Pronominal reference & inferred explanations: a Bayesian account

Hannah Rohde & Andrew Kehler

RefNet, 31 August 2014

When is a pronoun felicitous?

- Common wisdom: When referring to an entity that is salient, accessible, in focus, or the center of attention
  (Ariel, 1990; Gundel et al., 1993; Grosz et al., 1995; Arnold, 2001, inter alia)
- Production and interpretation cast as mirror images
- Both influenced by same factors

This talk:

- Contexts that appear to uphold this generalization
- Contexts that don't
- Bayesian account of pronoun use
- Psycholinguistics study

Implicit Causality (IC) contexts

- Implicit causality (IC) verbs favor re-mention of one referent in subsequent Explanations (Garvey & Caramazza, 1974; Caramazza, et al., 1977; Brown & Fish, 1983; McKoon et al., 1993; Kehler et al., 2008)

John <u>amused</u> Bob. He was riding a unicycle blindfolded.	**IC1**
John <u>noticed</u> Bob. He was riding a unicycle blindfolded.	**IC2**

IC interpretation & production

- Story continuation tasks (Fukumura & van Gompel, 2010, Rohde, 2008, Rohde & Kehler, 2014, Stevenson et al., 1994)
- Production choices with IC1 verbs

John amused Bob. *He was riding a unicycle blindfolded*

→ subject bias for re-mention
→ subject bias for pronominalization

- Interpretation choices with IC1 verbs

John amused Bob. He *was riding a unicycle blindfolded*

→ subject bias for pronoun interpretation

- Interpretation/production biases point in same direction.

Asymmetry

- Contexts with IC2 verbs (Rohde 2008, Fukumura & van Gompel 2010, Rohde & Kehler, 2014)

John noticed Bob. *Bob was riding a unicycle blindfolded*

→ object bias for re-mention
→ no object bias for pronominalization (names instead)

John noticed Bob. He *was riding a unicycle blindfolded*

→ object bias for pronoun interpretation

John noticed Bob. *He applauded*

→ subject bias for pronominalization

- Asymmetry between interpretation and production

Bayesian account (Kehler et al. 2008)

Interpretation — Prior — Production

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

John noticed Bob. _____

P(Bob)=.83 P(pronoun | Bob)=.4

P(John) =.17 P(pronoun | John)=1.0

John noticed Bob. He _____

P(Bob | pronoun) = .6

(Rohde & Kehler, LCP 2014)

Bayes' estimate $P(Bob \mid pronoun) = \frac{.83 * .4}{.83*.4 + .17*1.0} = .66$

Bayesian account of pronoun use

P(referent | pronoun) ~ P(referent) P(pronoun | referent)

Proposal

- *P(referent)* reflects semantic factors (e.g., coherence) (Hobbs 1979)
- *P(pronoun| referent)* reflects information structure (e.g., subjects as topics) (Grosz et al. 1995)

Prediction

- Manipulate coherence to change *P(referent)* while leaving *P(pronoun | referent)* the same.
- Together, these biases should account for the resulting pattern of pronoun interpretation, as per Bayes' Rule.

Inferring coherence

P(referent | pronoun) ~ P(referent) P(pronoun | referent)

The doctor reproached the patient who came in at 3pm. *He kept forgetting to take his medicine.*

The doctor reproached the patient who never takes his medicine. *He then prescribed a new medication.*

Control RC

Explanation RC

→ Explanation RC will reduce bias to explain
(Simner & Pickering, 2005, Bott & Solstad, 2012)

→ Explanation RC will reduce bias to mention object

→ RC manipulation will not impact pronominalization

→ Given Bayes' Rule, pronoun interpretation will reflect RC manipulation via the prior.

