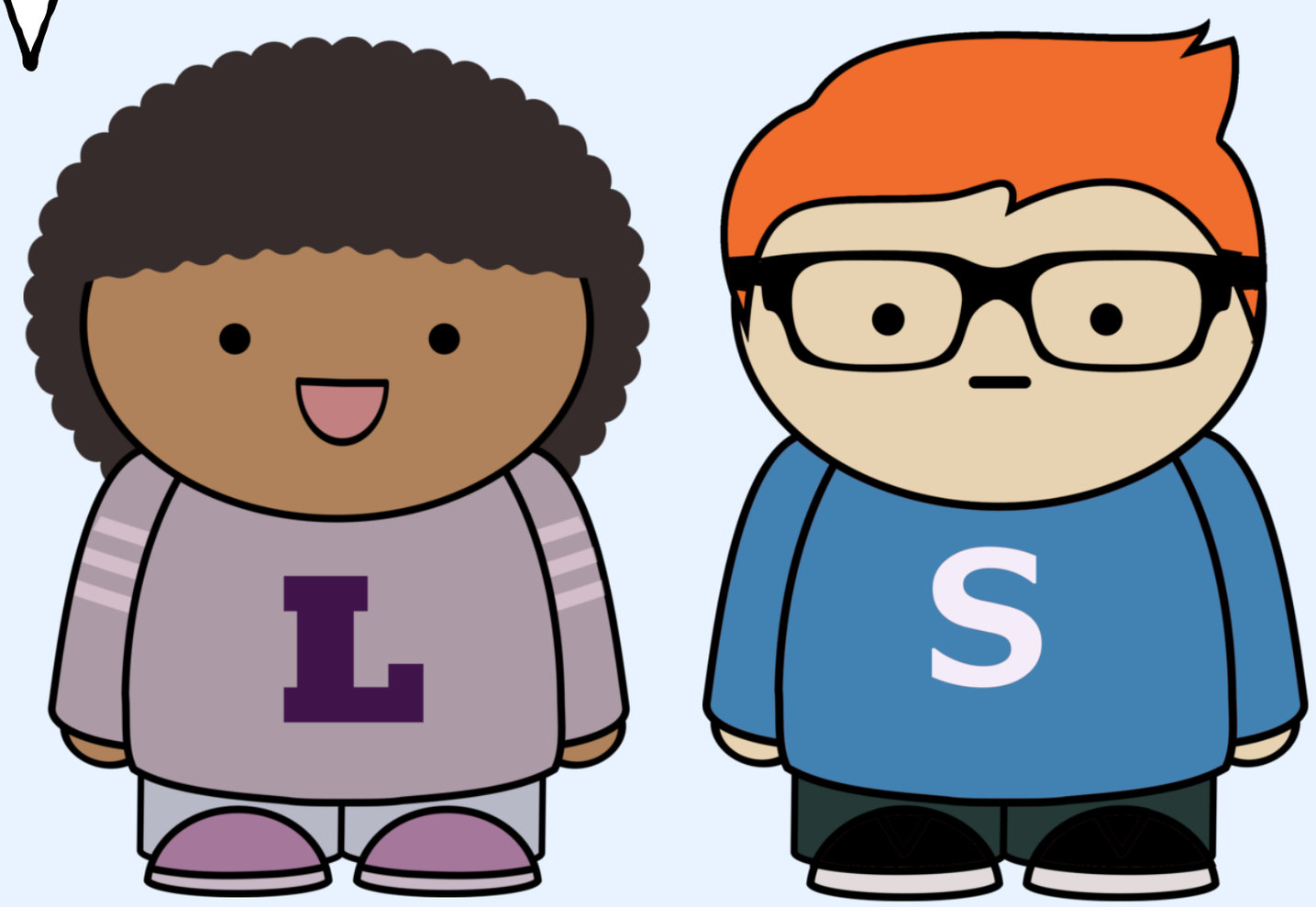
Guess Who's Coming to Dinner:
a Bayesian account of perspectival reasoning

Carolyn Jane Anderson
University of Massachusetts, Amherst

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!

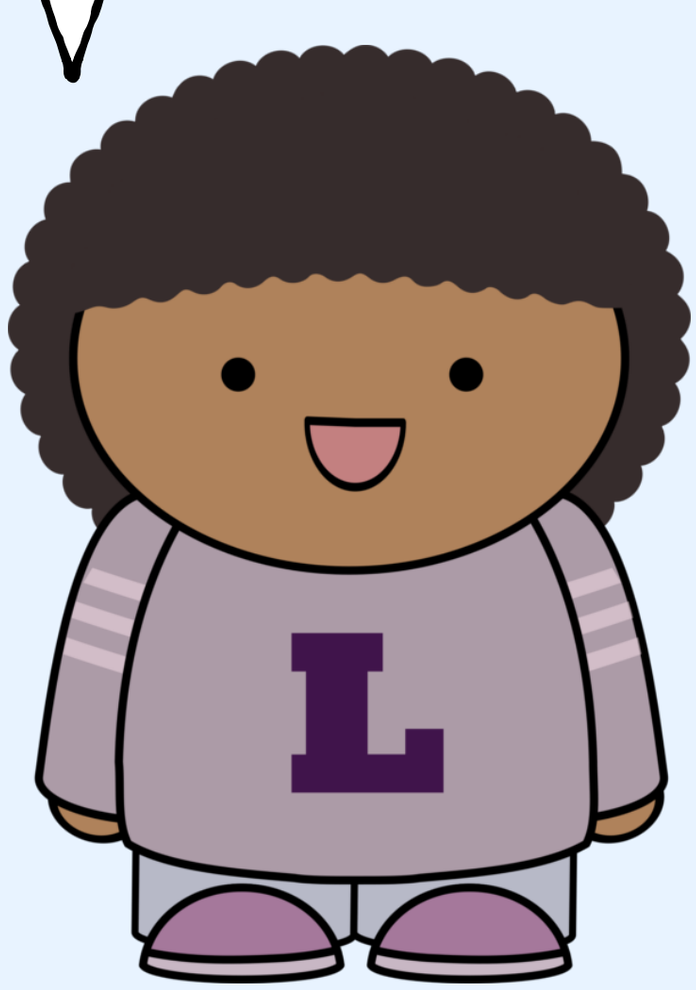


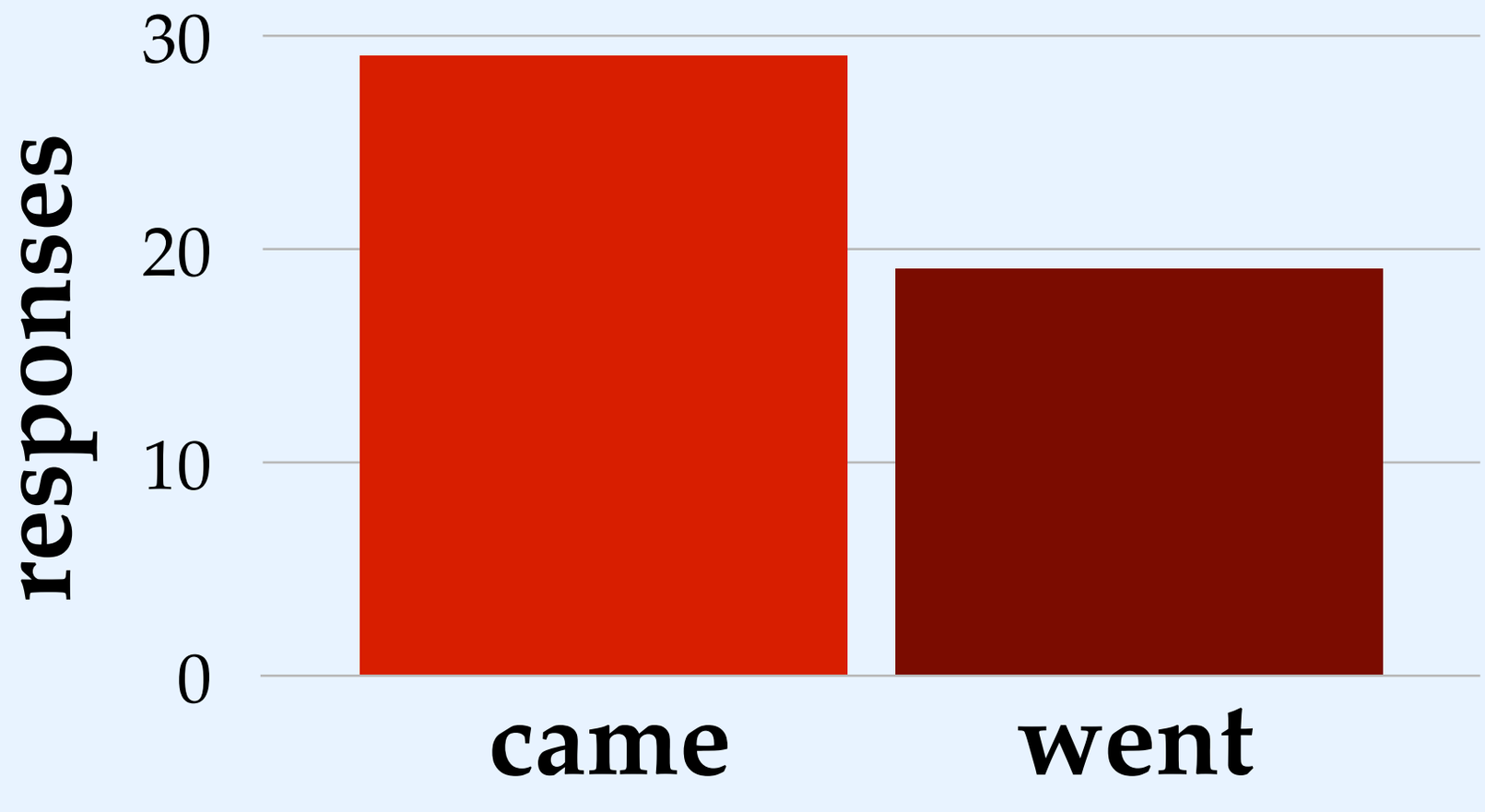
Fill in the blank:

(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!

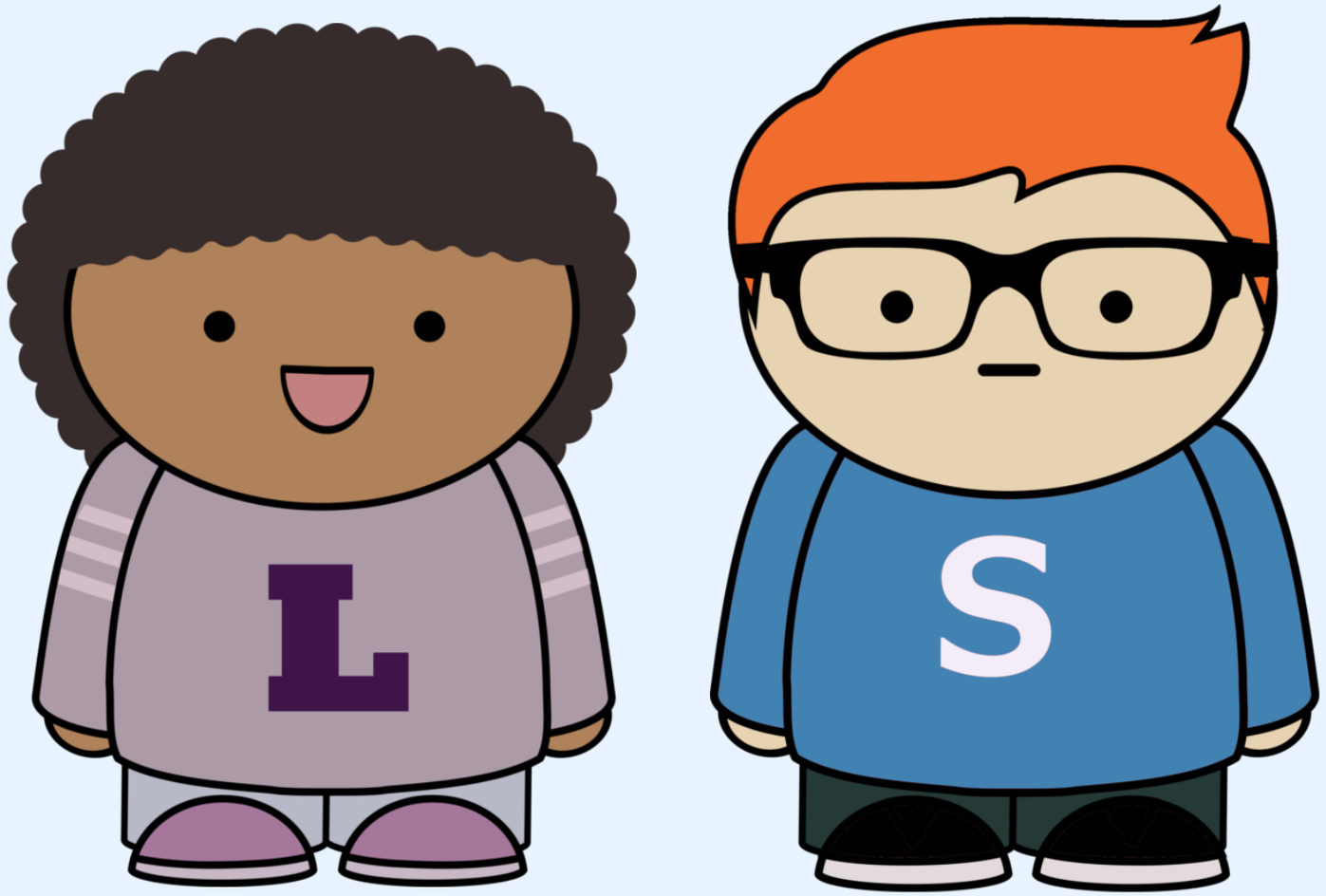




Showing my exchange student Thelma the US has been fun!

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.



