to describe a picture with a prepositional dative after hearing a sentence fragment that started as a prepositional dative but was corrected to a transitive (e.g., *The mechanic is giving the new part… uh… is recognizing the new part*) than after a fragment starting as a double-object dative and corrected to a transitive (Slevc & Ferreira 2013). Similarly, participants completed more sentences as transitives after temporarily ambiguous sentences such as *While the man was visiting the children who were surprisingly pleasant and funny played outside* than after identical sentences disambiguated by a comma (van Gompel et al. 2006). Importantly, the priming effects observed in these studies reflected temporarily suboptimal or erroneous parses that arose from the processing demands of online parsing.

Of course, structural priming reflects active processing even in the absence of errors or temporary ambiguity. Thus, the role of underlying cognitive faculties is a second aspect of structural priming that deserves more investigation. One such faculty is attention: Priming effects are greater when primes are directly attended (e.g., when a comprehender is addressed directly in a dialogue) compared to when she or he is not directly addressed (e.g., when a comprehender simply overhears a conversation). This result suggests that the depth of processing of a prime sentence directly affects the magnitude of priming (Branigan et al. 2007). A second relevant faculty is working memory. Although there is evidence that structural priming effects can be long lasting (e.g., Bock & Griffin 2000; Kaschak et al. 2011b) and may reflect implicit learning rather than short-term maintenance (e.g., Chang et al. 2006), working memory nonetheless has been implicated in structural priming. Ivanova et al. (2013) found that increased working memory demands during the *production* of target picture descriptions reduced priming (at least for datives in a picture description paradigm; note that priming the presence/absence of the complementizer *that* in a recall paradigm was unaffected by memory load). These data suggest that attention and memory demands can influence priming effects, although we still know very little about these influences.

These examples illustrate that the relationships among linguistic input, syntactic knowledge, and structural priming are indirect and mediated by processing constraints. This does not undermine the usefulness of structural priming to shed light on linguistic representations. It does suggest, however, that more work is needed to understand how specific task and stimuli details affect both parsing and structural priming. Of course, processing dynamics influence metalinguistic acceptability judgments as well (e.g., Lau & Ferreira 2005), and so it will be important to compare how processing demands affect these different paradigms. More generally, structural priming (like any method) has both advantages and limitations. We agree with Branigan and Pickering that it can be a useful tool to investigate the nature of linguistic representations, but we also caution that this tool still requires careful work to unpack the processes underlying our tendency to reuse recently experienced structure.

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

doi:10.1017/S0140525X17000590, e311

Jon Sprouse[a] and Diogo Almeida[b]

[a]*Department of Linguistics, University of Connecticut, Storrs, CT 06269-1145;* [b]*Department of Psychology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates.*

jon.sprouse@uconn.edu www.sprouse.uconn.edu
diogo@nyu.edu
http://nyuad.nyu.edu/en/academics/faculty/diogo-almeida.html

Abstract: Branigan & Pickering (B&P) advocate the use of syntactic priming to investigate linguistic representations and argue that 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 that the empirical evidence supports B&P's claims about acceptability judgments. We present the relevant evidence.

Branigan & 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 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 (sect. 1.2, para. 2). Claim 1 is a caricature of linguistic methodology that, to our knowledge, has never been supported by evidence. Nonetheless, a charitable interpretation of this claim reveals two separate concerns: (1) the routine use of small sample sizes, and (2) 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 & Carlin 2014: I, II, Sign, and Magnitude). By performing a large-scale comparison of the published results in linguistics with retests 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) historically have introduced little error to the empirical record for any combination of the following reasons: (1) the samples are larger than what critics claim; (2) the effect sizes are so large that small samples still yield good statistical power; or (3) AJ results are highly replicated before and after publication (e.g., Phillips 2009).

Claim 2: Acceptability judgments are highly susceptible to theoretical cognitive bias because linguists tend to use professional linguists as participants (sect. 1.2, para. 3). 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) reported a 1–3% sign reversal rate in the journal data. Mahowald et al. (2016a) and Häussler et al. (2016) have replicated the latter without reporting an increased sign reversal rate (0–6%). Comparisons of naïve and expert populations also were conducted by Culbertson and Gross (2009), who report high inter- and intra-group correlations on 73 sentence types, and by Dąbrowska (2010). The latter 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 (sect. 1.2, para. 3). 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 (sect. 1.2, para. 5). 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 & Kutas 1993; Hofmeister & Sag 2010), but in that case, a number of experimental studies have disproven the reductionist predictions (Phillips 2006; Sprouse et al. 2012; Yoshida et al. 2014). 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 reveal only set membership (sect. 1.2, para. 7). 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 judgments have yielded no consensus theory among linguists (sect. 1.2, para. 9). 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. We caution, however, that advocacy for one method should not be bolstered by misleading comparisons, especially with methods such as AJs, which yield data that are demonstrably robust, highly replicable, and comparatively convenient and inexpensive to collect.