Experiment

- Materials: RC type x prompt type

[ExplRC,free] **The doctor reproached the patient who never takes his medicine. _____**
[Control,free] **The doctor reproached the patient who came in at 3pm. _____**

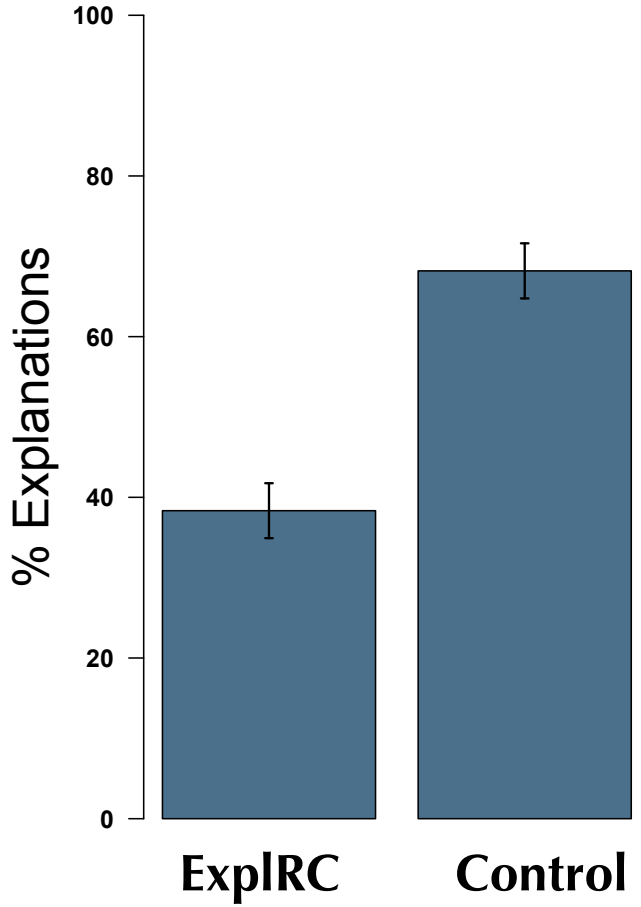
[ExplRC,pro] **The doctor reproached the patient who never takes his medicine. He __**
[Control,pro] **The doctor reproached the patient who came in at 3pm. He _____**

- Methods:
  N=40, 24 targets, 36 fillers,
  pictures to indicate gender of referents

- Annotation
  Coherence relations (Explanation or Other)
  Next-mentioned referent (Subject or Object)
  Form of Reference (Free prompt only; Pronoun or Other)

Results: Coherence relations

- Fewer Explanation continuations following Explanation RCs than Control RCs ($p<.001$)

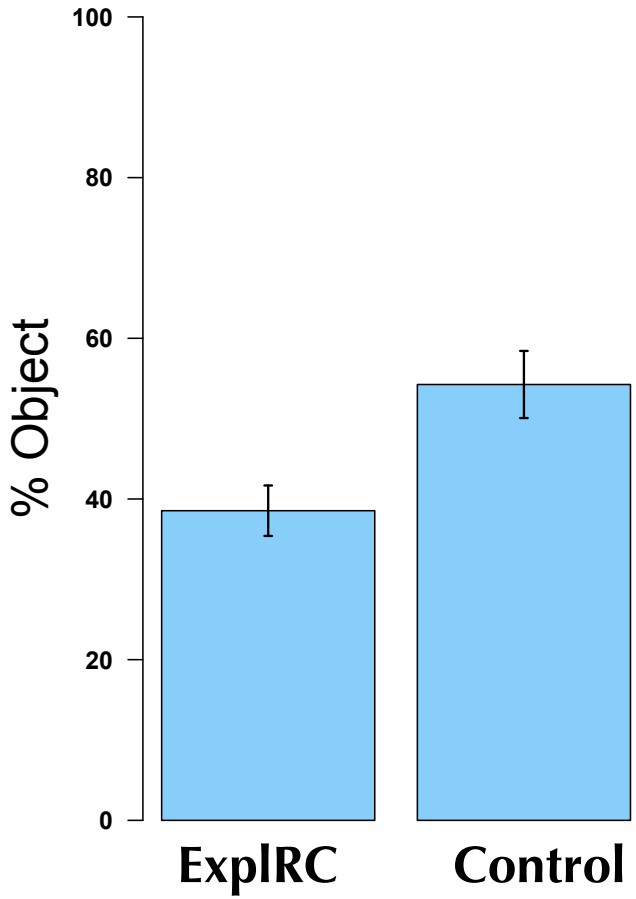


[ExplRC] The doctor reproached the patient who never takes his medicine.
[Control] The doctor reproached the patient who came in at 3pm.

