

## When errors aren't: How comprehenders selectively violate Binding Theory

Shayne Sloggett & Brian Dillon (University of Massachusetts Amherst)

A substantial body of work in sentence processing has shown that comprehenders rapidly employ Principle A to constrain the process of reflexive interpretation (Dillon, Mishler, Sloggett, & Phillips, 2013; Nicol & Swinney, 1989; Sturt, 2003, i.a.). This work demonstrates that, for the most part, even early stages of comprehension selectively attend to referents which would be compatible with Principle A (*targets*) and largely ignore Principle A incompatible referents (*lures*). However, recent work suggests that Principle A may be a fallible constraint on the on-line interpretation of reflexives, finding that reflexives which mismatch the phi-features of their target antecedents are read more quickly if they match the features of a lure referent (Parker, 2014; Patil, Vasishth, & Lewis, 2016). Descriptively, this looks like a *grammatical illusion*: comprehenders (erroneously) perceive ungrammatical reflexives as well-formed in the presence of a feature-matched lure.

One prominent explanation of this finding comes from memory models which posit an error-prone, cue-based retrieval mechanism underlying antecedent resolution in sentence comprehension (Lewis, Vasishth, & Van Dyke, 2006; McElree, 2006). These models posit that anaphoric interpretation routinely uses both binding constraints and morphosyntactic features to search for an antecedent, and that this search process is prone to making mistakes on the basis of partial feature matching. Critically, this explanation holds that sensitivity to lure referents is an error of the processing system: Principle A fallibility surfaces when a reflexive accidentally retrieves a feature matched lure, rather than the intended target antecedent (Engelmann, Jäger, & Vasishth, Submitted; Parker, 2014; Patil et al., 2016). The present work challenges this characterization of Principle A fallibility in reflexive comprehension, suggesting that alternative grammatical constraints, not errors in memory retrieval, underlie this behavior. We present the results of two eye-tracking while reading studies and three acceptability judgment studies which show that reflexive processing only accesses lures which can act as logophoric antecedents for the reflexive (the “logophoric reflexives” hypothesis).

Logophors are pronouns which necessarily refer to the entity whose speech, thoughts, or feelings are reported in an utterance (Clements, 1975). Cross-linguistically, they are most likely to refer to the subjects of speech verbs (e.g. *say*, *mention*), and least likely to refer to the subjects of perception verbs (e.g. *hear*, *see*) (Culy, 1994). The logophoric reflexives hypothesis thus predicts that comprehenders should be more sensitive to lures which are *speakers* than to those which are *perceivers*. Experiment 1 tested this prediction in separate acceptability judgment ( $n=64$ ) and eye-tracking ( $n=37$ ) studies by manipulating sentence as in (1). Reflexives were embedded under either a speech verb (e.g. *say*) or a perception verb (e.g. *hear*). We then manipulated the embedded (*target*) and matrix (*lure*) subjects independently, so that each either matched, or mismatched the morphosyntactic features of the embedded reflexive.

- (1)  $\left\{ \begin{array}{l} \text{Susan} \\ \text{Steven} \end{array} \right\} \left\{ \begin{array}{l} \text{said} \\ \text{heard} \end{array} \right\}$  that the obnoxious  $\left\{ \begin{array}{l} \text{schoolgirl} \\ \text{schoolboys} \end{array} \right\}$  embarrassed herself at lunch today.

Target	Lure	Verb Type	
		Speech	Perception
+match	+match	4.99(.12)	5.11(.12)
	-match	5.04(.12)	5.20(.10)
-match	+match	3.85(.15)	3.68(.16)
	-match	3.56(.17)	3.60(.16)

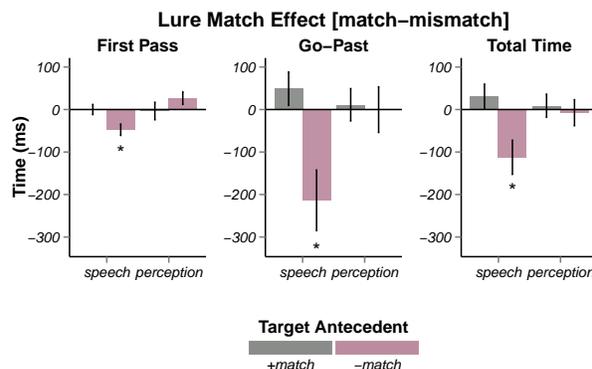


Table 1: *By-subject mean naturalness ratings in Experiment 1a (standard error in parentheses).*

The results of Experiment 1 are given in the table and figure above. In the judgment study, a feature-matched lure slightly improved acceptability judgments for target-mismatch reflexives embedded under

speech verbs (3.85 vs. 3.56), but not for those under perception verbs (3.68 vs. 3.60). In the eye tracking study, we observed substantial lure match effects (the difference between *Susan* and *Steven* conditions) in three separate reading measures, but only when: (1) the target antecedent mismatched the reflexive, and (2) the lure was the subject of a speech verb. These findings demonstrate that comprehenders selectively access lure referents which are the subjects of speech verbs, as predicted by the logophoric reflexives hypothesis.

