

References

- Davidson, K. (2020). Is “experimental” a gradable predicate? *Proceedings of NELS 50*.
- Jasbi, M., Waldon, B., & Degen, J. (2019). Linking Hypothesis and Number of Response Options Modulate Inferred Scalar Implicature Rate. *Frontiers in Psychology*, 10, 189. <https://doi.org/10.3389/fpsyg.2019.00189>
- Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1), 67–81. <https://doi.org/10.1016/j.cognition.2011.02.015>
- Marty, P., Chemla, E., & Sprouse, J. (2020). The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: A Journal of General Linguistics*, 5(1), 72. <https://doi.org/10.5334/gjgl.980>
- Sprouse, J., & Almeida, D. (2017). Setting the empirical record straight: Acceptability judgments appear to be reliable, robust, and replicable. *Behavioral and Brain Sciences*, 40, e311.
- Tonhauser, J., & Matthewson, L. (2015). *Empirical evidence in research on meaning*.
- Waldon, B., & Degen, J. (2020). Modeling Behavior in Truth Value Judgment Task Experiments. <https://doi.org/10.7275/SG32-AQ30>