Results: Next-mention biases

P(referent | pronoun) ~ P(referent) P(pronoun | referent)

- With free prompts, fewer object continuations following Explanation RCs than Control RCs ($p<.05$)



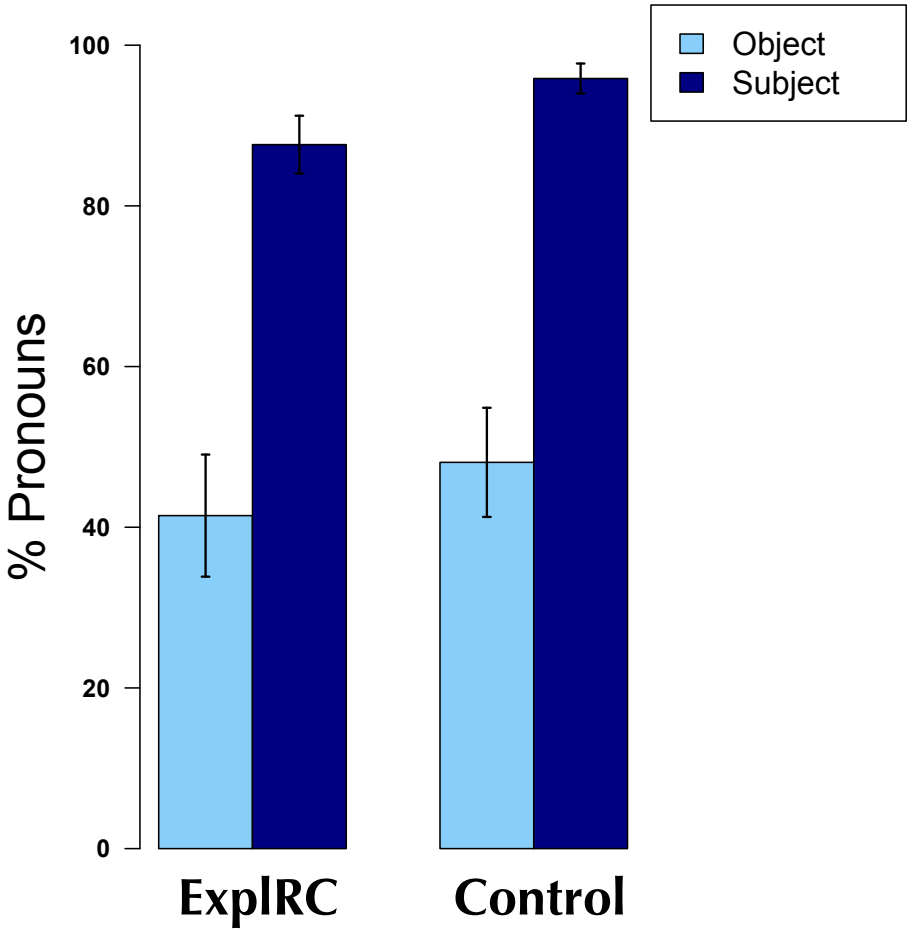
[ExplRC,free] **The doctor reproached the patient who never takes his medicine. __**

[Control,free] **The doctor reproached the patient who came in at 3pm. _____**

Results: Rate of pronominalization

P(referent | pronoun) ~ P(referent) P(pronoun | referent)

- In free prompts, more pronouns for subject referents (p<.001)…
- …regardless of RC type (no RC type X grammatical role interaction, p=.92)

