Back to the Future: Do Influential Results from 1980s Psycholinguistics Replicate?

Fernanda Ferreira (fferreira@ucdavis.edu), Gwendolyn Rehrig, Madison Barker, Eleonora Beier, Suphasiree Chantavaran, Beverly Cotter, Zhuang Qiu (UC Davis), Matthew Lowder (U of Richmond), Hossein Karimi (Mississippi State University)

Background. In the 1980s, a prominent research question concerned the effects of discourse context on parsing decisions. Two highly influential and widely cited studies reported contradictory results: Ferreira and Clifton (1986; F&C86) conducted two experiments, one using eyetracking and the other using self-paced reading, in which minimal attachment (MA; syntactically easy) or nonminimal attachment (NMA; syntactically difficult) sentences were presented either in biased or neutral contexts, and they reported that helpful context affected later processing stages but not the parser's initial attachment decisions. In contrast, Altmann and Steedman (1988; A&S88) conducted a self-paced reading study in which MA and NMA sentences were embedded in appropriately or inappropriately biasing contexts, and they reported that context did drive the parser's initial structure-building operations.

Recently, experimental psychologists have been concerned with issues of replicability, with several reports of failures to replicate well-known findings (e.g. Stack et al. 2018). Replication has received less attention in psycholinguistics, which is a lost opportunity since our field is uniquely positioned to highlight the opportunities and challenges associated with conducting replication studies, particularly regarding issues of direct versus conceptual replication. Because research practices change, analysis techniques advance, and language evolves so that past stimuli may no longer appropriately instantiate key linguistic manipulations, direct replications are often difficult in psycholinguistics. It is important to ascertain whether past findings replicate given that some past studies may not conform to current best practices.

Method. The study was conducted as a single eye movement experiment and designed as a conceptual replication of F&C1986 and A&S1988. We view the replication as conceptual because, although the same design was used as in the original studies, a few essential changes were made: (a) the *N* was increased to 60; (b) the stimuli were normed; (c) sentences were updated to fit current cultural norms (e.g., sexist items were changed); and (d) analyses were conducted according to current approaches. The eyetracking measures included for analyses were those reported in F&C86: first-pass reading time, probability of a first-pass regression out of a region, and second-pass reading time. Norming data and accuracy were also analyzed.

Results. Behavioral results were as follows: First, analyses of norms suggest the contexts from both studies were less effective than assumed by the original investigators. For the F&C86 stimuli, context had no effect on offline ratings of the appropriateness of either the MA or NMA sentences; instead, overall, subjects rated MA sentences as better than NMA sentences regardless of context bias. For A&S, the NMA-biased contexts did support the NMA form, but raters given MA-biased contexts had no preference for either the MA or the NMA sentence. Question-answering accuracy did not differ across conditions either for F&C86 or A&S88 (contrary to F&C86). Eyetracking results for regressions and first-pass reading times are shown on the following page (Fig. 1). The F&C86 replication showed no clear pattern of results for first-pass reading times, and the likelihood of a first-pass regression was overall greater for NMA than for MA structures, regardless of context. For A&S88 stimuli, regression probability was higher for VP-attached (MA) than for NP-attached (NMA) forms, with no effect of context. First-pass reading times for A&S88 did not differ for either structure given NP-biased contexts and were faster for VP-attached (MA) sentences given VP-biased contexts.

Conclusions. The results of this replication study differed substantially from the findings reported in F&C86 and A&S88. The discrepancies are due to numerous factors including lack of norming data for contexts and low statistical power. Overall, replicability is an important issue in psycholinguistics, and we would suggest that psycholinguistics has much to contribute to discussions concerning how to conduct and evaluate replication studies.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*(3), 191-238.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348-368.

Stack, C. M. H., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & cognition*, *46*(6), 864-877.

