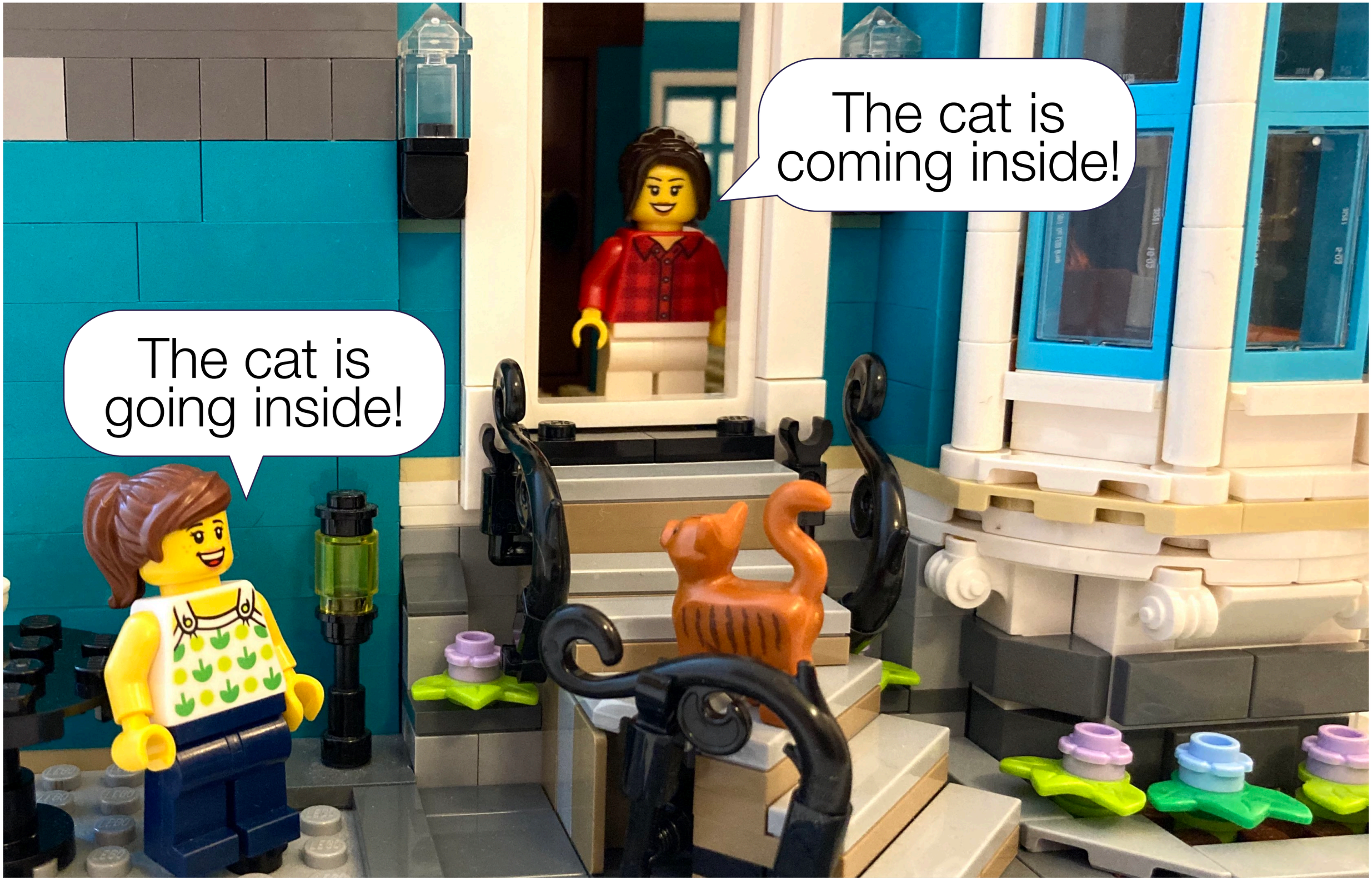
A Bayesian Approach to Modeling Grammatical Perspective-Taking

Carolyn Jane Anderson
Wellesley College

Perspective:

- the point-of-view from which an event or object is seen
- a person's beliefs and attitudes
- a person's self-identified spatial and temporal location

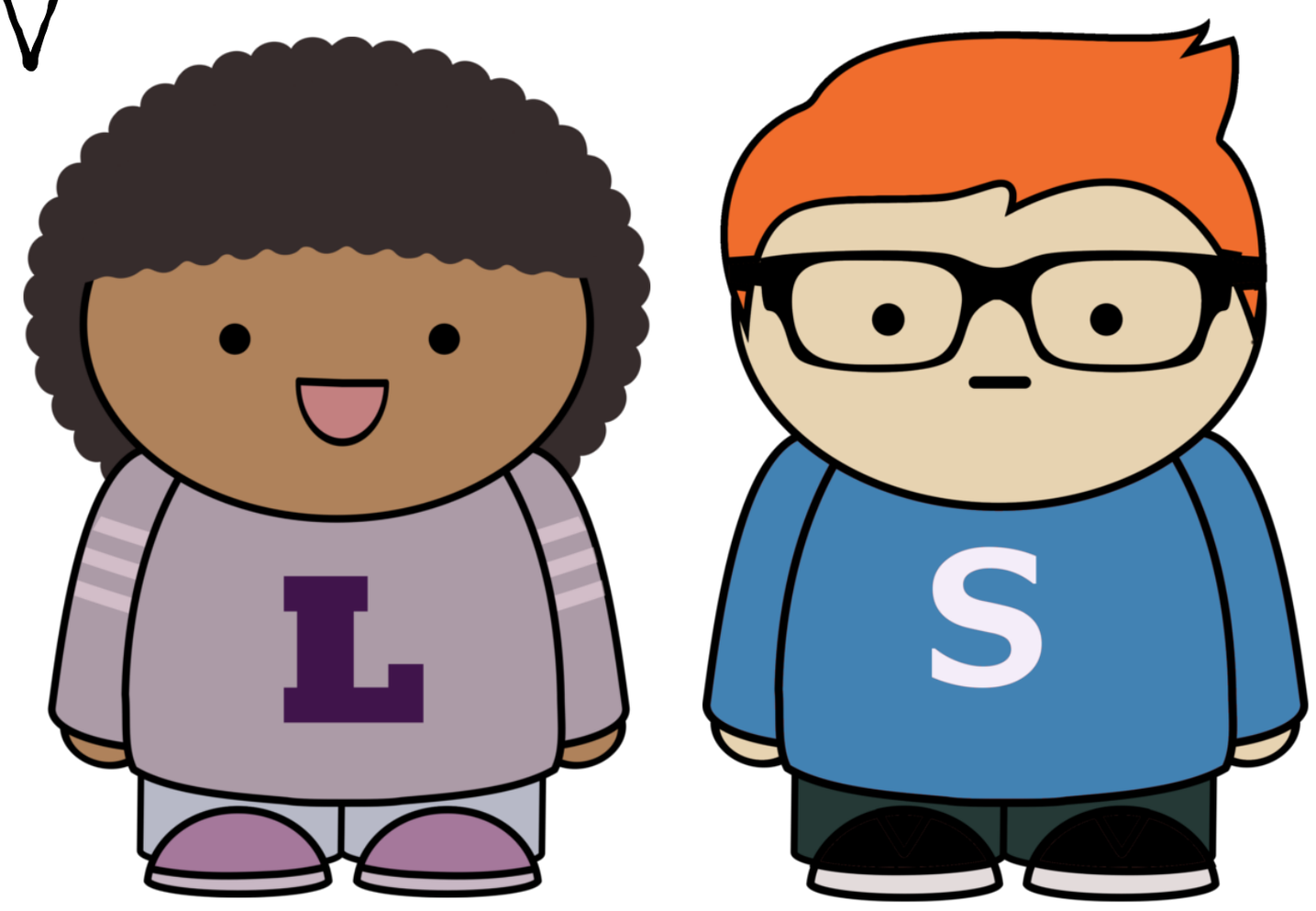

The cat is going inside!
The cat is coming inside!


What a magnificent creature!
I'd say it's average at best.

Perspectival Expression:

an expression whose meaning depends on the beliefs, perception, or location of a prominent individual chosen by the speaker.

Last weekend Thelma _____ with me to New York. We saw the Statue of Liberty, took a ferry to Staten Island, and ate pizza. We even asked a New Yorker for directions just to see him get mad!

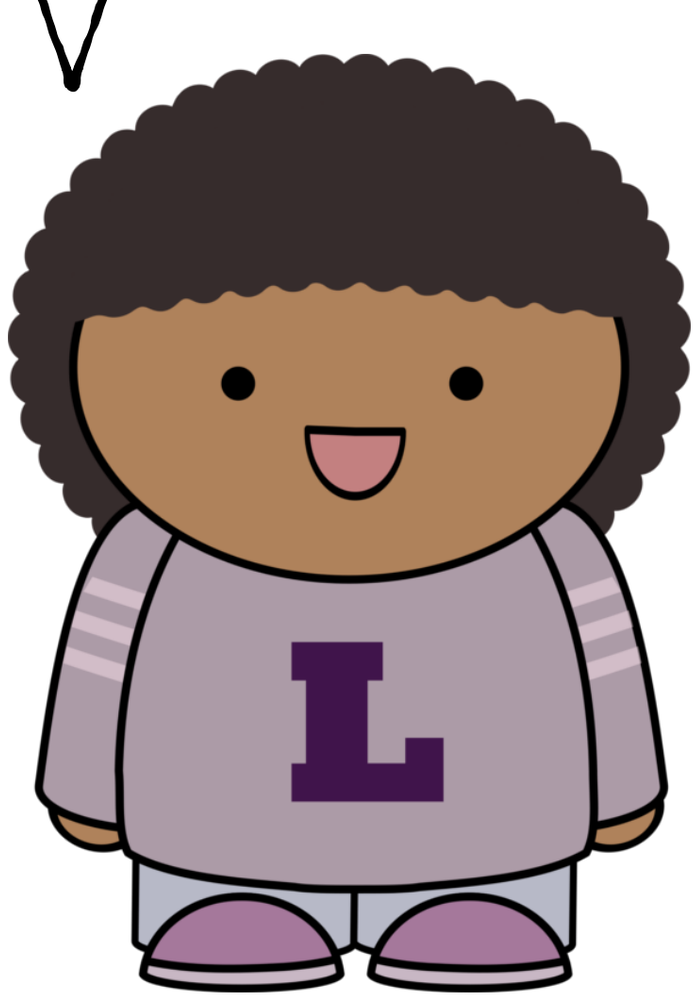


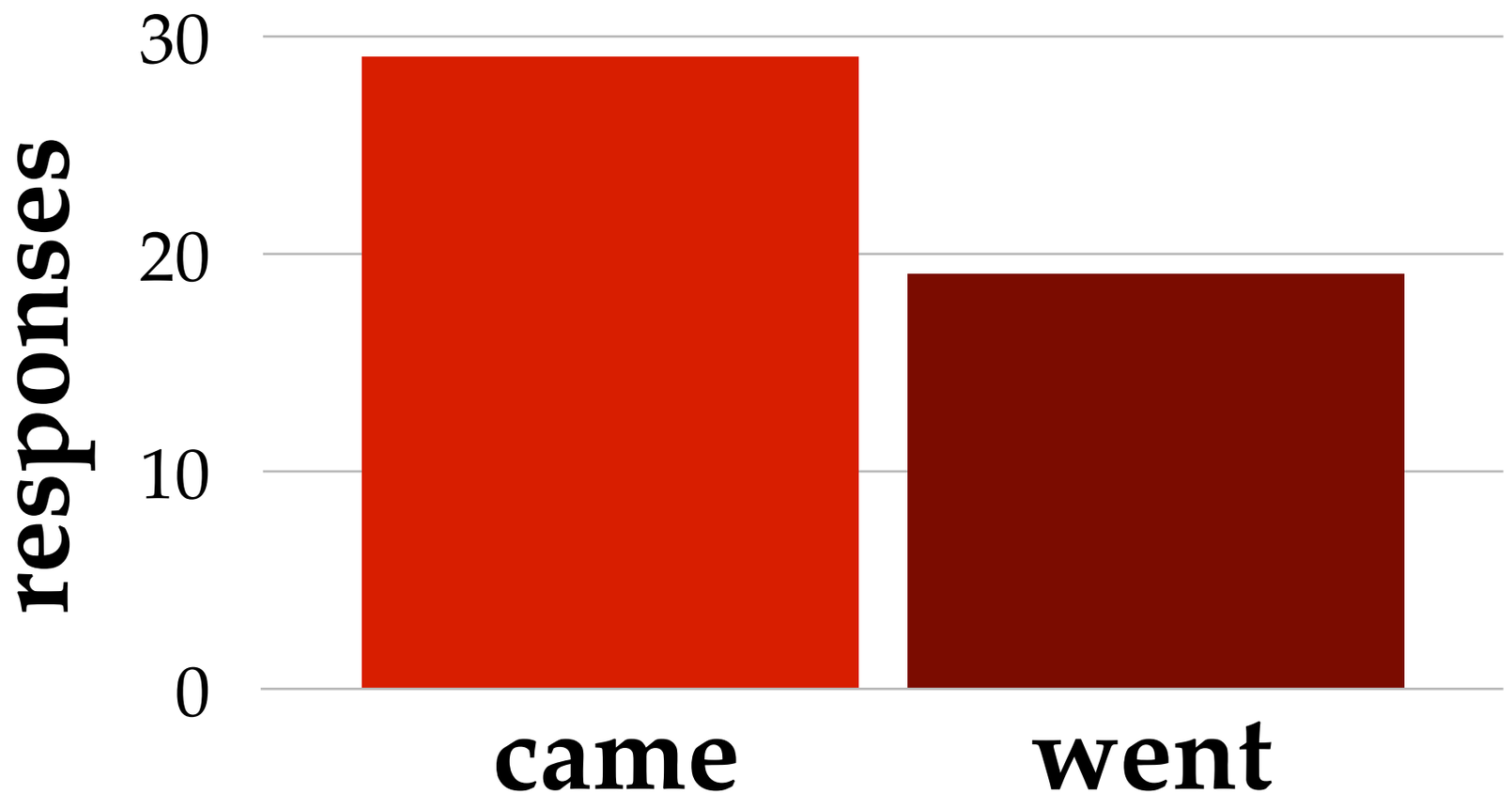
Fill in the blank:

(a) came

(b) went

Last weekend Thelma _____ with me to New York. We saw the Statue of Liberty, took a ferry to Staten Island, and ate pizza. We even asked a New Yorker for directions just to see him get mad!

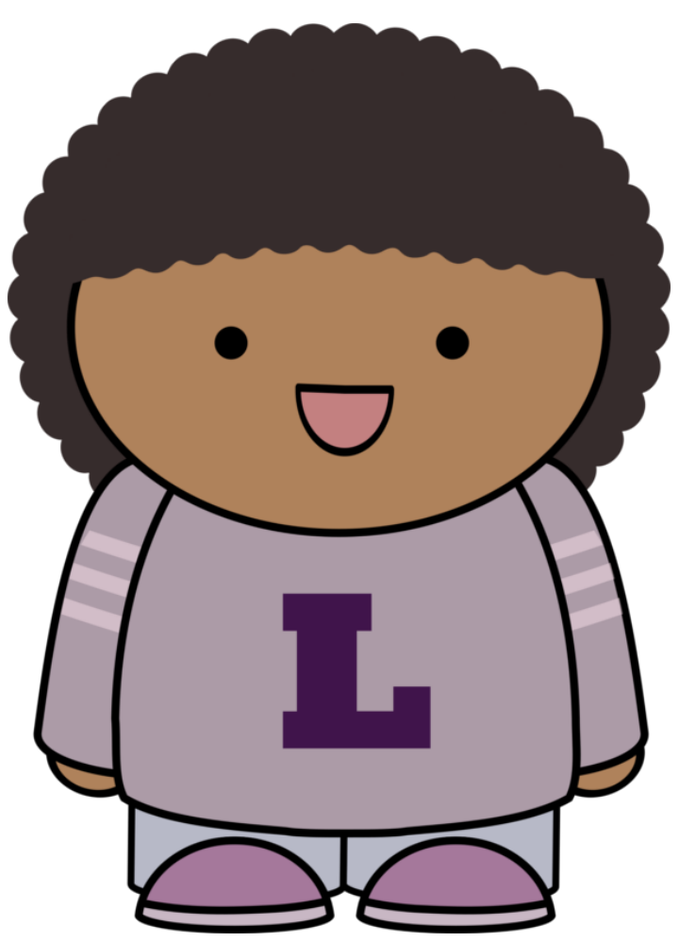


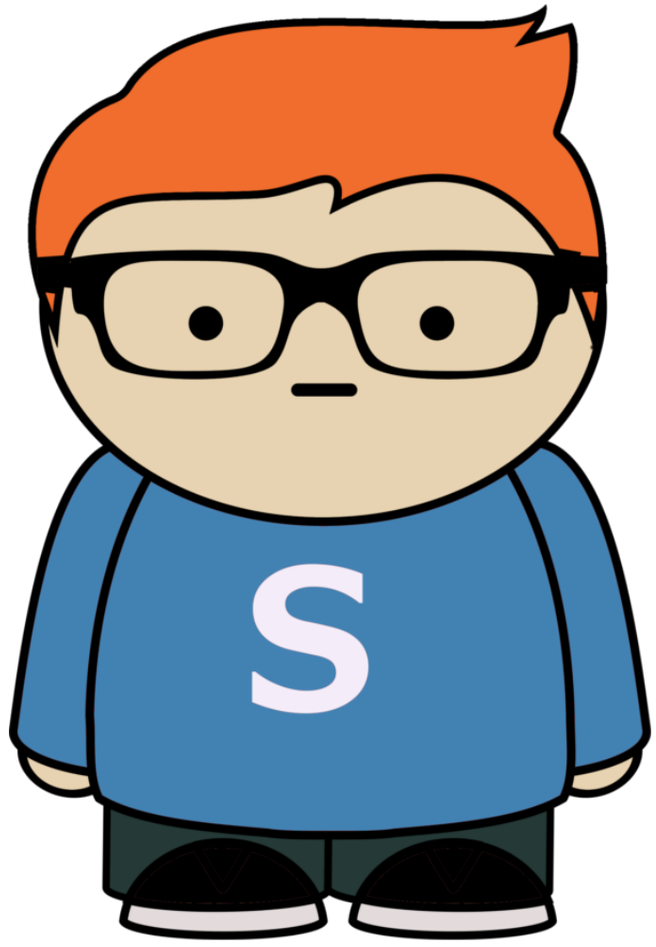


Showing my exchange student Thelma the US has been fun!

