# Demanding an Explanation: Implicit Causality Biases in Discourse Interpretation

Previous passage-completion studies using implicit causality (IC) contexts (Au 1986, Stevenson et al. 1994, inter alia) report strong next-mention biases toward a particular referent with *because*-prompts (1a). However, this bias is reduced or eliminated with a full-stop prompt (1b).

(1) a. John scolded Mary because _____    (strong bias toward *Mary*)

    b. John scolded Mary. _____    (mixed biases)

These results have led to confusion in the literature about the role of *because* in such examples (see e.g., Au 1986); Stevenson et al. (2000), for instance, argue that connectives constrain pronoun interpretation by directly modifying the salience of entities, in a role distinct from that of constraining coherence relations (Hobbs 1979; *because* signals an Explanation relation).

However, based on two passage completion studies using transfer-of-possession contexts, Rohde et al. (2007) argue that two types of *coherence-driven* expectations influence hearers' biases toward resolving a pronoun to a particular referent (equation 2): expectations about how the discourse will be continued with respect to coherence ($P(CR)$), and biases towards particular referents as conditioned on those relations ($P(pronoun{=}referent|CR)$).

(2) $P(pronoun{=}referent) = \sum_{CR \in CRs} P(CR) * P(pronoun{=}referent|CR)$

(3) $P(nextmention{=}referent) = \sum_{CR \in CRs} P(CR) * P(nextmention{=}referent|CR)$

A generalization of this equation to all next-mentioned referents (equation 3) predicts that the difference in biases revealed by (1a-b) results only from the fact that *because* drives the probability $P(CR{=}Explanation)$ to 1.

To test this prediction, we asked whether the IC biases found in sentence completions using a *because* prompt (1a) are statistically indistinguishable from those found in a full-stop condition (1b) when *only those passages that participate in an Explanation relation are analyzed* (in other words, whether $P(nextmention{=}referent|CR{=}Explanation)$ is consistent across prompt conditions). A study (n=75) was performed with a 2x3 design that crossed verb type (IC verb vs. non-IC verb) with continuation type (full stop vs. *because* vs. dialogue prompt; discussion of the dialogue condition omitted). 40 IC verbs (20 NP1-biased, and 20 NP2-biased) and 40 non-IC verbs taken from McKoon et al. (1993) were used in context sentences that mention one male and one female. Judges categorized the corpus of completions by coherence relation.

The predictions were borne out: In all three conditions (IC-NP1, IC-NP2, non-IC), the next mention biases were (i) significantly different between the two prompt conditions, but were nonetheless (ii) statistically indistinguishable when only the full-stop passages categorized as Explanation relations were analyzed. While (ii) can only be coincidental on Stevenson et al.'s analysis, the results pattern with Rohde et al. in that the overall statistics conceal a consistent system of stronger biases once coherence relations are conditioned on, and thus offer an explanation of the mixed results of Au and of Stevenson et al.

Finally, the results revealed an additional IC bias with respect to the term $P(CR)$ in equation (3): IC verbs were significantly more likely to evoke Explanation continuations (60%) than non-IC verbs (24%) in the full-stop condition, regardless of which referent gets mentioned first. Whereas the strong next-mention biases for IC verbs have long been the bedrock observation in the area, this latter bias has never been demonstrated because previous studies have not categorized their data by coherence.