

Supplementary files

The four tasks in Sprouse and Almeida (2017)

Forced-Choice task

In the (two-alternative) FC task, participants are presented with pairs of test sentences and asked, for each pair, to compare both sentences and decide which of them is more acceptable or natural-sounding. In formal studies like S&A's, test sentences are presented in vertically arranged pairs, with each sentence in the pair followed by a single radio button, as shown in (1). For each pair, participants report their judgment by selecting the radio button next to the sentence they believe to be more acceptable. Crucially, each pair of sentence types is designed to be as structurally and lexically similar as possible so as to form a syntactic minimal pair that varies only by the structural property of interest. For instance, in the case of (1), the structural property of interest is the ordering of the two *wh*-phrases.

$$(1) \quad \text{Item} = \left\{ \begin{array}{ll} \text{Which table did Peter put what on?} & \circ \\ \text{What did Peter put on which table?} & \circ \end{array} \right\}$$

Magnitude Estimation task

In the ME task (Stevens, 1957; Bard, Robertson, & Sorace, 1996), participants are presented with a single test sentence presented together with a reference sentence, called the standard, which is pre-assigned an acceptability rating called the modulus (set at 100 in S&A). The standard is generally chosen such that it is in the middle range of acceptability and theoretically unrelated to the test sentence. Participants are asked to use the standard to estimate the acceptability of the test sentence by providing a numerical score that is a multiple of the modulus. For example, in the case of (2), if the participant believes that the test sentence is twice as acceptable as the standard, then she will rate the test sentence as 200.

$$(2) \quad \text{Item} = \left\{ \begin{array}{ll} \text{Who was kept tabs on by the FBI?} & 100 \\ \text{What did Peter put on which table?} & \square \end{array} \right\}$$

Yes-No task

In the standard YN task, each experimental item is a single test sentence presented together with a pair of response options labelled 'yes' and 'no', as shown in (3). Participants are asked to use these two options to indicate whether the test sentence is acceptable or not.

$$(3) \quad \text{Item} = \left\{ \text{What did Peter put on which table?} \quad \text{No } \circ \quad \text{Yes } \circ \right\}$$

Likert Scale task

In the standard LS task, each experimental item is a single test sentence presented together with a series of response options forming a graded response scale. Participants are asked to use these options to indicate their acceptability judgments. In S&A's study, the response scale consisted of 7 response options labeled 1-7, with 1 labeled 'least acceptable' and 7 labeled 'most acceptable', as shown in (4).

$$(4) \quad \text{Item} = \left\{ \text{What did Peter put on which table?} \quad \text{least acceptable} \quad \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ \end{array} \quad \text{most acceptable} \right\}$$

References

- Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72(1), 32--68.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological review*, 64(3), 153.