Prominence and Coherence in a Bayesian Theory of Pronoun Interpretation

Andrew Kehler
UC San Diego

(Joint work with Hannah Rohde, University of Edinburgh)

SLE 2015, Leiden, Netherlands, September 2, 2015

Pronoun Interpretation and Production

- Common wisdom: There is a unified notion of prominence that determines when…
  - a speaker will produce a pronoun to mention a referent, and hence
  - a comprehender will successfully interpret the reference
- The task is to identify what factors affect prominence (grammatical role, parallelism, thematic role, information structural, semantic, etc)
- I will try to disabuse you of this, and argue instead for a different model

Implicit Causality

- Previous work has shown that so-called *implicit causality* verbs are associated with strong pronoun biases (Garvey and Caramazza, 1974 and many others)

  Amanda amazes Brittany because she _____ [subject-biased]

  Amanda detests Brittany because she _____ [object-biased]

- The connective *because* indicates an Explanation coherence relation: the second sentence describes a cause or reason for the eventuality described by the first

- For free prompts, IC verbs result in a greater number of Explanation continuations (60%) than non-IC controls (24%) (Kehler et al. 2008)

Background

- A study by Stevenson et al (1994) compared pronoun prompts with free prompts:

  Amanda detests Brittany. *She* _____

  Amanda detests Brittany. _____

- Two results:
  - Interpretation: Greater number of subject references in the pronoun-prompt condition than the free-prompt condition
  - Production: In the free-prompt condition, a strong tendency to use a pronoun to refer to the subject and a name to refer to a non-subject

Bayesian Interpretation (Kehler et al. 2008)

- Bayesian formulation:

$$P(\text{referent} \mid \text{pronoun}) = \frac{P(\text{pronoun} \mid \text{referent})\, P(\text{referent})}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun} \mid \text{referent})\, P(\text{referent})}$$

Interpretation

Production (Subject Bias)

Prior Expectation (Semantics/Coherence)

- Data is consistent with a scenario in which semantics/coherence-driven biases primary affect probability of *next-mention,* whereas grammatical biases affect *choice of referential form*

- Results in the counterintuitive prediction that production biases are insensitive to a set of factors that affect the ultimate interpretation bias

Implicit Causality (Ambiguous Contexts)
(Rohde, 2008; Fukumura & van Gompel 2010; Rohde & Kehler 2014)

Measure next mention bias P(referent) and production bias P(pronoun | referent)

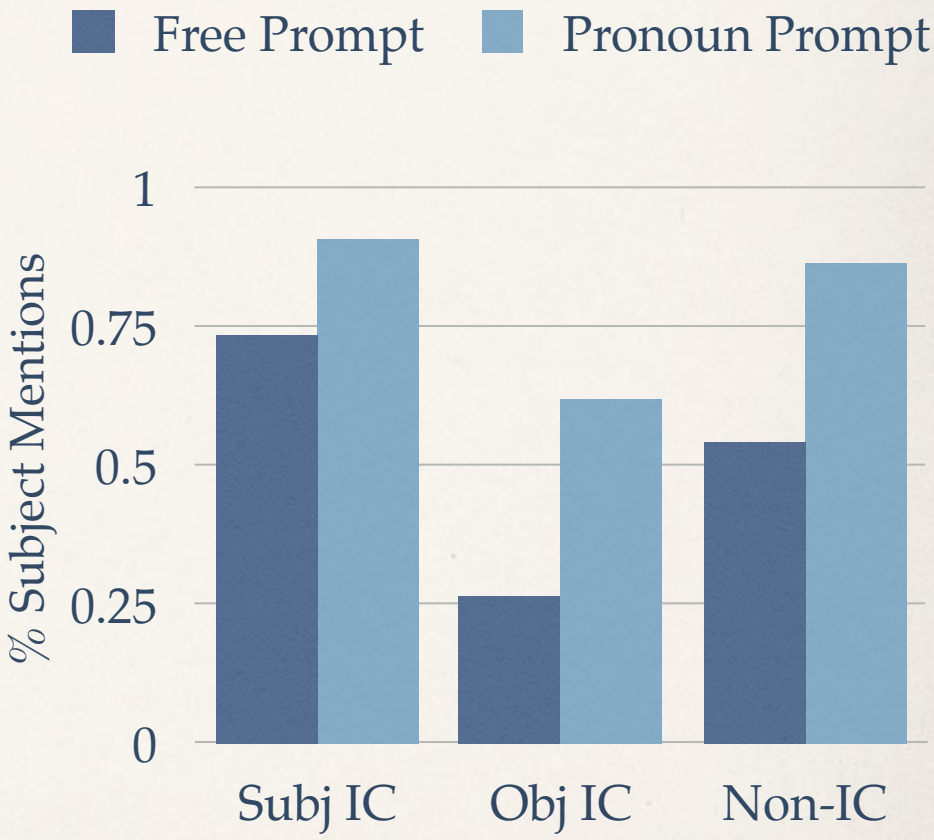
- Free prompts:
  - *Amanda amazed Brittany.* _____ [IC, subject-biased]
  - *Amanda detested Brittany.* _____ [IC, object-biased]
  - *Amanda chatted with Brittany.* _____ [non-IC]