Priming is swell, but it's far from simple

doi:10.1017/S0140525X17000607, e312

Jayden Ziegler,[a] Jesse Snedeker,[a] and Eva Wittenberg[b]
[a]*Department of Psychology, Harvard University, Cambridge, MA 02138;* [b]*Department of Linguistics, University of California, La Jolla, CA 92093.*
ziegler@g.harvard.edu **http://www.jaydenziegler.com**
snedeker@wjh.harvard.edu **https://software.rc.fas.harvard.edu/lds/research/snedeker/jesse-snedeker/** **ewittenberg@ucsd.edu**
http://www.evawittenberg.com/i/start.html

Abstract: Clearly, structural priming is a valuable tool for probing linguistic representation. But we don't think that the existing results provide strong support for Branigan & Pickering's (B&P's) model, largely because the priming effects are more confusing and diverse than their theory would suggest. Fortunately, there are a number of other experimental tools available, and linguists are increasingly making use of them.

Branigan & Pickering (B&P) tell a straightforward tale. Linguists rely on grammaticality judgments to uncover representations. Judgments have limitations, but no other psycholinguistic methods systematically reveal linguistic structure. Fortunately, priming offers a direct window onto representation, providing evidence for two distinct levels: a surface syntactic form, independent of meaning and void of lexical content, and a semantic form that includes information about thematic roles, quantifier scope, and information structure.

We are fond of priming ourselves, but this elegant story is misleading in several ways. First, the priming literature does not strongly support the theory that B&P propose. As they dive deeper, the loose ends and contradictions emerge, but their final conclusions bypass this complexity. If we rearrange the evidence a bit, the theoretical ambiguity becomes clearer.

The primary evidence for syntactic representations comes from studies of argument alternations (dative or active-passive) that perfectly confound *surface syntax* with *thematic mappings*. B&P note that a few foundational studies demonstrated that syntax *can* be primed independent of thematic mappings (sect. 2.1). Thus, they privilege syntax in their theory. But there is now an equally robust literature demonstrating that thematic mappings can be primed independent of syntax (e.g., Cai et al. 2012; Chang et al. 2003; Cho-Reyes et al. 2016; Hare & Goldberg 1999; Salamoura & Williams 2007; Ziegler & Snedeker 2016b). B&P acknowledge this work (sect. 2.4) but treat it as a secondary, interface phenomenon: Thematic information remains separate from syntax (Fig. 1).

Similarly, the observation that priming *can* occur in the absence of lexical overlap motivates a theory in which the syntactic skeleton is separate from the lexical content. To account for the lexical boost, B&P must complicate their story, by linking lemmas to structures (sect. 2.3). But perhaps we should revisit the claim that the syntactic structure lacks lexical nodes. Indeed, function words *can* be a locus of priming (Bencini et al. 2002; Ferreira 2003). We know that only partial overlap in the syntactic skeleton is needed for structural priming (sect. 2.1), but we don't assume that the unnecessary pieces are removed from the syntactic representation. Lexical content may be similar: always present and sometimes contributing to priming via overlap.

The evidence for their semantic level is also sparse. We know: (1) Quantifier scope can be primed, (2) this priming is isolated to the particular quantifier used (e.g., *each* does not prime *every*), and (3) it abstracts away from the nouns and verbs in a sentence (Feiman & Snedeker 2016; Raffray & Pickering 2010). However, B&P's claim that scopal priming is bound to thematic roles and cannot be captured by an LF representation is controversial (Chemla & Bott 2015). It rests on a single null result with prime stimuli (*A boy climbed every tree*) that have not been shown to produce priming when thematic roles are the same. Furthermore, the manipulation used confounds verb-specific roles, thematic roles, and the notion of deep subject/object. It's just too early to conclude that scope and thematic roles are tightly coupled, or that LF isn't the locus of scopal priming.

It seems that, under the right conditions, almost any linguistic representation, mapping, or process can be primed. Consequently, evidence for priming is always interpretable to some degree (it demonstrates a commonality between prime and target). But the absence (or magnitude) of an effect is often less constraining, because there is so much variability across tasks and stimuli. In some comprehension tasks, there is no priming in the absence of verb overlap (Arai et al. 2007), while in others, abstract priming is robust (Thothathiri & Snedeker 2008a; 2008b). This problem isn't unique to comprehension. The pattern of effects in production can depend on how the sentences are elicited (stem completion vs. full sentence generation; Ziegler & Snedeker 2016a).

Understanding this instability is critical; we suspect that the answer lies in thinking through the processes involved in each task and how they engage both stored representations and representations that are constructed on the fly. To do this, we will have to move beyond the notion of priming as a static, atemporal