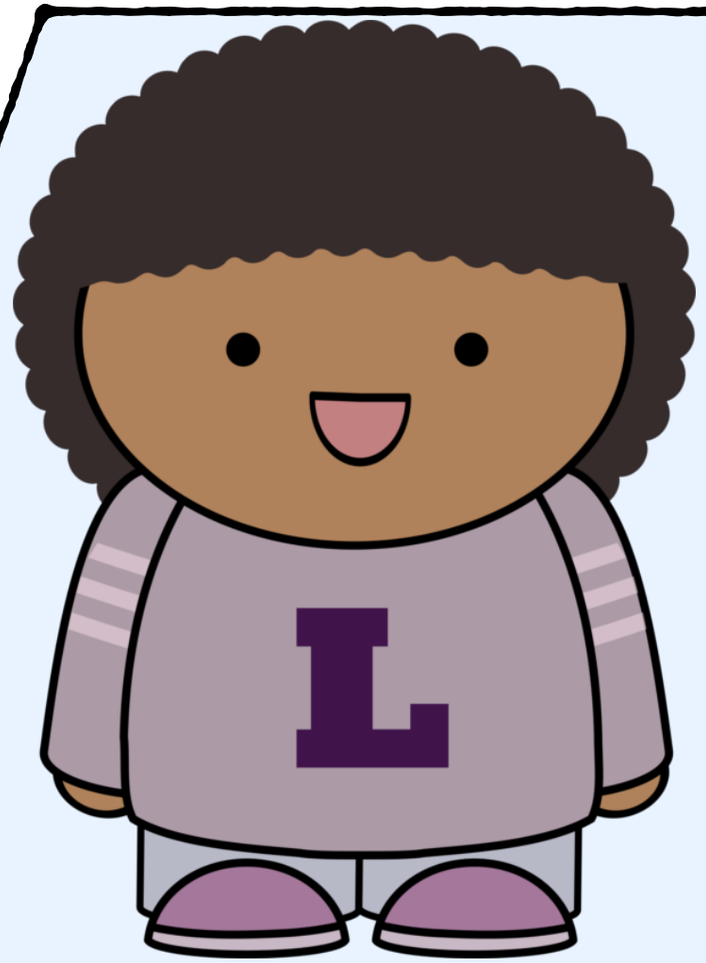
Fill in the blank:

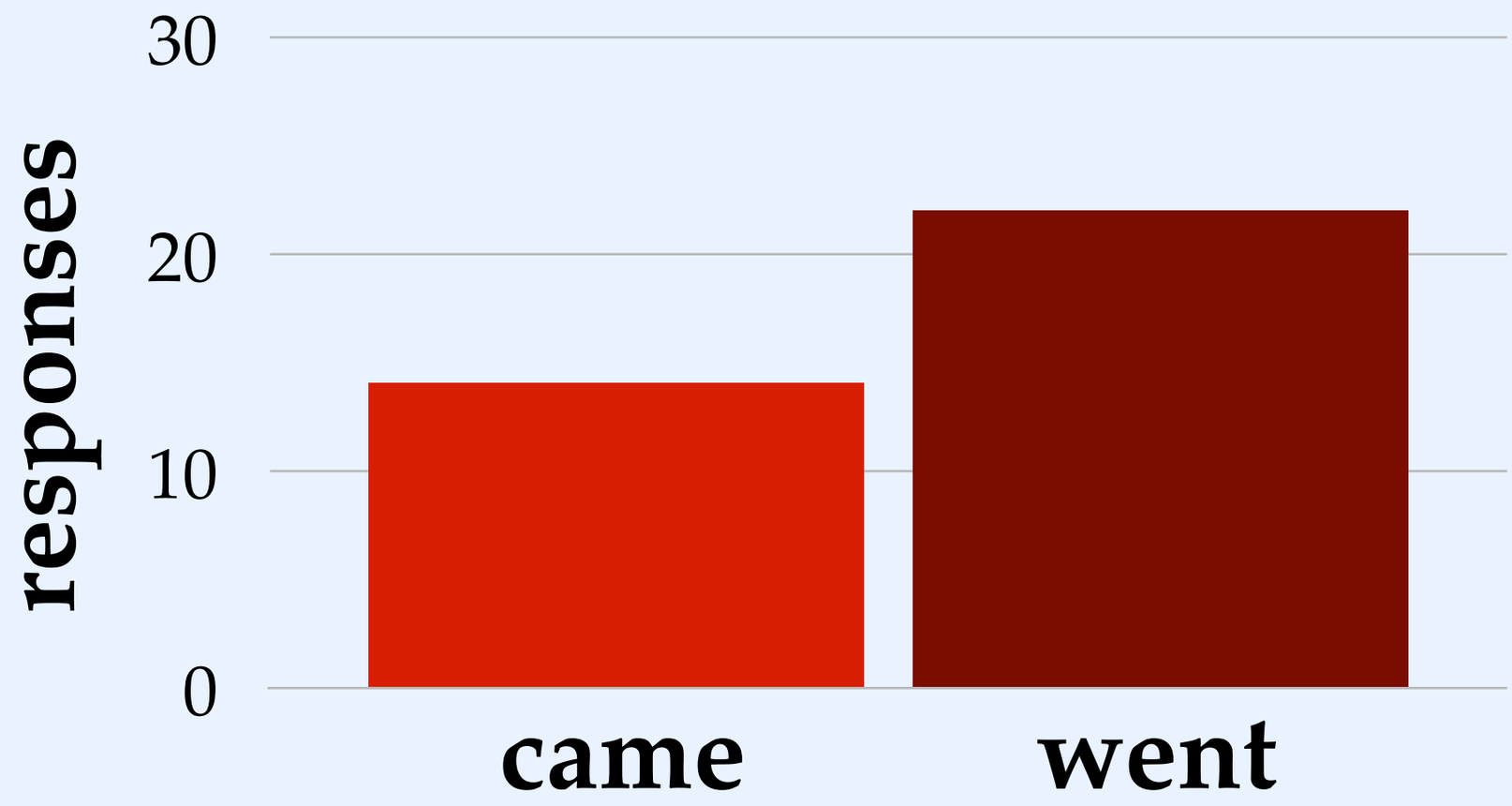
(a) came

(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?

Can we model how listeners reason about what the speaker is trying to say?

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

Perspective in language

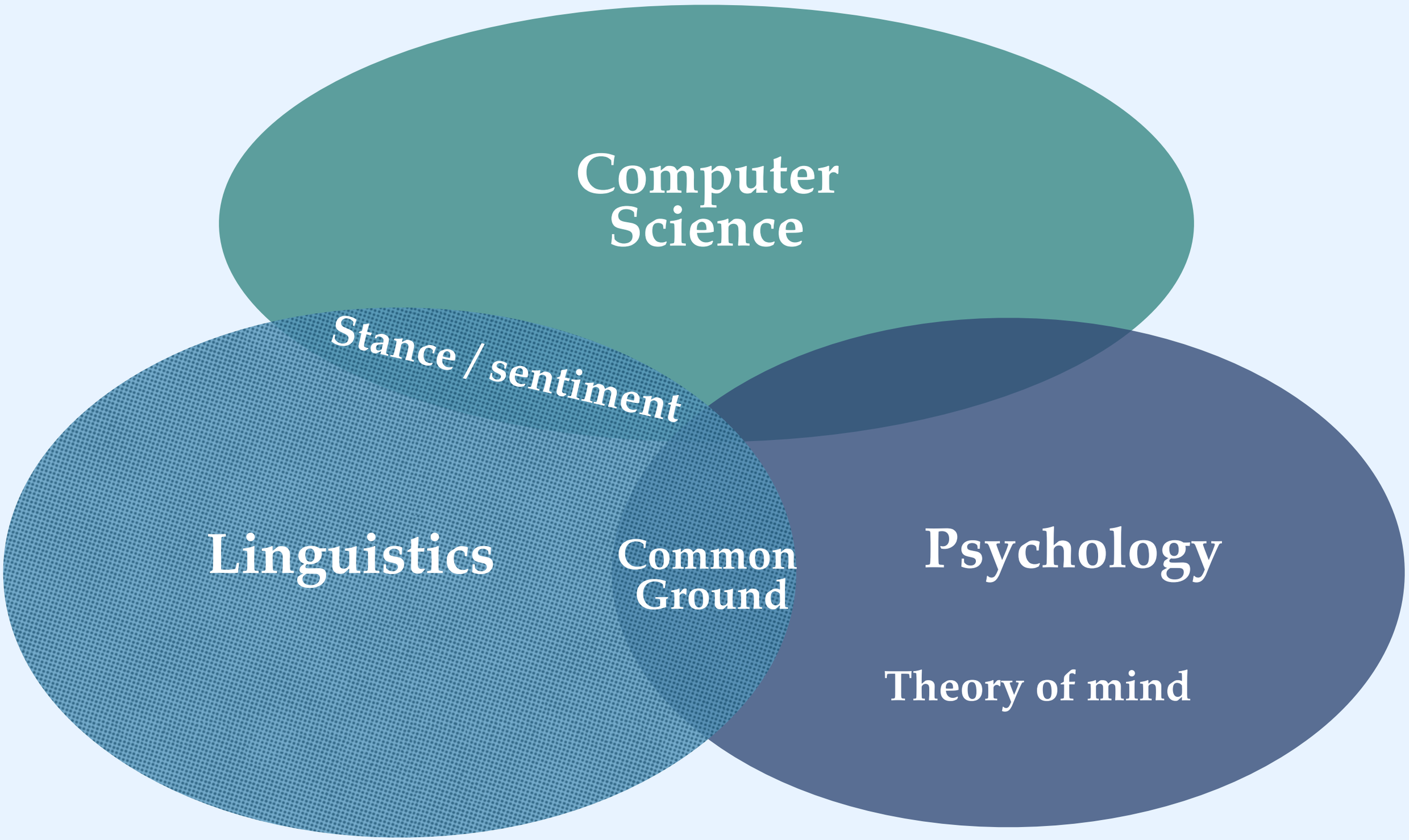
Spatial descriptions (*on the right*)

Epithets (*that jerk*)

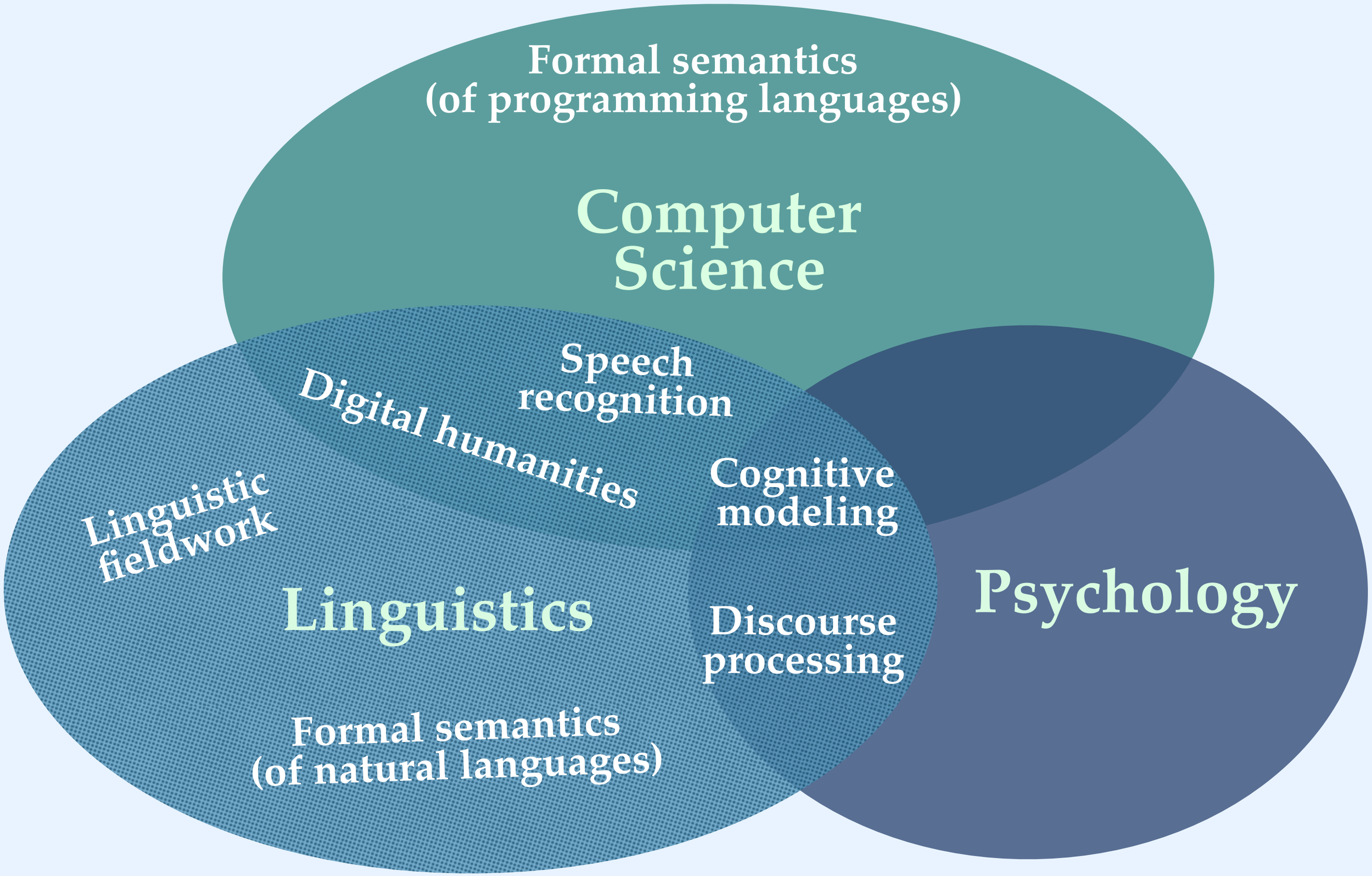
Predicates of personal taste (*tasty*)

Perspectival motion verbs (*come* and *go*)