Experiment 2 tested a separate prediction of the logophoric reflexives hypothesis. Logophors preferentially refer to the perspective center of an utterance (Sells, 1987). This predicts that indexical pronouns (e.g. *I, you*), which act as automatic perspective centers, should prevent other, third person referents from anteceding a logophor. This kind of “person blocking effect” is observed with the Mandarin reflexive *ziji*: long-distance interpretations of *ziji* are blocked in the presence of a first or second person pronoun, as predicted if long-distance reflexives are actually logophoric (Huang & Liu, 2001). The logophoric reflexives hypothesis thus predicts that first and second person pronouns should similarly reduce sensitivity to lures in English reflexive processing. Experiment 2 tested this prediction in two acceptability rating studies ( $n=54$  each), and one eye-tracking while reading study ( $n=40$ ) by manipulating sentences as in (2). Again, the embedded reflexive either matched, or mismatched a lure referent (*actress* vs. *actor*). In addition, the embedded subject was either a proper name which matched the reflexive (*Joanna*), a third-person, inanimate pronoun (*it*), a first person pronoun (*I*, expts 2a and 2b), or a second person pronoun (*you*, expt 2c).

(2) The  $\left\{ \begin{matrix} \text{actress} \\ \text{actor} \end{matrix} \right\}$  said that  $\left\{ \begin{matrix} \text{Joanna} \\ \text{it} \\ \text{I} \end{matrix} \right\}$  horribly misrepresented herself in the article...

	Target	Lure	
		+match	-match
Expt 2a	Name	5.23(.15)	5.26(.15)
	<i>it</i>	3.81(.19)	3.04(.18)
	<i>I</i>	3.23(.19)	2.77(.18)
Expt 2c	Name	5.45(.13)	5.54(.11)
	<i>it</i>	4.03(.20)	3.16(.19)
	<i>you</i>	3.35(.20)	3.02(.19)

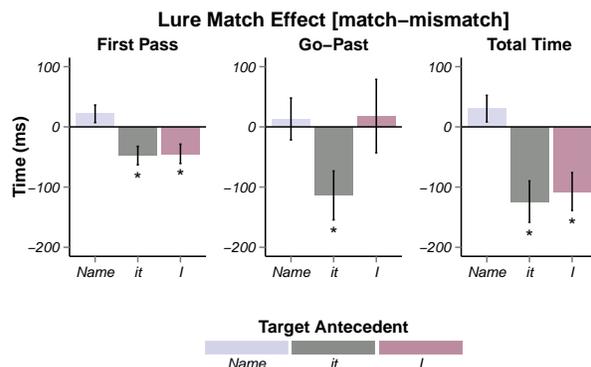


Table 2: By-subject mean naturalness ratings in Experiments 2a and 2c (standard error in parentheses).

The results of Experiment 2 are given in the table and figure above. In the judgment studies (expts 2a, 2c), lures which matched the reflexive improved acceptability ratings for both *it* and *I/you* sentences. However, lures had a greater effect on *it* sentences than on either *I* or *you* sentences. Similarly, in the eye-tracking study (expt 2b), we observed substantial lure-match effects in go-past reading times for *it* sentences, but not for *I* sentences. For *I* sentences, sensitivity to lures was mostly delayed until later reading measures (e.g. total reading time). These findings demonstrate that comprehenders access lure referents to a lesser extent (expts 2a, 2c) and at a delay (expt 2b) in the presence of an indexical pronoun. This pattern mirrors the person blocking effects observed in Mandarin, further supporting the logophoric reflexives hypothesis.

Taken together, experiments 1 and 2 provide strong support for the logophoric reflexives hypothesis. Comprehenders are not *generally* sensitive to lure referents, as predicted by cue-based parsing models, but rather selectively access lures which could act as logophoric antecedents for the reflexive. This suggests an importantly different interpretation of Principle A fallibility than comes from cue-based models of memory retrieval. Apparent violations of Binding Theory in reflexive comprehension aren’t actually errors of the processing system, but instead arise from competing grammatical constraints on reflexive interpretation.