

## Both Principle B and Competition Are Necessary to Explain Disjoint Reference Effects

**Introduction.** Many languages exhibit a restriction against pronouns expressing local coreference [1]. It remains debated whether this is due to an explicit grammatical constraint against local pronominal coreference (classical *Principle B*) [1,6], or if it instead reflects *Competition*, a pragmatic reasoning process selecting between competing alternative forms [3-5]. To evaluate these approaches, we conducted two experiments using *Evans Sentences* as in (1) [2]. These apparent violations of Principle B have been critical to the development of *Competition* theories emphasizing distinctness of meaning in context [3-5] and taken to indicate that Principle B governs bound variable anaphora rather than coreference [3].

- (1) Sarah said that everyone voted for Michael, but she lied. Only Michael<sub>i</sub> voted for him<sub>i</sub>.

*Competition* claims coreference is possible when the context makes bound variable anaphora unavailable. For example, coreference in (1) should be available because the context distinguishes a bound variable interpretation (*Only Michael(x voted for x)*) from coreference (*Only Michael(x voted for Michael)*). *Competition* then expresses a requirement to use a reflexive form when the meaning is indistinguishable from a bound variable interpretation (e.g. *Rule 1* or similar [3, 5]). *Principle B* [1,6] and *Competition* [3,4,5] make differing predictions about the production and comprehension of these sentences. In production, *Principle B* prohibits pronouns for local coreference, and so producers should always find some other way to express local coreference in any context. *Competition* allows producers to use pronouns in contexts that prohibit bound variable anaphora. Since pronouns are better than alternative possibilities like repeated names, we expect to see pronouns selectively in these contexts. In comprehension, *Principle B* predicts rejection of pronouns with a local antecedent, but if participants do accept it, there should be no correlation with context. *Competition* predicts comprehenders to allow coreference when they can associate it with contexts that prohibit bound variable anaphora. Our experiments support both predictions, revealing the need for both an explicit constraint against local coreference and Competition in deriving Principle B effects. In short, we found an overall preference for reflexives for both coreferent and bound meanings, as predicted by *Principle B* [1,6], but when participants did accept the pronoun form in the comprehension experiment, they preferred a coreferential interpretation, as predicted by *Competition* [3].

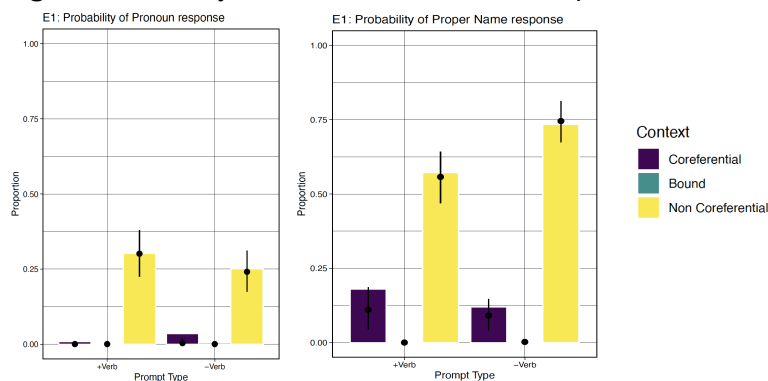
**Experiment 1 ( $N_{\text{subj}} = 36$ ).** *Which forms do participants produce, given a meaning?* Participants completed natural SMS exchanges [7]. We manipulated the context so that participants had to choose a form to express a locally bound, locally coreferential, or locally non-coreferential dependency (Table 1); we further manipulated whether the prompt contained the repeated verb or not (+/- verb). We created 48 critical items, distributed via Latin square and randomized with 48 filler items.

**Table 1.** Sample item, manipulating CONTEXT

A: By the way, Zachary said that everyone listened to...  
 (Bound:  $x$  listened to  $x$ ) ...**themselves.**  
 (Coreferential:  $x$  listened to Ashley) ... **Ashley.**  
 (NonCoreferential:  $x$  listened to Jacob) ...**Jacob.**  
 B: He lied! I overheard that only Ashley listened to \_\_\_\_\_.

we further manipulated whether the prompt contained the repeated verb or not (+/- verb). We created 48 critical items, distributed via Latin square and randomized with 48 filler items.

