WHEN CIRCUMSTANCES CHANGE, UPDATE YOUR PRONOUNS
Joshua K. Hartshorne¹, Mariela V. Jennings¹, Tobias Gerstenberg², Joshua Tenenbaum³
¹Department of Psychology Boston College, ²Department of Psychology Stanford University, ³Department of Brain and Cognitive Sciences Massachusetts Institute of Technology
Poster available at l3atbc.org

Background & Motivation
Language is frequently ambiguous, with the same sentence having several possible interpretations. One particularly prevalent example is third-person pronouns:

(1) Al beat Bart at tug-of-war because he is strong.
(2) Al beat Bart at tug-of-war because he is weak.

Most people interpret the pronoun in (1) as referring to Al; and the pronoun in (2) as referring to Bart.

Problem: Different theories of pronoun interpretation pose that pronoun interpretation is a function of verb biases¹⁰ and that pronouns refer to the most causally-connected entity⁹. They cannot explain (1) and (2). Handling such sentences requires recourse to world knowledge¹⁰⁻¹²

Question: What is the format of world knowledge?

Model
• Language interpretation = inference over a generative model of the speaker (Figure 1)
• World knowledge = inference over a generative model of the world¹³
• Informative speakers say things that are true; the listener uses world knowledge to fill in any gaps
• Listeners interpret he as referring to Al in (1) and Bart in (2) because those interpretations result in the most probable worlds

![Figure 1](image)

References

Acknowledgements
Funding source: NIH NRSA award to Joshua K. Hartshorne (5F32HD072748)

Testing the Model
(1) Create a simplified world in which we can fully characterize world knowledge
(2) Compare our model’s predictions to judgements provided by participants recruited via Amazon Mechanical Turk

Results

Conclusions
• Pronoun interpretation is heavily influenced by dynamically updated beliefs about the world
• The generative intuitive theory approach provides a flexible definition of world knowledge which is crucial for accurately capturing human judgements