Assessing Validity: Instructions and the Grammaticality Judgment Task

Parker T. Robbins
Northeastern University (IRB Approval # 18-12-20)

Cornell Undergraduate Linguistics Colloquium 13
Sunday, 28 April 2019

Background: The Grammaticality Judgment Task

All speakers have intuitions about what sounds *natural*.

With the grammaticality judgment task, we can:

- Systematically ask participants to judge sentences.
- Use this information to test our hypotheses about syntax.

Sample Grammaticality Judgment Task:

Hypothesis: Size adjectives must precede color adjectives.

Test: Which sentence sounds more natural?

(1) The big black clouds loomed over the castle.

(2) The black big clouds loomed over the castle.

Assessing the Validity of the Judgment Task: Previous Research

As the subfield of experimental syntax began to develop, researchers:

- Began to investigate how experimental control would increase the reliability of judgments.
- Found a number of task- and participant-related factors that can affect judgments and some that do not (see Schütze, 2016, for a review).
    - Bader and Haeussler (2010): Judgments are relatively stable across judgment task type.
    - Snyder (2000): Evidence for syntactic satiation, or repetition effects.

Prescriptive Rules: An Understudied Threat to Reliability

Exclusive use of prescriptive rules threatens the reliability of judgments.

> Thou shalt not begin a sentence with a conjunction.
> Thou shalt not split infinitives (e.g., to *boldly* go).
> Thou shalt not end a sentence with a preposition.

Research question

Can task instructions affect participants' use of prescriptive grammar rules in judgment studies?

Previous Research: Cowart (1997)

Only one previous study: Cowart (1997).

Conditions

Intuitive and *Prescriptive* instruction types.

Task

Rate stimuli on a five-point Likert-type scale.

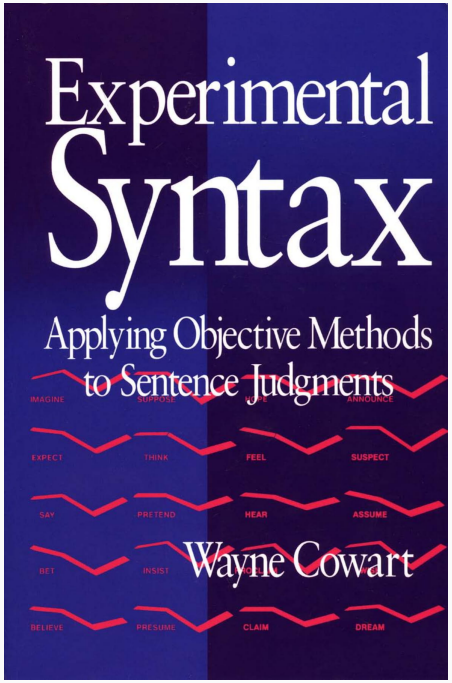
Stimuli

Binding with local and remote antecedents.

(3) a. Cathy's parents require that Paul support himself.

b. *Paul requires that Cathy's parents support himself.

(Cowart, 1997, p. 20)



Source: Google Books

Cowart (1997): Results and Discussion

Finding

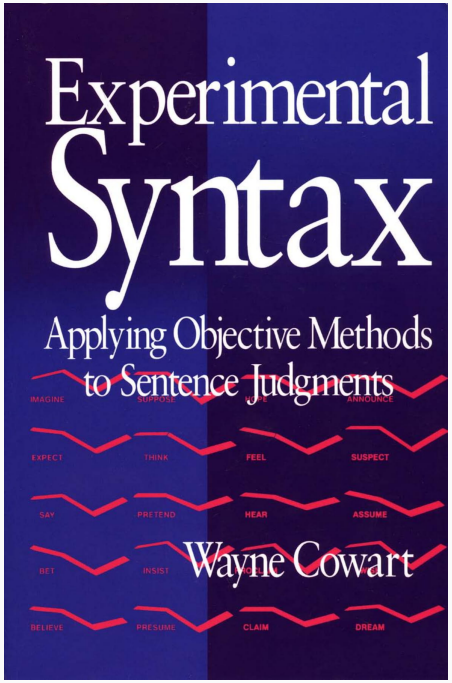
No effect of instruction type.