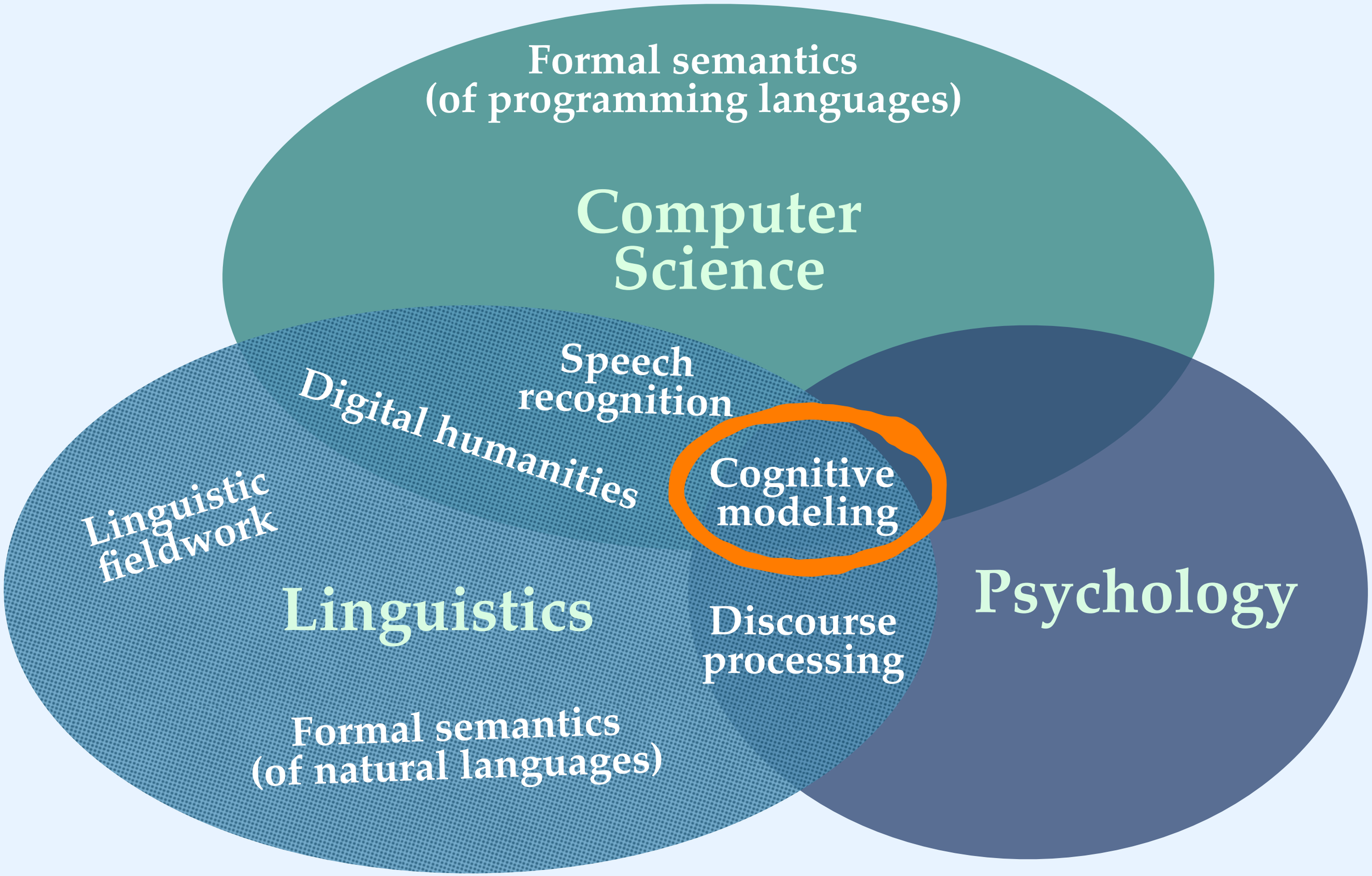
Perspective across disciplines



Research interests



Research interests



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

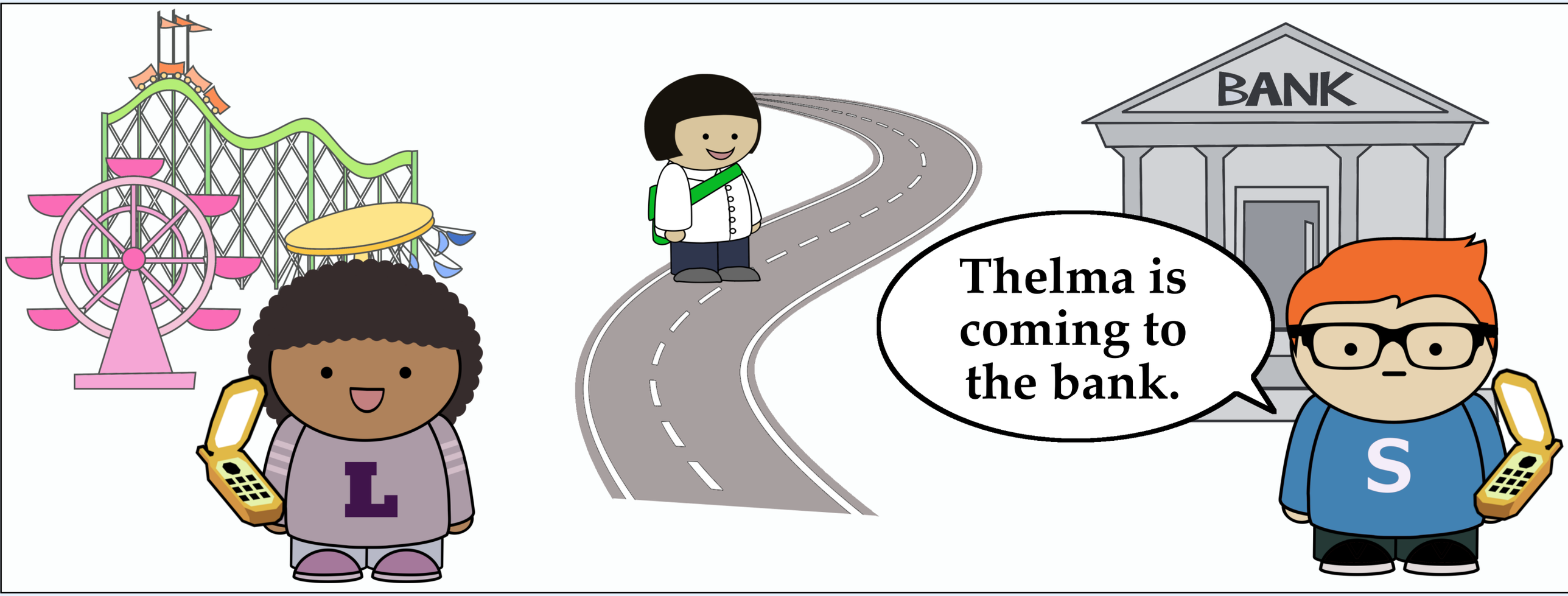
Perspectival motion verbs

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

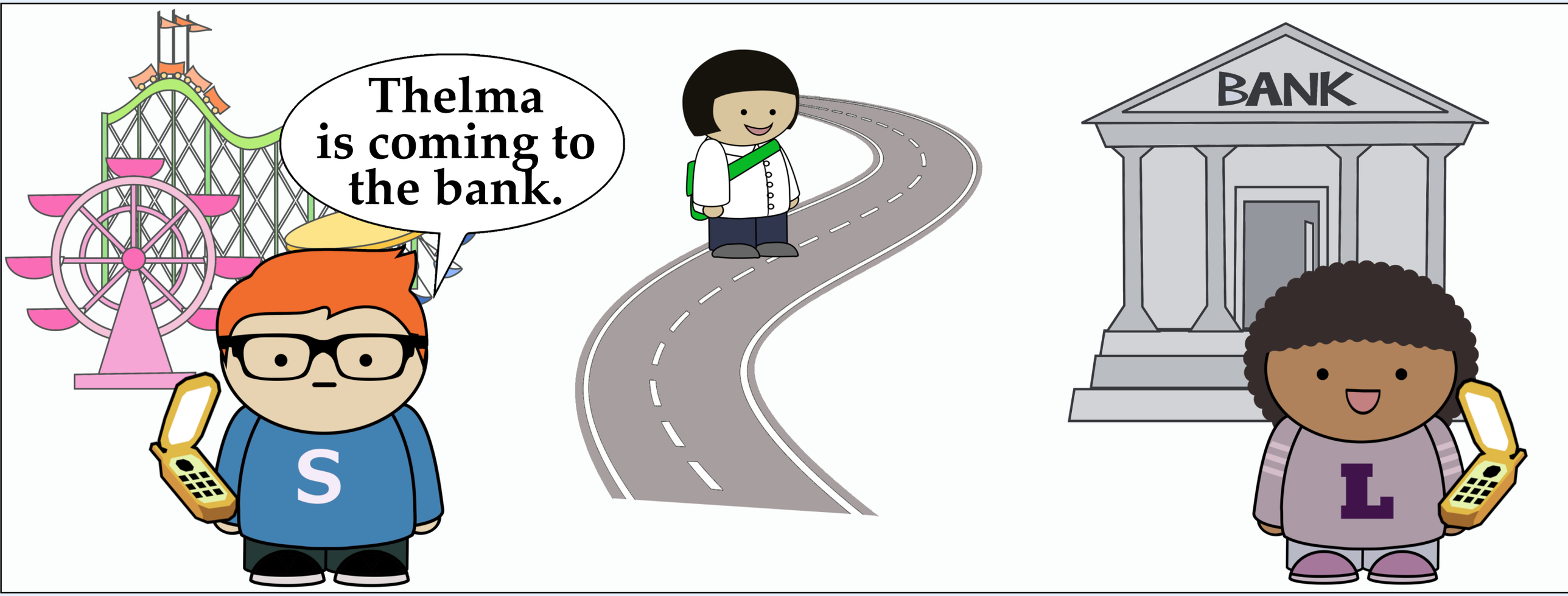
Perspective holders:

- The **listener**
- The **speaker**

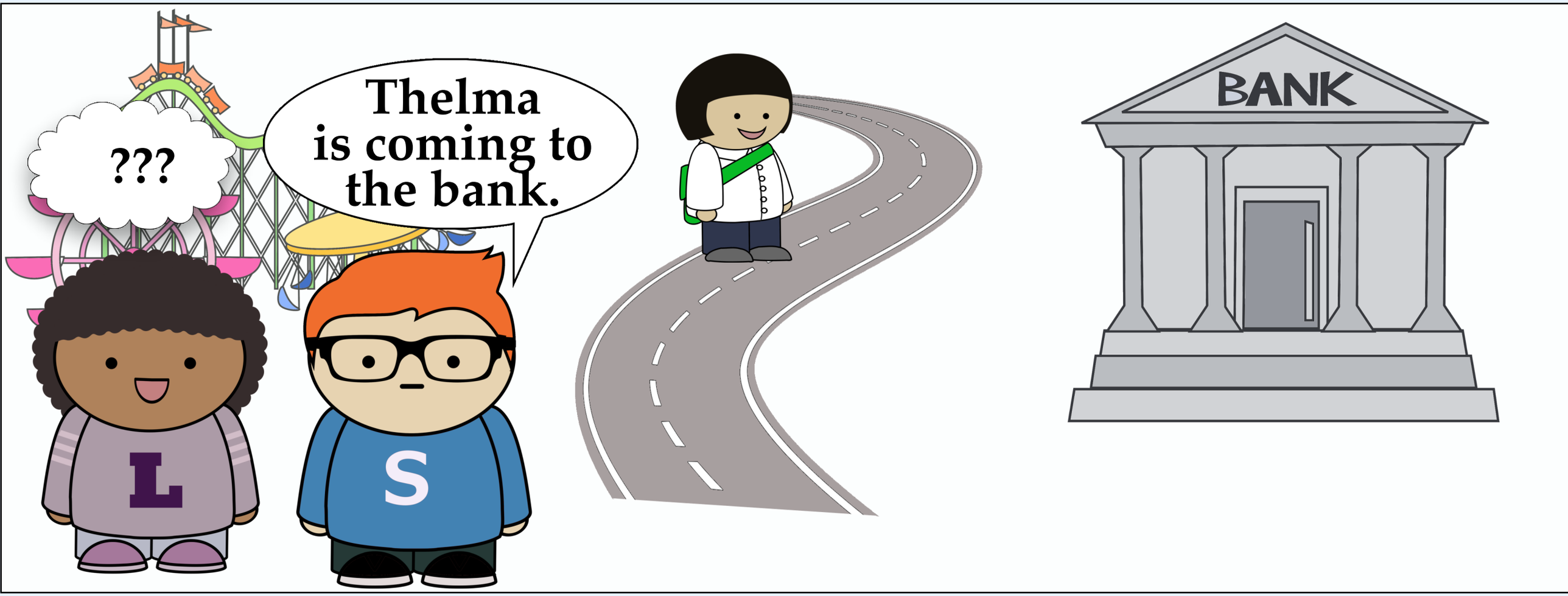
Speaker perspective



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**

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

p(perspective | utterance)

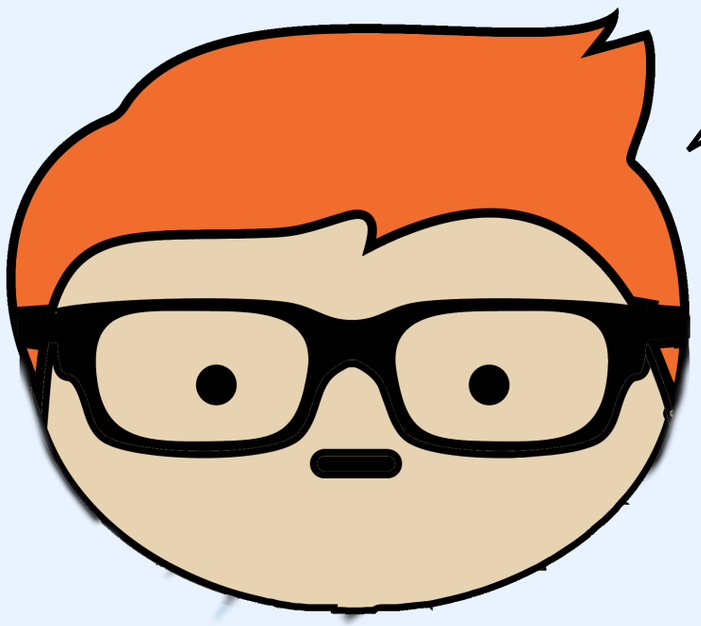
How do speakers pick a perspective?

p(perspective | meaning)

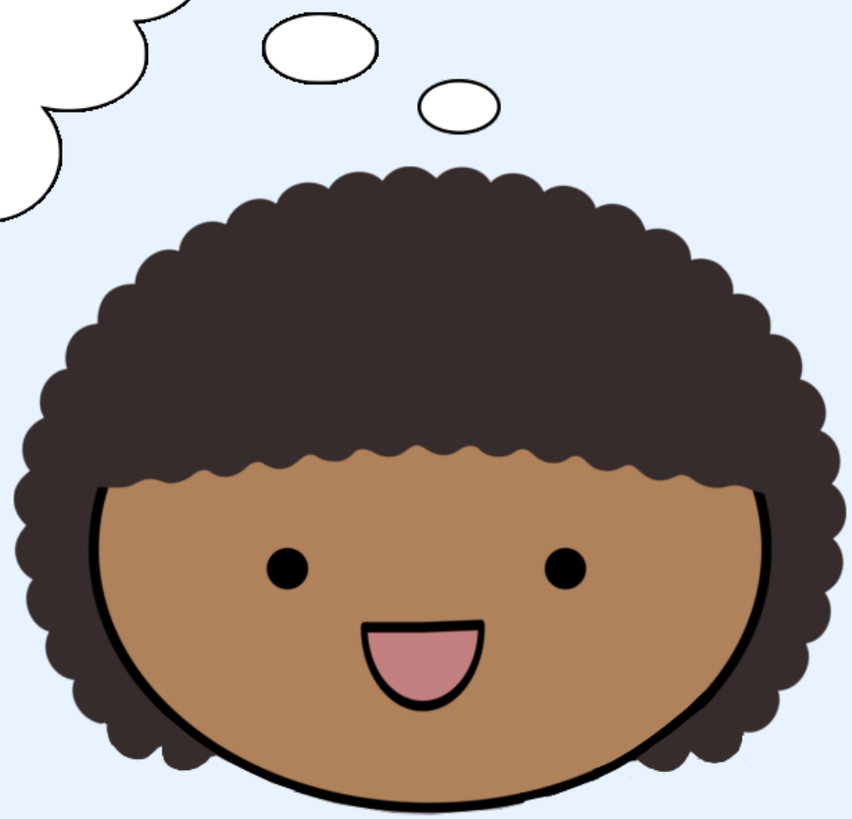
How do contextual factors license or lead to perspective shift?

p(perspective | context)

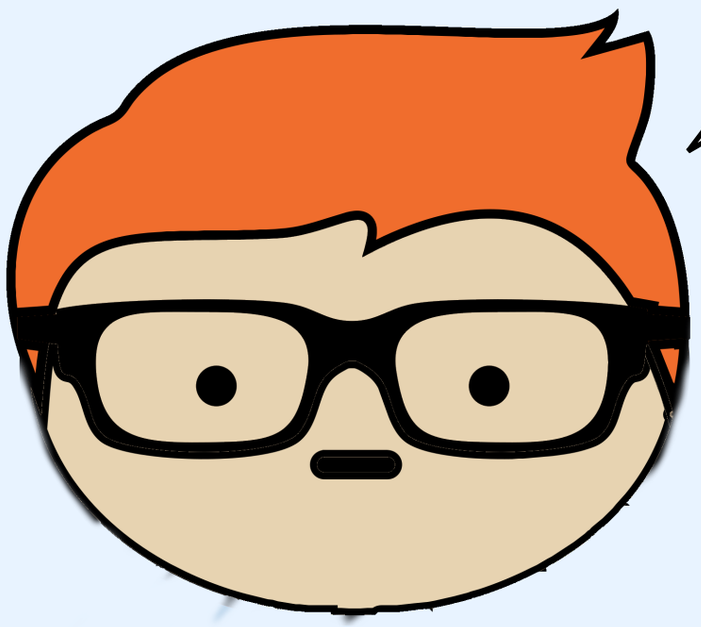
Thelma is coming to the zoo.