How many states has she been to so far?

She's been to Connecticut, Maine, and New Hampshire. And last weekend Thelma _____ with me to New York. So, counting Vermont, five.





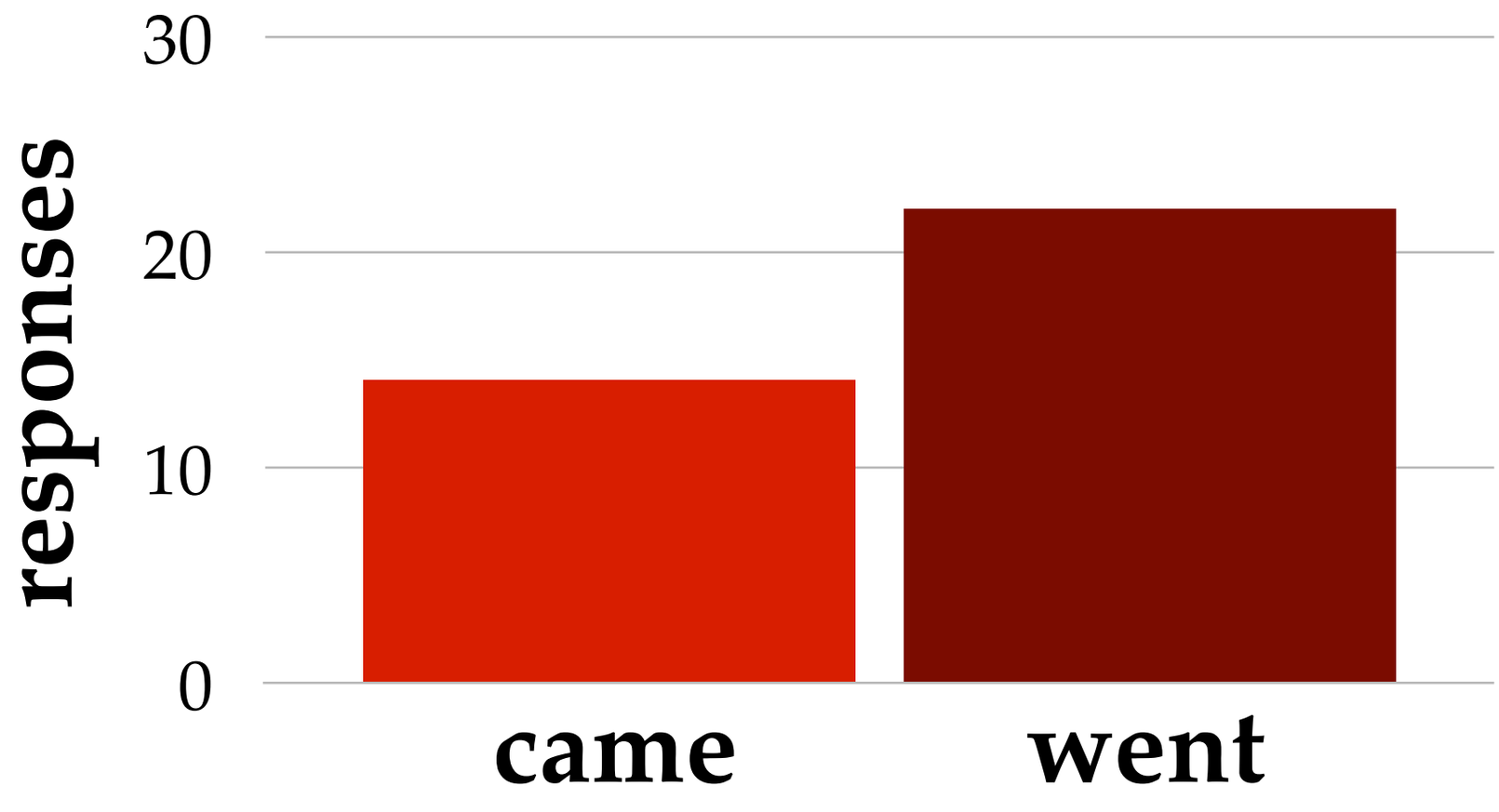
Fill in the blank:

(a) came

(b) went

Showing my exchange student Thelma the US has been fun!

She's been to Connecticut, Maine, and New Hampshire. And last weekend Thelma _____ with me to New York. So, counting Vermont, five.



Last weekend Thelma _____ with me to New York.

How does context affect our preference for *come* over *go*?

How do speakers decide which verb to use?

How do listeners reason about what the speaker is trying to say?

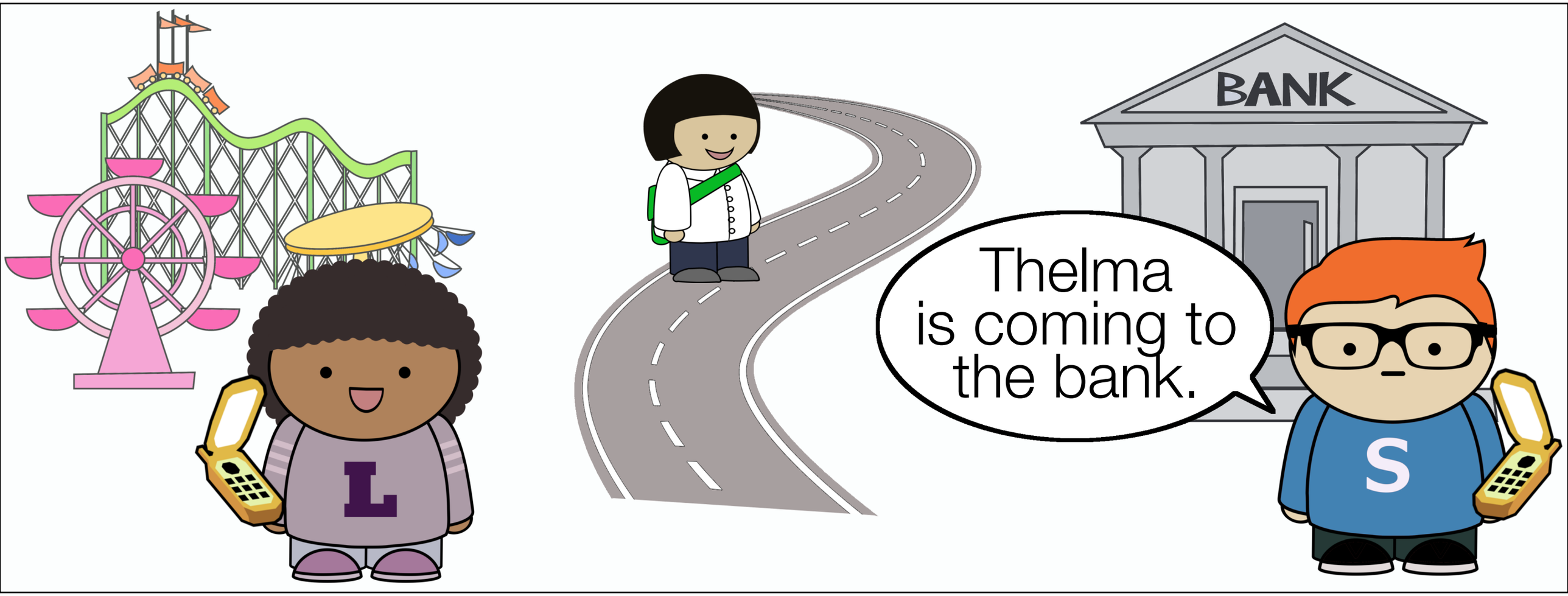
Talk Outline

1. Develop a computational model of perspectival reasoning
2. Run simulations to predict speaker and listener behavior
3. Compare model predictions against evidence from crowdsourced behavioral experiments
    - How do listeners understand perspectival expressions?
    - How do speakers choose perspectival expressions?
4. Open questions

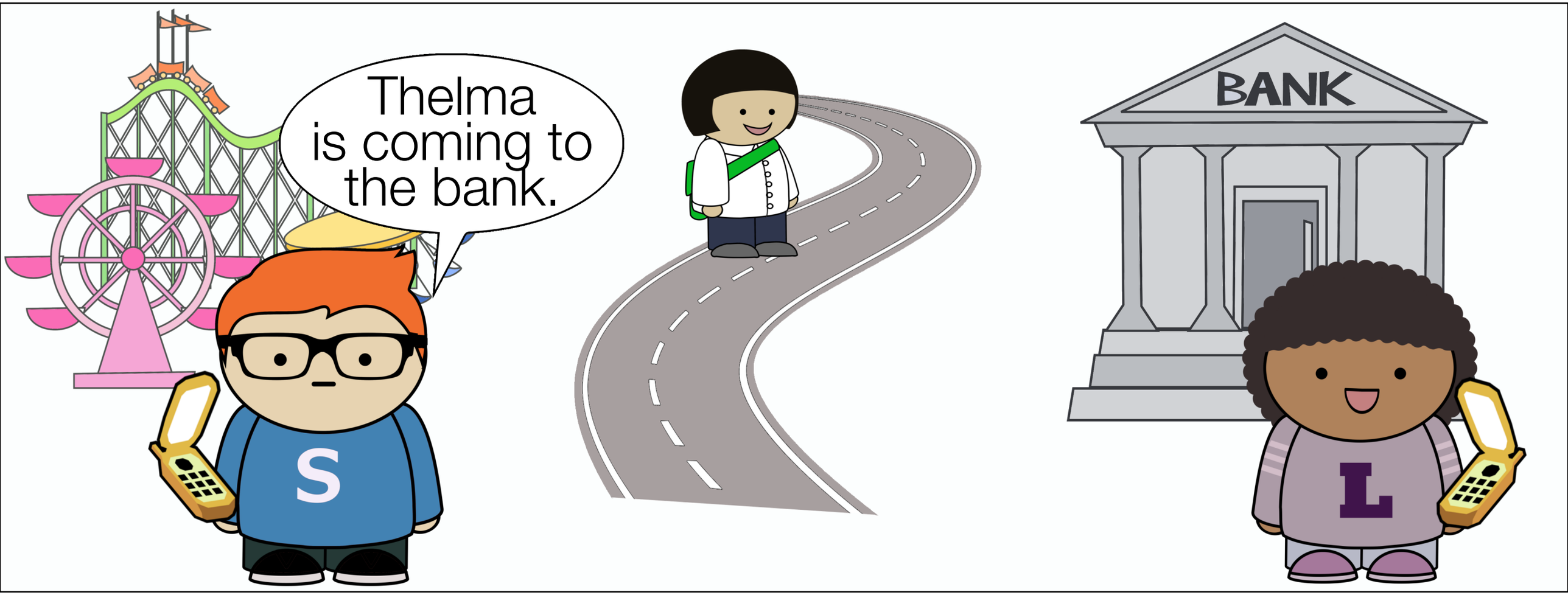
Perspectival motion verbs

Come requires the **perspective holder** to be located at the destination of motion.

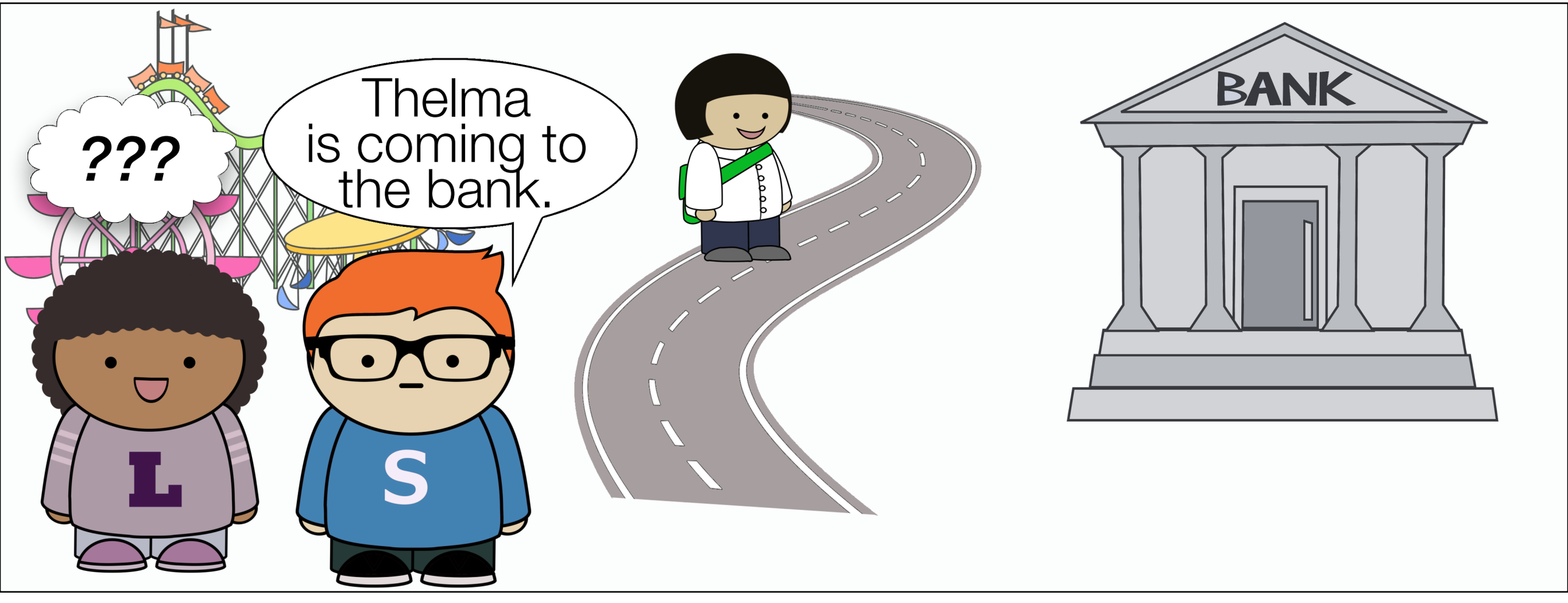
Speaker perspective



Listener perspective



No available perspective



Perspectival motion verbs

Come requires the **perspective holder** to be located at the destination of motion.

Perspective holders:

- The **listener**
- The **speaker**
- Attitude holders
- Protagonists

The Semantics of Perspectival Expressions

Indexical Family:
(Taylor 1988, Oshima 2006, Sudo 2018, Korotkova 2016)

$[[\text{come}]]^{C,g} =$
$\lambda x.\lambda e.\textsc{move}(e) \wedge \textsc{agent}(e,x) \wedge \textsc{dest}(e,\textsc{loc}(C_{speaker}))$

Logophoric Binding Family:
(Nishigauchi 2014, Charnavel 2018, Sundaresan 2018, Charnavel 2019)

$[[\text{come}]]^{C,g} =$
$\lambda a.\lambda x.\lambda e.\textsc{move}(e) \wedge \textsc{agent}(e,x) \wedge \textsc{dest}(e,\textsc{loc}(a))$

Anaphoric Family:
(Barlew 2017, Roberts 2015)

$[[\text{come}]]^{C,g} =$
$\lambda x.\lambda e.\textsc{move}(e) \wedge \textsc{agent}(e,x) \wedge \textsc{dest}(e,\textsc{loc}(a))$

How do **speakers** pick a perspective?

p(perspective | meaning)

How do **listeners** reason about whose perspective the speaker is using?

p(perspective | utterance)

How do **contextual factors** influence perspective identification and selection?

p(perspective | context)

Talk Outline

1. **Develop a computational model of perspectival reasoning as a joint inference problem**
2. Run simulations to predict speaker and listener behavior
3. Compare model predictions against evidence from crowdsourced behavioral experiments
   - How do listeners understand perspectival expressions?
   - How do speakers choose perspectival expressions?
4. Open questions

Cognitive Biases in Perspectival Processing

- **Bias towards the speaker's perspective**
  - Speakers are **egocentric**: they prefer to use their own perspectives (Keysar & Barr 2003, Epley et al. 2004, Heller et al. 2012)
  - Accessing other perspectives is **cognitively costly** (Lin et al. 2010)
  - Speaker perspective appears to be **most common cross-linguistically** (Gathercole 1987, Nakazawa 2007, Korotkova 2016, Barlew 2017)
  - **Individual differences** in ability to access other perspectives (Pratt et al. 1996, Brown-Schmidt 2009, Wardlow 2013, Köder & Maier 2016)

- **Bias against perspective shift**
  - **Perspective shift is costly** and language **users avoid it** (Millis 1995, Harris 2012, Köder et al. 2015, Ferguson et al. 2017)
  - Perspective shift is a **risky conversational move** (Harris & Potts 2009)

Joint perspectival reasoning

Speaker:

$p(\text{perspective}, \text{utterance} \mid \text{meaning})$

Listener:

$p(\text{perspective}, \text{meaning} \mid \text{utterance})$

The Rational Speech Acts model (Frank & Goodman 2012)

- Captures **reasoning over alternatives** in a **rational approach**: agents maximize communicative success
- Speakers and listeners use **Bayesian reasoning**: they reason using recursive mental models of each other's behavior
- Easy to **implement computationally** to generate experimentally testable **quantitative predictions**

The Rational Speech Acts model (Frank & Goodman 2012)

Applied to a variety of phenomena:

scalar implicatures (Bergen et al. 2012, Degen et al. 2015, Potts et al. 2016, Brochhagen et al. 2016 …)

politeness (Yoon et al. 2016, Yoon et al. 2017)

irony (Cohn-Gordon & Bergen 2019)

hyperbole (Kao et al. 2014)

social meaning (Qing & Cohn-Gordon 2018)

visual perspective (Hawkins et al. 2021)

grammatical perspective (Anderson & Dillon 2018; Anderson & Dillon forthcoming)

Basic Rational Speech Acts Model

Speaker's goal: p(utterance | meaning)

Guess which sentence is most likely to communicate the intended meaning to the listener.

Listener's goal: p(meaning | utterance)

Guess which meaning the speaker's sentence is supposed to communicate.

A meaning is represented as a possible world sampled from the world set.