Significance

His results have largely been taken as definitive (Schütze & Sprouse, 2014).

Limitations

- Too few participants for a Likert-type scale task.
- Stimuli did not test the application of prescriptive rules.
- *Intuitive* instructions specifically referenced “school grammar.”



Source: Google Books

The Present Study: Objective and Procedure

Objective

Reproduce Cowart (1997) with:

- More participants.
- Improved stimuli.
- New instruction conditions.

Procedure

1. Instructions given.
2. Forced choice between sentence pairs.
3. Demographic survey.

The Present Study: Instruction Conditions

Instruction Types

Professor
Imagine you are an English professor...

Tutor
Imagine you are tutoring a friend learning English...

Intuitive Plain
Give your "gut reaction."

Intuitive School
Give your "gut reaction" and do not use "school grammar."

The Present Study: Sentence Pair Stimuli

Both sentences are grammatical:

Prepositional verb sentence pairs.

(4) The master violinist was a musician whom few professionals could really compare to.

(5) The master violinist was a musician to whom few professionals could really compare.

Prescriptive rule creates ungrammatical sentence:

Particle-prepositional verb sentence pairs.

(6) Your strange cousin is a person whom the family must reluctantly put up with.

(7) * Your strange cousin is a person up with whom the family must reluctantly put.

The Present Study: Deployment and Participants

Participants

200 workers on Amazon Mechanical Turk.

Amazon Mechanical Turk

- Platform for workers to pick up small tasks.
- Validated for linguistic research (Schnoebelen & Kuperman, 2010).
- Validated for acceptability judgments (Sprouse, 2011).
- Quick and cheap data collection.

Source: Amazon

The Present Study: Deployment and Participants

Measures to ensure data quality

- Participant recruitment measures
  - Workers who have a 95 percent acceptance rate on past tasks.
  - Only recruited native English speakers located in the US.
  - Participants excluded if they had reported having taken a linguistics course.
- Task design measures
  - One instruction comprehension check.
  - Two attention check questions.
  - Pseudorandomized design.

Source: Amazon

The Present Study: Hypotheses and Analysis

Hypotheses

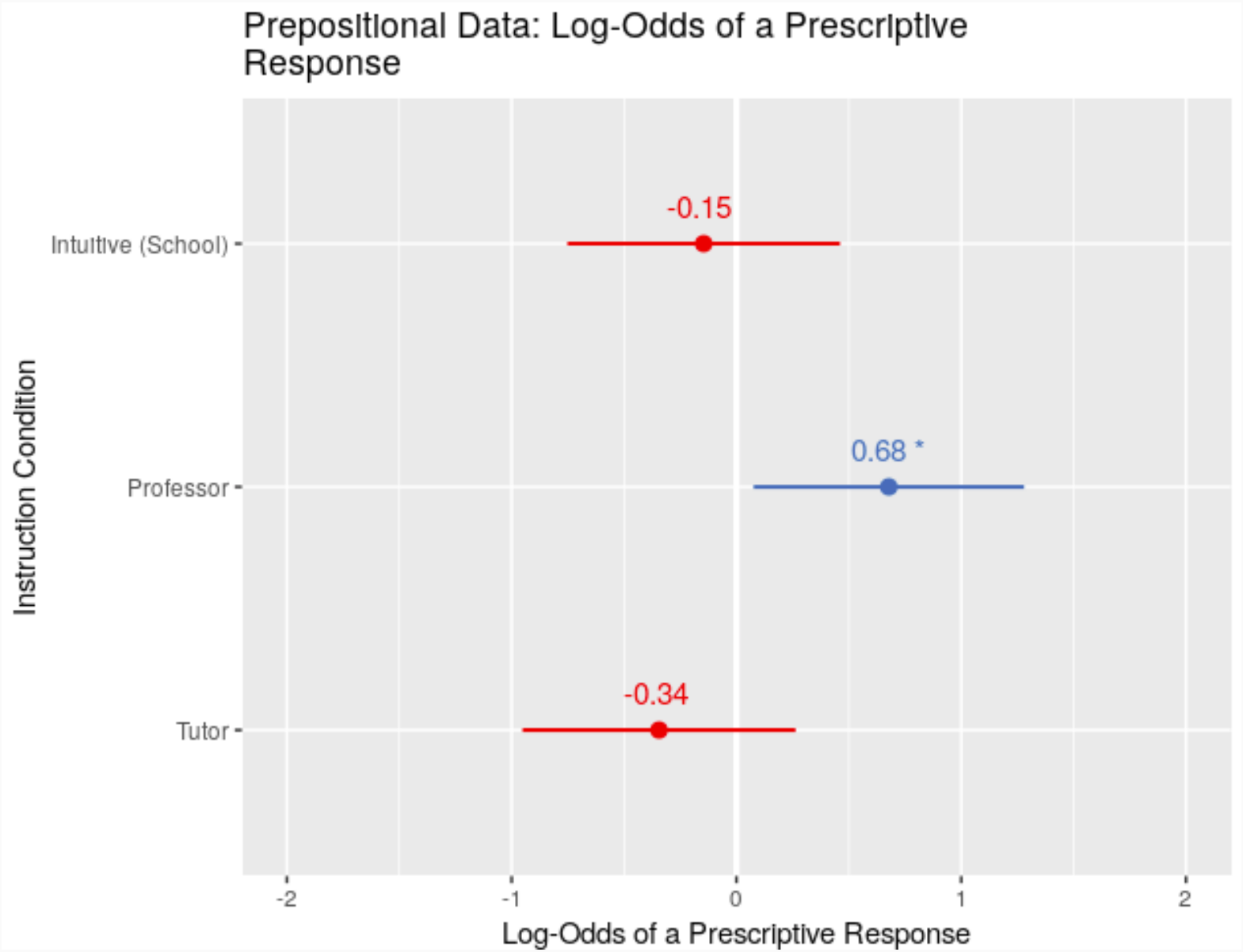
Participants in the professor condition are more likely to choose prescriptively-correct sentences when:

1. Both sentences are grammatical (i.e., prepositional verbs).
2. Applying a prescriptive rule would make a sentence ungrammatical (i.e., particle-prepositional verbs).

