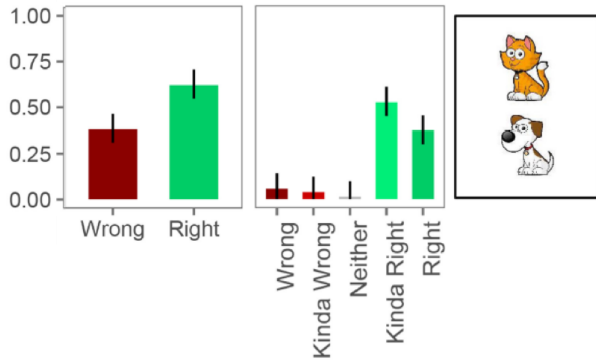# How many response options in a TVJT? It depends

Yuhan Zhang, Giuseppe Ricciardi & Kathryn Davidson (*Harvard University*)

Investigations of compositional semantics and pragmatics rely on quantitative data collection, and there is increasing awareness that basic task features affect participant behavior in ways that crucially bear on theoretical conclusions (e.g. Katsos & Bishop 2011; Sprouse & Almeida 2017; Jasbi et al. 2019; Davidson, 2020; Marty et al. 2020; Waldon & Degen 2020). Here, we focus on the effect of the number of response options in sentence evaluation tasks with a context (i.e., TVJTs/"truth value judgment tasks") by comparing adult participant behavior with two and five options across five different semantic phenomena: three data points come from Jasbi et al. 2019 (scalar implicature of 'or', *ad hoc* scalar implicature, conjunction) and two from novel experiments (*de re* and *de dicto* definite DPs). We argue that when it comes to deciding how many response options to offer in a TVJT, the most insightful practice is to manipulate their number.
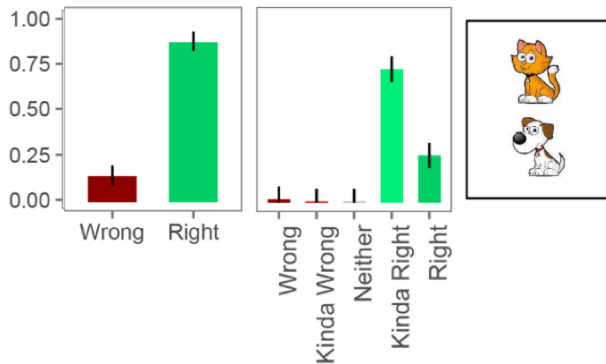
A: lexical scalar implicature. In Jasbi et al. 2019, participants evaluated the guesses of a blindfolded character about the content of a card. When the guess was "There is a cat or a dog" and the card contained a cat and a dog (Fig 1), 38% of the participants selected "wrong" in the binary condition, but in the quinary almost everyone chose either "kinda right" or "right". This suggests that choosing binary "wrong" was a resistance to choosing "right"; in the quinary, this judgment was realized instead as "kinda right". **B: *ad hoc* scalar implicature.** When the character guessed "There is a cat" when the card had a cat and a dog (Fig 2), most participants selected "right" in the binary condition, while in the quinary, instead, the majority selected "kinda right". The binary pattern (falsely) suggests that many English speakers are not sensitive to the pragmatic oddity of the sentence; in contrast, in the quinary condition participants are clearly aware that it is not entirely "right". **C: conjunction.** Here, the character guessed "There is a cat and a dog" when only a cat was pictured (Fig 3). Participants overwhelmingly judged this "wrong" in the binary condition, as expected based on the semantics of the boolean connective; however, in the quinary condition many participants chose "kinda wrong" and "kinda right", perhaps induced by the presence of intermediate options to analyze the conjunctive statement as a sequence of independent statements, one being true ("there is a cat") and the other false ("there is a dog"). In this case, intermediate options create a task demand not fitting the immediate theoretical goal. **D: *de re* definite DP.** In the novel task, participants were asked to evaluate a belief statement where the subject of the embedded clause was a definite DP interpretable as *de re* or *de dicto* depending on the context (Table 1). As in Jasbi et al. 2019, we manipulated between-subjects the number of response options (binary vs. quinary). In the binary condition of *de re* trials (Fig 4), one can observe a bimodal pattern, similar to that of case A. However unlike case A, this pattern persists in the quinary condition, suggesting that in the *de re* case what is underlying participants behavior is inherent disagreement about the truth/falsity of the sentence, rather than sensitivity to the pragmatic oddity of the sentence as in case A. **E: *de dicto* definite DP.** In the *de dicto* trials (Fig 5), most participants accepted the sentence in the binary condition, superficially similar to the binary condition of case B; however in this quinary condition, participants did not shift to the intermediate option, which we take to be due to *de dicto* sentences being judged as both true and pragmatically felicitous.