Who is at the zoo?

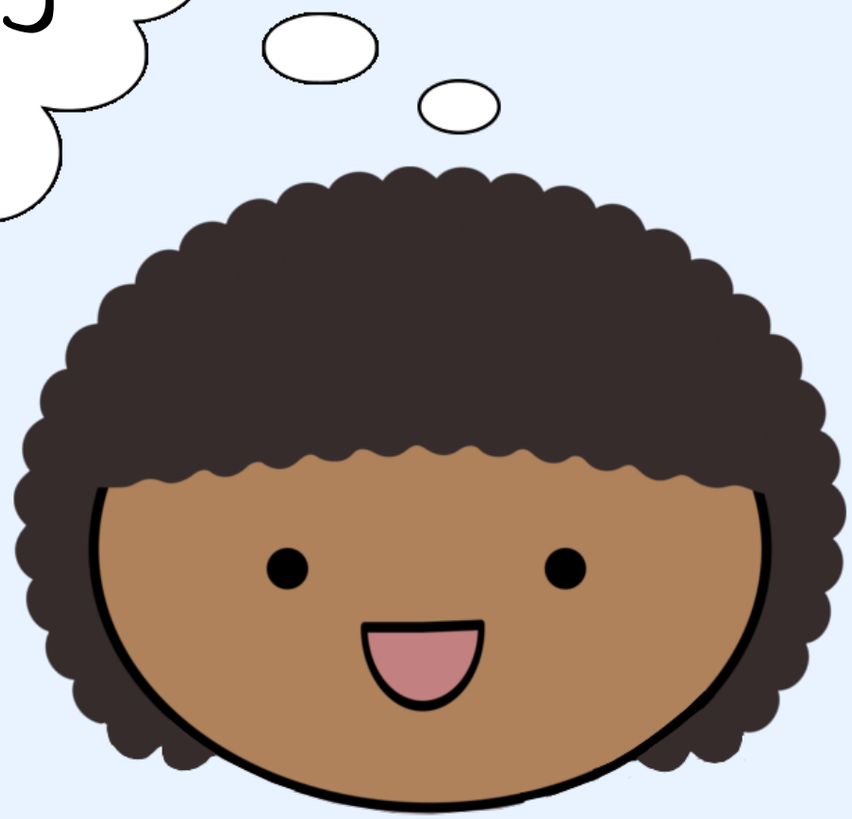


p(perspective | utterance)



Thelma
is coming.

Where is
Thelma going?



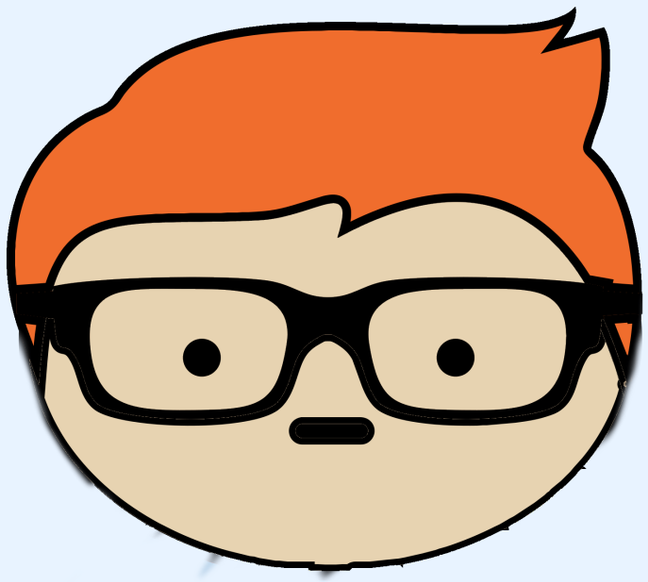
p(meaning | utterance)

Thelma is coming to the park.

Sam must be at the park.

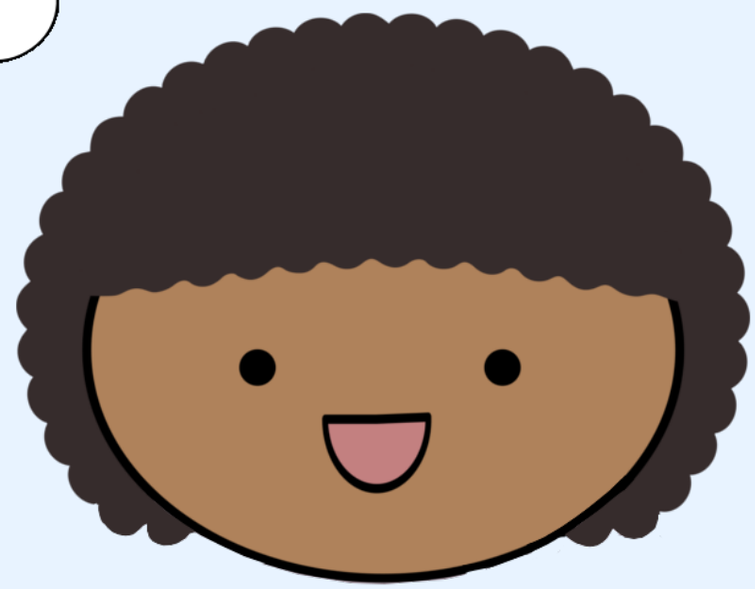
ZOO

Thelma is coming.



What is Sam trying to communicate?

Whose perspective is Sam using?



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

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

Joint perspectival reasoning

Speaker:

p(perspective,utterance | meaning)

Listener:

p(perspective, meaning | utterance)

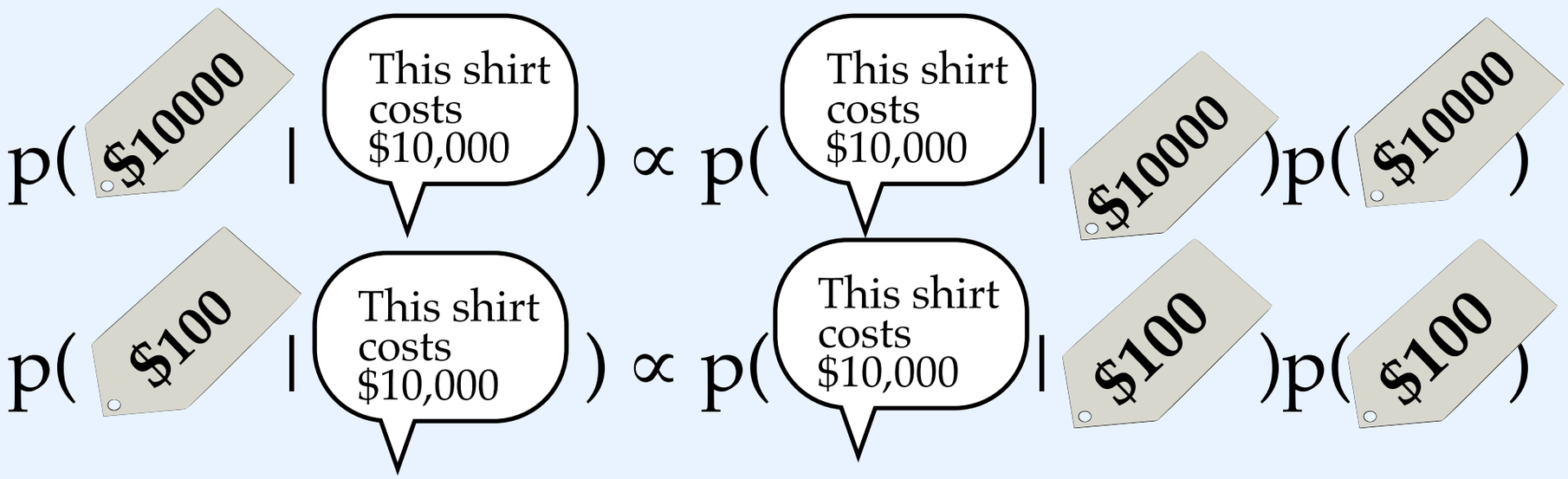
The Rational Speech Acts model

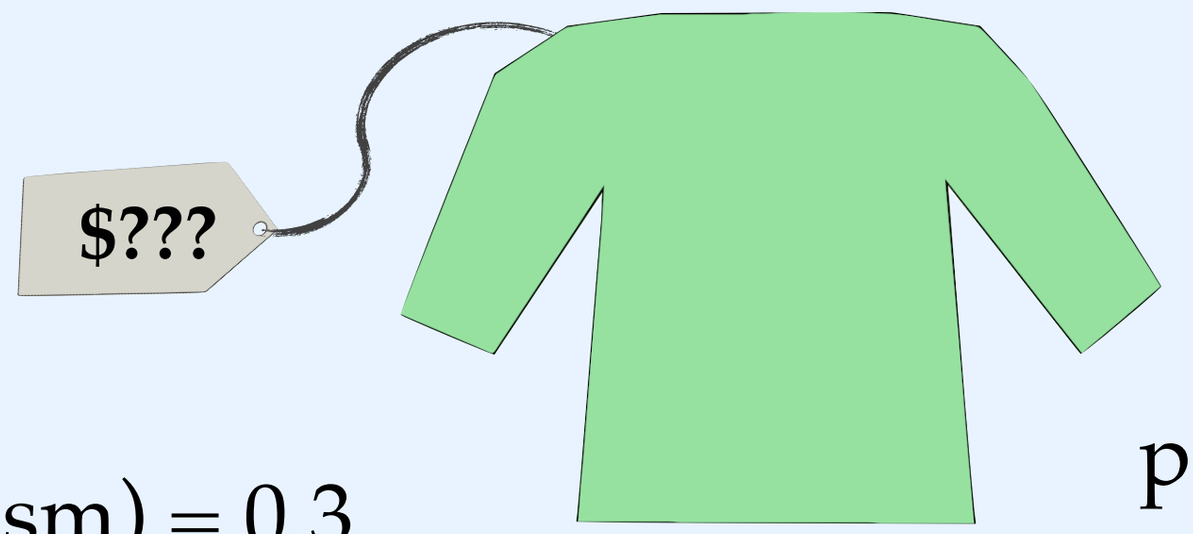
Frank & Goodman (2012)

Listener: p(meaning | utterance)

$$p(\text{meaning} \mid \text{utterance}) \propto p(\text{utterance} \mid \text{meaning}) p(\text{meaning})$$

Literal Truth is Discounted by Prior Probability





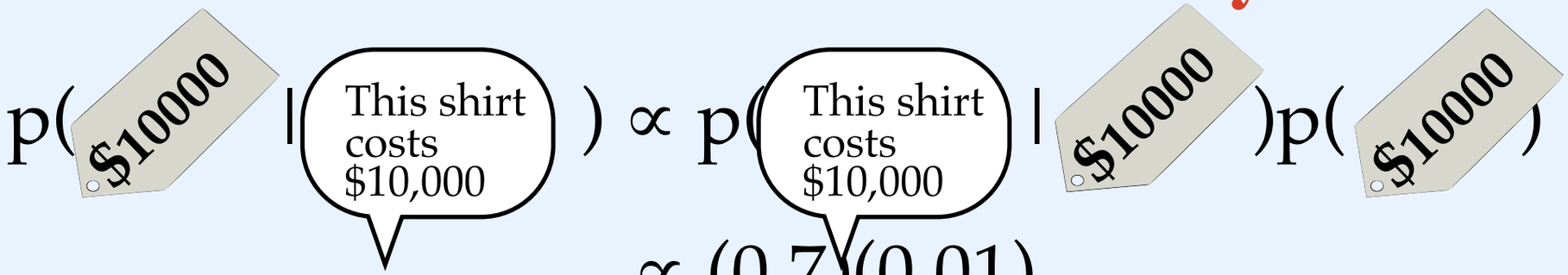
p(sarcasm) = 0.3

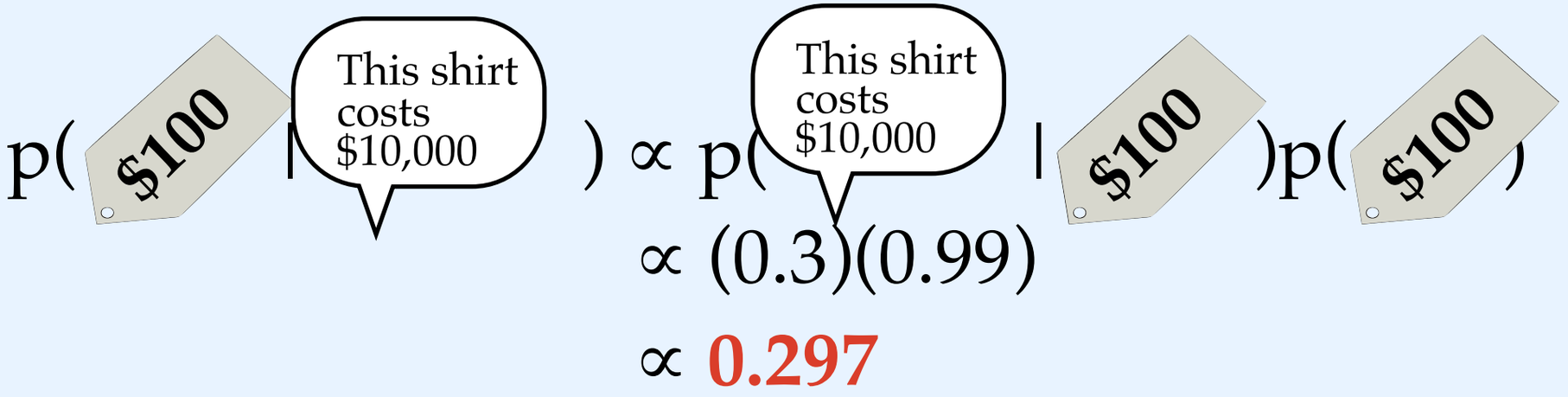
p(earnest) = 0.7

p($100) = 0.99

p($10000) = 0.01

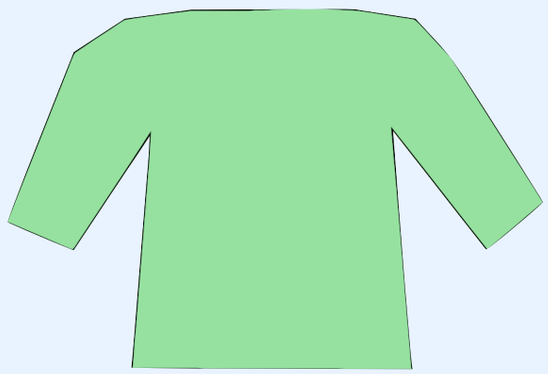
Literal Truth is Discounted by Prior Probability