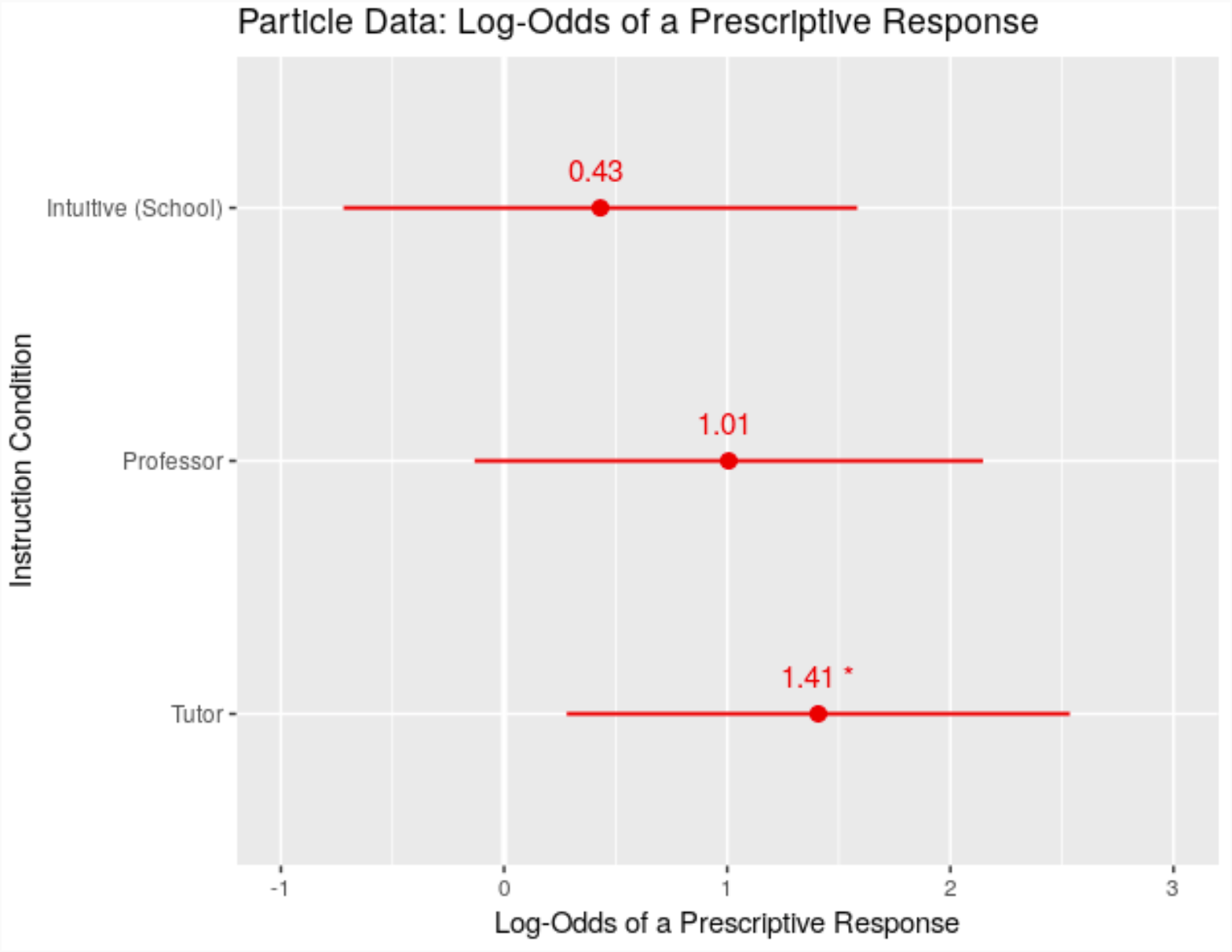
Analysis.

- Mixed-effects logistic regression:
  - Participant was set as a random factor.
  - Instruction conditions were fixed factors.
  - Reference condition (baseline): Intuitive Plain.
- Conducted using the `lme4` package for R.

The Present Study: Prepositional Verb Results



The Present Study: Particle-Prepositional Verb Results



Summary of Results and Implications

Results

- Prepositional verb pairs: Professor condition participants were more likely to respond prescriptively.
- Particle-prepositional verb pairs: Tutor condition participants were more likely to choose an ungrammatical prescriptive response.

Implications

- More care should be taken in designing judgment task instructions.
- Participants *can* modify judgment criteria to some extent.

Directions for Future Research

1. Prescriptive rules are not relevant to all judgment tasks, but in minority language or dialect contexts, they may be quite important.
2. How is age associated with use of prescriptive judgments?
3. How might level of education play a role?
4. Why did the tutor condition prefer ungrammatical sentences?

References

Bader, M., & Haeussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, *46*, 273–330.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. London: Sage.

Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, *43*(4), 441–464.

Schütze, C. T. (2016). *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.

Schütze, C. T., & Sprouse, J. (2014). Judgment data. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). Cambridge: Cambridge University Press.

Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, *31*(3), 575–582.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, *43*(1), 155–167.

Acknowledgments

This project would not be possible without:

Northeastern's Office of Undergraduate Research and Fellowships, for funding data collection.

Northeastern Honors Program, for travel funding.

Professor Heather Littlefield, my project Advisor.

Professor Adam Cooper, my project's Second Reader.

Thank you so much!

Questions?