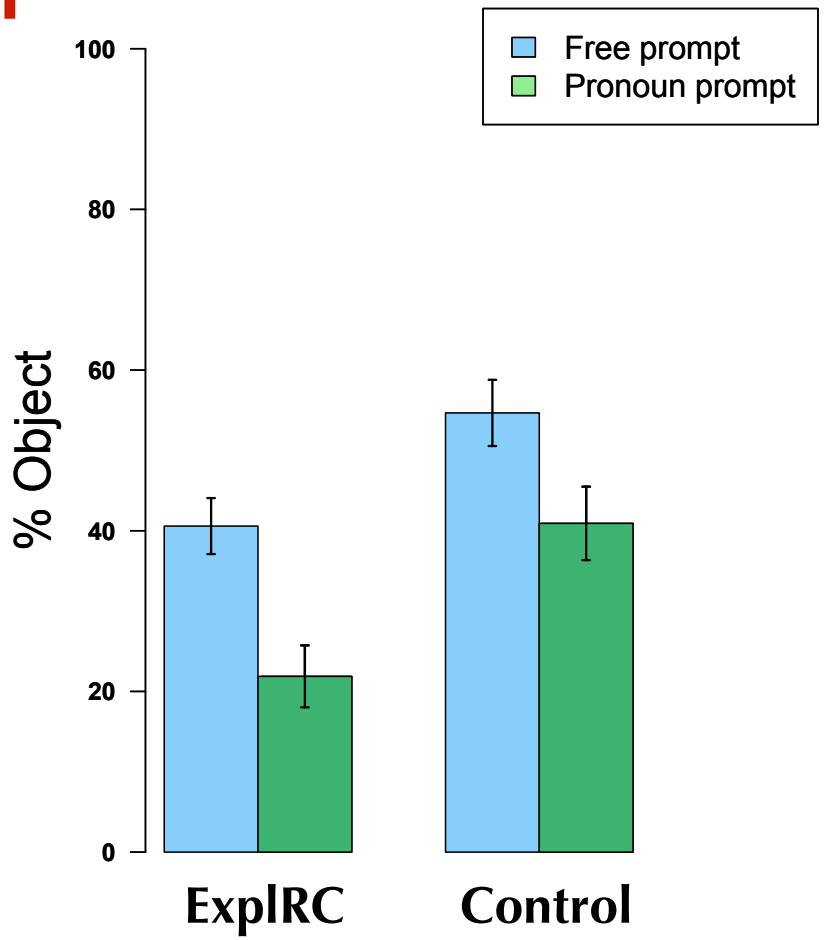


[ExplRC,free] The doctor reproached the patient who never takes his medicine. __
[Control,free] The doctor reproached the patient who came in at 3pm. _____

Results: Pronoun interpretation

P(referent | pronoun) ~ P(referent) P(pronoun | referent)

- With Pronoun prompts, fewer object continuations for Explanation RCs than Control RCs (p<.005)…
- …and more subject continuations for Pronoun than Free prompts (p<.001)
- Marginal interaction between RC type and prompt type (p=.078)



[ExplRC,free] The doctor reproached the patient who never takes his medicine. _____
[Control,free] The doctor reproached the patient who came in at 3pm. _____
[ExplRC,pro] The doctor reproached the patient who never takes his medicine. He __
[Control,pro] The doctor reproached the patient who came in at 3pm. He _____

Model evaluation

- Estimating prior and likelihood from data in the free prompt condition to calculate a Bayes' derived pronoun interpretation bias
- Compare that to the observed pronoun interpretation bias in the pronoun prompt condition

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

Competing model: mirror model

- A common assumption is that the factors that interpreters use to interpret pronouns are those that speakers use when choosing to use one.
- That is, speakers use pronouns when they think the hearer's model will be biased to the intended referent.

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

Competing Model: Expectancy Model

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

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

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
NoExplRC	0.41	0.373	0.334	0.542
		R^2=.48/.49	R^2=.34/.42	R^2=.14/.12

P(referent | pronoun) ~ P(referent) P(pronoun | referent)

Conclusion

- Pronoun interpretation is sensitive to a coherence-driven factor regarding the inference of an explanation.
- Pronoun production is not.
- This shows the asymmetry between interpretation and production predicted by the Bayesian analysis.

Thanks!

IC1 contexts

John amused Bob. _____

P(John)=.7 P(pronoun | John)=.9

P(Bob) =.3 P(pronoun | Bob) =0.0

John amused Bob. He _____

P(John | pronoun) = 1.0

(Rohde & Kehler, LCP 2014)

Bayes' estimate $P(John \mid pronoun) = \frac{.7 * .9}{.7*.9 + .3*0.0} = 1.0$