Basic Rational Speech Acts Model

Speaker's goal: p(utterance | world)

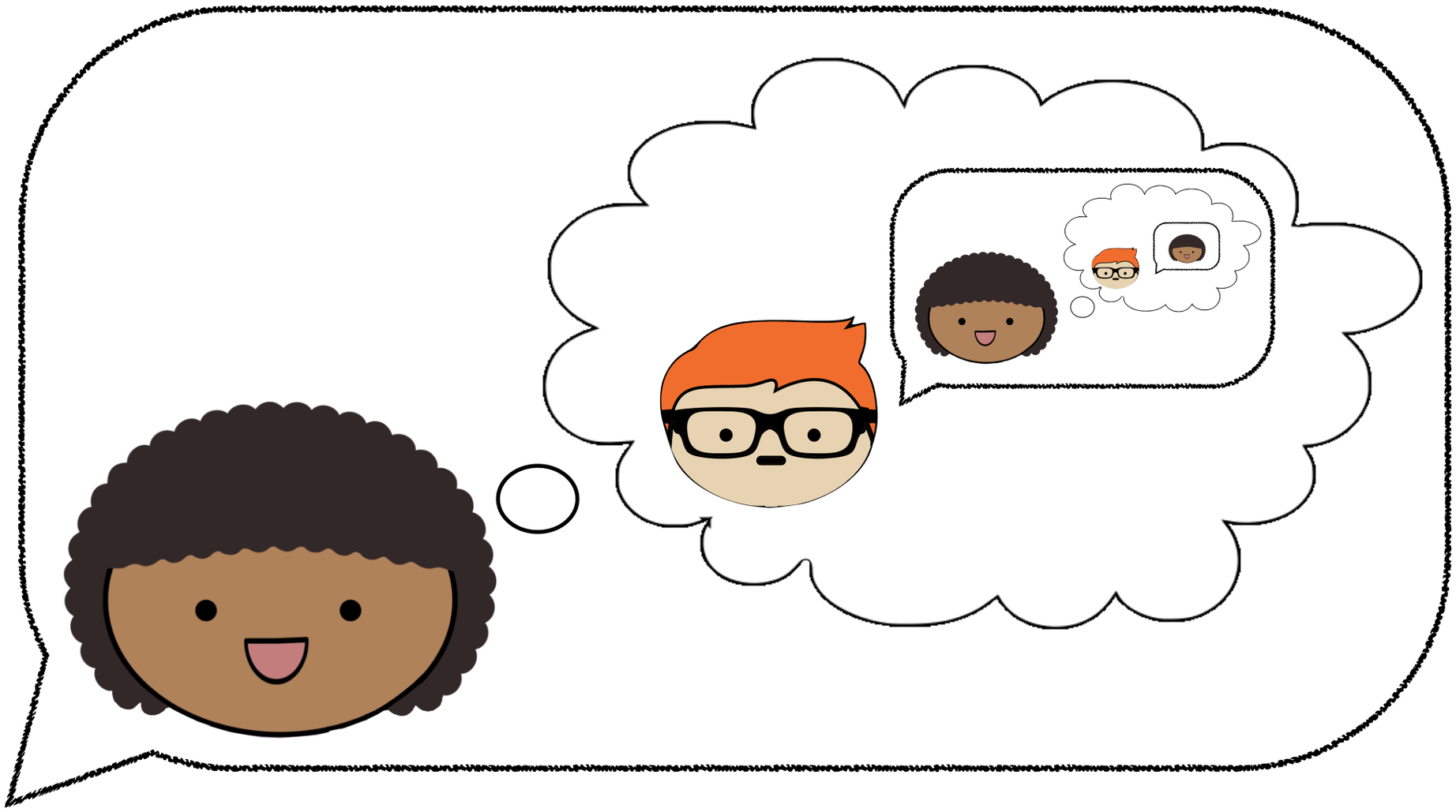
$$p(\text{utterance} \mid \text{world}) = p(\underline{\text{world} \mid \text{utterance}})\ p(\text{utterance})$$
$$p(\text{world})$$

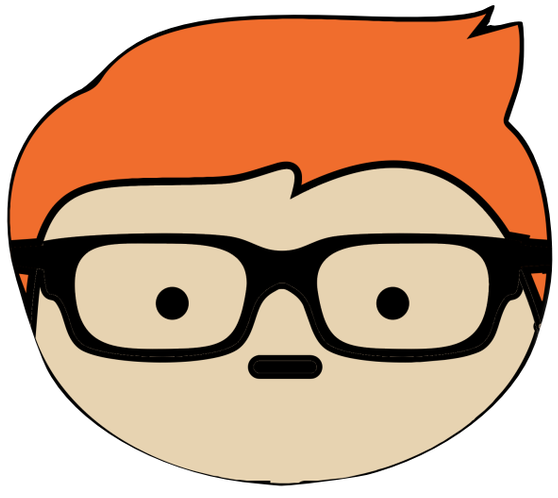
Listener's goal: p(world | utterance)

$$p(\text{world} \mid \text{utterance}) = \frac{p(\text{utterance} \mid \text{world})\ p(\text{world})}{p(\text{utterance})}$$

$$\text{by Bayes' Rule: } p(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

A Recursive Model





Speaker: $p(u \mid w) \propto p(w \mid u)\, p(u)$

Listener: $p(w \mid u) \propto p(u \mid w)\, p(w)$

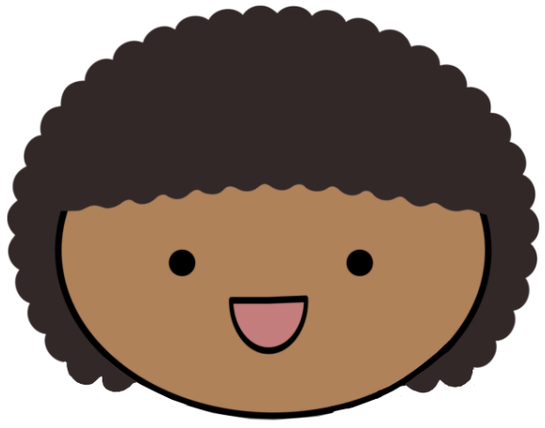
Basic Rational Speech Acts model

Pragmatic Listener: p(world | utterance)

$$p(w \mid u) \propto p(u \mid w)p(w)$$

$$L_1(w \mid u) \propto \text{Speaker}(u \mid w)\, p(w)$$

where w = world and u = utterance



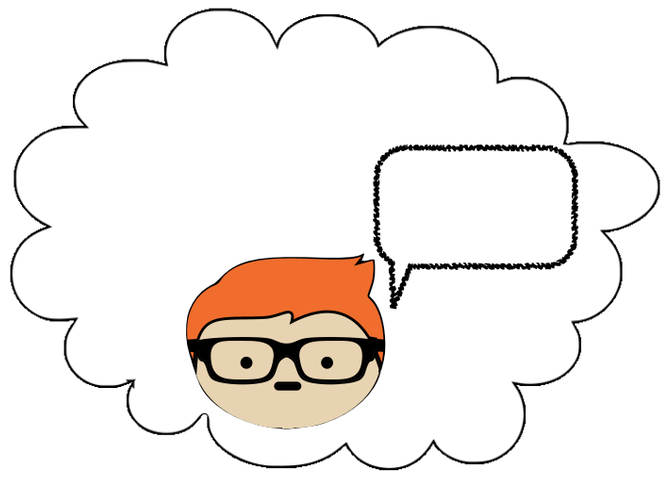
Basic Rational Speech Acts model

Pragmatic Speaker: p(utterance | world)

$$p(u \mid w) \propto p(w \mid u)\, p(u)$$

$$S_1(u \mid w) \propto \text{Max}(\text{Listener}(w \mid u)\, p(u))$$

where w = world and u = utterance

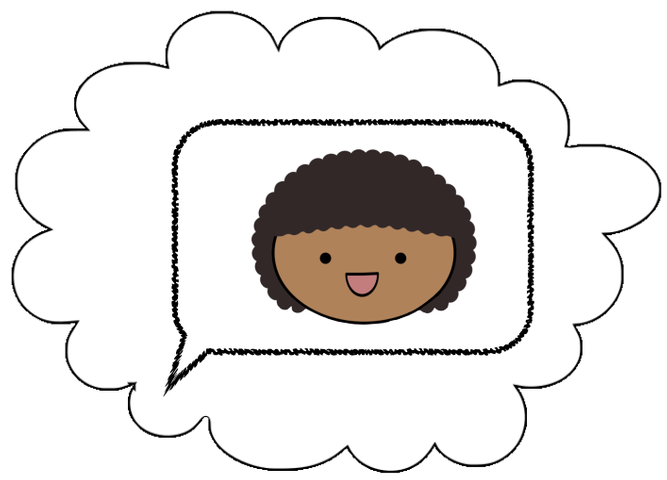


Basic Rational Speech Acts model

Literal Listener: p(world | utterance)

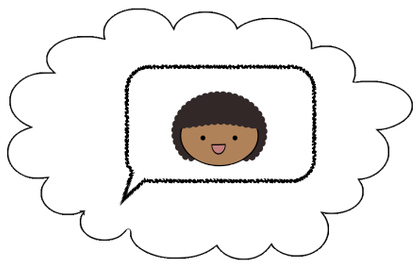
$$p(w \mid u) \propto p(u \mid w)p(w)$$

$$L_0(w \mid u) \propto [[u]]^w p(w)$$



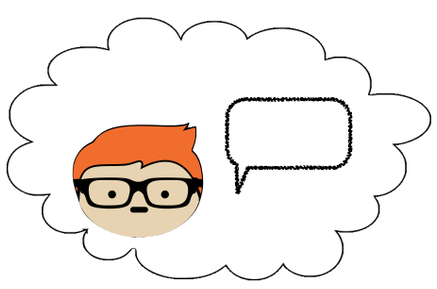
where w = world and u = utterance

Basic Rational Speech Acts model



Literal Listener: p(world | utterance)

$$L_0(w \mid u) \propto [[u]]^w p(w)$$



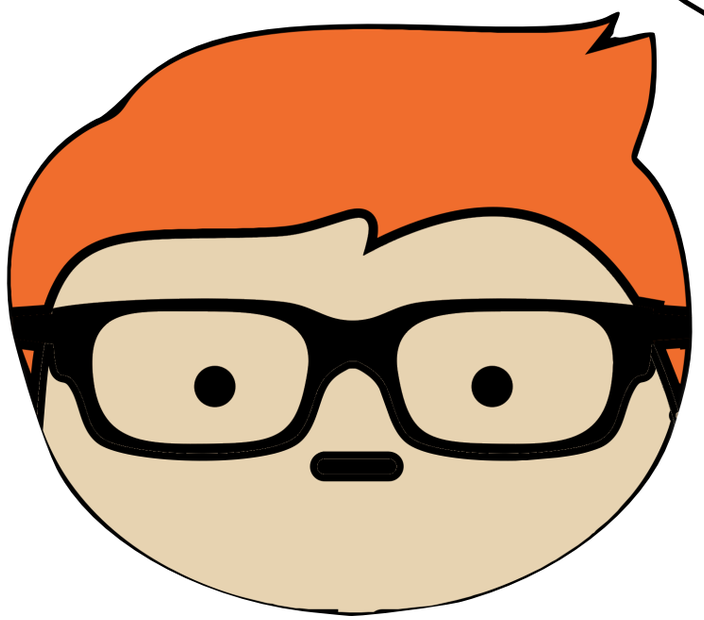
Pragmatic Speaker: p(utterance | world)

$$S_1(u \mid w) \propto \text{Max}(\text{Listener}(w \mid u)\, p(u))$$

Pragmatic Listener: p(world | utterance)

$$L_1(w \mid u) \propto \text{Speaker}(u \mid w)\, p(w)$$

Thelma is coming.



p(world | utterance)

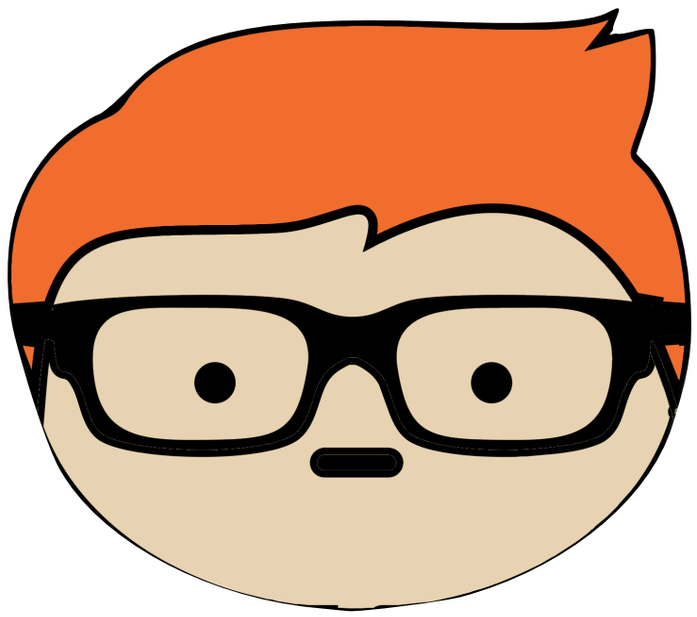
What is Sam trying to communicate?

Whose perspective is Sam using?

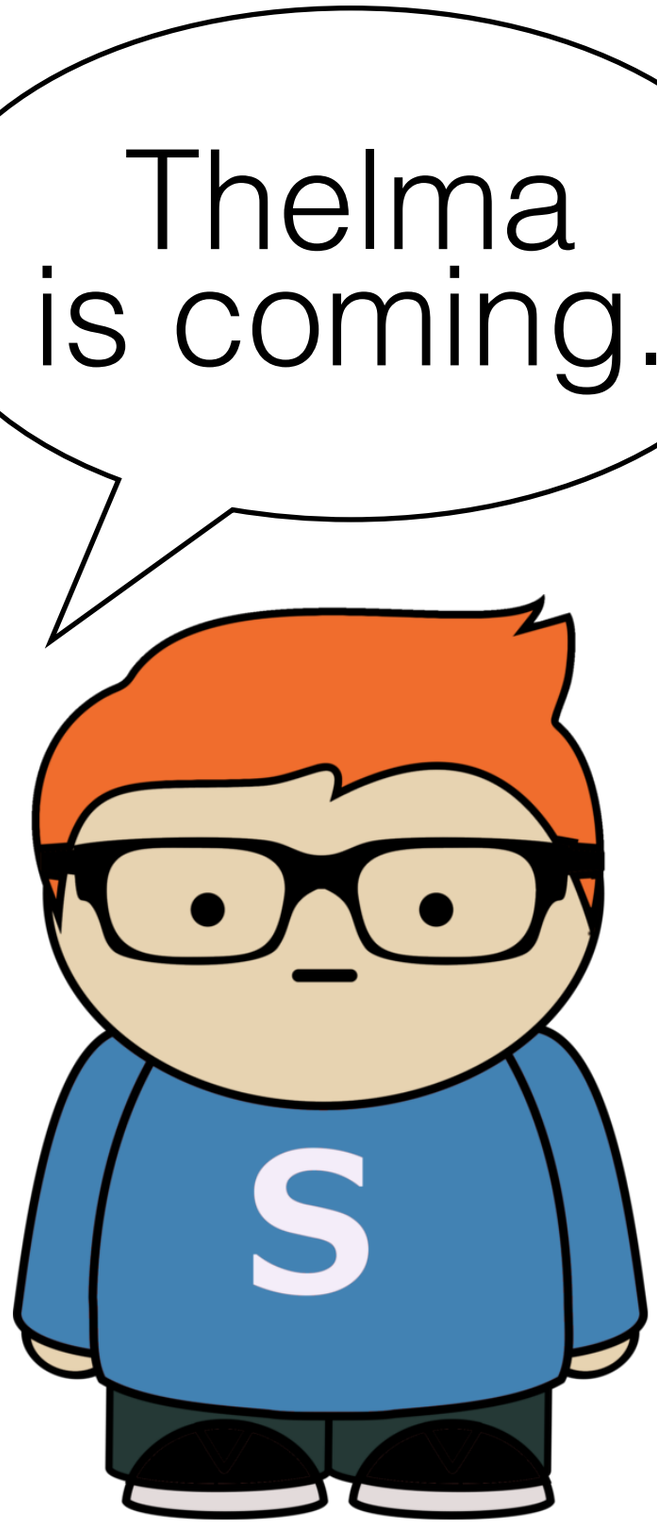
p(perspective | utterance)

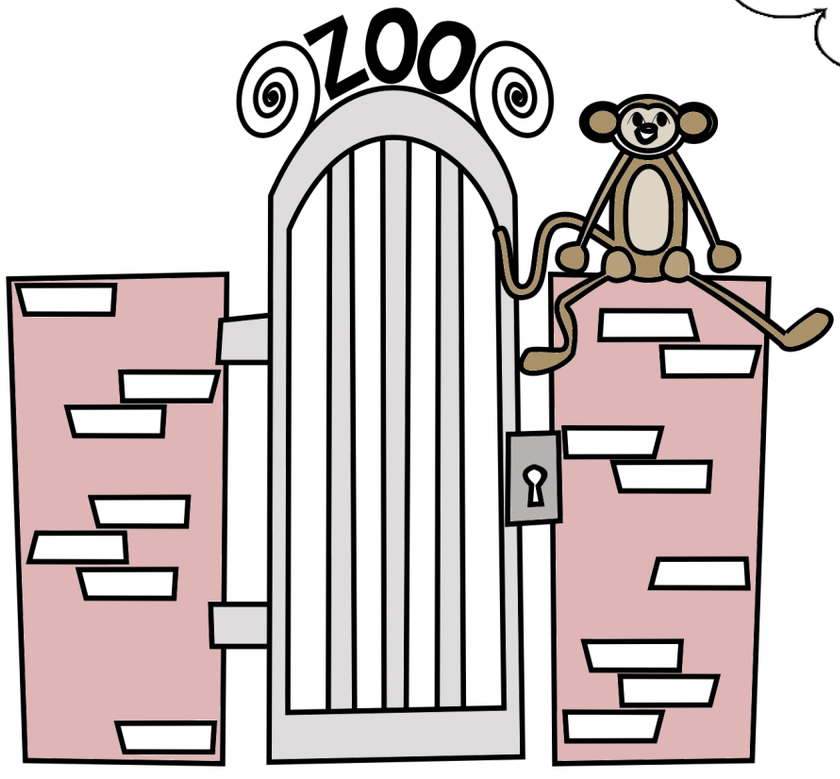
Thelma is coming to the park.

Sam must be at the park.