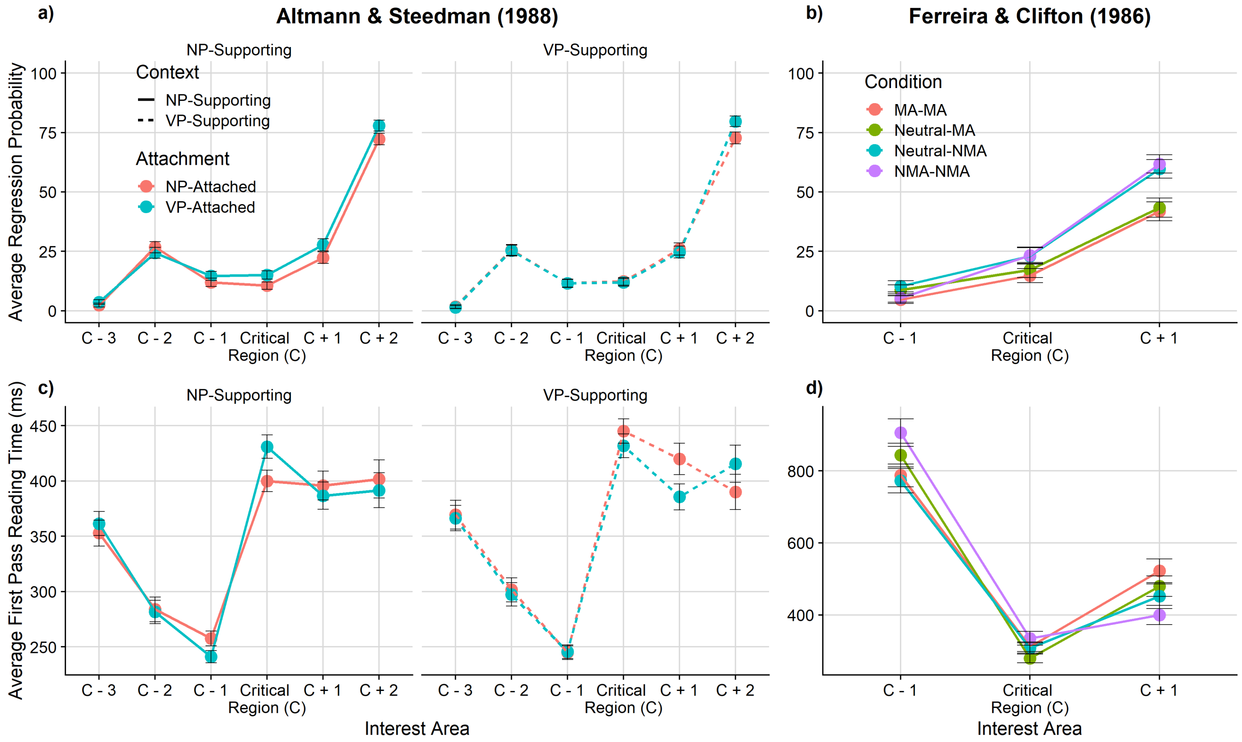

Figure 1. Average of eyetracking measures at each interest area for A&S88 (left) and F&C86 (right). The interest area relative to the critical region (C) is indicated on the x-axis. Panels a) and b) show average regression probability (y-axis) for each interest area and each condition. Panels c) and d) show average first pass reading times (y-axis) for each interest area (note that the y-axis range for average reading time differs for A&S88 and F&C86).

Table 1. *Generalized Mixed-Effects Model Analysis Summary for F&C86 and A&S88*

Experiment	*DV*	*Region*	*Summary*
F&C86	Regression probability	C - 1	Neutral-MA (*p*=.04), NMA-NMA (*p*=.02)
	First pass reading	C - 1	Neutral-MA (*p*=.002)
	First pass reading	C	Neutral-MA (*p*=.03)
	First pass reading	C + 1	Neutral-NMA (*p*=.002), NMA-NMA (*p*<.001)
A&S88	Regression probability	C + 1	Attachment (*p*=.004)
	First pass reading	C	Context (*p*=.049), Context x Attachment (*p*=.003)