p(sarcasm) = 0.3

p(earnest) = 0.7



p($100) = 0.99

p($10000) = 0.01

The Rational Speech Acts model

Frank & Goodman (2012)

Listener: p(meaning | utterance)

$$p(\text{meaning} \mid \text{utterance}) \propto p(\text{utterance} \mid \text{meaning}) p(\text{meaning})$$

The Rational Speech Acts model

Frank & Goodman (2012)

Listener: p(meaning | utterance)

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

where w = world and u = utterance

The Rational Speech Acts model

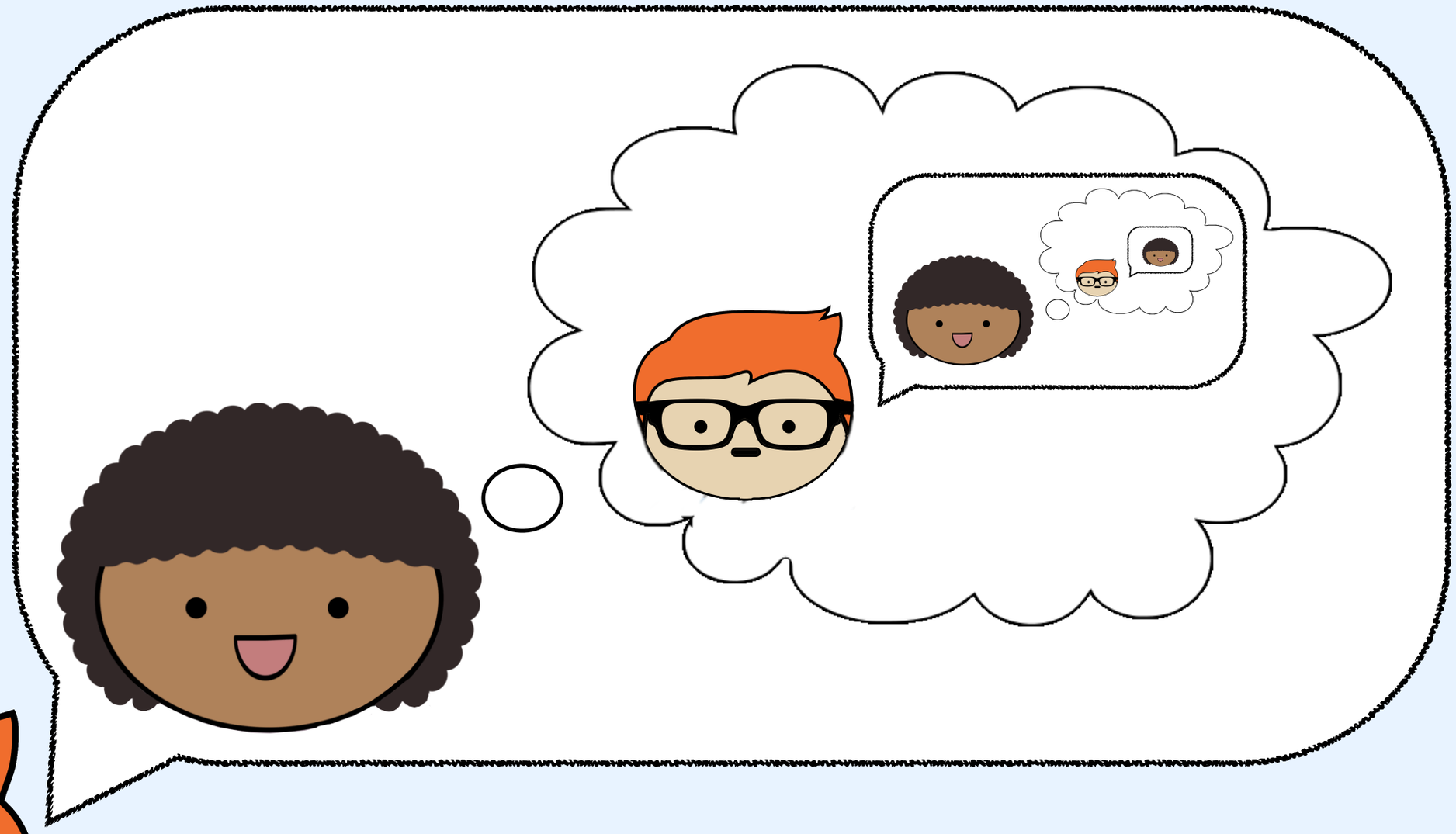
Frank & Goodman (2012)

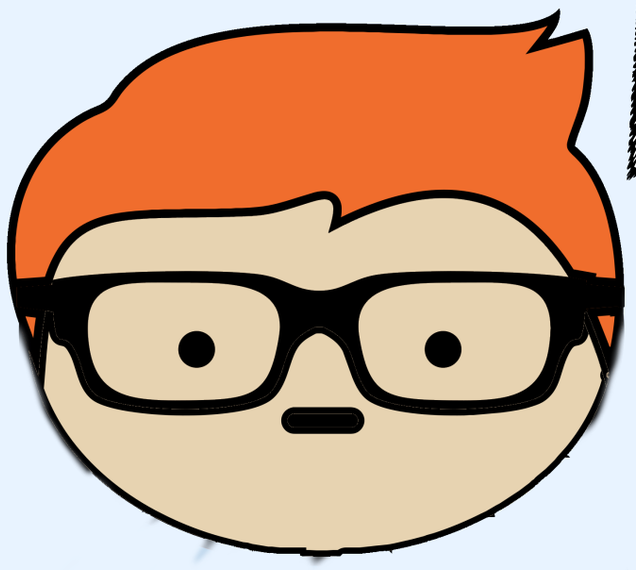
Speaker: p(utterance | meaning)

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

where w = world and u = utterance

A Recursive Model





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

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

The Rational Speech Acts model

Frank & Goodman (2012)

Literal Listener: p(meaning | utterance)

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

where w = world and u = utterance

The Rational Speech Acts model

Frank & Goodman (2012)

Pragmatic Speaker: p(utterance | meaning)

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

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

where w = world and u = utterance

The Rational Speech Acts model

Frank & Goodman (2012)

Pragmatic Listener: p(meaning | 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

Perspectival RSA model

Anderson & Dillon (2018)

Literal listener:

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

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

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

Perspectival RSA model

Anderson & Dillon (2018)

Pragmatic speaker:

p(utterance, perspective | meaning)

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

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

Perspectival RSA model

Anderson & Dillon (2018)

Pragmatic listener:

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

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

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

Perspectival RSA model

Anderson & Dillon (2018)

Literal listener:

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

Pragmatic speaker:

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

Pragmatic listener:

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

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

Model Functions

Interpretation function: $[[u]]^{w,a}$

Returns 0 or 1 depending on whether the utterance is true in world w according to perspective a.

Perspective cost: cost(a)

Penalizes non-speaker perspectives.

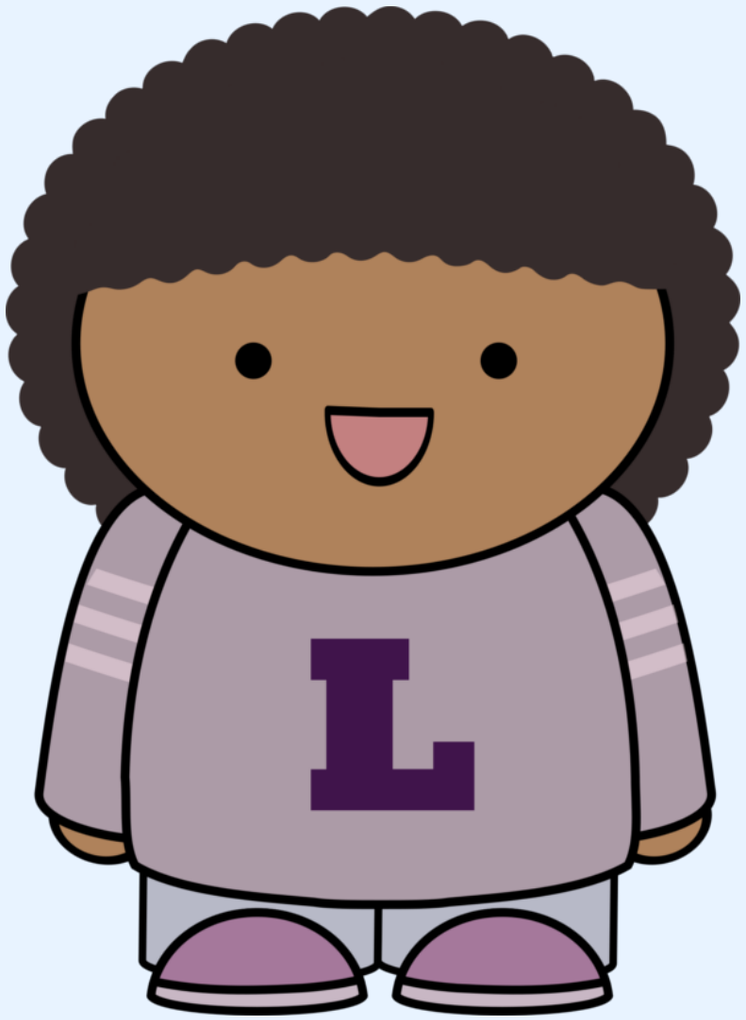
Motivated by ego-centricity findings (Epley et al. 2004) and by speaker-default analyses (Harris 2012).

Talk Outline

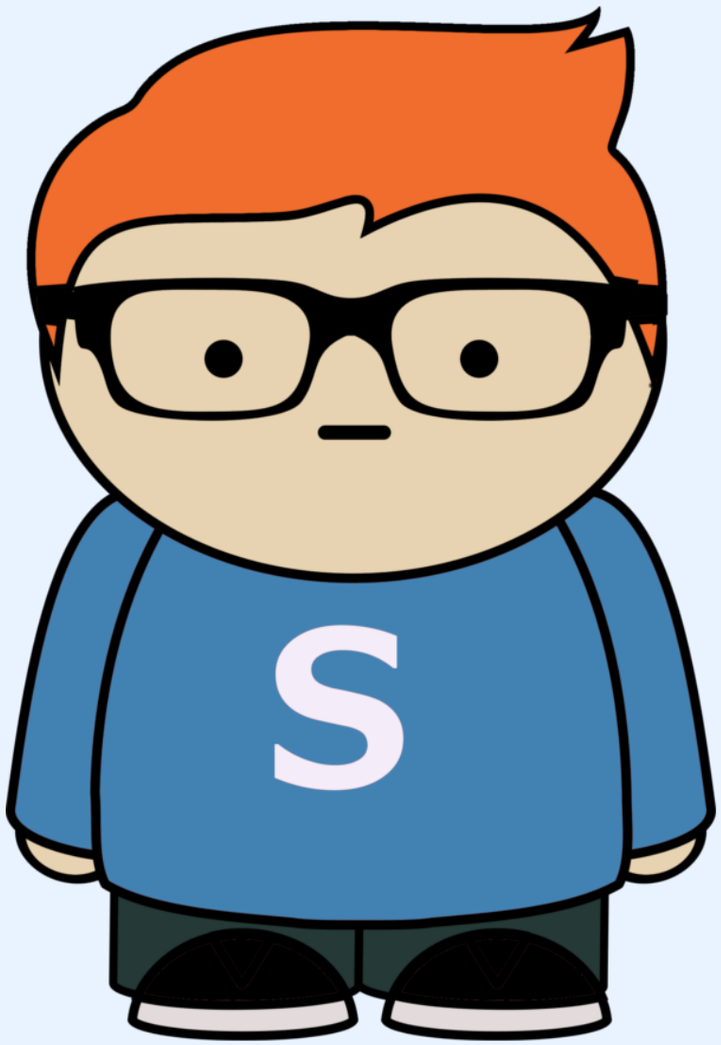
1. Develop a computational model of perspectival reasoning as a joint inference problem
2. Run simulations to predict speaker and listener behavior
3. Compare model predictions against evidence from crowdsourced behavioral experiments
    - How do listeners understand perspectival expressions?
    -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

Perspective 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

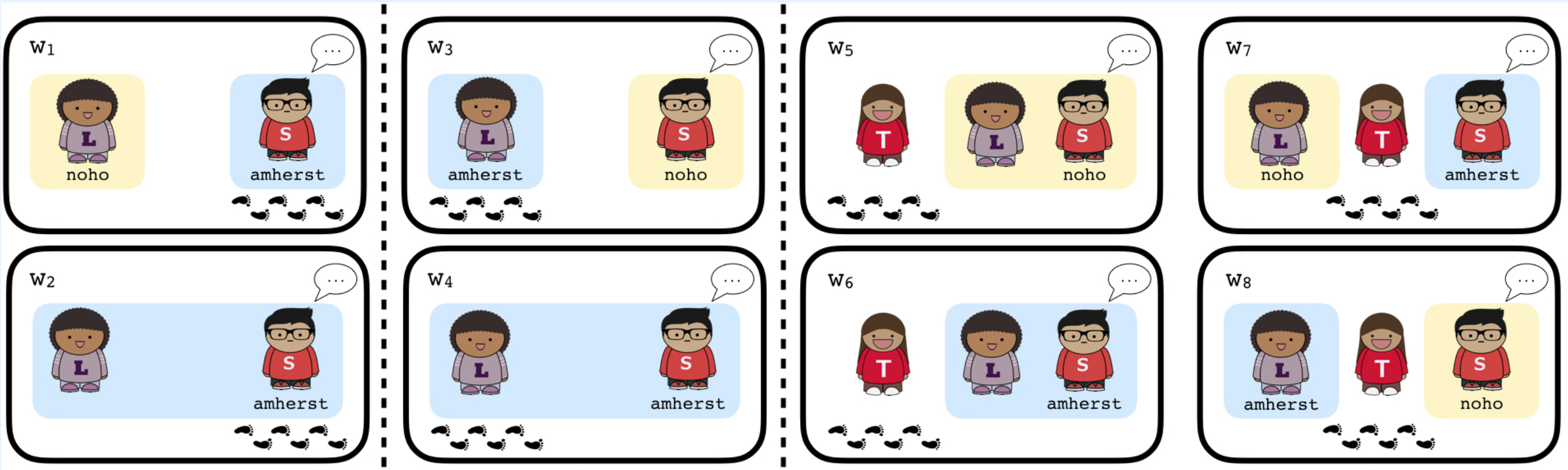
The literal listener looks up the literal meaning of the message according to the perspective and world.

