Setting the empirical record straight: Acceptability judgments appear to be reliable, robust, and replicable.

Jon Sprouse  
Diogo Almeida

Department of Linguistics, University of Connecticut  
Department of Psychology, New York University Abu Dhabi

Department of Linguistics  
University of Connecticut  
365 Fairfield Way, Unit 1145  
Storrs, CT 06269-1145

167, Computational Research Building,  
New York University Abu Dhabi  
P.O. Box 129188, Abu Dhabi  
United Arab Emirates

+1.860.486.6864  
+971.2.628.5040

jon.sprouse@uconn.edu  
diogo@nyu.edu

www.sprouse.uconn.edu  

Abstract

Branigan and Pickering advocate the use of syntactic priming to investigate linguistic representations, and argue it overcomes several purported deficiencies of acceptability judgments. While we recognize the merit of drawing attention to a potentially underexplored experimental methodology in language science, we do not believe the empirical evidence supports Branigan and Pickering’s claims about acceptability judgments. We present the relevant evidence.

Main Text

Branigan and Pickering (B&P) advocate the use of syntactic priming to investigate linguistic representations. We support the use of any data types that scientists find relevant for specific research questions, including syntactic priming. We regret then that B&P appear to repeat unsubstantiated claims that paint a relatively misleading picture of acceptability judgments (AJs),
a data type that linguists have been using fruitfully for decades. From our perspective, much of the literature criticizing AJs has repeatedly focused on logically possible concerns about their use without investigating whether those concerns are empirically attested. This risks a vicious circle: articles can cite each other for support, giving the illusion of empirical support. In this commentary we would like to highlight a number of studies that have pursued this issue head on, which we leverage to examine six of B&P’s claims about AJs in detail.

Claim 1: Linguists standardly ask a single informant about the acceptability of a few sentences. (p12)

Claim 1 is a caricature of linguistic methodology that has never, to our knowledge, been supported by evidence. Nonetheless, a charitable interpretation of this claim reveals two separate concerns: (i) the routine use of small sample sizes and (ii) the susceptibility of AJs to investigator bias (claim 2 below). An obvious consequence of using small samples sizes in research is an increase in errors (probably of all four types identified by Gelman and Carlin 2014: I, II, Sign, and Magnitude). By performing a large-scale comparison of the published results in linguistics with re-tests of those results using large samples of naïve participants, one can evaluate the quality of their convergence rate. This cannot identify specific errors, but it can tell us whether the differences between methods actually produce different results.

Sprouse and Almeida 2012 tested every English data point from a popular syntax textbook (Adger 2003) using large samples of naïve participants. Out of 365 phenomena, they conservatively estimate a minimum convergence rate of 98%. Sprouse et al. 2013 randomly sampled 148 phenomena from a leading linguistics journal (Linguistic Inquiry), and conservatively estimate a convergence rate of 95% (± 5% because of the random sampling). These high (conservative) convergence rates suggest that the sample sizes used by linguists (whatever they are) have historically introduced little error to the empirical record, either because (i) the samples are larger than what critics claim, (ii) the effect sizes are so large that small samples still yield good statistical power, or (iii) AJ results are highly replicated before and after publication (e.g., Phillips, 2009), or any combination of (i-iii).

Claim 2: Acceptability judgments are highly susceptible to theoretical cognitive bias because linguists tend to use professional linguists as participants. (p13)

This can also be addressed by the studies discussed above. Cognitive bias should predict sign reversals between naïve and expert populations. Sprouse and Almeida 2012 found no sign reversals from the textbook data. Sprouse et al. 2013 report a 1%-3% sign reversal rate in the journal data. Mahowald et al. 2016 and Häussler et al. 2016 have replicated the latter without reporting an increased sign reversal rate (0%-6%). Comparisons of naïve and expert populations were also conducted by Culbertson and Gross 2009, who report high inter and intra-group correlations on 73 sentence types, and by Dąbrowska 2010, who found that while experts gave less variable ratings than naïve participants on several sentence types, experts rated certain theoretically interesting syntactic violations as more acceptable than naïve participants, in apparent conflict with their theoretical commitments. Taken together, these results are not what one would expect if AJs were highly susceptible to cognitive bias.

Claim 3: Acceptability judgments are susceptible to differences in instructions. (p13)
Claim 3 has been directly investigated by Cowart 1997, who reports that the systematic manipulation of instructions does not change the pattern of acceptability judgments for factorial designs.

Claim 4: Acceptability judgments are impacted by sentence processing effects. (p13)

Claim 4 is technically true, but B&P exaggerate its consequences. First, many classic lexical and sentence processing effects have relatively small or negligible effects on acceptability (e.g., Featherston 2009, Phillips 2009, Sprouse 2008, Sprouse et al. 2012). Second, very few syntactic phenomena have been proposed to be fully reducible to sentence processing effects. The lone exceptions to this appear to be constraints on long-distance dependencies (e.g., Kluender and Kutas 1993, Hofmeister and Sag 2010), but in that case a number of experimental studies have disproven the reductionist predictions (Phillips 2007, Sprouse et al. 2012, Yoshida et al. 2013). Thus, to the extent that AJs are impacted by sentence processing, it appears as though the effects can be dealt with like any other source of noise in an experimental setting.

Claim 5: Acceptability judgments only reveal set membership. (p14)

Claim 5 is confusing. It is false in the sense that, if one is interested in set membership, this property still needs to be inferred from acceptability data, using a logic that maps that data type back to the relevant cognitive computations. In this, AJs are like any other data type in cognitive science: no data types, including priming, directly reveal the underlying computations of the human brain, and all data types require a linking hypothesis between the observable data and the unobservable cognitive process.

Claim 6: Acceptability judgment have yielded no consensus theory among linguists. (p15)

Claim 6 is a strange criticism to make of any data type, especially AJs. First, the beliefs of scientists are a subjective issue based on how they weigh different kinds of evidence. Second, AJs are by all accounts a robust and replicable data type. Whatever disagreements there are in linguistics literature, they appear to obtain mostly at the level of interpreting, not establishing, the data (e.g., Phillips, 2009).

In conclusion, we support B&P’s desire to bring new evidence to bear on questions about linguistic representation, but we caution that advocacy for one method should not be bolstered by misleading comparisons, especially with methods like AJs, which yield data that is demonstrably robust, highly replicable, and comparatively convenient and inexpensive to collect.

References