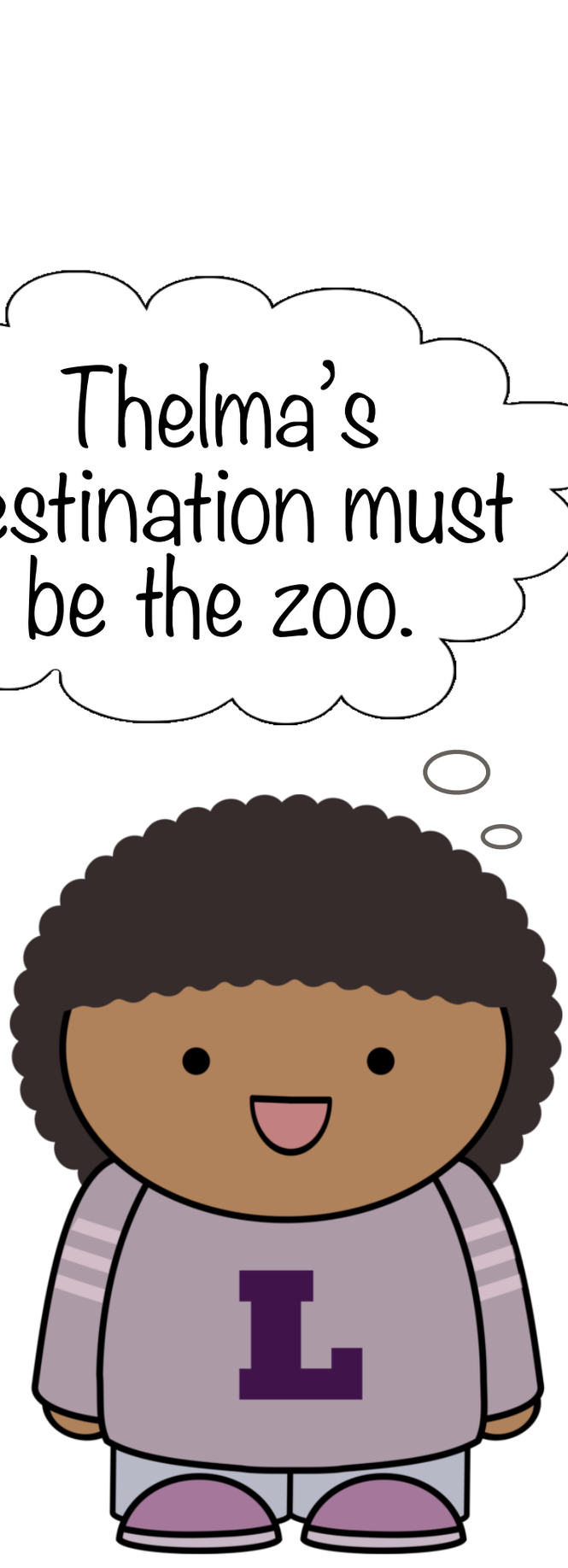


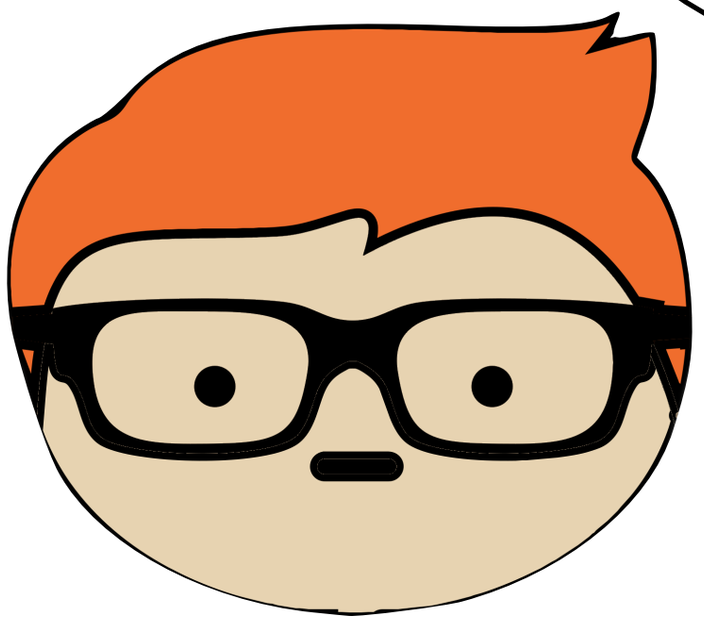
Thelma is coming.





Thelma's destination must be the zoo.





Thelma is coming.

What is Sam trying to communicate?

Whose perspective is Sam using?

$$p(\text{perspective}, \text{world} \mid \text{utterance})$$

Joint perspectival reasoning

Listener:

p(perspective, world | utterance)

Speaker:

p(perspective, utterance | world)

Perspectival RSA model

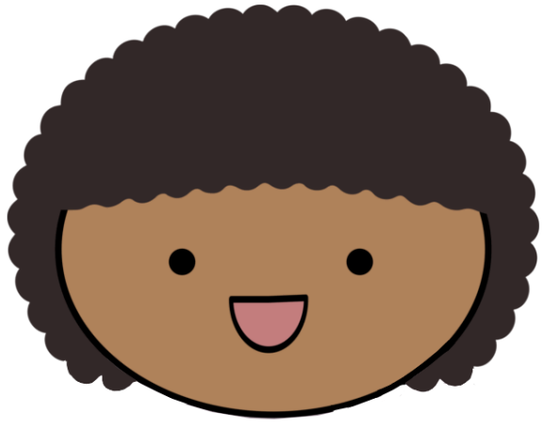
Pragmatic listener:

p(world, **perspective** | utterance) ∝

p(utterance, **perspective** | world) p(world)

$$L_1(w, \mathbf{a} \mid u) \propto S_1(u, \mathbf{a} \mid w)\, p(w)$$

where w = world, u = utterance, a = perspective



Perspectival RSA model

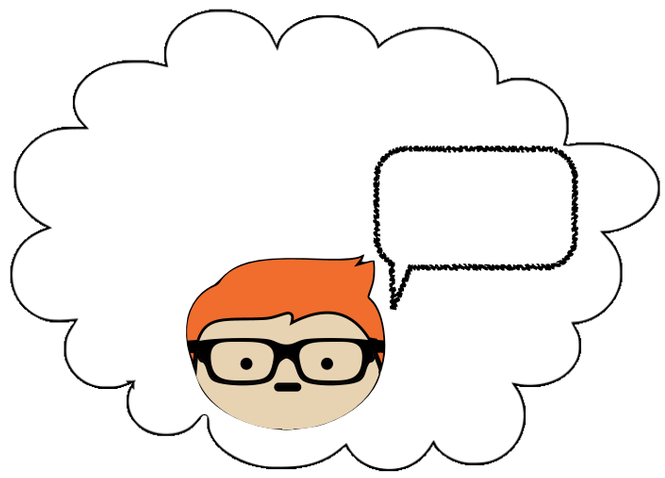
Pragmatic speaker:

p(utterance, **perspective** | world) ∝

p(world | utterance, **perspective**)p(utterance, **perspective**)

$$S_1(u, \mathbf{a} \mid w) \propto \text{Max}(L_0(w \mid u, \mathbf{a})\ p(u \mid \mathbf{a})\ p(\mathbf{a}))$$

where w = world, u = utterance, a = perspective



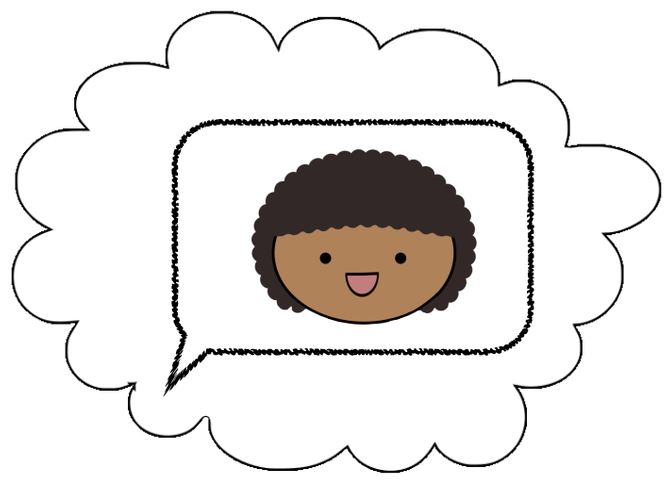
Perspectival RSA model

Literal listener:

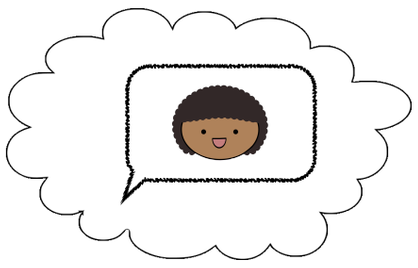
$$p(\text{world} \mid \textbf{perspective}, \text{utterance}) \propto p(\textbf{perspective}, \text{utterance} \mid \text{world})\, p(\text{world})$$

$$L_0(w \mid u, \textbf{a}) \propto [[u]]^{w,\textbf{a}}\, p(w)$$

where w = world, u = utterance, a = perspective

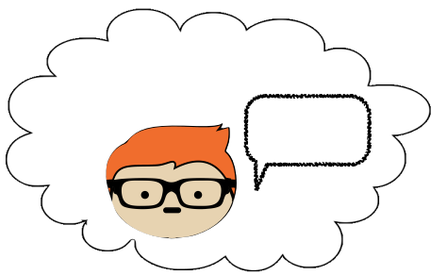


Perspectival RSA model



Literal listener:

$L_0(w \mid u, a) \propto [[u]]^{w,a}\, p(w)$



Pragmatic speaker:

$S_1(u, a \mid w) \propto \text{Max}(L_0(w \mid u, a)\, p(u \mid a)\, p(a))$

Pragmatic listener:

$L_1(w, a \mid u) \propto S_1(u, a \mid w)\, p(w)$

where w = world, u = utterance, a = perspective

Perspectival RSA model

Perspective Cost Function: Penalizes non-speaker perspectives. Motivated by egocentricity biases (Lin & Epley 2010, Köder et al. 2015) and Harris (2012)'s Speaker Default heuristic.

Literal Listener:

$$L_0(w \mid u, a) \propto [[u]]^{w,a}\, p(w)$$

Pragmatic Speaker:

$$S_1(u, a \mid w) \propto \text{Max}(L_0(w \mid u, a)\, p(u \mid a)\, p(a) - \textbf{Cost(a)})$$

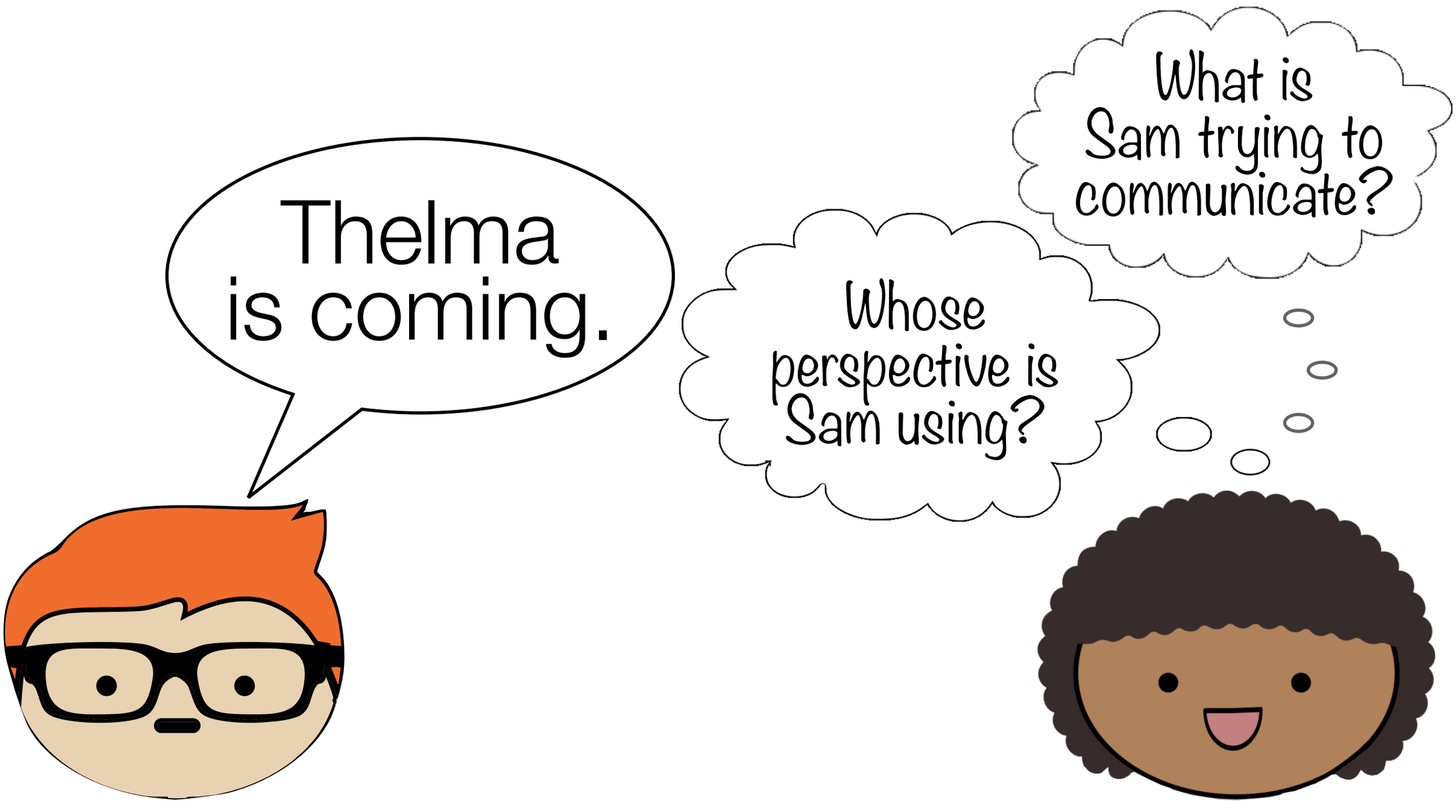
Pragmatic Listener:

$$L_1(w, a \mid u) \propto S_1(u, a \mid w)\, p(w)$$

Talk Outline

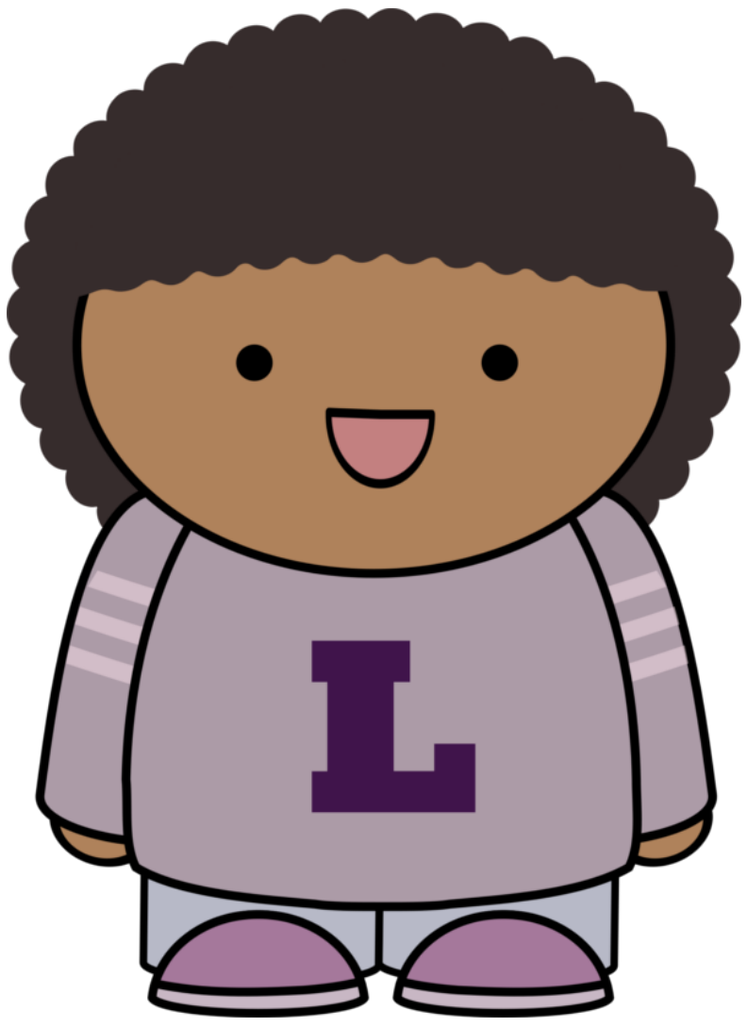
1. Develop a computational model of perspectival reasoning as a joint inference problem
2. Run simulations to predict speaker and listener behavior
3. Compare model predictions against evidence from crowdsourced behavioral experiments
   - How do listeners understand perspectival expressions?
   - How do speakers choose perspectival expressions?
4. Open questions

Perspective Identification

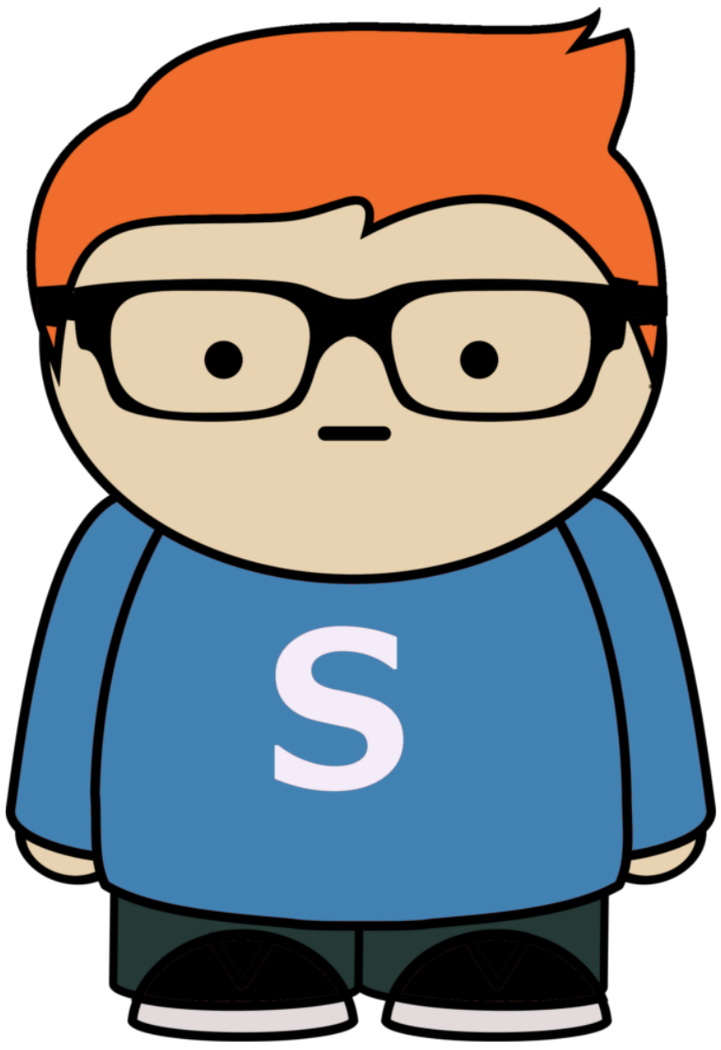


Perspective Set

Lucy Listener



Sam Speaker



Utterance Set

$$\left\{\begin{array}{c}\text{Thelma is}\\ \text{I am}\\ \text{You are}\end{array}\right\}\left\{\left\{\begin{array}{c}\text{coming}\\ \text{going}\\ \text{walking}\end{array}\right\}\right\}\text{to Northampton.}$$

Utterance semantics

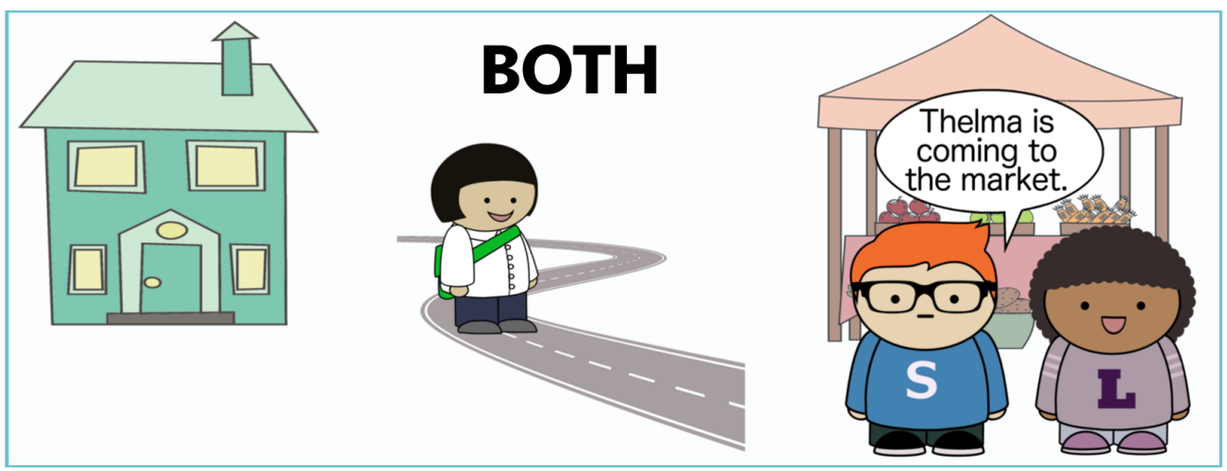
The literal listener looks up the literal meaning of the utterance according to the perspective and world.