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


w1
L
noho
S
amherst
w3
L
amherst
S
noho
w5
T
L
S
noho
w7
L
noho
T
S
amherst
w2
L
S
amherst
w4
L
S
amherst
w6
T
L
S
amherst
w8
L
amherst
T
S
noho

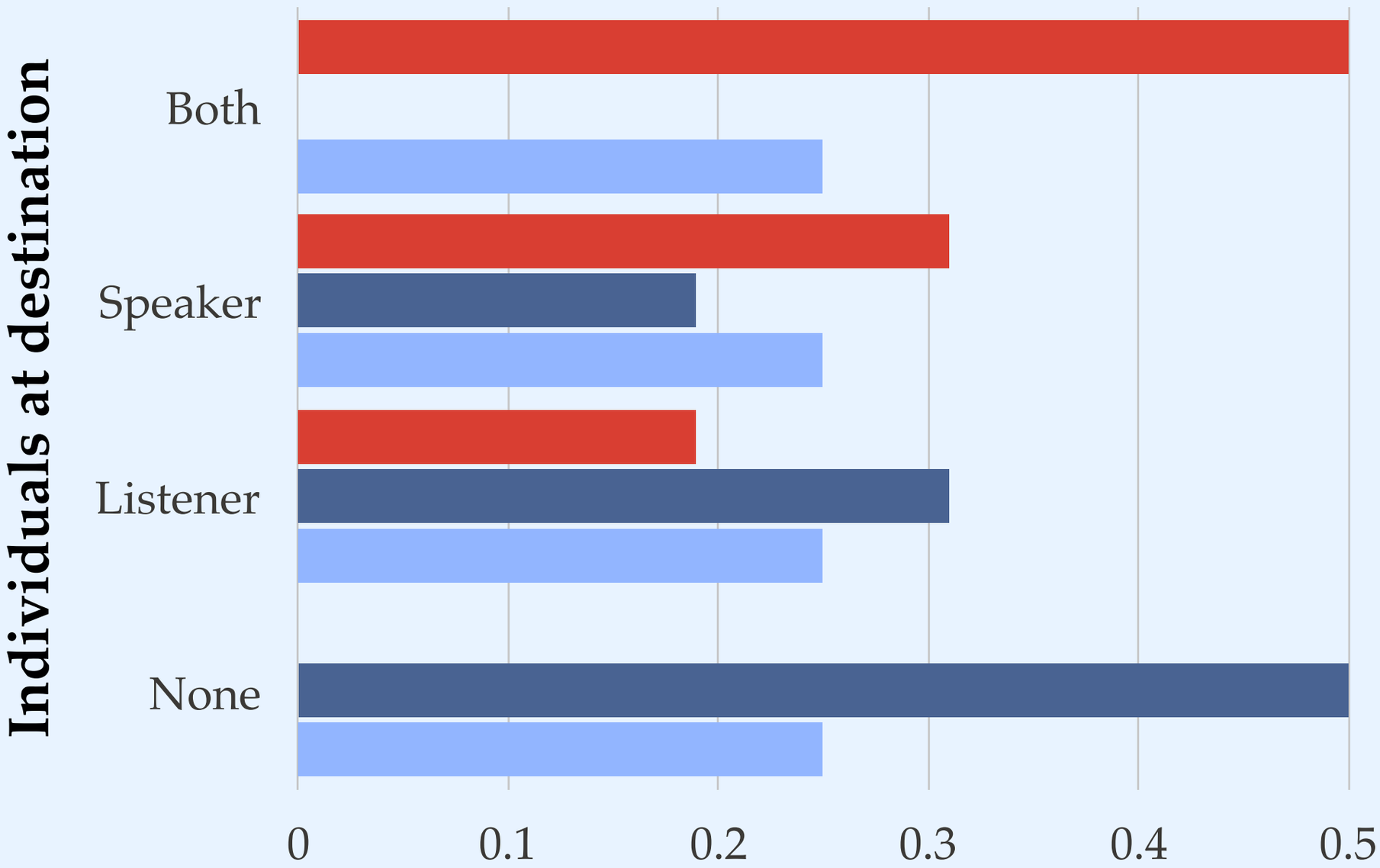
Model Predictions

Generating predictions

- Model is implemented in the WebPPL probabilistic programming language
- Simulations run using Markov Chain Monte Carlo sampling (100000 iterations, burn-in of 1000)

Pragmatic Listener: marginal probabilities

p(w | *Thelma is ___ to Northampton*) with cost=0.5



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

Pragmatic Listener: marginal probabilities

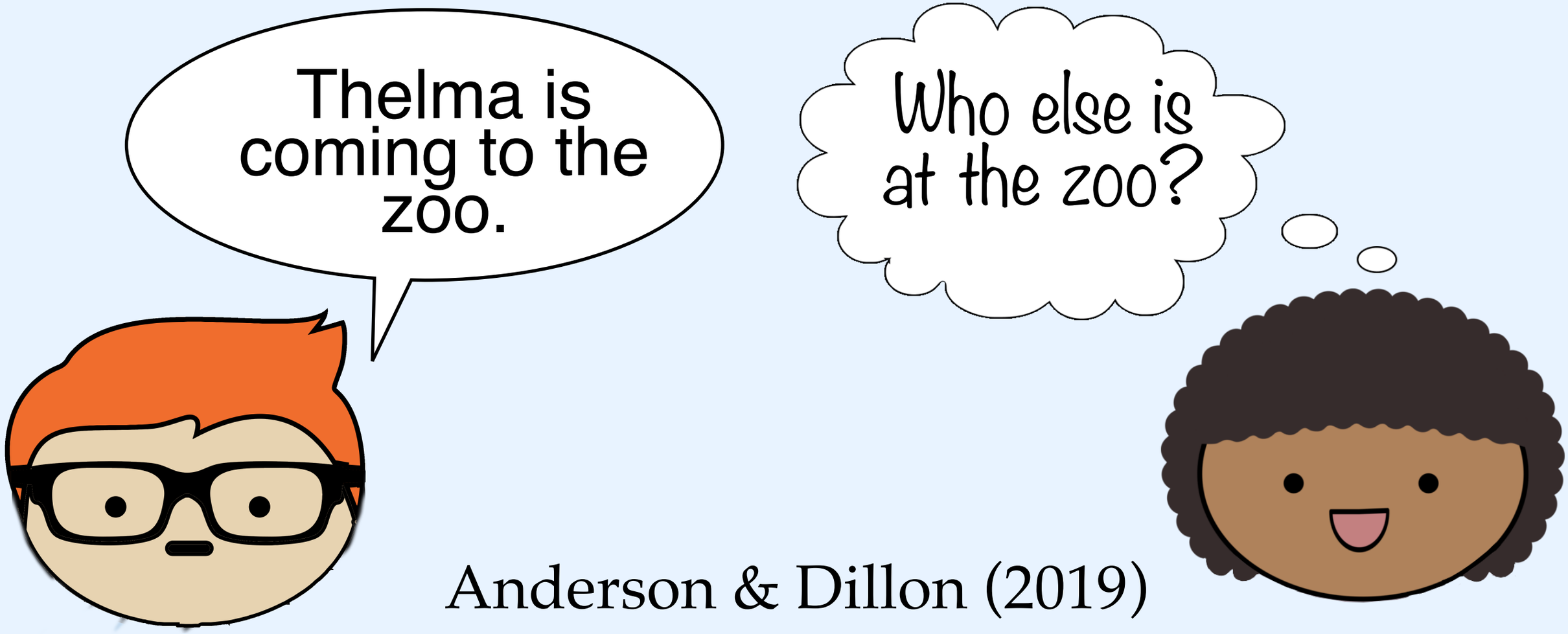
The model predicts highest marginal probability for the world in which both speaker and listener are located at the destination.

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

Comprehension experiment

p(w | u)



Anderson & Dillon (2019)

Thelma is coming hypotheses

Multiple perspectives 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.

Perspective heuristics

Speaker-default (Roberts 2012, Harris 2012): listeners should assume that the speaker is the perspective holder (unless contextual factors rule this out)

p(perspective=speaker | utterance) = 1 if context allows

Thelma is coming hypotheses

Multiple perspectives 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.

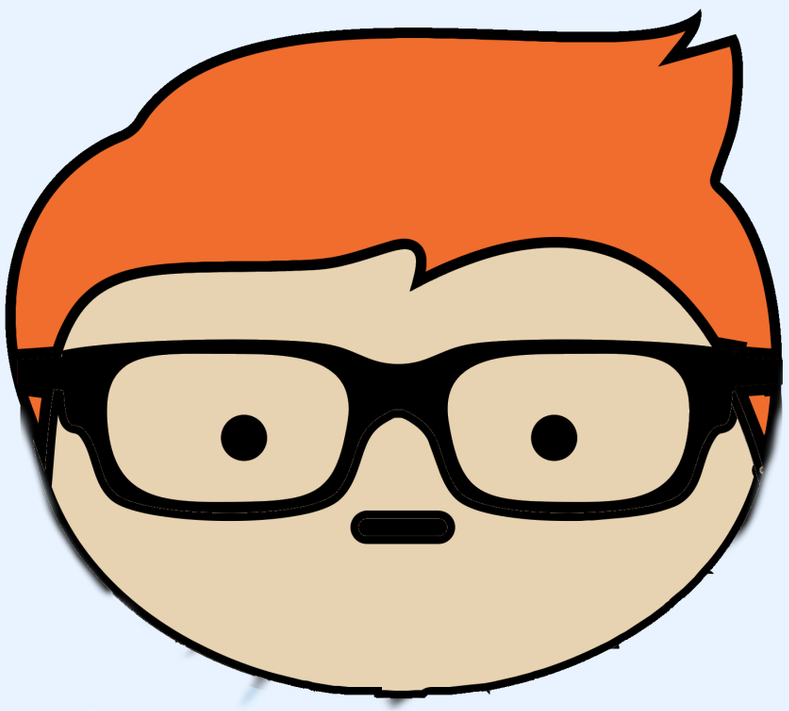
Speaker default heuristic (Roberts 2012, Harris 2012): If listeners assume that the speaker is the perspective holder, they should assign equal marginal posterior probability to all worlds where the speaker is at the destination.

Experimental design

- Participants see the speaker saying a sentence
- After a pause, they see an image depicting the sentence
- They indicate whether or not the scene matches the sentence.
- We measure reaction times and accuracy.

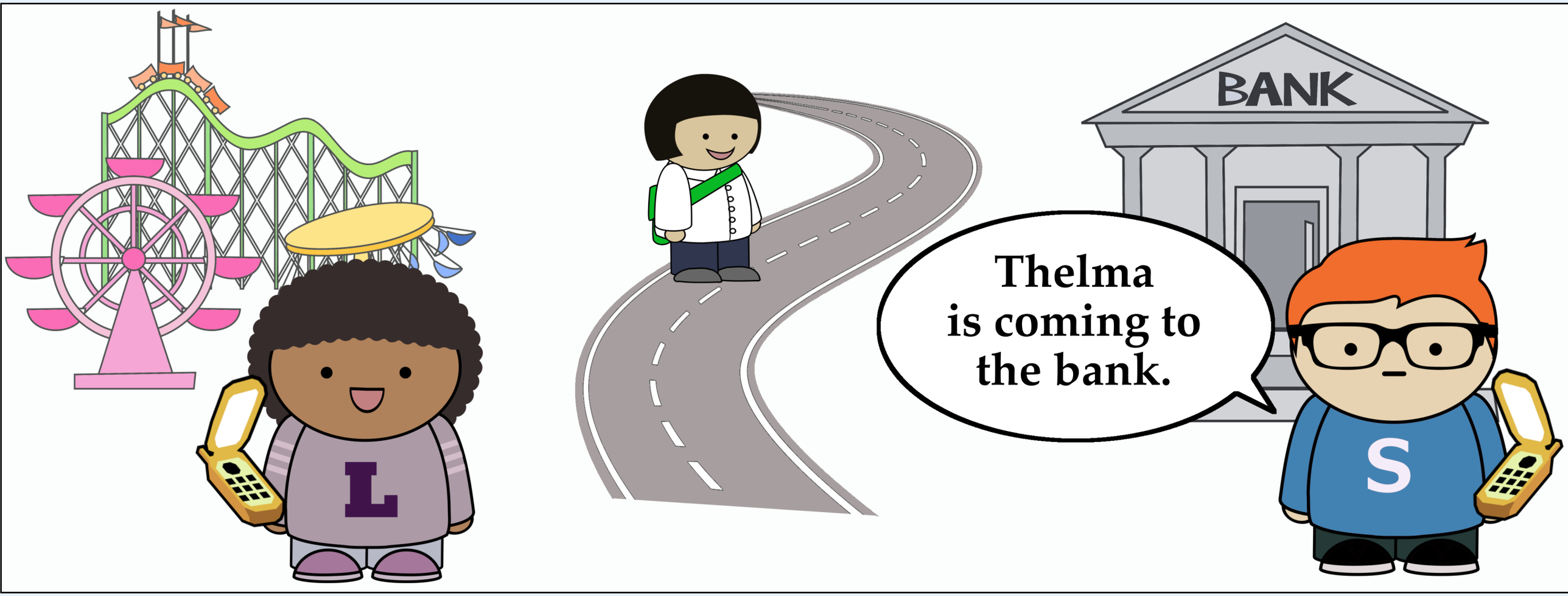
Prompt

Thelma is coming to the bank.



Perspectival, both speaker and listener at destination

Perspectival, only speaker at destination



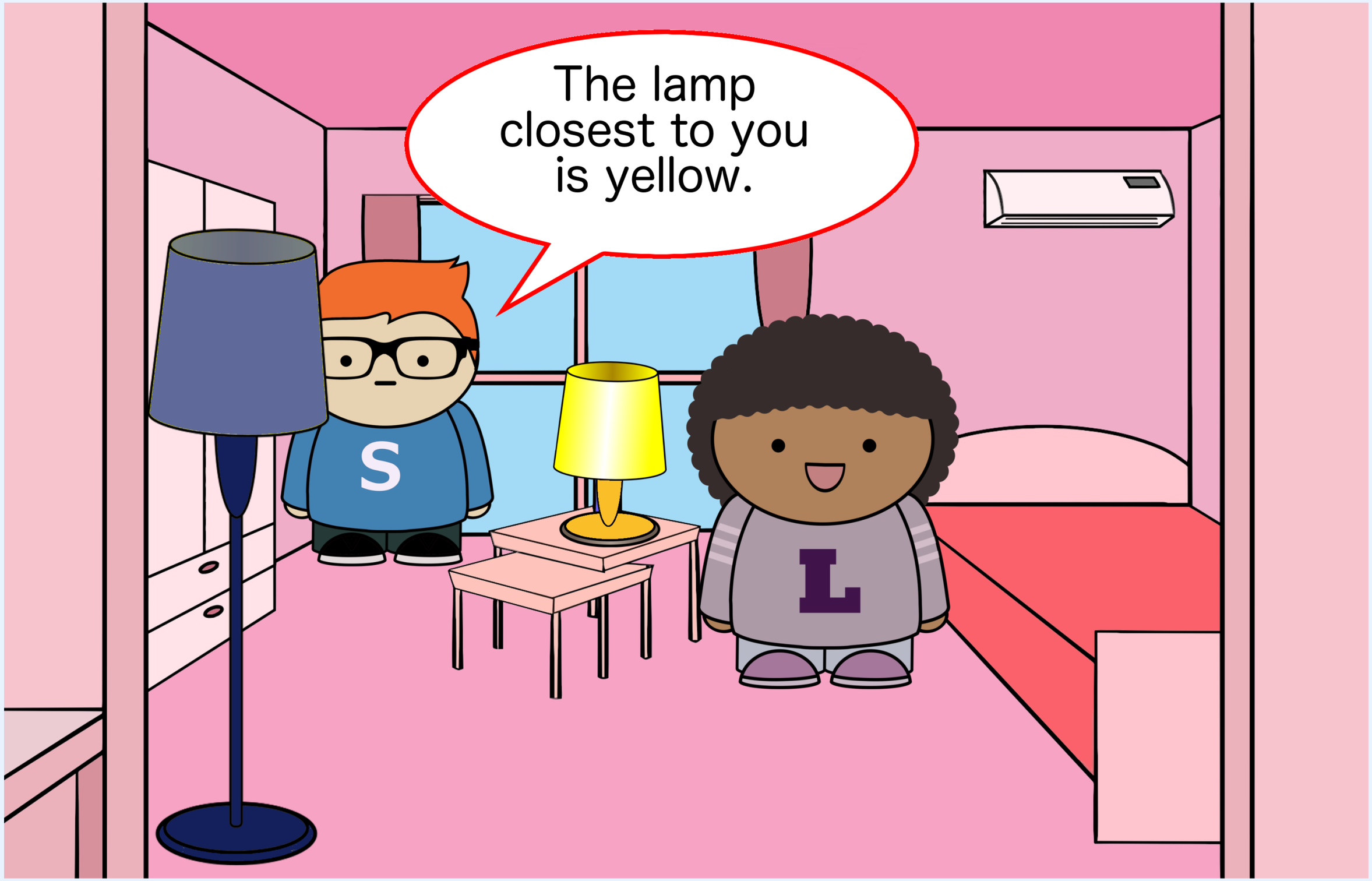
Experimental design

- Eight conditions: 4 worlds x 2 sentences
- Perspectival condition:

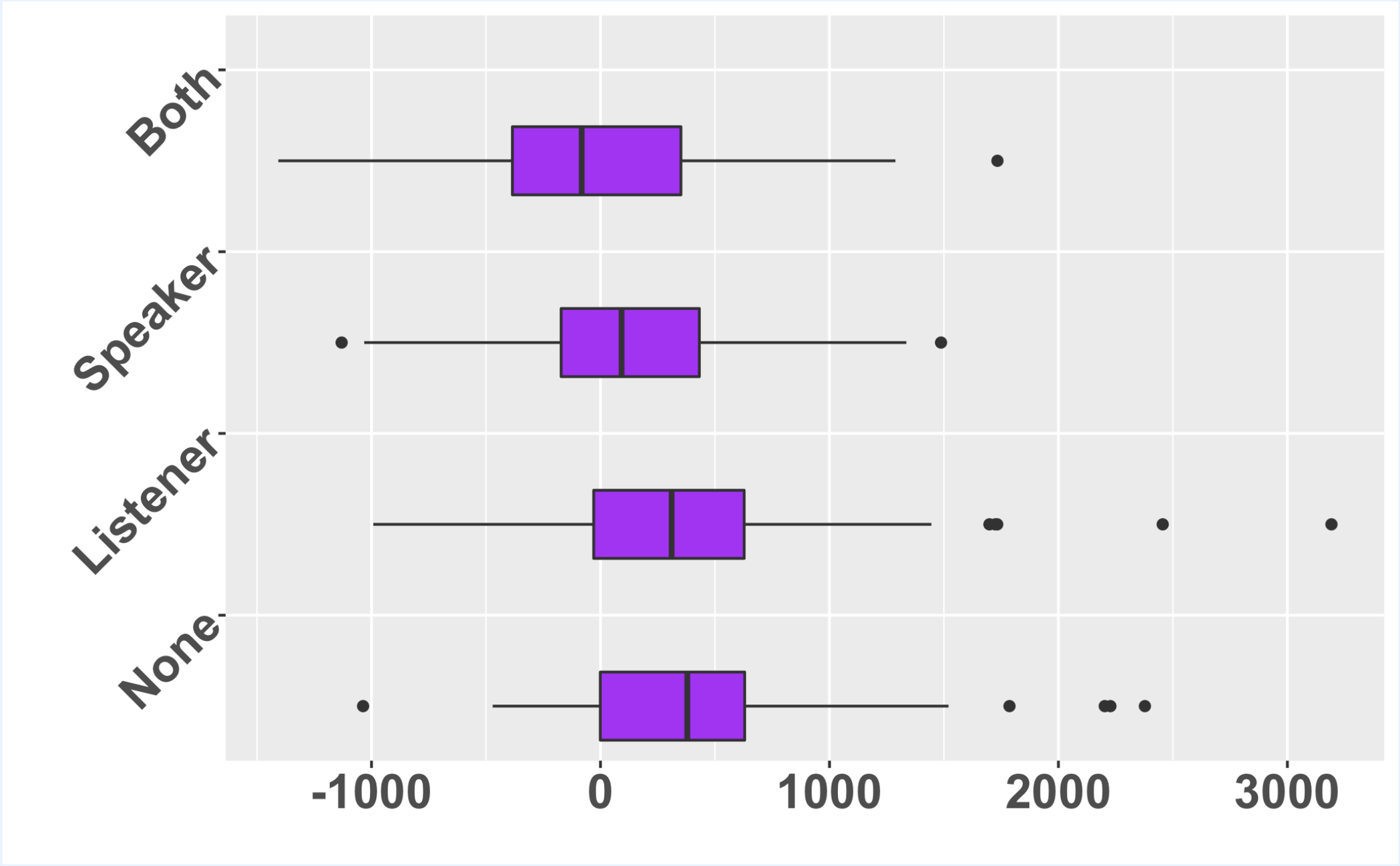
  "Thelma is coming to the zoo."
- Plain condition:

  "Thelma is driving to the zoo."
- 80 monolingual American English-speaking participants recruited on Prolific
- 6 items in each condition

Spatial items



Reaction time results

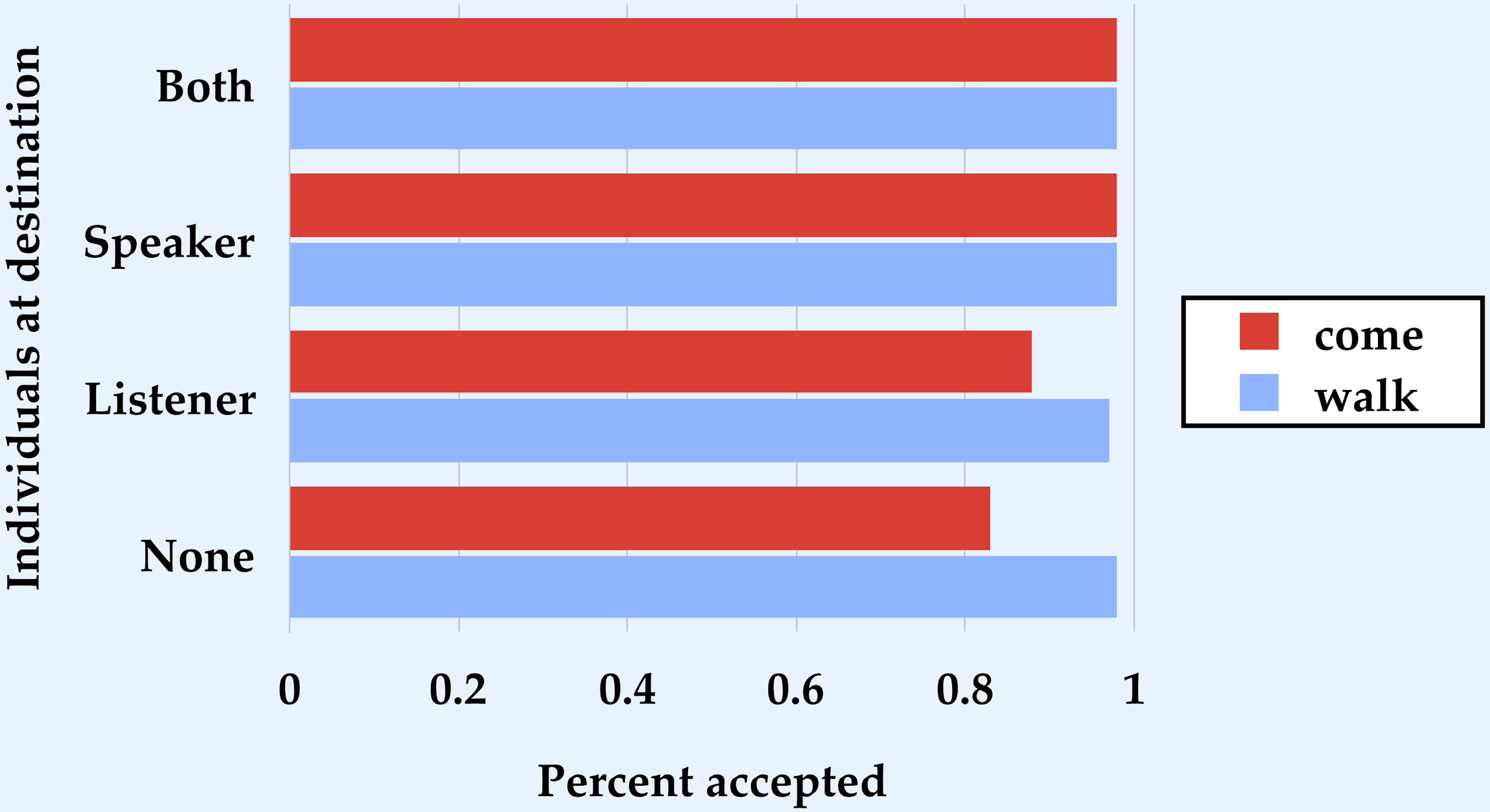


Difference in mean Cosineau-corrected RTs between *walk* and *come*

Mixed effects model on RTs

Condition	beta	*z*	*p*
(Intercept)	7.65 (+/-0.04)	179.5	**<0.0001**
come	0.06 (+/-0.02)	3.2	**0.002**
both	0.03 (+/-0.02)	1.4	0.15
listener	-0.03 (+/-0.02)	-1.3	0.2
none	-0.08 (+/-0.02)	-4.1	**<0.0001**
come:both	-0.06 (+/-0.03)	-2.2	**0.027**
come:listener	0.03(+/-0.03)	1.1	0.3
come:none	0.05(+/-0.03)	1.7	0.1

Accuracy results



Comprehension Results: p(w | u)