Measure interpretation bias P(referent | pronoun)

- Pronoun prompts:
  - *Amanda amazed Brittany. She* _____ [IC, subject-biased]
  - *Amanda detested Brittany. She* _____ [IC, object-biased]
  - *Amanda chatted with Brittany. She* _____ [non-IC]

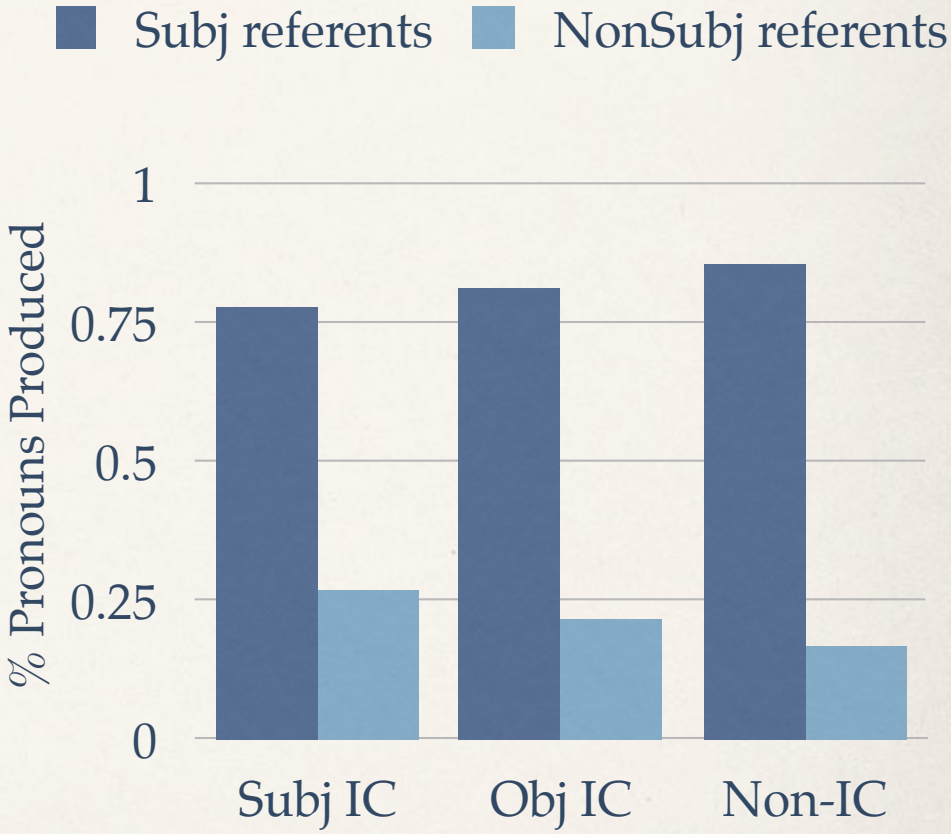
Implicit Causality (Ambiguous Contexts)
(Rohde, 2008; Fukumura & van Gompel 2010; Rohde & Kehler 2014)

- Rohde (2008), Rohde & Kehler (2014): IC affects interpretation
  - *Amanda amazed Brittany.*
    (She) _____ [IC, subject-biased]
  - *Amanda detested Brittany.*
    (She) _____ [IC, object-biased]
  - *Amanda chatted with Brittany.*
    (She) _____ [non-IC]
- Result: IC bias affects next-mention (prior) and pronoun interpretation



Production Biases (Ambiguous Contexts)
(Rohde, 2008; Fukumura & van Gompel 2010; Rohde & Kehler 2014)

- Rohde (2008), Rohde & Kehler (2014): IC doesn't affect production
  - *John amazed Mary.* _____ [IC, subject-biased]
  - *John detested Mary.* _____ [IC, object-biased]
  - *John chatted with Mary.* _____ [non-IC]
- Result: grammatical role matters, but semantic bias does not



Testing the Theory: Inferred Causes

- Passage completion study:

  The boss fired the employee who was hired in 2002. He _____ [Control]

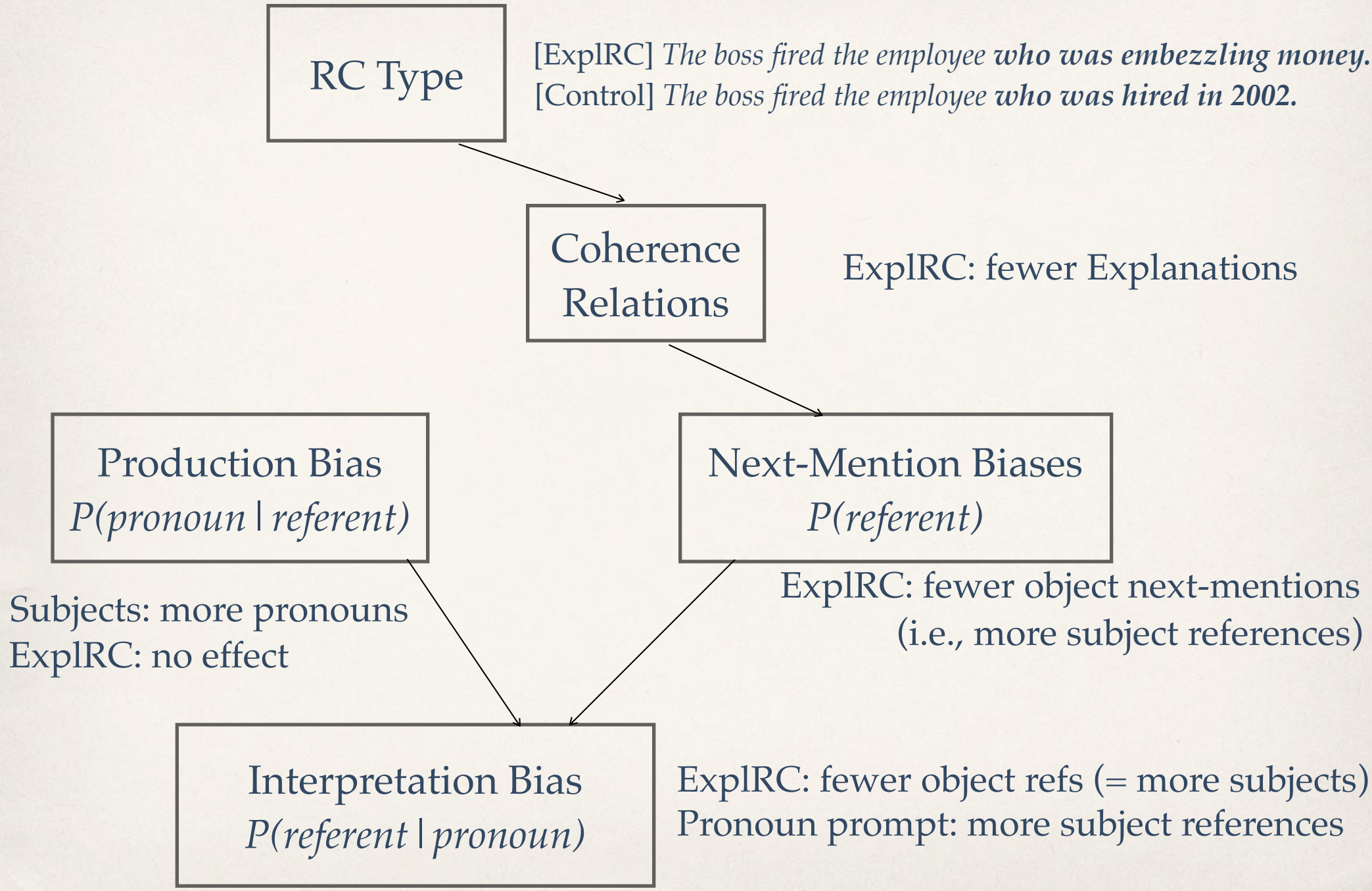
  The boss fired the employee who was embezzling money. He _____ [ExplRC]

  The boss fired the employee who was hired in 2002. _____ [Control]

  The boss fired the employee who was embezzling money. _____ [ExplRC]

- Analyze:
  - Coherence relations (Explanation or Other)
  - Next-mentioned referent (Subject or Object)
  - Form of Reference (free-prompt condition; Pronoun or Other)

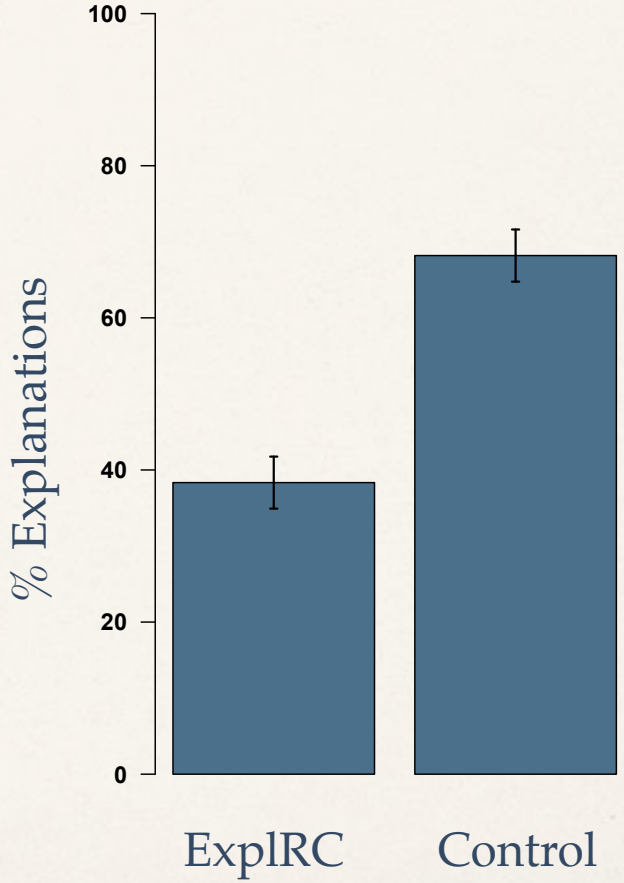
Predictions



Prediction 1: Coherence Relations

- Predict a smaller percentage of Explanation relations in the ExplRC condition than the Control condition

- Confirmed: (β=2.06; p<.001)



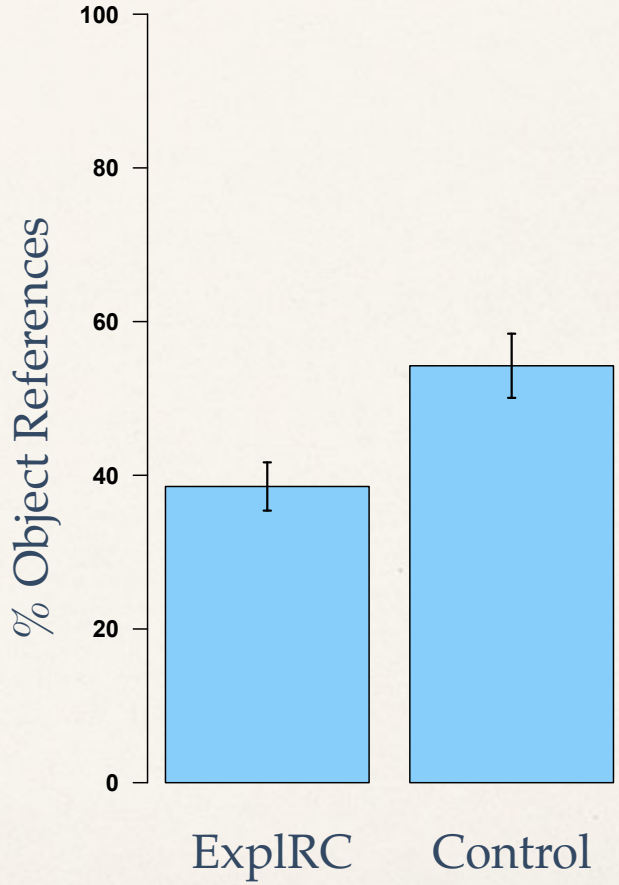
*[ExplRC] The boss fired the employee **who was embezzling money.***
*[Control] The boss fired the employee **who was hired in 2002.***

Prediction 2: Next-Mention Biases

P(referent | pronoun) ∝ *P(pronoun | referent)* ***P(referent)***

- For free-prompt condition, predict a smaller percentage of next mentions of the object in ExplRC condition than the Control condition

- Confirmed: (β=.720; p<.05)

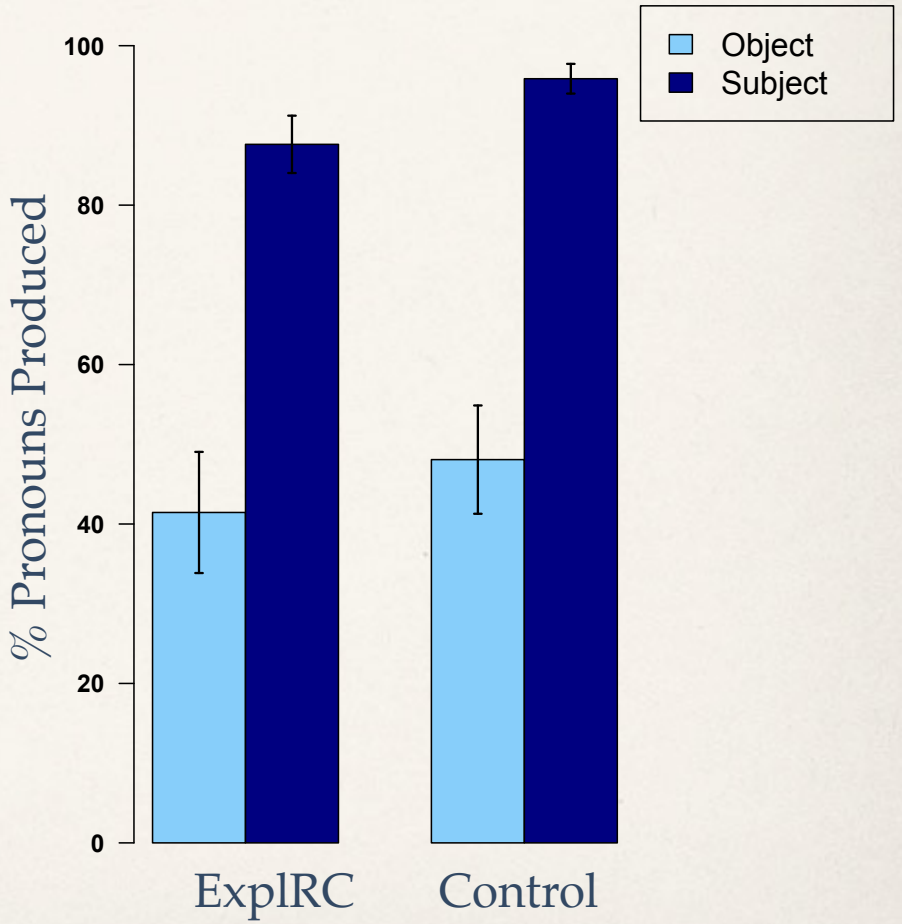


[ExplRC] The boss fired the employee ***who was embezzling money.***
[Control] The boss fired the employee ***who was hired in 2002.***

Prediction 3: Rate of Pronominalization

$$P(referent \mid pronoun) \propto \mathbf{P(pronoun \mid referent)}\, P(referent)$$

- Predict an effect of grammatical role on pronominalization rate (favoring subjects; free prompt condition)
  - Confirmed: (β=4.11; p<.001)
- But no interaction with RC condition
  - Confirmed (β=0.12; p=.92)
  - Marginal effect of RC condition (β=0.94; p=.078)

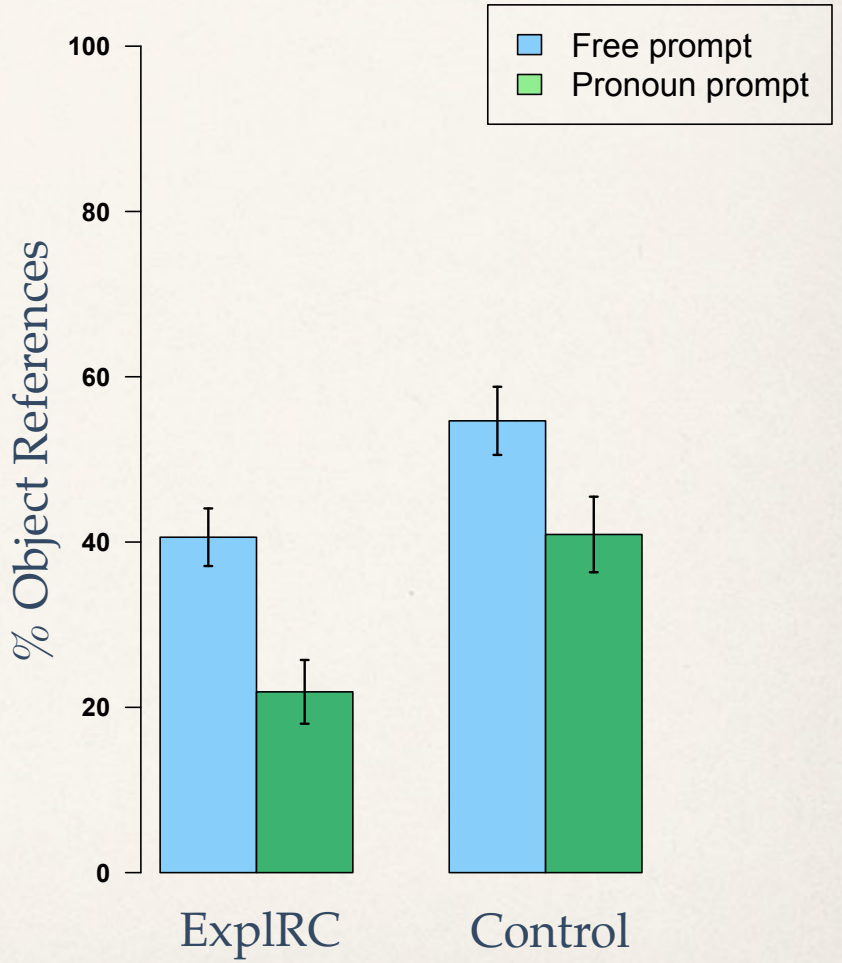


[ExplRC] The boss fired the employee ***who was embezzling money.***
[Control] The boss fired the employee ***who was hired in 2002.***

Predictions 4 & 5: Pronoun Interpretation

P(referent | pronoun) $\propto$ *P(pronoun | referent) P(referent)*

- Predict a smaller percentage of object mentions in the ExplRC condition than the Control condition...
  - Confirmed: (β=1.17; p<.005)
- ...and in the free-prompt condition than the pronoun-prompt condition
  - Confirmed (β=-1.27; p=.001)
- Marginal interaction (β=0.85; p=.078)
- Effect in Pronoun subset only (β=1.46; p<.005)



[ExplRC] The boss fired the employee ***who was embezzling money.***
[Control] The boss fired the employee ***who was hired in 2002.***

Model Comparison

- We can evaluate the predictions of the model by estimating the likelihood and prior from the data in the free prompt condition to generate a *predicted* pronoun interpretation bias
- We then compare that to the *actual* pronoun interpretation bias estimated from the data in the pronoun-prompt condition

$$P(\text{referent} \mid \text{pronoun}) = \frac{P(\text{pronoun} \mid \text{referent})\, P(\text{referent})}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun} \mid \text{referent})\, P(\text{referent})}$$

Competing Model: Mirror Model

- Recall the common wisdom: the factors that comprehenders use to interpret pronouns are those that speakers use when choosing to use one.
- That means the interpreter's biases will be proportional to (their estimates of) the speaker's production biases

$$P(\text{referent} \mid \text{pronoun}) = \frac{P(\text{pronoun} \mid \text{referent})\ \cancel{P(\text{referent})}}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun} \mid \text{referent})\ \cancel{P(\text{referent})}}$$

Competing Model: Expectancy Model

- According to Arnold's Expectancy Hypothesis (1998, 2001, inter alia), comprehenders will interpret a pronoun to refer to whatever referent they expect to be mentioned next

$$P(\text{referent} \mid \text{pronoun}) = \frac{\cancel{P(\text{pronoun} \mid \text{referent})}\ P(\text{referent})}{\sum_{\text{referent} \in \text{referents}} \cancel{P(\text{pronoun} \mid \text{referent})}\ P(\text{referent})}$$

Model Comparison: Results

- Comparison of actual rates of pronominal reference to object (pronoun-prompt condition) to the predicted rates for three competing models (using estimates from free-prompt condition)

	Actual	Bayesian	Mirror	Expectancy
ExplRC	0.215	0.229	0.321	0.385
Control	0.410	0.373	0.334	0.542
		R^2=.48/.49	R^2=.34/.42	R^2=.14/.12

Conclusions

- Pronoun interpretation is sensitive to coherence factors, in this case the invited inference of an explanation
- Pronoun production, however, is not
- The data thus demonstrate precisely the asymmetry predicted by the Bayesian analysis
- A corollary is that there is no unified notion of prominence that drives interpretation and production
- Indeed, perhaps the best *independent* measure of prominence is provided by next-mention expectations, but pronoun biases are not the same

Thank you!