**Fig. 1.** Probability of Pronoun and Name responses in E1



Participants (i) overwhelmingly preferred the reflexive form in both the Coreferential (88.4%) and Bound (100%) contexts, and (ii) produced almost no pronouns coreferential with the local subject (see Fig. 1). In production, we see a strong preference for abiding by Principle B, without influence from

the greater discourse context. However, Competition may reflect an interpretive strategy [4]. We address this in Exp. 2.

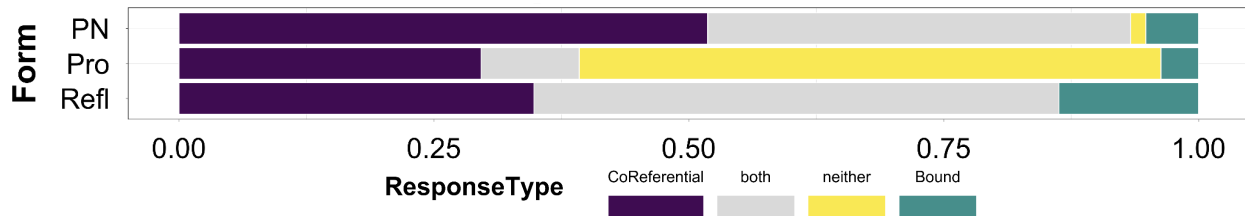
**Experiment 2** ( $N_{\text{subj}} = 54$ ). *Do comprehenders accept pronouns with local antecedents, and if so, in what contexts?* Participants were shown an SMS exchange, using Experiment 1’s stimuli. The final sentence was complete, but the critical context sentence was blanked out. We manipulated the form that participants saw (Table 2). They were asked to choose the best sentence to fit in the context blank, and could choose the Bound sentence, Coreferential sentence (Table 1), *Both*, or *Neither*. Principle B predicts that the pronoun form should be unacceptable with a local antecedent, and thus we should expect only *Neither* responses in the *Pronoun* condition, since both contexts force an interpretation with a local antecedent. Competition predicts a strong preference for coreference in the Pronoun condition.

Table 2. Sample item, manipulating FORM	
A: By the way... _____.	
B: He lied. I overheard that only Ashley listened to...	
(Reflexive)	...herself.
(Pronoun)	...her.
(ProperName)	...Ashley.

Our results support both predictions (Fig. 2). Consistent with Principle B, *Pronoun* was the only condition where participants rejected both contexts at a high rate (57%; PN=1.5%, Refl=0%), supporting the dispreference found in Exp 1. In an analysis that excluded 'neither' responses, participants endorsed the coreferential

context at a higher rate for Pronouns (91.4%) than for Reflexives (86.3%); for the purposes of this analysis, we treated 'Both' and 'Coreferential' responses as endorsement of the latter.

**Fig. 2.** Exp. 2: Distribution of participants’ responses



**Discussion.** In both of our experiments, we find strong avoidance of local antecedents for pronouns no matter the context, suggesting a grammatical constraint against local pronominal coreference (*Principle B*). At the same time, when comprehenders do assign an interpretation to pronouns with local antecedents, they systematically associate them with coreferential, not bound readings, as predicted by *Competition*. Together, our results suggest that syntactic context has more influence on availability of local coreference than discourse context, but both are required for a complete theory. Our results also suggest an asymmetry between production and comprehension of these sentences, as in Experiment 1 participants almost categorically avoided the pronoun form, but in Experiment 2, they were able to systematically interpret it. This may mean that Competition reflects comprehender-side Gricean [8] or Bayesian [9] reasoning processes that complement, but do not fully explain, the constraint against locally coreferent pronominals.

**References.** [1] Chomsky, N. (1981). *Lectures on government and binding*. [2] Evans, G. (1980). Pronouns. [3] Grodzinsky, Y., & Reinhart, T. (1993). *The innateness of binding and coreference*. [4] Reinhart, T. (2006). Interface strategies: Optimal and costly computations. [5] Roelofsen, F. (2010). *Condition B effects in two simple steps*. [6] Heim, I. (2007). *Forks in the Road to Rule I*. [7] Kroll, M.I. (2020). *Comprehending ellipsis*. [8] Marty, P. P. (2017). Implicatures in the DP domain. [9] Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference.