A Bayesian Model of Pronoun Production and Interpretation

Andrew Kehler
UCSD Linguistics

(Joint work with Hannah Rohde)

CORBON 2016, San Diego, CA, June 16, 2016

What's the Problem?

Subject Assignment (Crawley et al, 1990)

a. Donald narrowly defeated Ted, and the press promptly followed him to the next primary state. [him = Donald]

b. Ted was narrowly defeated by Donald, and the press promptly followed him to the next primary state. [him = Ted]

c. Donald narrowly defeated Ted, and Marco absolutely trounced him. [him = Ted]

d. Donald narrowly defeated Ted, and he quickly demanded a recount. [he = Ted]

Grammatical Role Parallelism
(Kamayama, 1986; Smyth, 1994)

Reasoning / World Knowledge
(Hobbs, 1979)

The SMASH Approach

- Search: Collect possible referents (within some contextual window)
- Match: Filter out those referents that fail 'hard' morphosyntactic constraints (number, gender, person, binding)
- And Select using Heuristics: Select a referent based on some combination of 'soft' constraints (grammatical role, grammatical parallelism, thematic role, referential form, ...)

The Big Question

- Why would anybody ever use a pronoun?
  - Speaker elects to use an ambiguous expression in lieu of an unambiguous one, seemingly without hindering interpretation
  - A theory should tell us why we find evidence for different 'preferences', and why they prevail in different contextual circumstances
  - We ask: *What would the discourse processing architecture have to look like to allow for a simple theory of pronoun interpretation?*

Two Approaches to Discourse Coherence

- Centering Theory (Grosz et al. 1986; 1995):

  "Certain entities in an utterance are more central than others and this property imposes constraints on a speaker's use of different types of referring expressions... The coherence of a discourse is affected by the compatibility between centering properties of an utterance and choice of referring expression."

- Define Centering constructs and rules:
  - A (single) backward-looking center (C_b; the 'topic')
  - A list of "forward-looking centers" (C_f; ranked by salience)
  - Constraints governing the pronominalization of the C_b
  - Ranking on transition types defined by the C_b and the C_f

Centering

- A Centering-driven approach could conceivably explain why linguistic form could affect pronoun biases:

  Donald narrowly defeated Ted, and the press promptly followed him to the next primary state. [him = Donald]

  Ted was narrowly defeated by Donald, and the press promptly followed him to the next primary state. [him = Ted]

- Semantics and world knowledge do not come into play

Coherence and Coreference

- Hobbs' (1979) Coherence-Driven Approach
  - Pronoun interpretation occurs as a by-product of general, semantically-driven reasoning processes
  - Pronouns are modeled as free variables which get bound during inferencing (e.g., coherence establishment)

    The city council denied the demonstrators a permit because

    a. they feared violence

    b. they advocated violence *(adapted from Winograd 1972)*

  - Choice of linguistic form does not come into play

Agenda

- Briefly outline the Hobbsian approach to discourse coherence
- Describe a series of experiments demonstrating that pronoun interpretation is influenced by coherence relations
- Present other evidence that suggests a role for a Centering-driven theory
- Present a model that integrates aspects of both approaches
- Describe experiments that examine predictions of the model
- Conclude with some potential ramifications for computational work

The Case for Coherence

- The meaning of a discourse is greater than the sum of the meanings of its parts
- Hearers will generally not interpret juxtaposed statements independently:

  I need to work tonight. I am presenting a talk at the CORBON meeting.

- Explanation: Infer P from the assertion of S_1, and Q from the assertion of S_2, where normally $Q \rightarrow P$.

  ?? *I need to work tonight. OntoNotes Release 5 became available in 2013.*

Selected Other Relations

- Occasion: Infer a change of state for a system of entities from the assertion of S_2, establishing the initial state for this system from the end state of S_1.

  Donald flew to San Diego. He took a stretch limo to his first campaign rally.

- Elaboration: Infer $p(a_1,a_2,...,a_n)$ from the assertions of S_1 and S_2.

  Donald flew to San Diego. He took his private jet into Lindbergh Field.

Transfer of Possession (Rohde, Kehler, and Elman 2006)

- Goal/Source preferences (Stevenson et al., 1994):

  Obama seized the speech from Biden. He... [Obama]
  Obama passed the speech to Biden. He... [Obama/Biden]

- Possible explanations:
  - Thematic role preferences (`superficial')
  - Focus on end states of events (`deep')

- Latter is what one would expect for Occasion relations

  Occasion: Infer a change of state for a system of entities from S_2, establishing the initial state for this system from the end state of S_1

Rohde, Kehler, and Elman (2006)

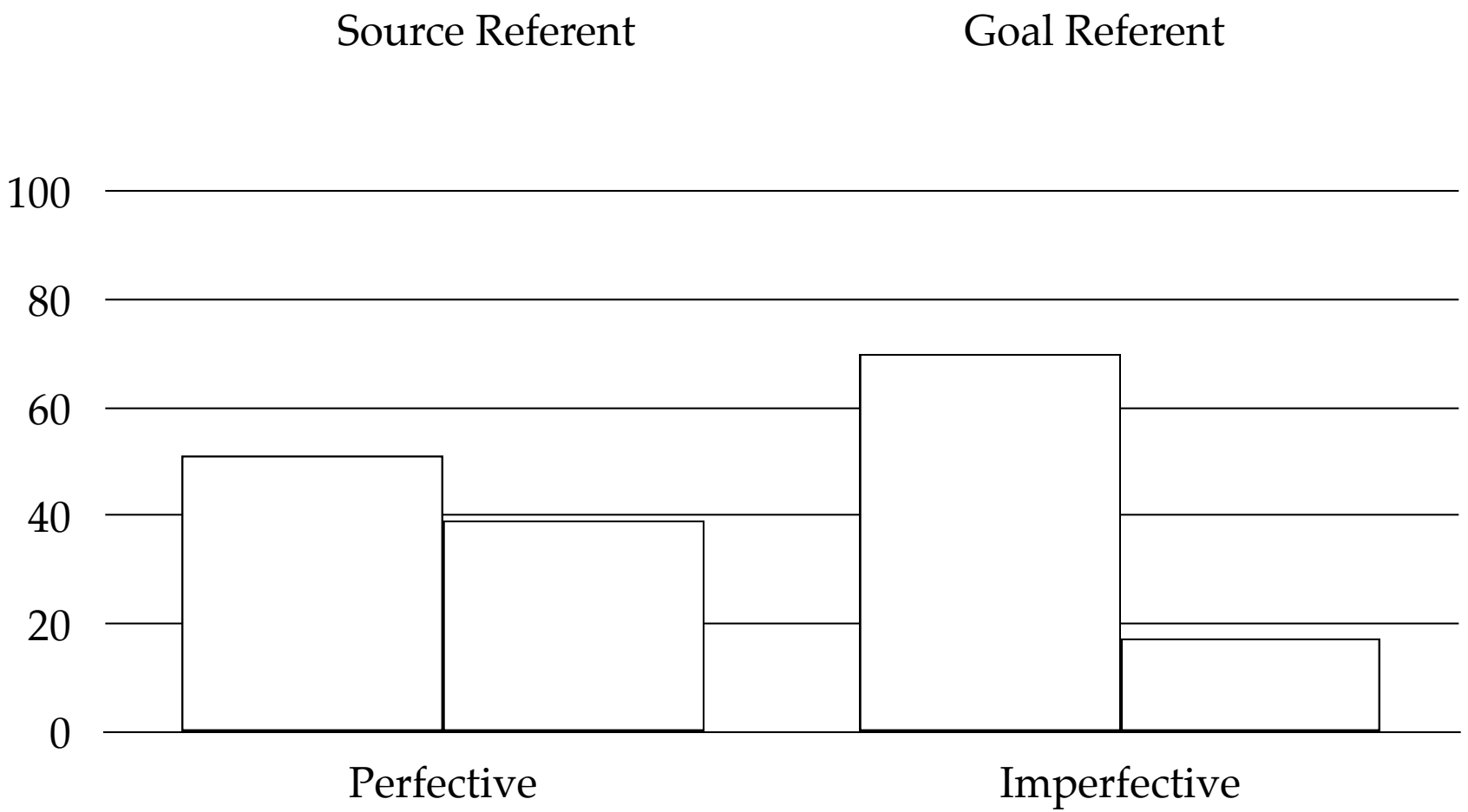
- Ran an experiment to distinguish these, comparing the perfective and imperfective forms for Source/Goal verbs

  Obama passed the speech to Biden. He...

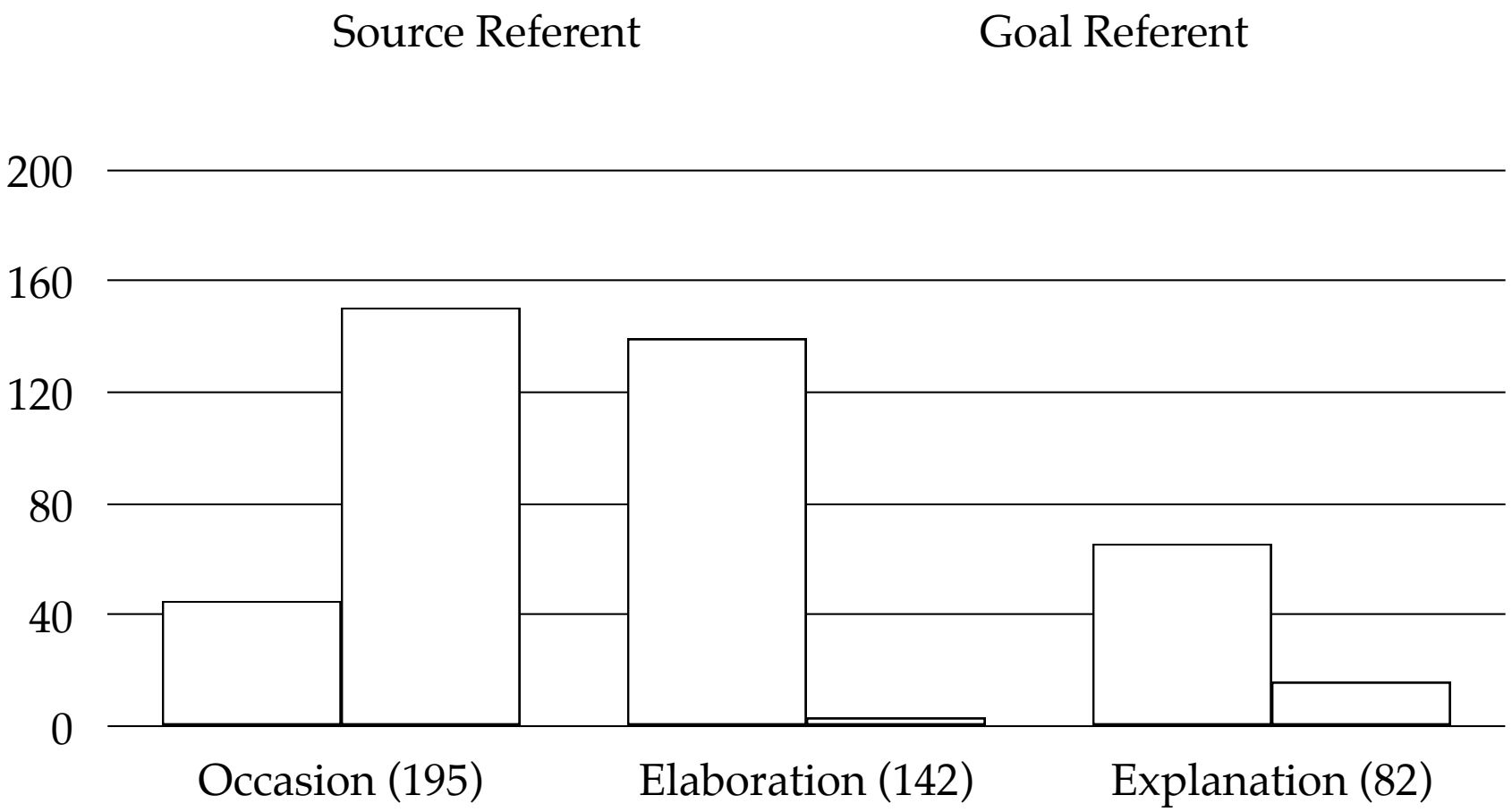
  Obama was passing the speech to Biden. He...

- More references to the Source/Subject in the imperfective case would support the event structure/coherence analysis

Results


Source Referent
Goal Referent
100
80
60
40
20
0
Perfective
Imperfective

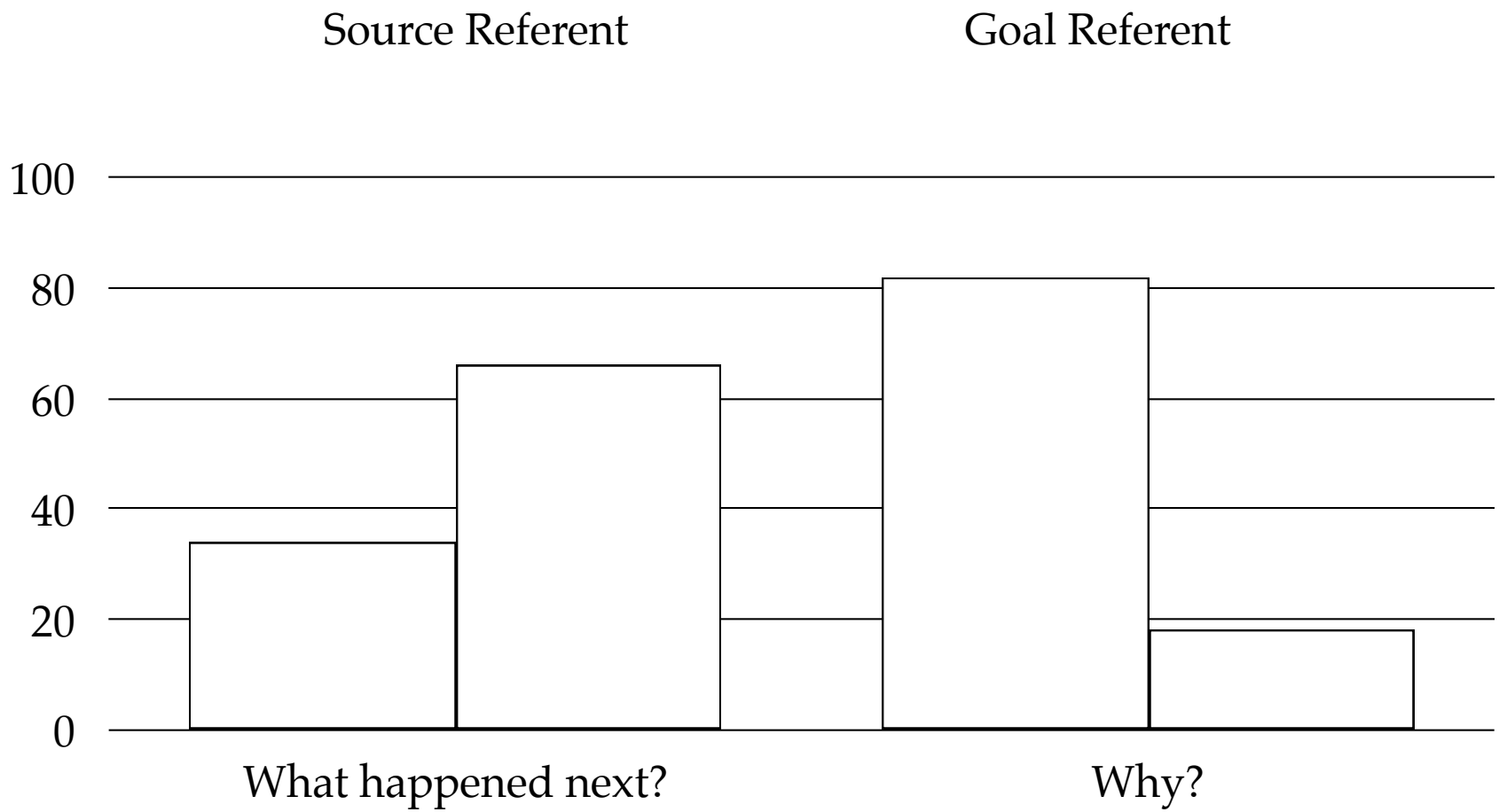
Breakdown by Coherence Type (Perfective Only)



Manipulating Coherence (Rohde, Kehler, and Elman 2007)

- If coherence matters, a shift in the distribution of coherence relations should induce a shift in the distribution of pronoun interpretations
- Run the previous experiment again, except with one difference in the instructions for how to continue the passage:
  - What happened next? (Occasion)
  - Why? (Explanation)
- Stimuli kept identical across conditions

Results


Source Referent
Goal Referent
100
80
60
40
20
0
What happened next?
Why?

The Subject Preference

- Stevenson et al's (1994) study paired their pronoun-prompt condition with a free prompt condition:

  Obama passed the speech to Biden. He _____

  Obama passed the speech to Biden. _____

- Always found more mentions of the subject in the pronoun condition than the free condition.

- They found a near 50/50 split in Source vs. Goal interpretations for pronouns in the prompt condition

- But in the no-prompt condition, they found a strong tendency to use a pronoun to refer to the subject and a name to refer to the object

Bayesian Interpretation (Kehler et al. 2008)

Production

Prior Expectation

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

Interpretation

Bayesian Interpretation (Kehler et al. 2008)

- Bayesian formulation:

$$\underbrace{P(\text{referent} \mid \text{pronoun})}_{\text{Interpretation}} = \frac{\overbrace{P(\text{pronoun} \mid \text{referent})}^{\text{Production (Subject Bias)}} \overbrace{P(\text{referent})}^{\text{Prior Expectation (Semantics/Coherence)}}}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun} \mid \text{referent}) P(\text{referent})}$$

- 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

- 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)

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]

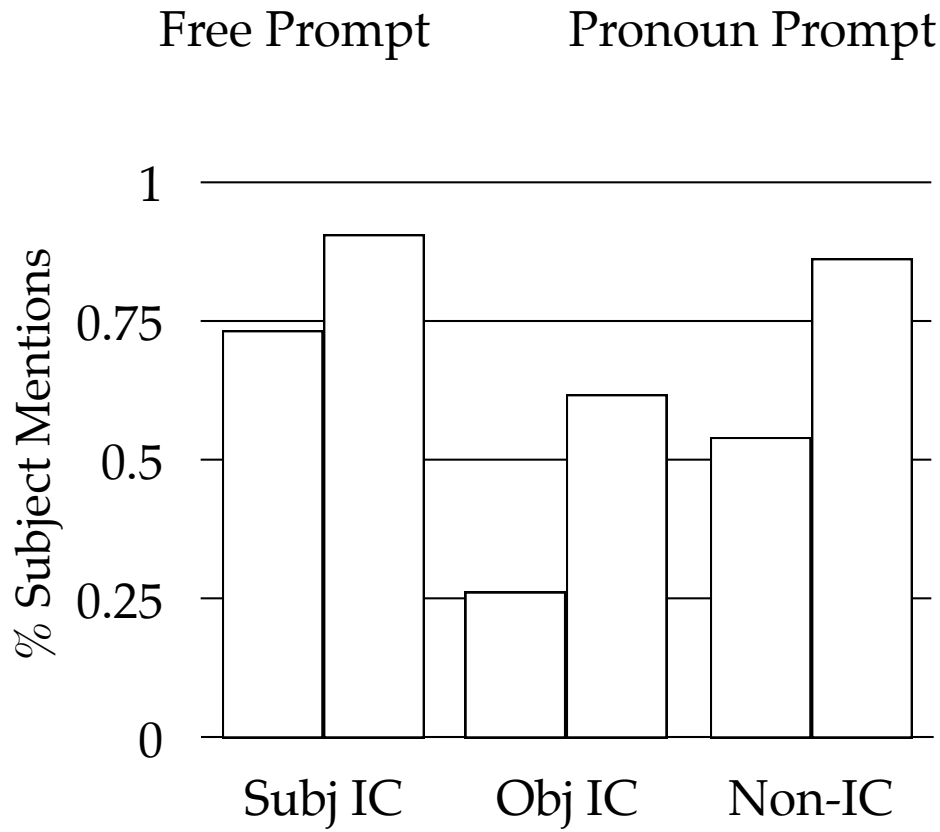
Measure interpretation bias P(referent | pronoun)

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

Production Biases (Ambiguous Contexts)

(Rohde, 2008; Fukumura & van Gompel 2010; Rohde & Kehler 2014)

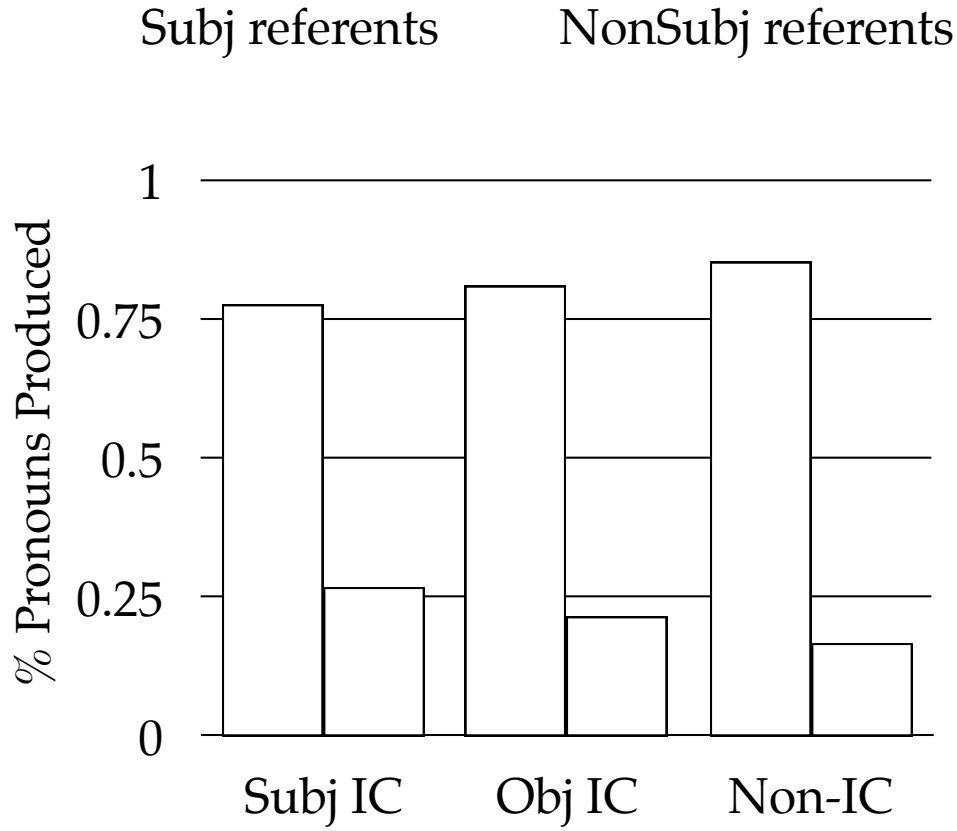
- Rohde (2008), Rohde & Kehler (2014): IC affects interpretation
  - *Amanda amazed Brittany.*
    (She) _____ [IC, subject-biased]
  - *Amanda detested Brittany.*
    (She) _____ [IC, object-biased]
  - *Amanda chatted with Brittany.*
    (She) _____ [non-IC]
- Result: IC bias affects next-mention (prior) 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
  - *Amanda amazed Brittany.* _____ [IC, subject-biased]
  - *Amanda detested Brittany.* _____ [IC, object-biased]
  - *Amanda chatted with Brittany.* _____ [non-IC]
- Result: grammatical role matters, but semantic bias does not



Testing the Theory: Inferred Causes
(Kehler & Rohde, CogSci 2015)

- Passage completion study:

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

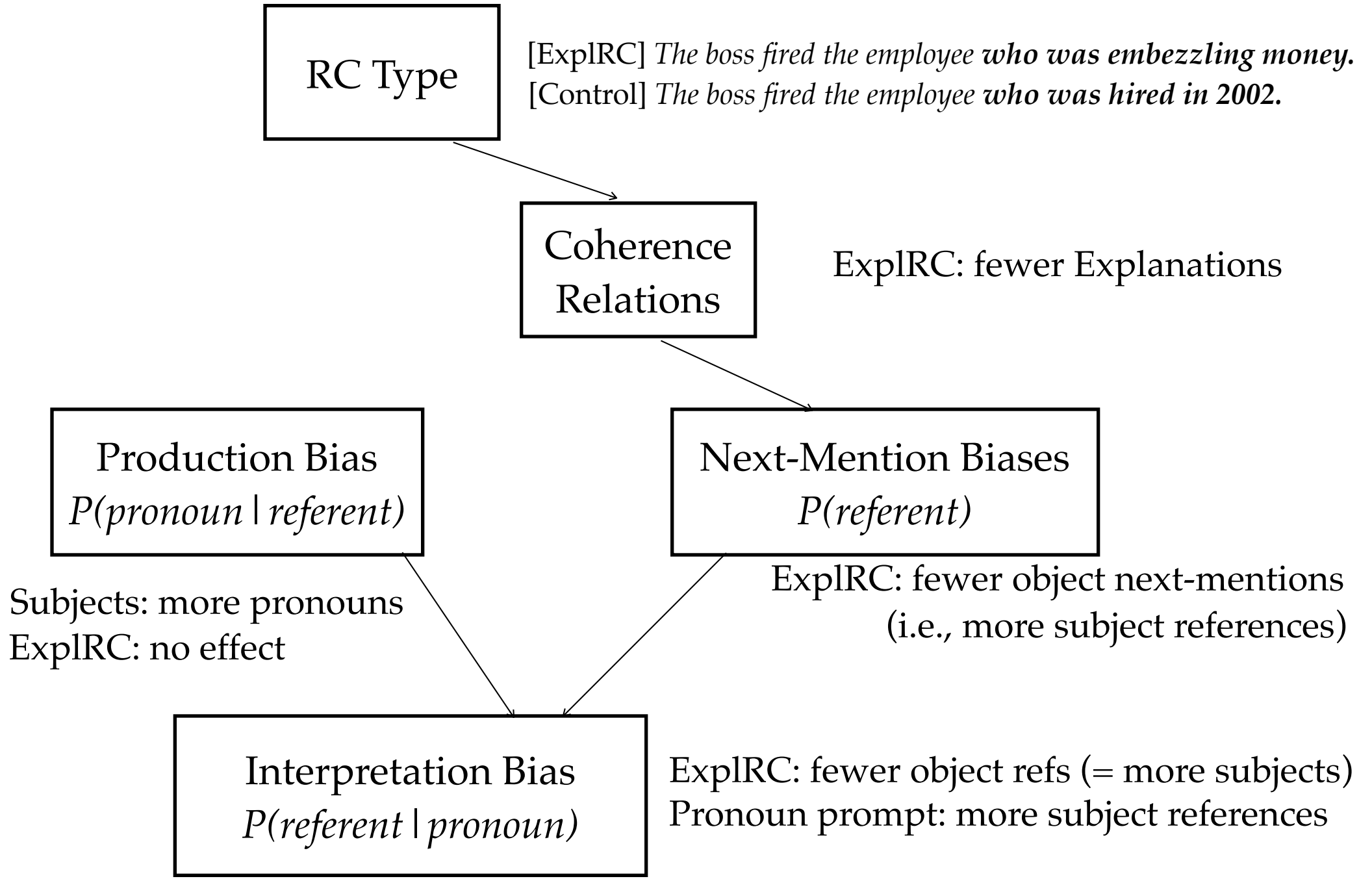
  The boss fired the employee who was embezzling money. He _____ [ExplRC]

  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)

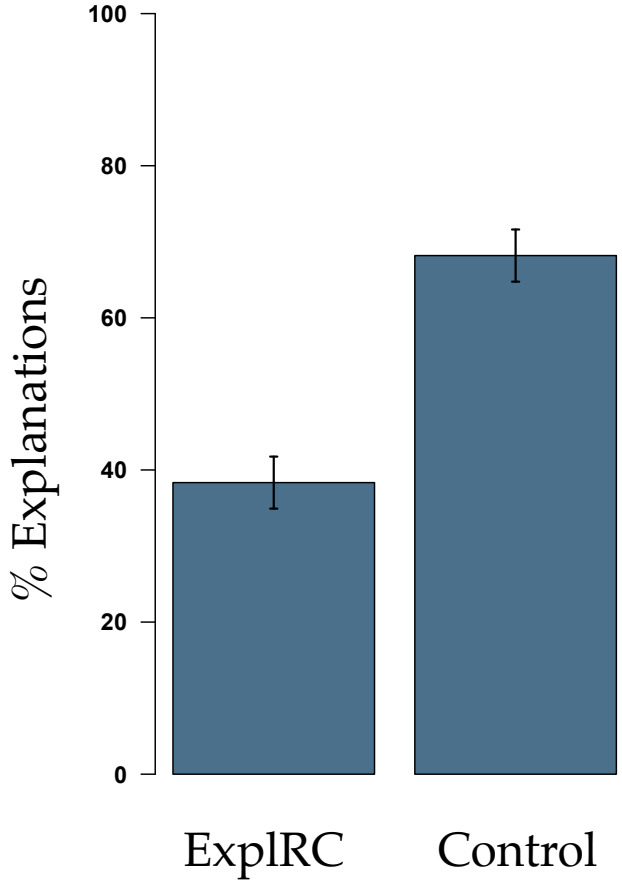
Predictions



Prediction 1: Coherence Relations

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

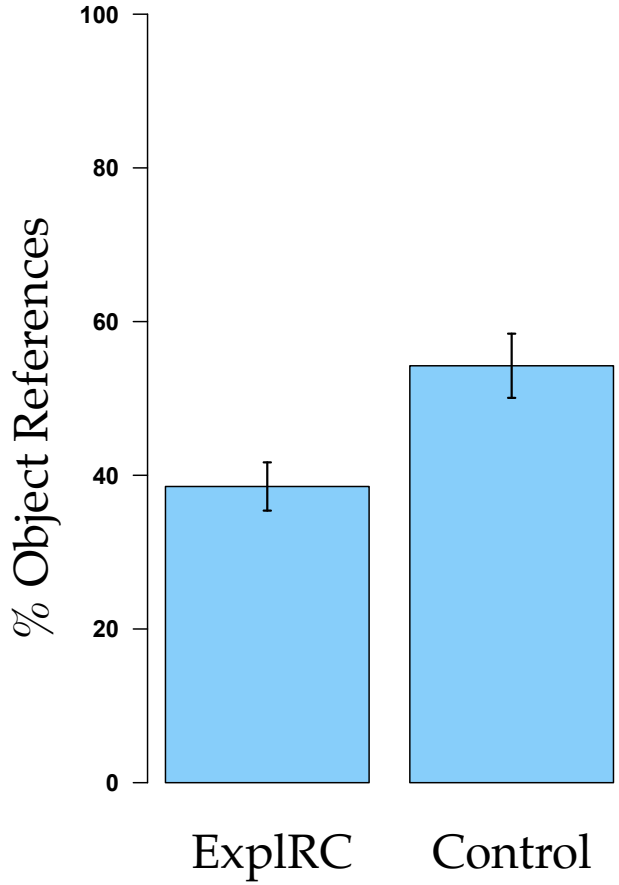
- Confirmed: (β=2.06; p<.001)



*[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

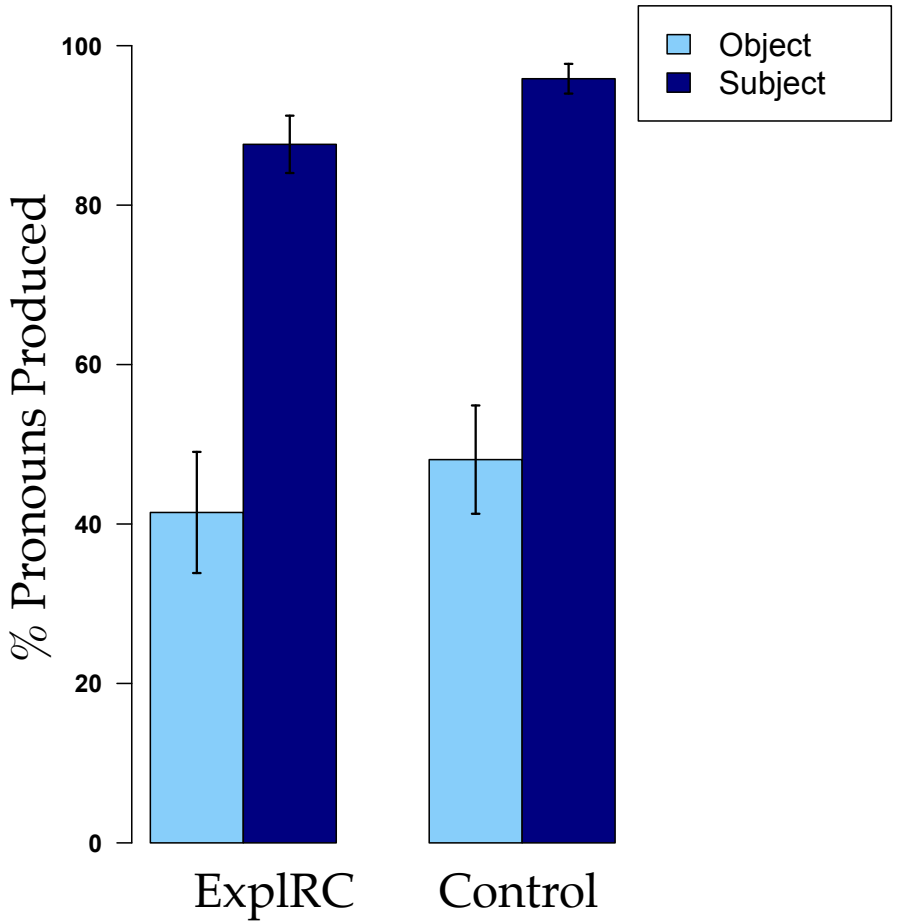
- For free-prompt condition, predict a smaller percentage of next mentions of the object in ExplRC condition than the Control condition
- Confirmed: (β=.720; $p<.05$)



*[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

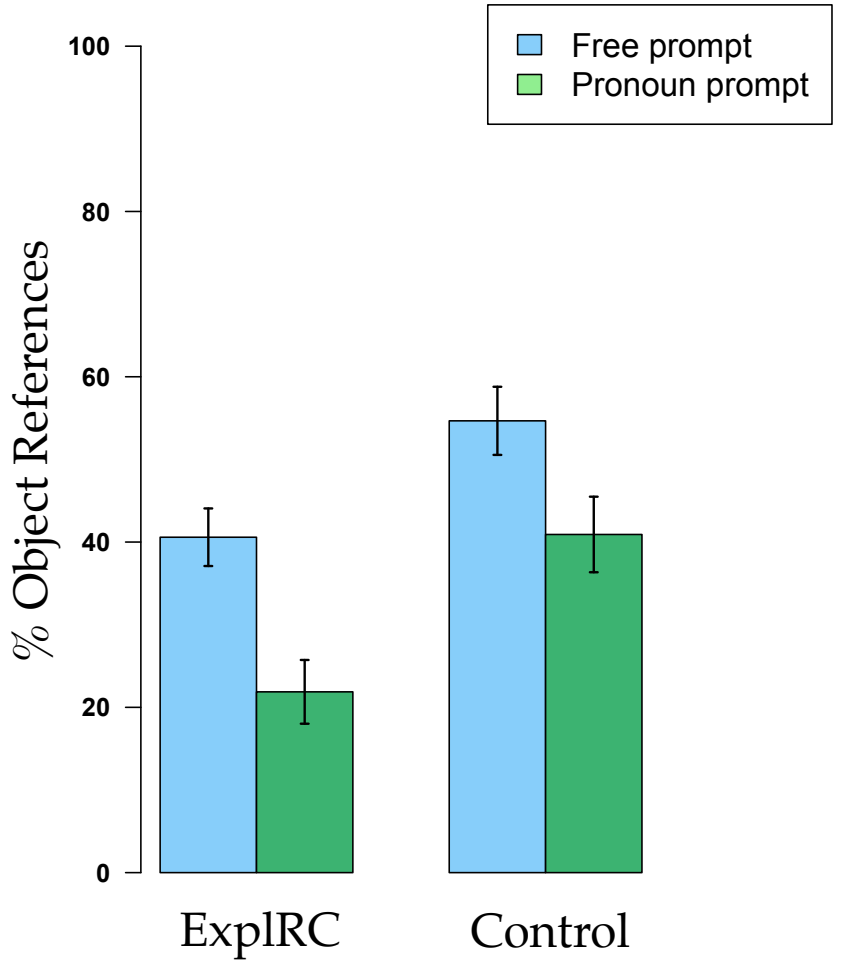
- Predict an effect of grammatical role on pronominalization rate (favoring subjects; free prompt condition)
  - Confirmed: (β=4.11; p<.001)
- But no interaction with RC condition
  - Confirmed (β=0.12; p=.92)
  - Marginal effect of RC condition (β=0.94; p=.078)



[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

- 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

- The common wisdom: there is a unified notion of entity salience that mediates between production and interpretation
- Hence, the factors that comprehenders use to interpret pronouns are the same ones 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

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

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

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

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

Experimental Summary

- Pronoun interpretation is sensitive to coherence factors, in this case the invited inference of an explanation
- Pronoun production, however, is not
- The data demonstrate precisely the asymmetry predicted by the Bayesian analysis
- A corollary is that there is no unified notion of salience that guides both interpretation and production
- Indeed, perhaps the best *independent* measure of salience is provided by next-mention expectations, but pronoun biases are not the same (Miltsakaki, 2007)

Lessons for Computational Approaches

- In recent computational work, advances in modeling have outpaced advances in feature engineering
- Basic cue-driven models are still fairly standard
- Lack of annotated training data is an impediment to using anything beyond the most general features (number, gender, distance, etc)
- Using fine-grained information about verb semantics and coherence is untenable without very large annotated data sets

Lessons for Computational Approaches

- But the Bayesian model suggests that we don't need them:
  - The likelihood (production model) can be trained on (limited amounts of) annotated data
  - The prior (next-mention model) can be trained on cases of unambiguous reference in large corpora

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

Pronoun Dependent

Pronoun Independent

Lessons for Computational Approaches

- The situation is analogous to the Bayesian approaches to other tasks, e.g. speech recognition:

$$P(\text{word} \mid \text{acoustic signal}) = \frac{P(\text{acoustic signal} \mid \text{word})\, P(\text{word})}{\sum_{\text{word} \in \text{words}} P(\text{acoustic signal} \mid \text{word})\, P(\text{word})}$$

- Pronouns are similarly underspecified linguistic signals that, while placing constraints on their interpretation, may be ambiguous and hence require reference to contextual information to fully resolve

Conclusions

- The data presented here suggests a potential reconciliation of coherence-relation-driven and Centering-driven theories:
  - Coherence relations create top-down expectations about next mention
  - Centering-style constraints yield bottom-up evidence specific to choice of referential form

$$P(\text{referent} \mid \text{pronoun}) = \frac{P(\text{pronoun} \mid \text{referent})\ P(\text{referent})}{P(\text{pronoun})}$$

Production (Centering-Driven) — Prior Expectation (Coherence-Driven)

- Fits within a modern view in psycholinguistics that casts interpretation as the interaction of "top-down" expectations and "bottom-up" linguistic evidence

Thank you!