The overall picture is that more intermediate response options can reveal multiple underlying patterns, which may be due to TVJTs relying sometimes on primarily pragmatic judgments (see cases A & B) and other times on semantic ones (see cases D & E). Based on this, one might be inclined to give up binary TVJTs altogether. However, as shown by case C, the presence of more response options can, depending on the properties of the phenomenon under investigation and the specific experimental setup, induce additional unintended inferences. Therefore, we conclude that, especially for understudied phenomena in semantics and pragmatics, it may be most informative to design TVJTs by manipulating the number of response options and draw conclusions based on a comparison across conditions.
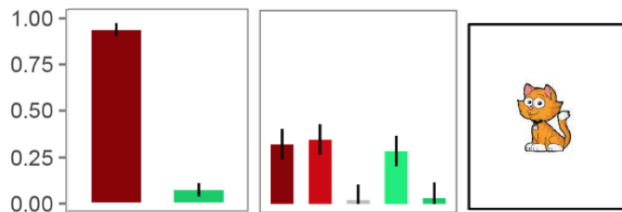
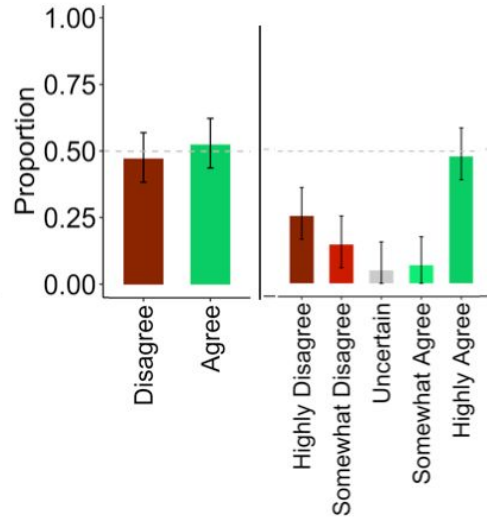**(A) Fig. 1** Lexical scalar implicature
"There is a cat or a dog."

(B) Fig. 2 *Ad hoc* scalar implicature
"There is a cat."

(C) Fig. 3 Conjunctive statement
"There is a cat and a dog."

(D) Fig. 4 *De re* reading of Definite DPs
"...believe...Elizabeth's poem..."

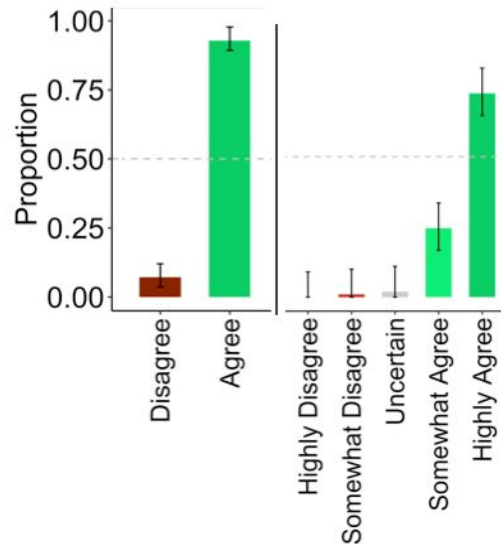**(E) Fig. 5** *De dicto* reading of Definite DPs
"...believe...Nicole's poem..."

Table 1 An example context of experiments testing *de re*/*de dicto* readings of definite DPs

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