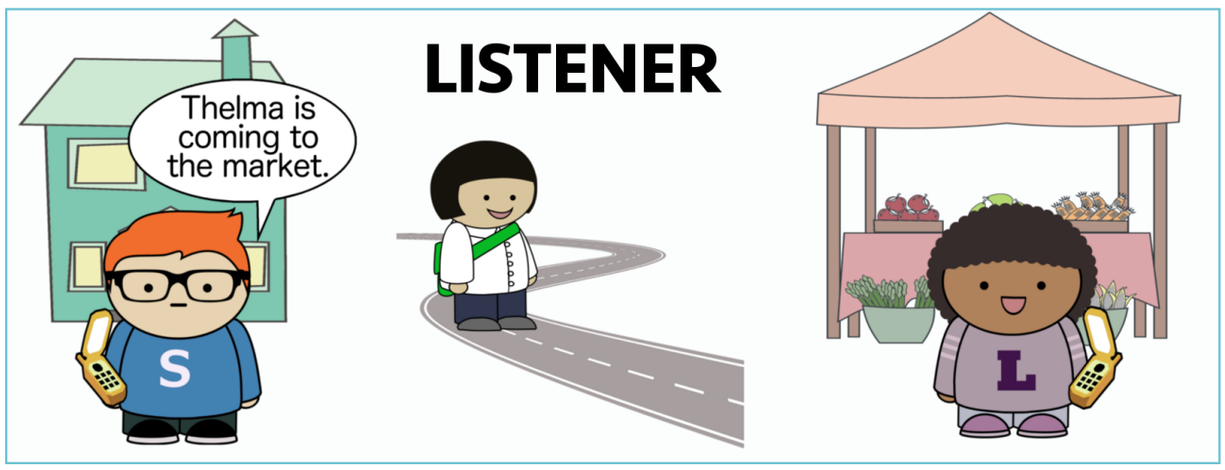
come: 1 if the perspective holder is at the destination of motion and the subject is in motion.

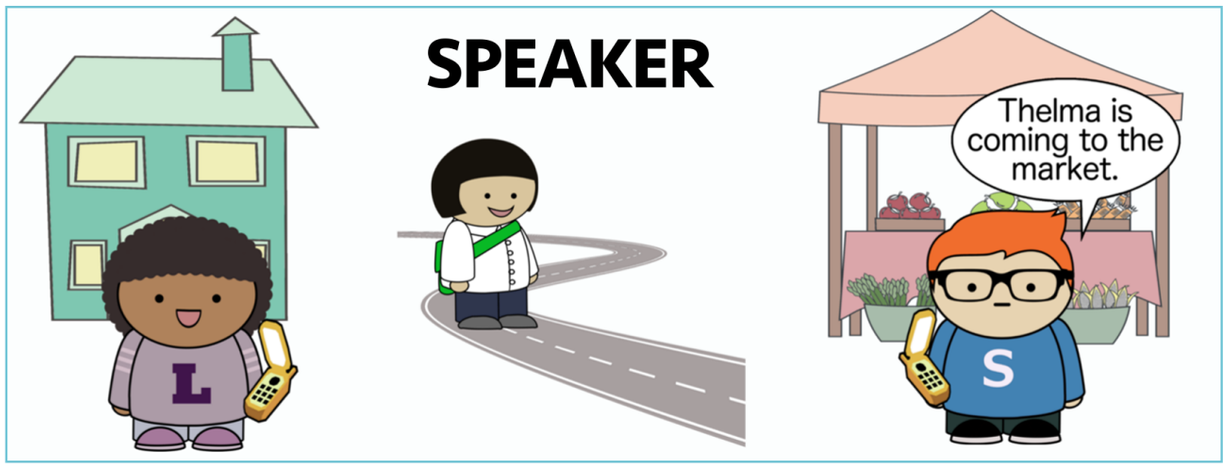
walk: 1 if the subject is in motion.

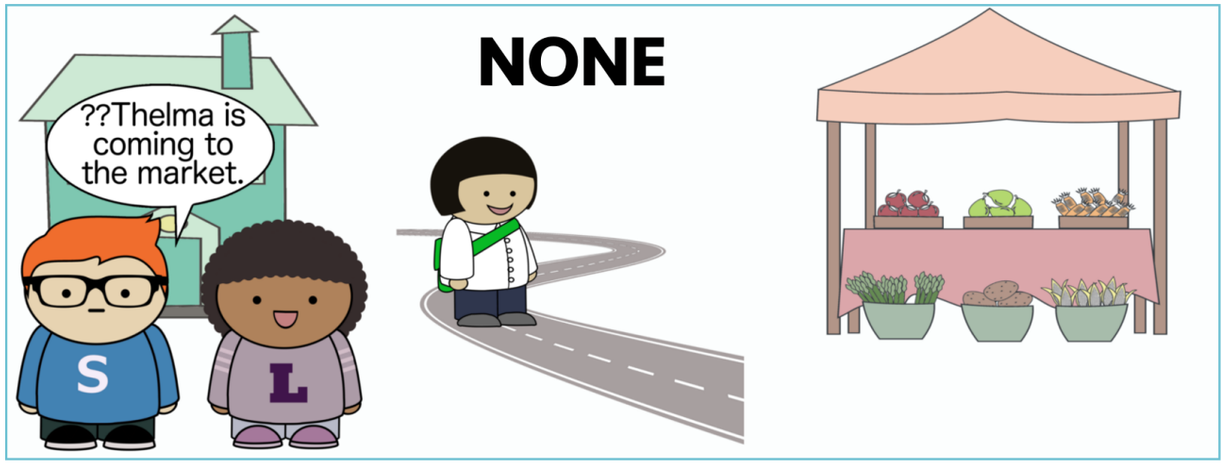
go: 1 if the subject is in motion and the perspective holder is not at the destination of motion.

World Set


BOTH
Thelma is coming to the market.
S
L


Thelma is coming to the market.
S
LISTENER
L


SPEAKER
L
Thelma is coming to the market.
S

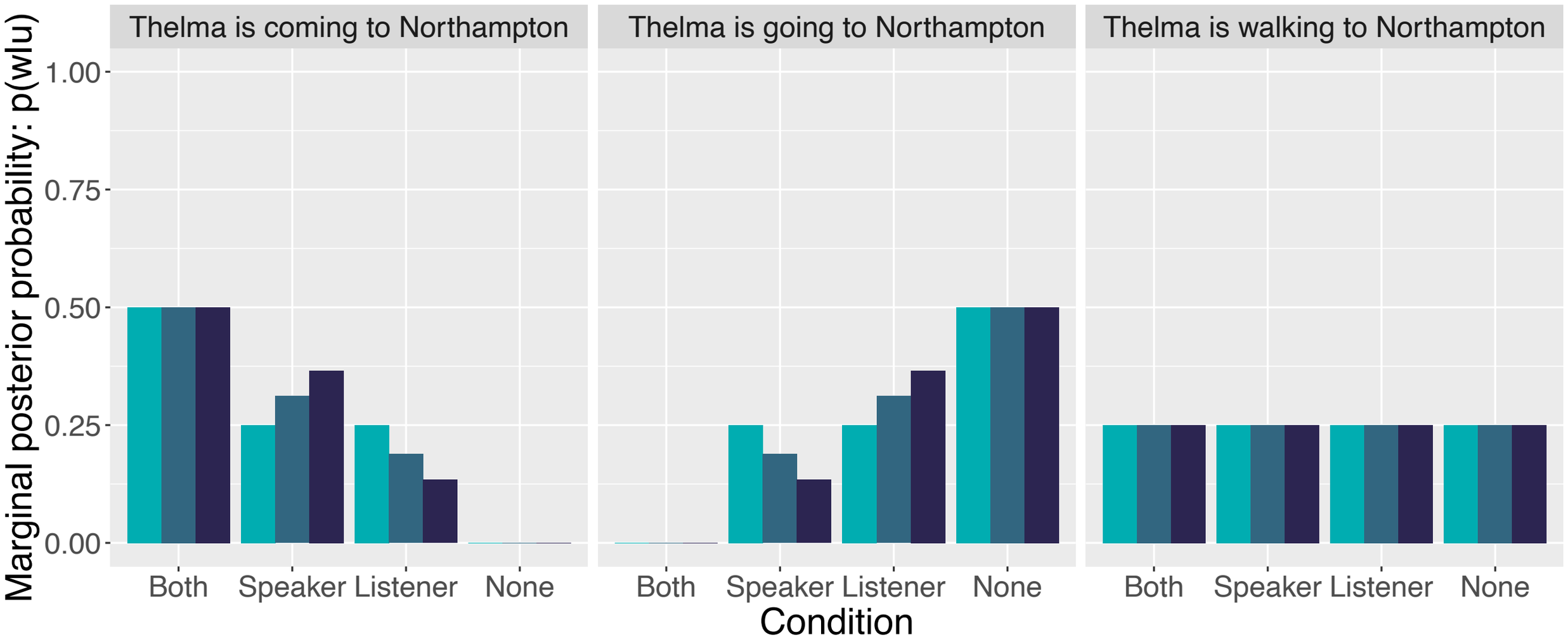

??Thelma is coming to the market.
S
L
NONE

Generating predictions

- Model is implemented in the **WebPPL** probabilistic programming language
- Simulations run using Markov Chain Monte Carlo sampling
- Tested perspective cost function settings from 0 - 1

RSA Listener Predictions

$$p(w \mid u) = \sum_a p(w,a \mid u)$$



Simulated with perspectival semantics for *go* and perspective cost settings of 0, 0.5, and 1

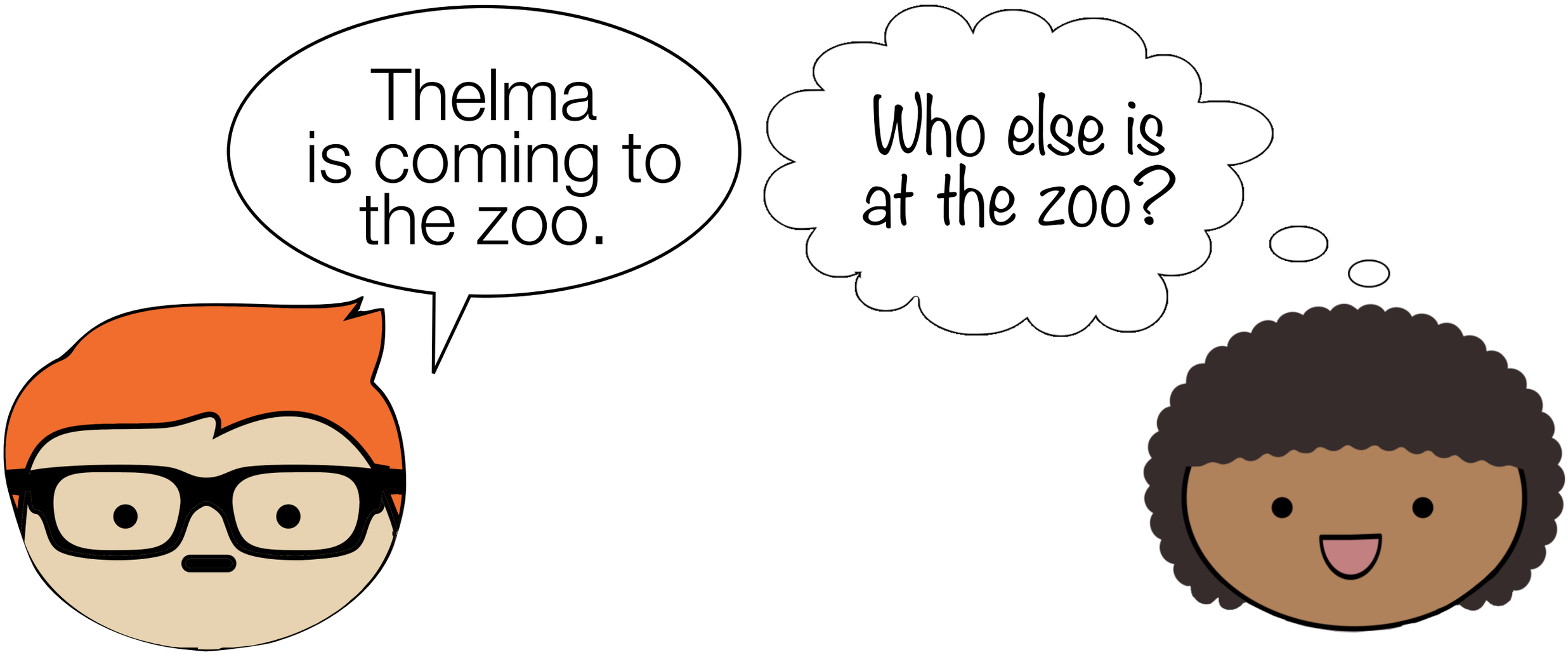
Convergent Perspective Boost

Given the sentence *Thelma is coming to Northampton,* the model predicts highest marginal probability for the world where **both speaker and listener** are located at the destination.

Talk Outline

- Part I: Encoding Perspectives
  - The semantic landscape of perspectival expressions
- Part II: Taking Perspectives
  - Perspective identification
    - A rational approach to perspectival reasoning
    - Model predictions
    - **Comprehension studies**
  - Perspective selection
    - Production study
- Concluding thoughts

Comprehension experiment: p(w | u)



Models of Grammatical Perspective Processing

- Heuristic approach: **listeners use simple rules**
  - **Speaker Default**: assume speakers use their own perspectives whenever possible
  - **Maintain Perspective**: assume speakers avoid perspective shift
- Perspectival reasoning approach: **listeners reason over multiple perspectives**
  - Listeners use a range of contextual evidence to infer the speaker's adopted perspective
- Two-stage system (Harris 2012): listeners rely on **heuristics when possible**, but **switch** to a **more costly reasoning system** when necessary

Thelma is coming hypotheses

Perspectival Rational Speech Acts Model: If listeners consider multiple perspectives simultaneously, they should assign highest marginal posterior probability to the **world where both listener and speaker are at the destination.**

Predicts a Convergent Perspective Boost.

Heuristic approach

Speaker default heuristic (Harris 2012): If listeners assume that the speaker is the perspective holder, they should assign **equal** marginal posterior probability to **all worlds where the speaker is at the destination**.

Predicts a Simple Speaker Advantage.

Thelma is coming hypotheses

PERSPECTIVAL RATIONAL SPEECH ACTS MODEL: If listeners consider multiple perspectives simultaneously, they should assign highest marginal posterior probability to the **world where both listener and speaker are at the destination.**

Predicts a Convergent Perspective Boost.

SPEAKER DEFAULT HEURISTIC (Harris 2012): If listeners assume that the speaker is the perspective holder, they should assign **equal** marginal posterior probability to **all worlds where the speaker is at the destination**.

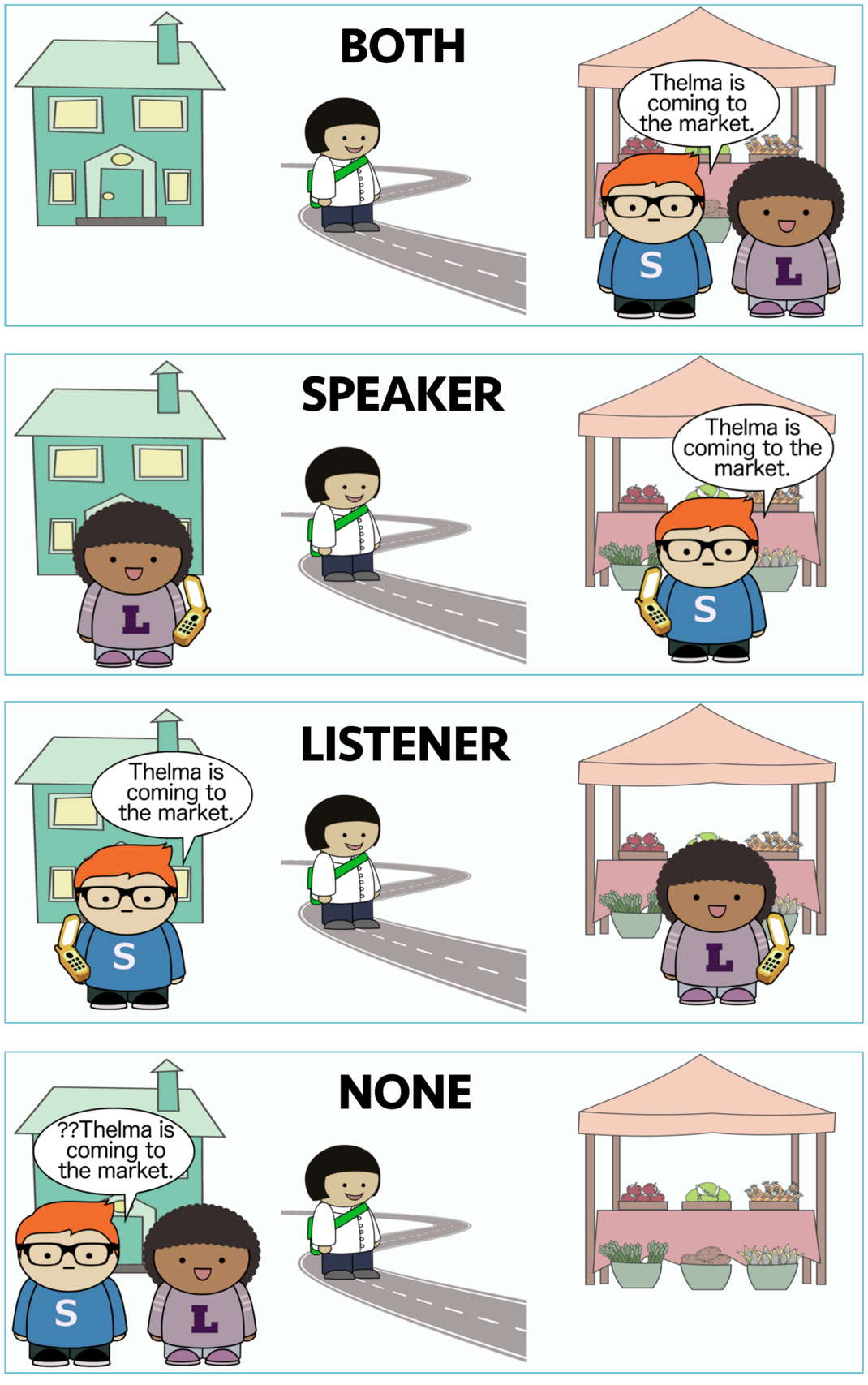
Predicts a Simple Speaker Advantage.

Speaker Default

Simple Speaker Advantage

Perspective Reselection

No Compatible Perspective



PRSA

Convergent Perspective Boost

Perspective Cost Determines Difference

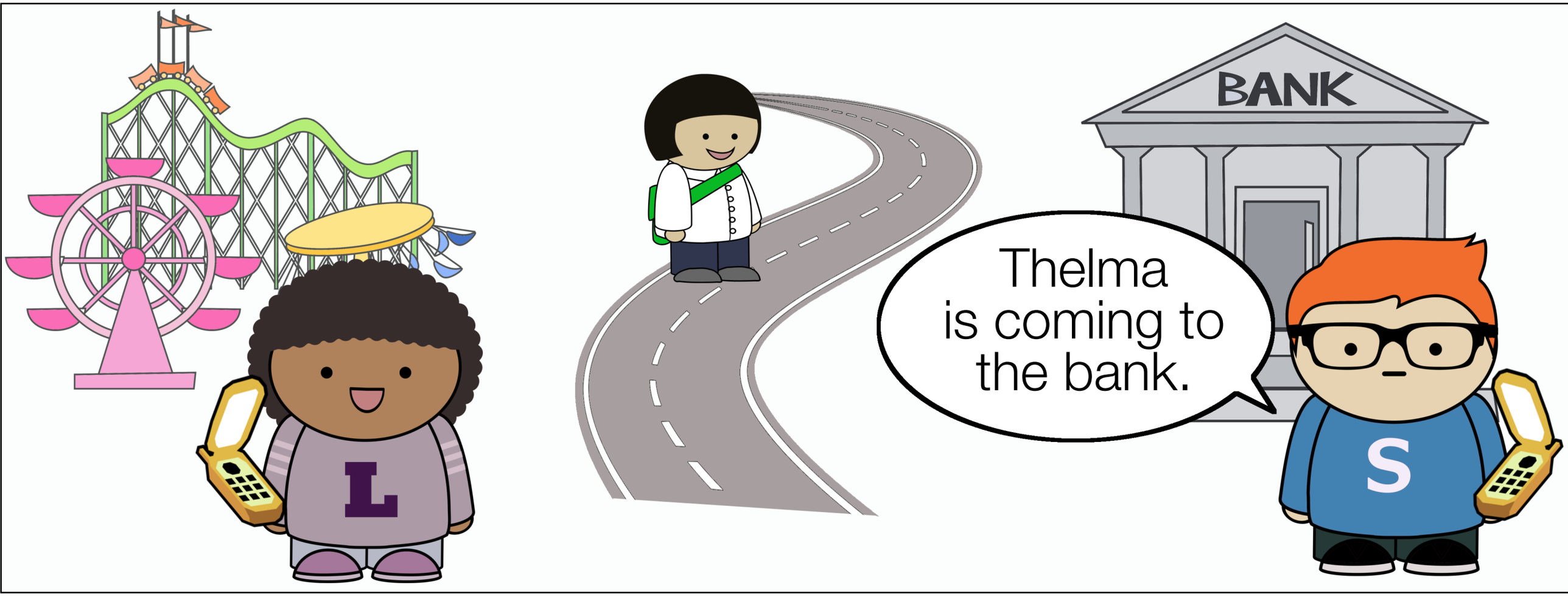
No Compatible Perspective

Experimental design

- Eight conditions: 4 worlds x 2 sentences
- **Perspectival condition**: *Thelma is coming to the zoo.*
- **Plain condition**: *Thelma is driving to the zoo.*

Prompt

Perspectival, only speaker at destination

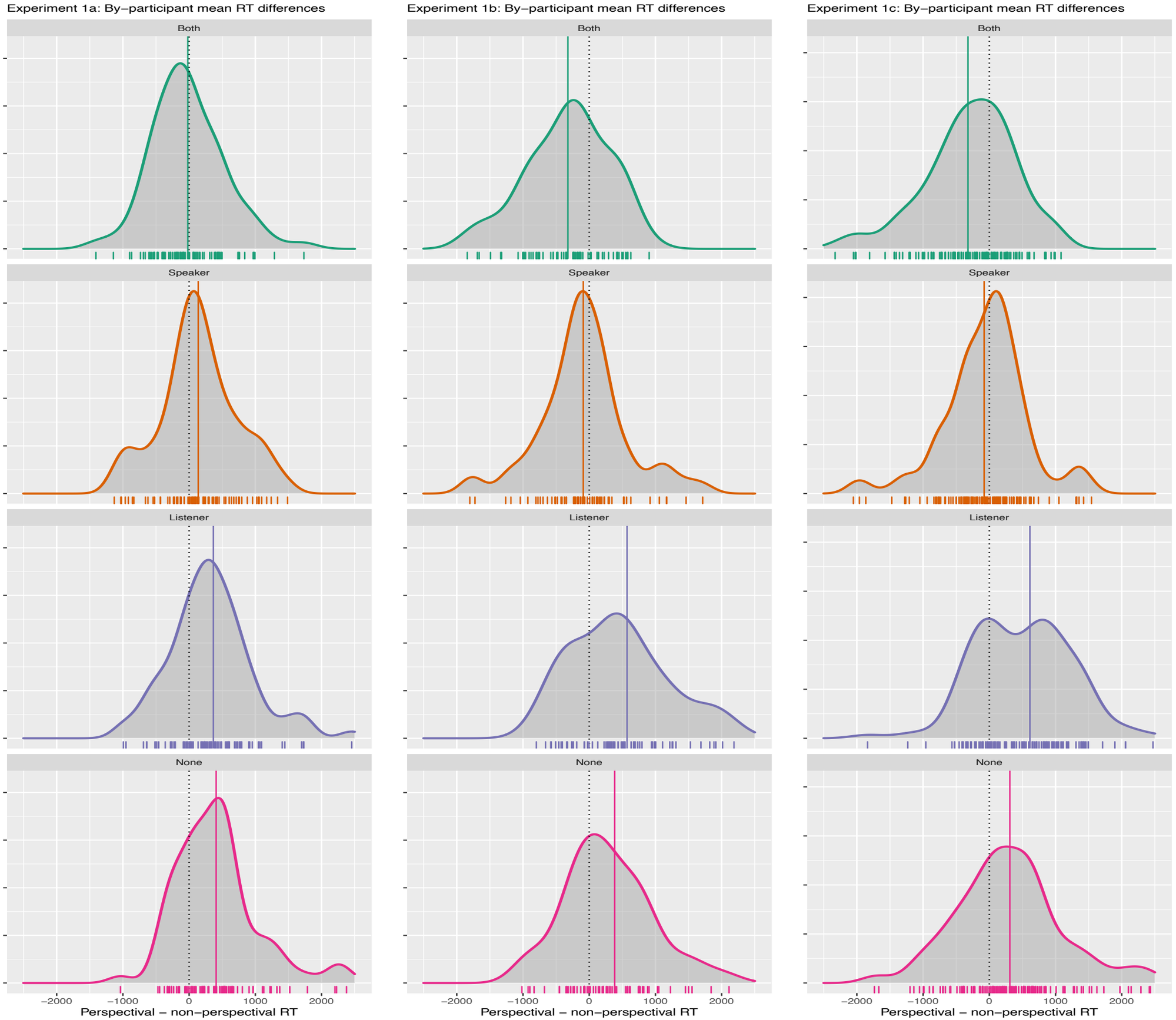


Experimental design

- Eight conditions: 4 worlds x 2 sentences
- 6 items in each condition
- **Perspectival condition**: *Thelma is coming to the zoo.*
- **Plain condition**: *Thelma is driving to the zoo.*
- Dependent variables: reaction times and acceptability.
- Monolingual American English-speaking participants recruited on Prolific
- 3 experiments:
  - Expt. 1A: n= 80, Expt. 1B: n=64, Expt. 1C: n=120

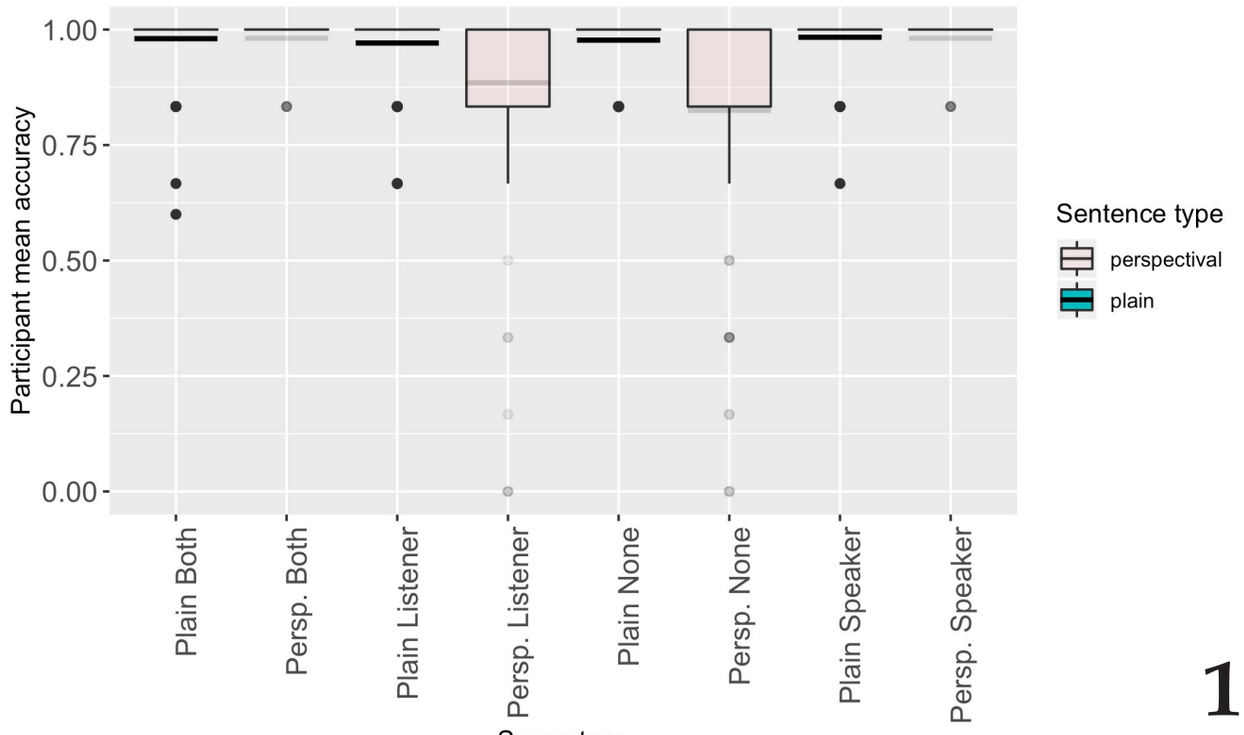
Spatial items

Reaction Time Results

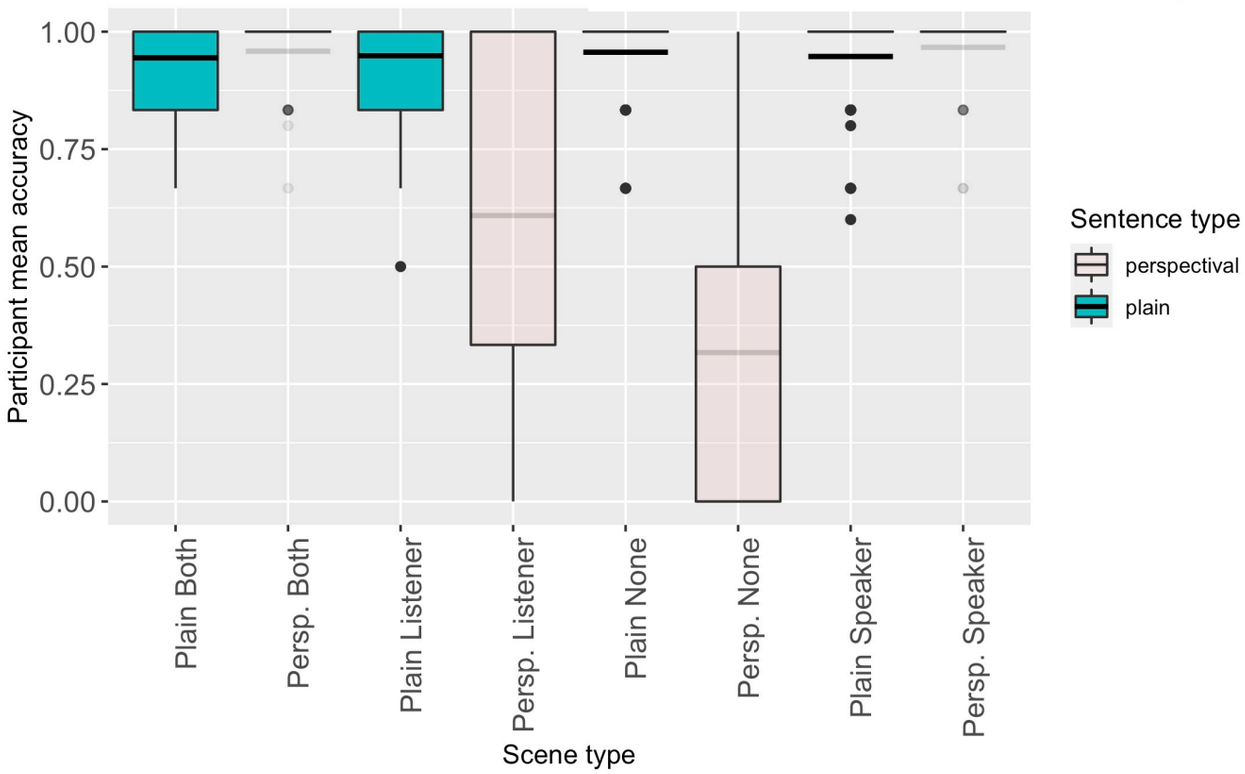


Acceptability Results

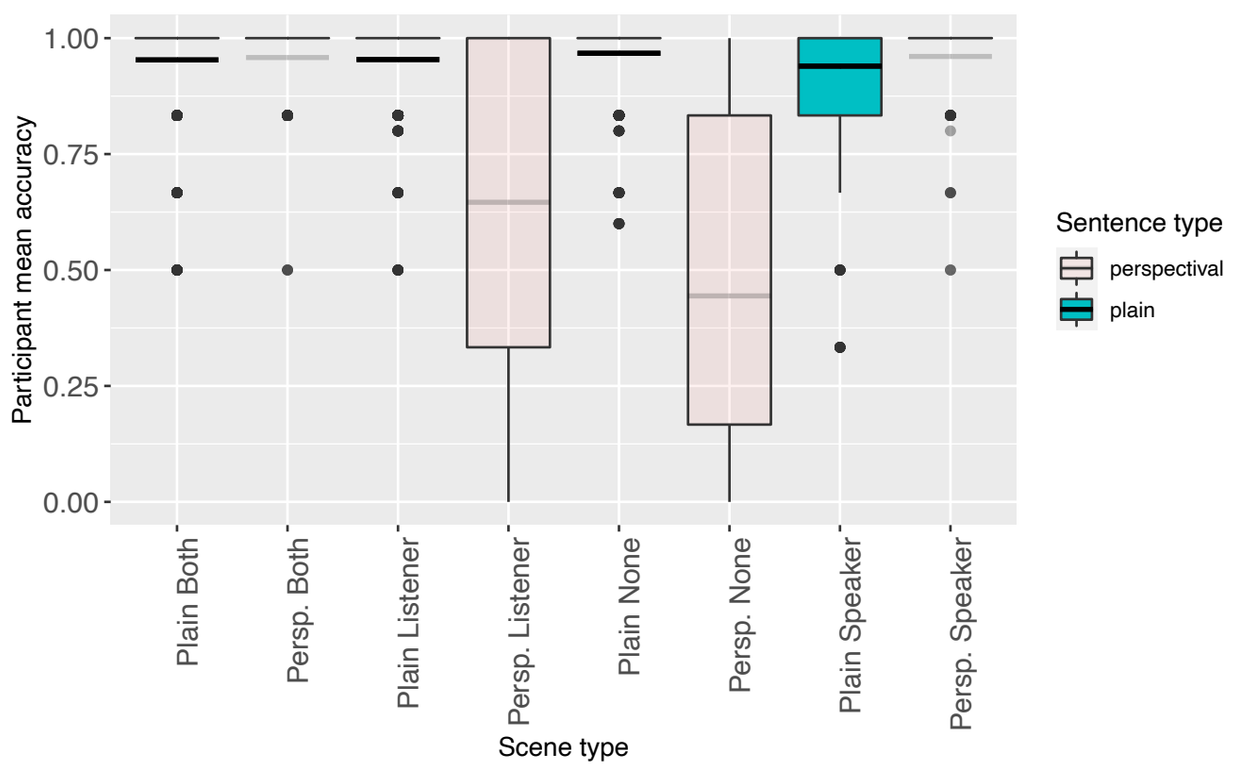
1A



1B



1C



Comprehension Results: p(w | u)

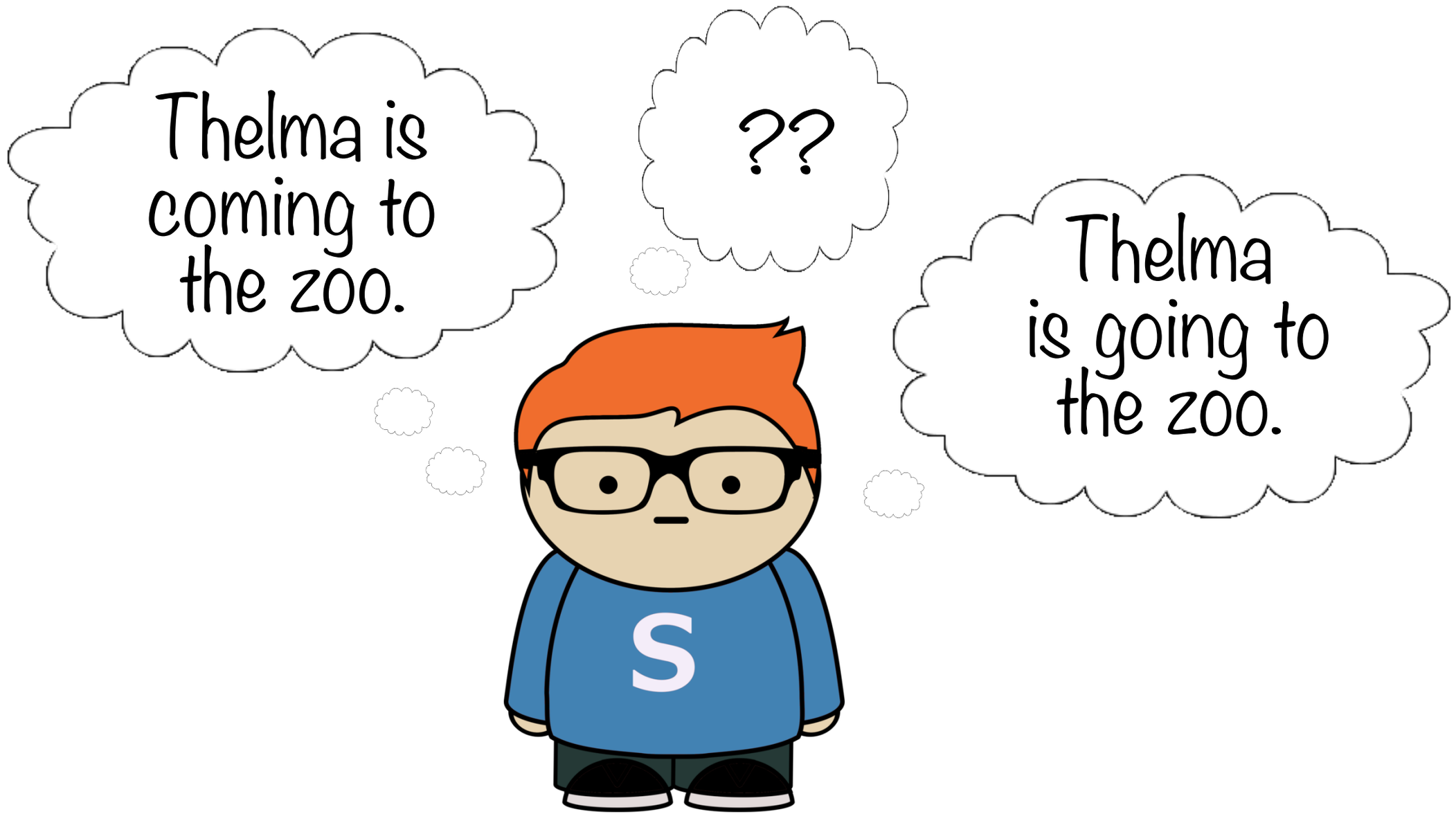
Summary:

- Participants were fastest to accept scenes following *come* in the both condition
- The acceptance rate in the listener *come* condition was low ⟶ high speaker bias

Talk Outline

1. Develop a computational model of perspectival reasoning as a joint inference problem
2. Run simulations to predict speaker and listener behavior
3. Compare model predictions against evidence from crowdsourced behavioral experiments
   - How do listeners understand perspectival expressions?
   - How do speakers choose perspectival expressions?
4. Open questions

Perspective Selection: p(a | w)



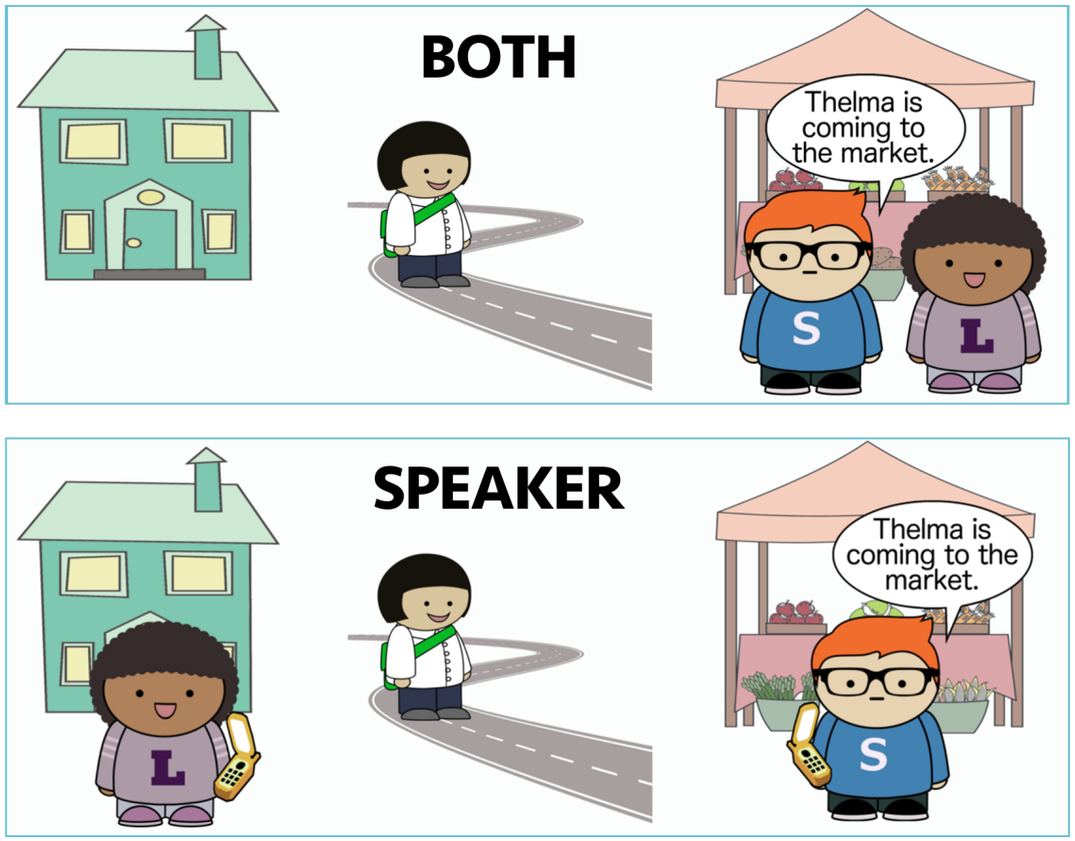
Experimental design

- Experiment 2A: 4 conditions (Both, Speaker, Listener, None)
- Experiment 2B: 6 conditions (same 4 + Speaker Moving Listener, Speaker Moving None)
- 12 items in each condition
- Participants completed the speaker's speech bubble in a text box.
- Prompt: *Thelma is...*
- Monolingual American English-speaking participants recruited on Prolific
- 2 experiments:
  - Expt. 2A: n= 40, Expt. 2B: n=56

Speaker Default

PRSA

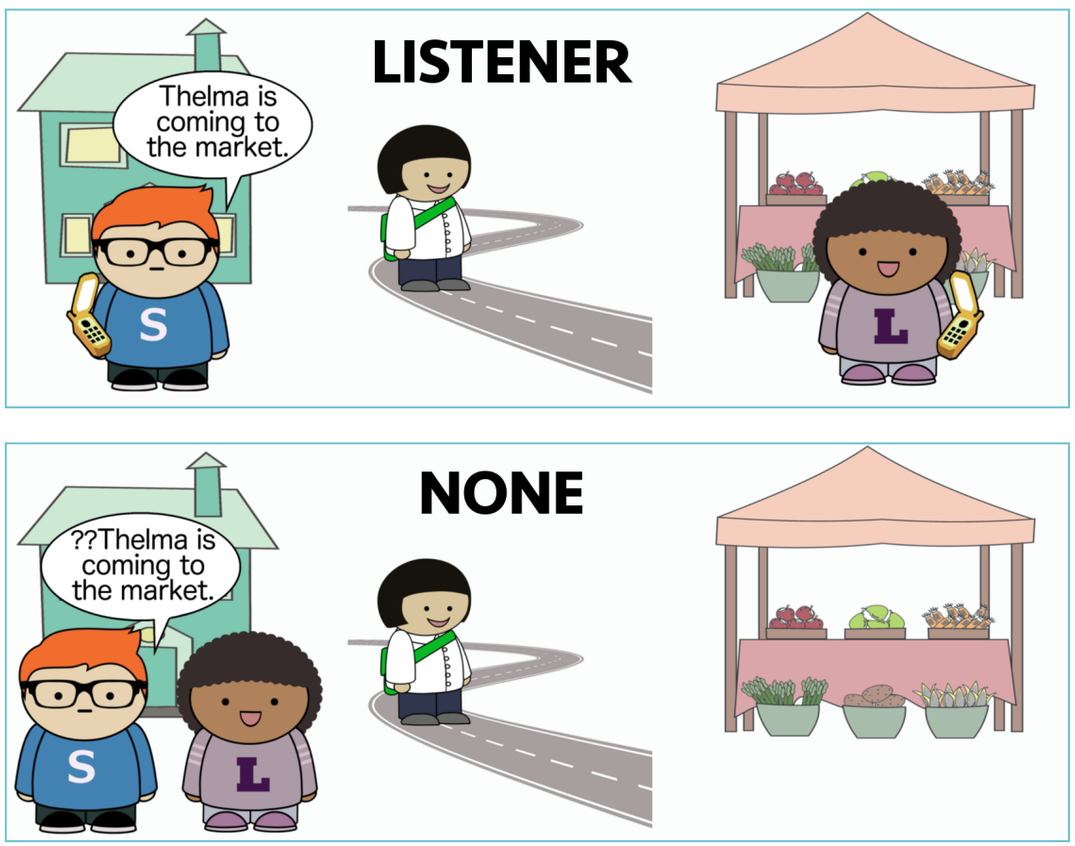
Speaker's Perspective Compatible



Convergent Perspective Boost

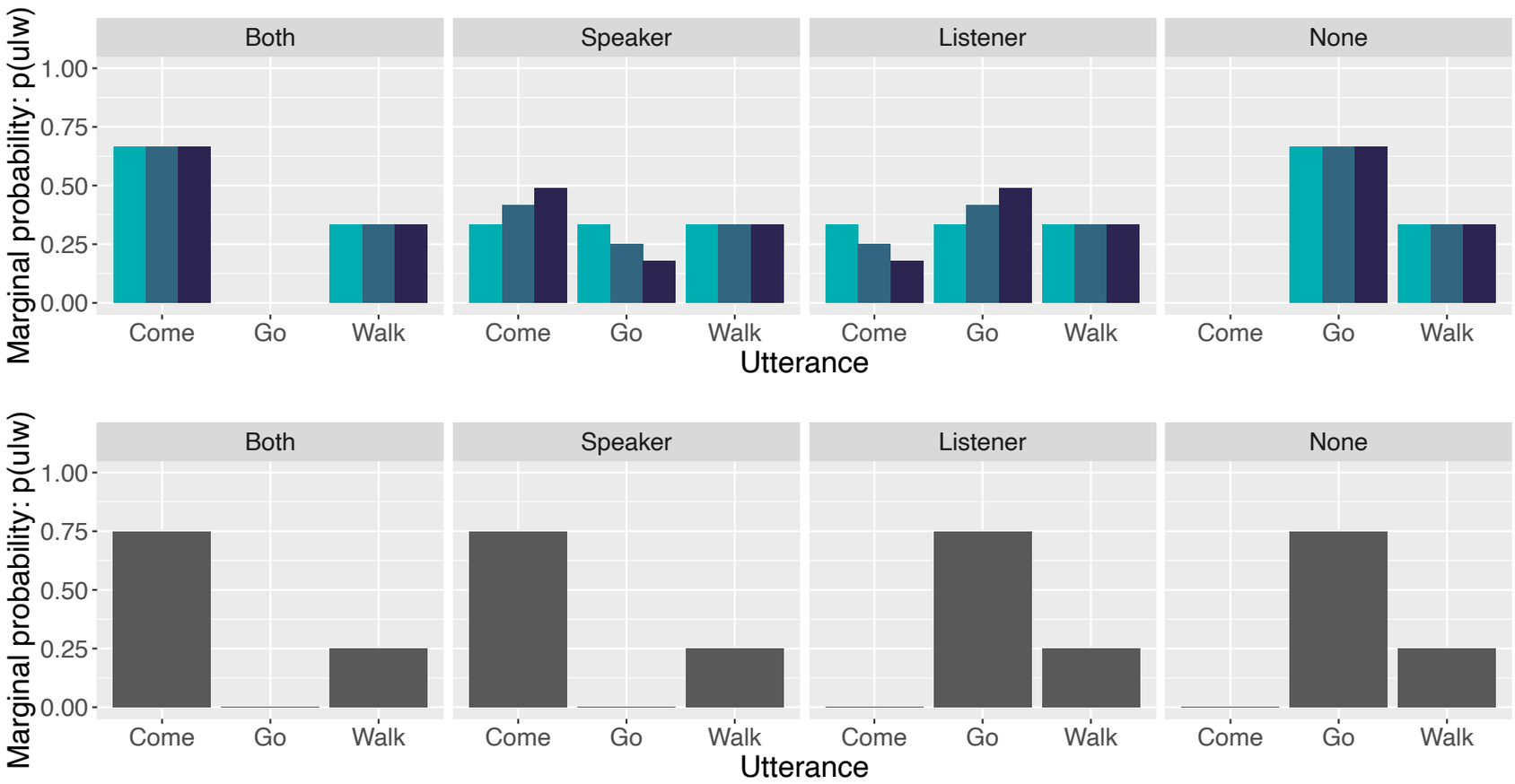
Perspective Cost Determines Difference

Speaker's Perspective Incompatible



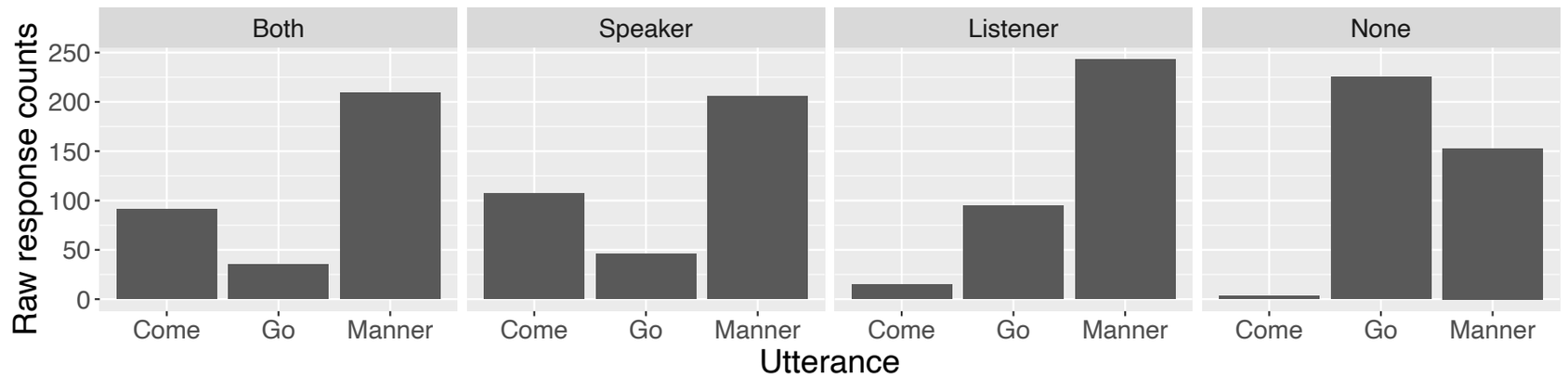
No Compatible Perspective

Production Results

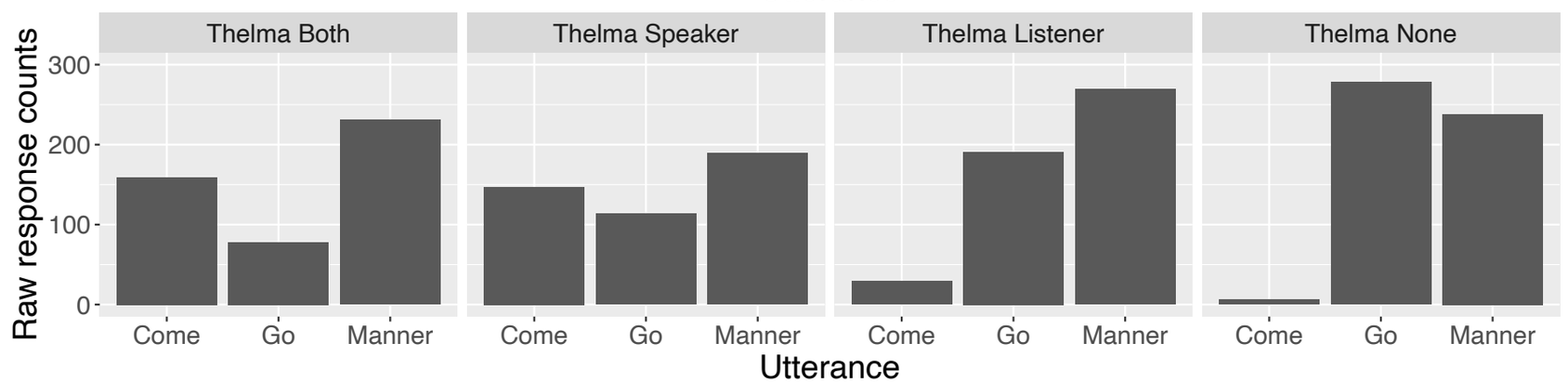


PRSA

Speaker Default



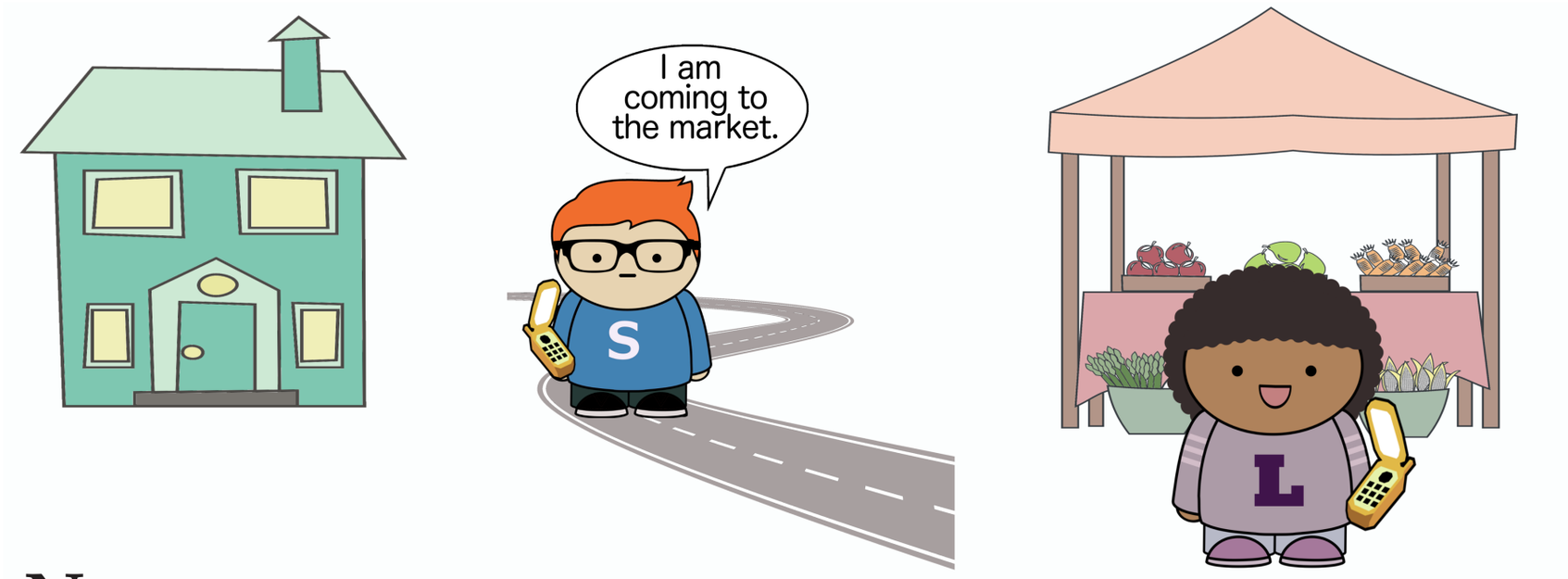
Experiment 2a Results



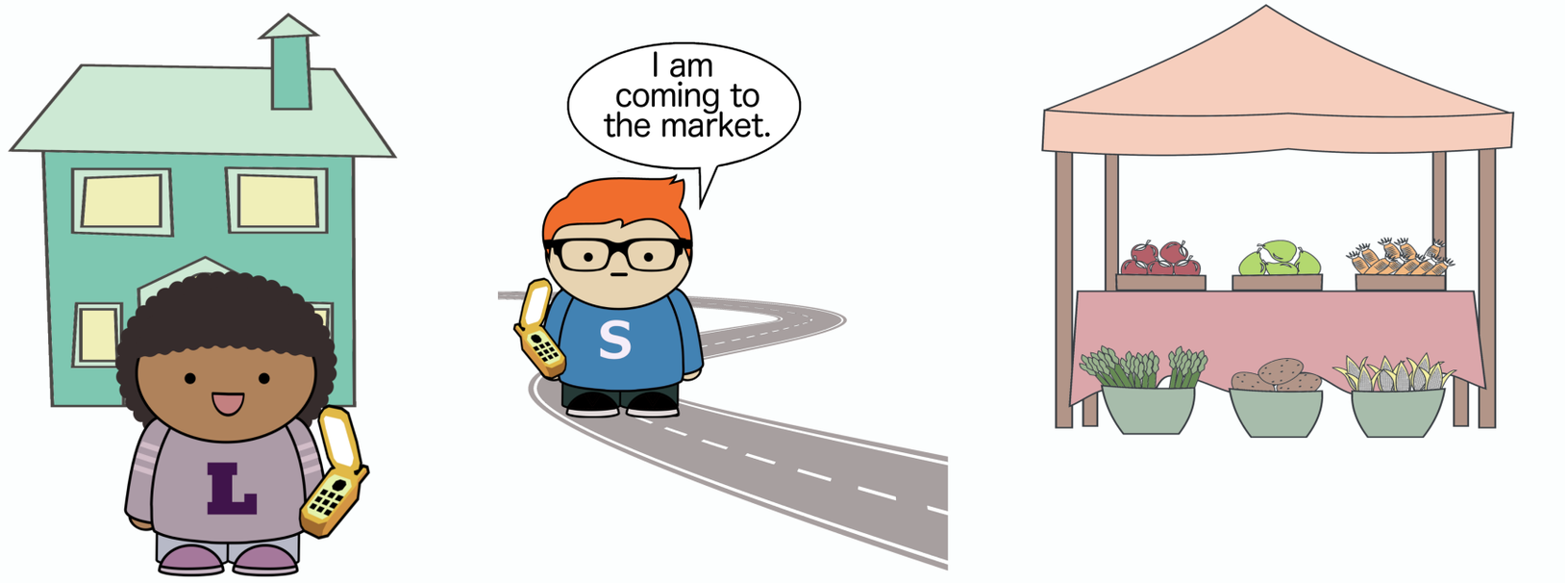
Experiment 2b Results

Speaker Moving Conditions

Speaker Moving Listener



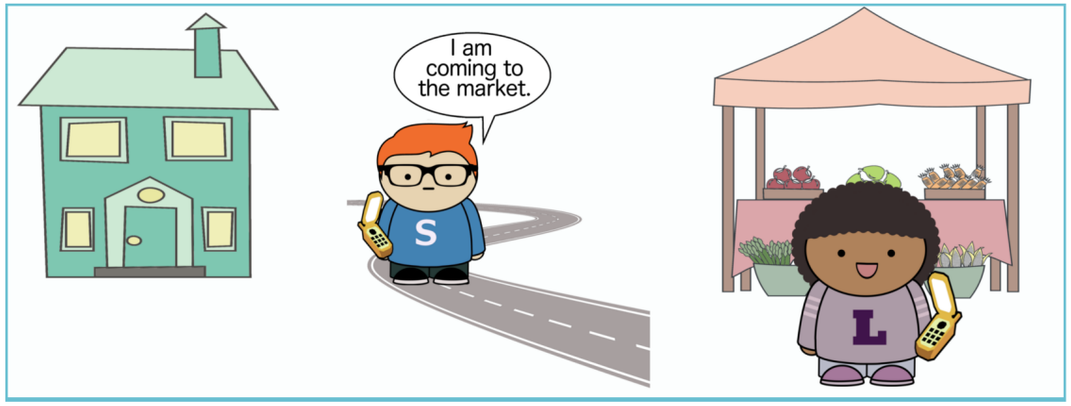
Speaker Moving None



Speaker Default

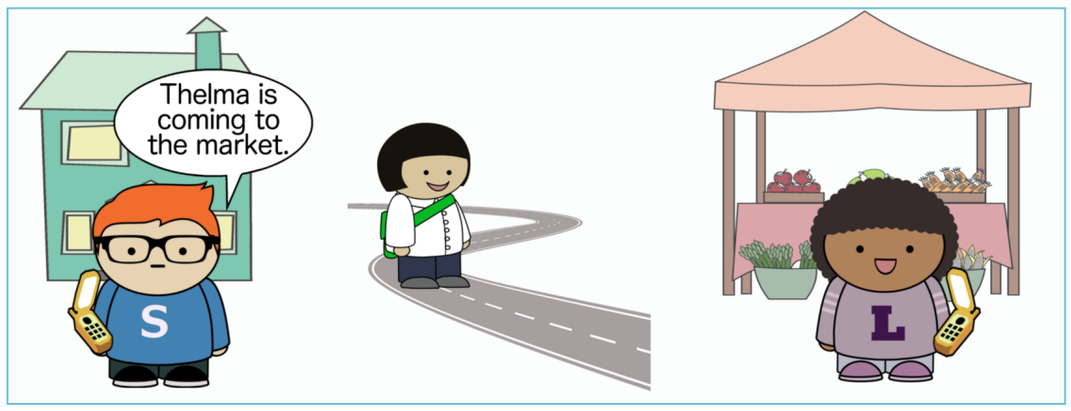
PRSA

SPEAKER MOVING LISTENER



Ambiguity Elimination Advantage

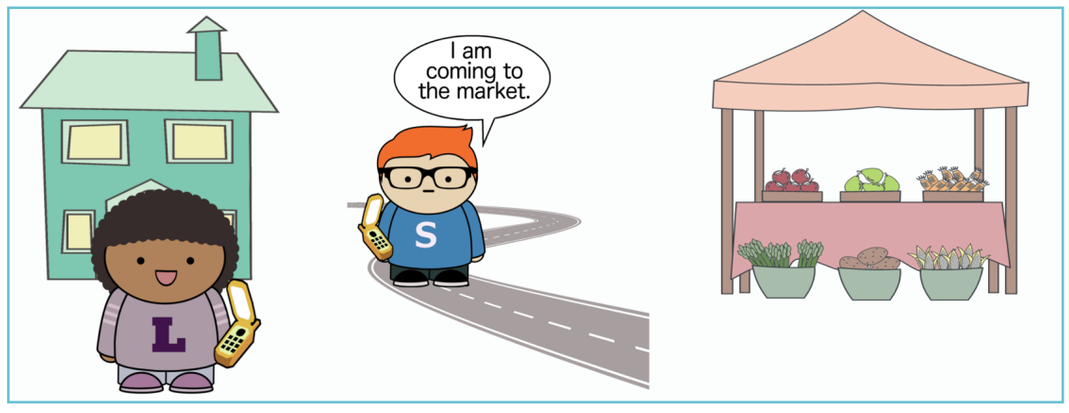
LISTENER



Compatible But Ambiguous

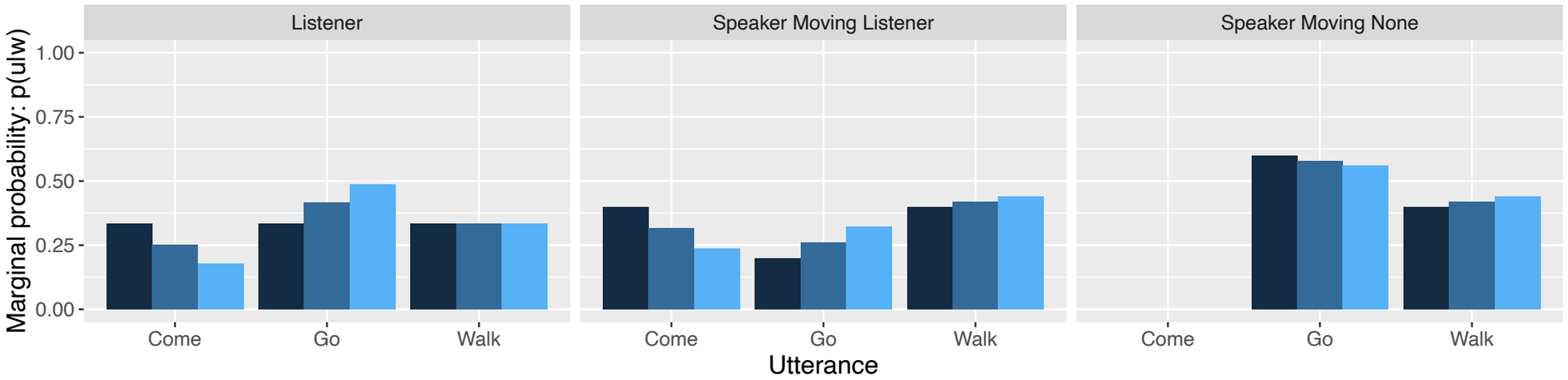
Speaker's Perspective Incompatible

SPEAKER MOVING NONE

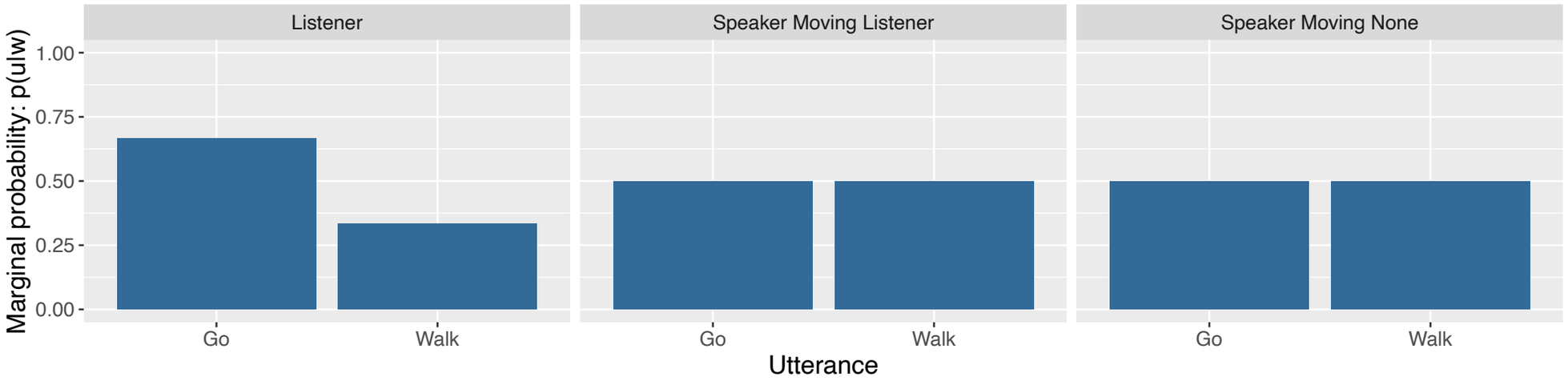


No Compatible Perspective

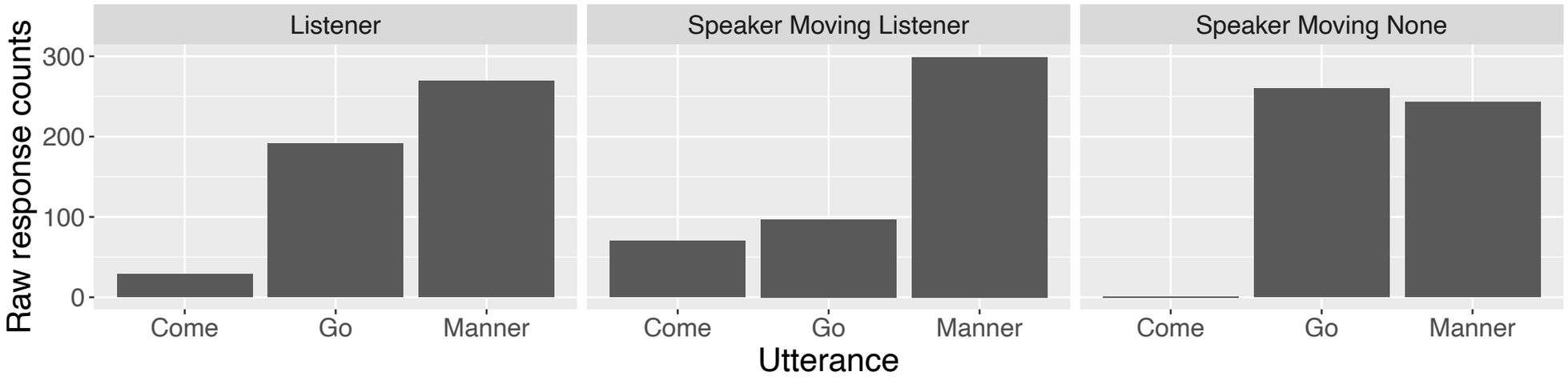
Production Results



PRSA



Speaker Default



Expt. 2b Results

Converging Evidence: Watson et al. (2021)

- Adapts a probabilistic model of visual perspective-taking from Heller et al. 2016 and Mozuraitis et al. 2018 to grammatical perspective-taking
- 4x2 online production experiment: same world conditions (Both, Speaker, Listener, None) x speaker certainty about the listener's location
- **Results support a Convergent Perspective Boost**

Watson, J., Kapron-King, A., Aggarwal, J., Beekhuizen, B., Heller, D., & Stevenson, S. (2021). Come Together: Integrating Perspective Taking and Perspectival Expressions. *CogSci* 43.

Summary

Since speakers do shift perspective, listeners must reason about the perspective they've adopted.

- The comprehension data supports the predicted reasoning behavior of rational listeners

Yet speakers seem to be guided by heuristics like Speaker Default more strongly than listeners expect.

- The production data suggests that speakers don't shift perspective at the rate predicted by listeners.

This goes against a fundamental assumption of the RSA framework: that speakers and listeners learn from and mirror each other's behavior.

But it mirrors Kehler & Rohde (2019)'s proposal for pronoun resolution.

It is also compatible with Harris (2012)'s proposed two-stage system: listeners may use the costly reasoning system more than speakers.

Acknowledgements

Brian Dillon
UMass Amherst
Co-author

Data Coders:

Tessa Masis
UMass Amherst

Carina Bolaños Lewen
Wesleyan College