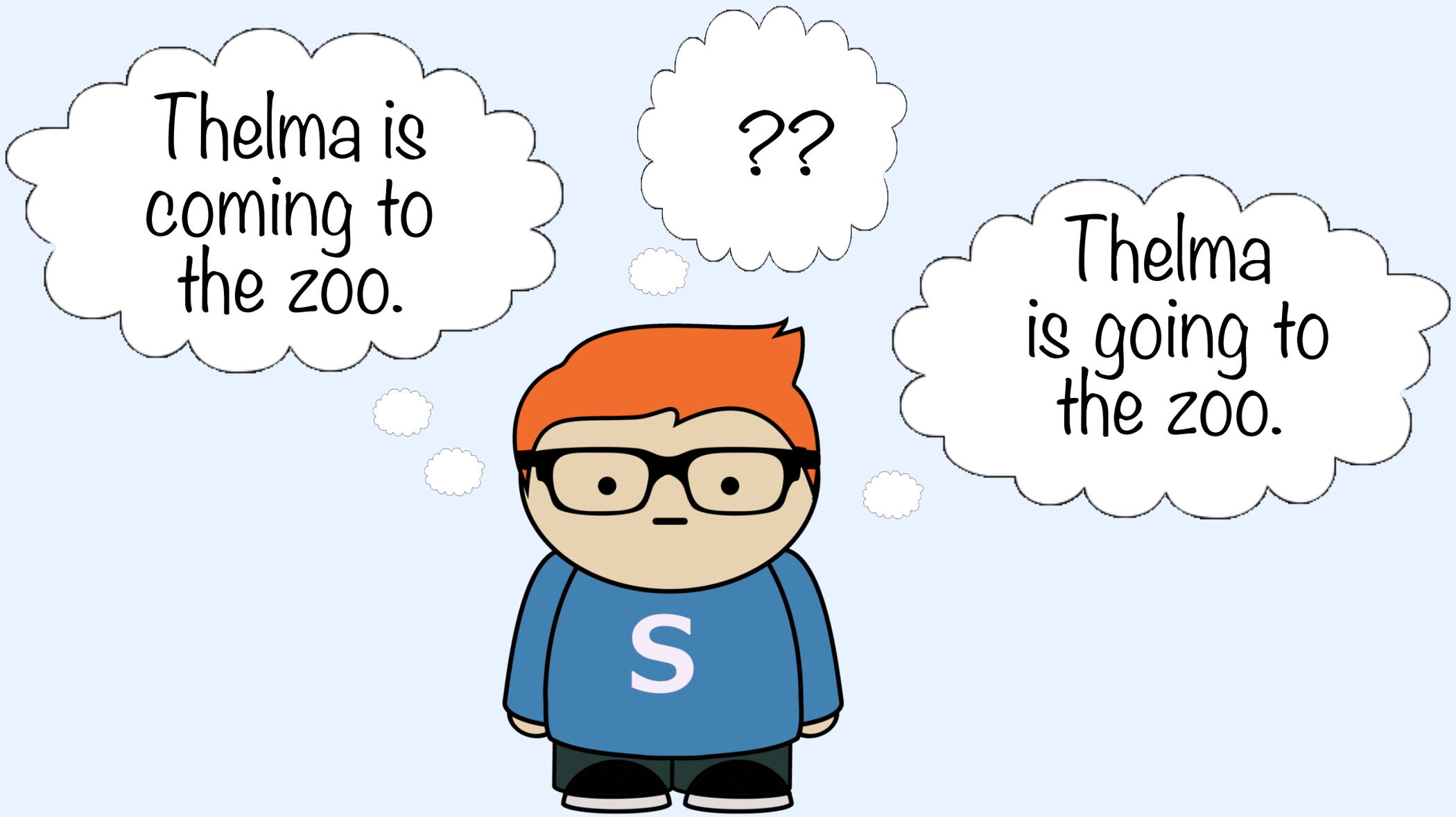
Summary:

- Participants were fastest to accept scenes following *come* in the both condition
- Accuracy was at ceiling for the speaker and both conditions
- There was a surprisingly small difference in accuracy between the none and listener *come* conditions

Talk Outline

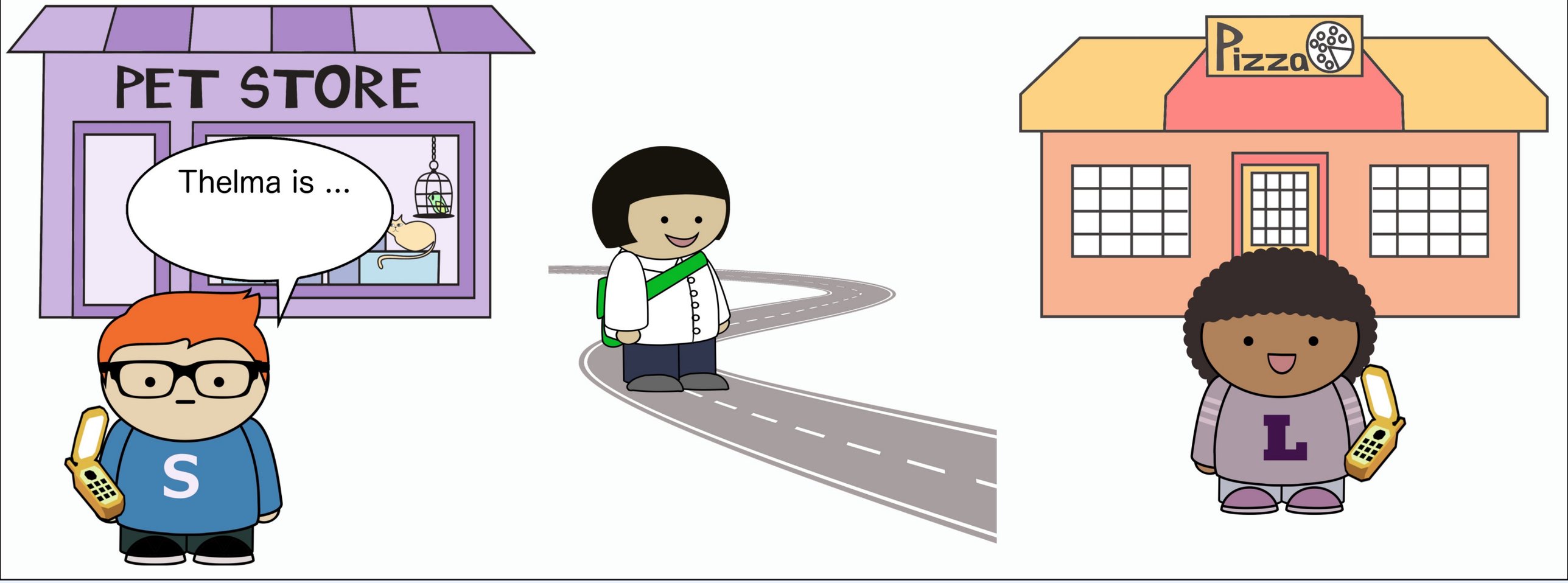
1. Develop a computational model of perspectival reasoning as a joint inference problem
2. Run simulations to predict speaker and listener behavior
3. Compare model predictions against evidence from crowdsourced behavioral experiments
   - How do listeners understand perspectival expressions?
   - How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

Production experiment: p(u | w)

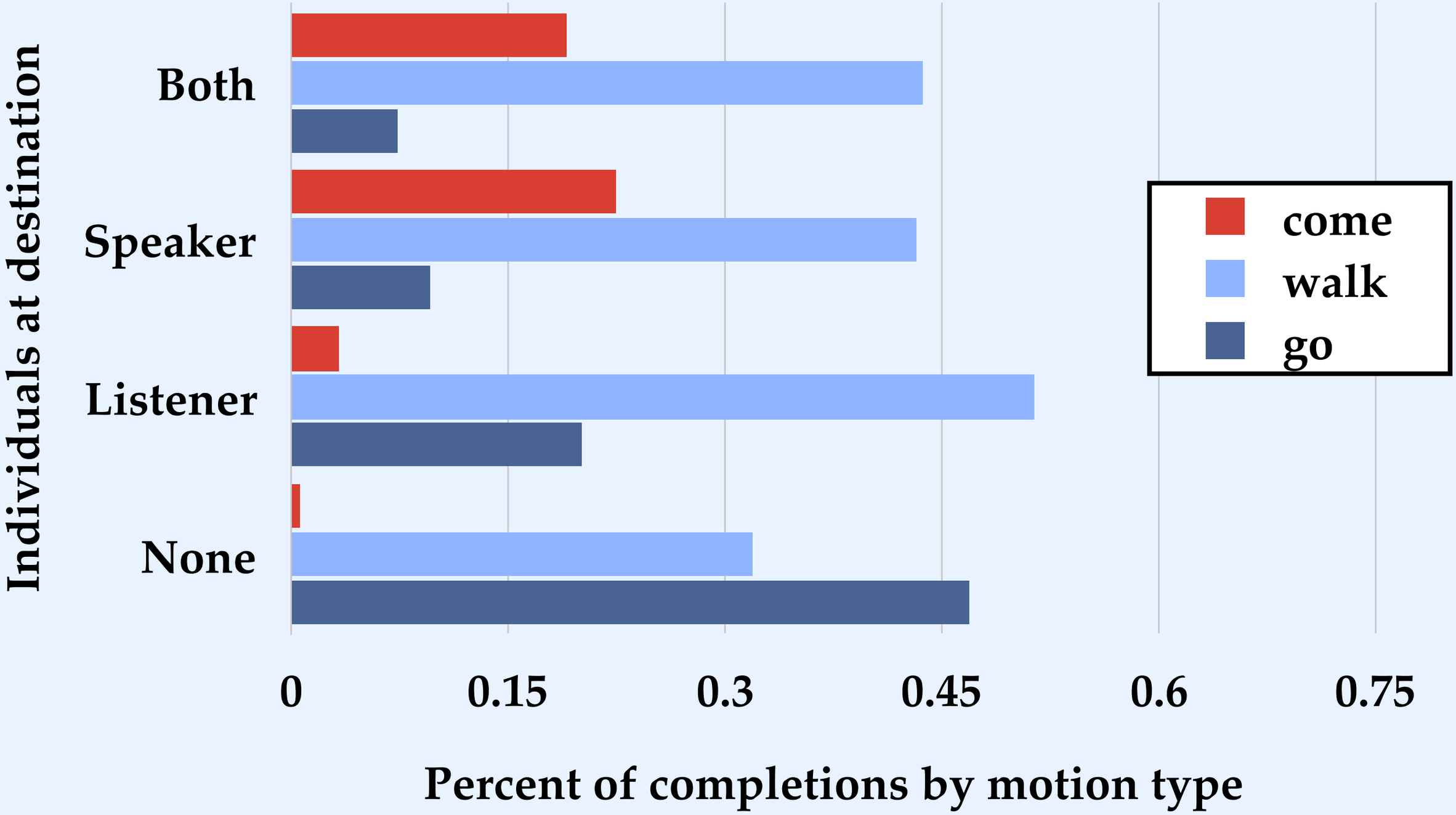


Anderson & Dillon (2019)

Production Experiment



Production results



Production Results: p(u | w)

Summary:

- Participants produced *come* more in the speaker condition than in the *both* condition
- Participants avoided both *come* and *go* in the listener condition: manner-of-motion completions were more common than in other conditions

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 have a strong bias towards their own perspective that isn't satisfactorily explained.

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

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

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

Using Bayesian inference based on a mental model of how the speaker selects utterances and perspectives!

How do speakers pick a perspective?

Less clear— maybe rationally (based on how listeners interpret sentences), maybe not.

How do contextual factors license or lead to perspective shift?

Other projects of mine explore this question!

Discourse Factors on Perspective Prominence

{One of the stage managers/Dave} purchased a sofa for the set of the new play.

A Subject: He got a 10% discount on it from a salesclerk.
B Subject: A salesclerk gave him a 10% discount on it.

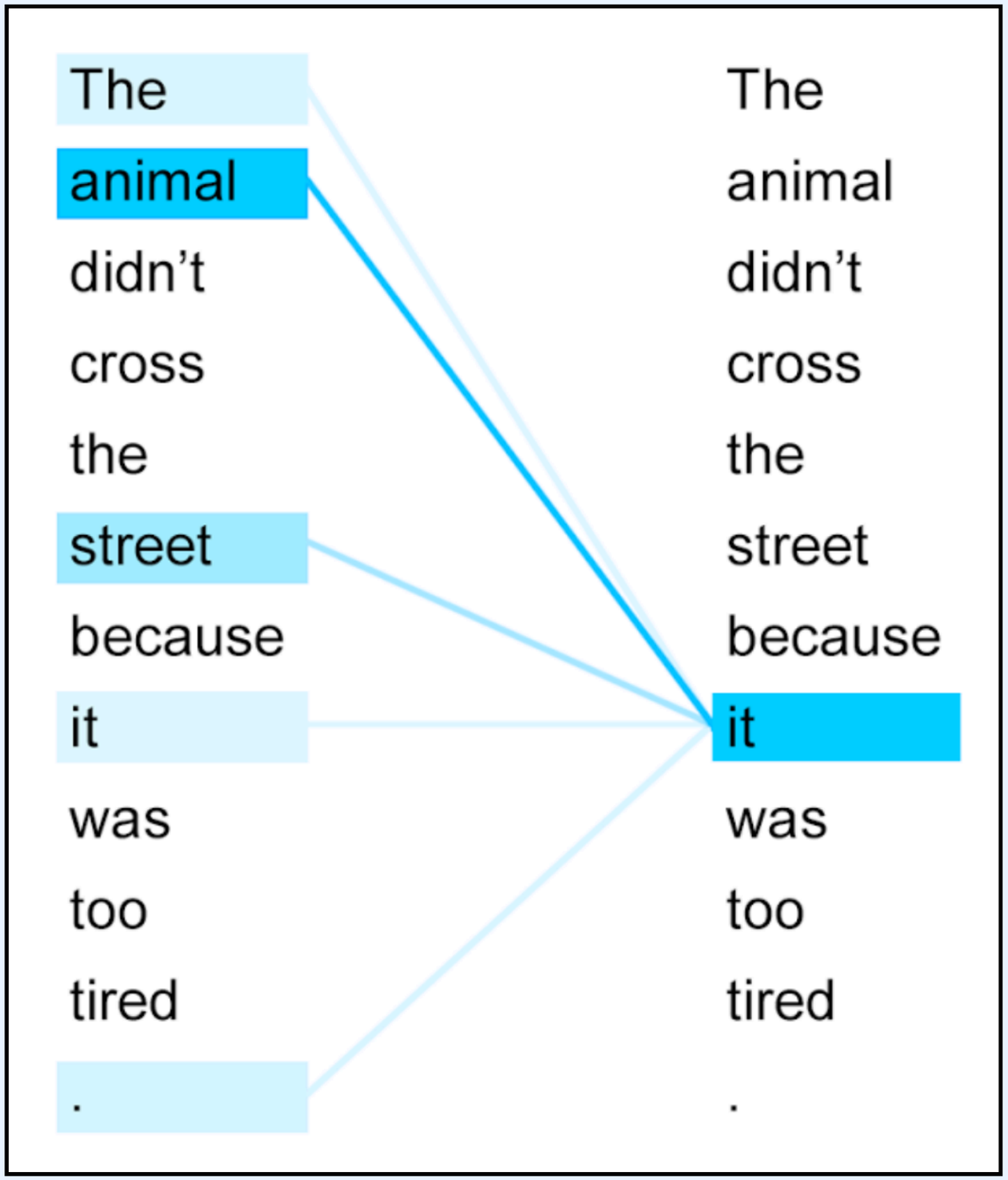
B perspective: When he to pick it up, he gave him free tickets for the show.
A perspective: When he _____ to deliver it, he got free tickets for the show.

Fill in the blank:

(a) came

(b) went

Context-Sensitive Neural Networks



- Models like BERT and GPT-2 are state-of-the-art because they are more **sensitive to context** than previous models.
- Many modern NLP applications are built on top of embeddings learned from bidirectional, contextual deep language models.
- Unlike humans, though, neural networks aren't **grounded** in a physical location.

ai.googleblog.com/2017/08/transformer-novel-neural-network.html

Google Search Meets Sesame Street

SEARCH

Understanding searches better than ever before

Pandu Nayak
Google Fellow and Vice President, Search

Published Oct 25, 2019

If there's one thing I've learned over the 15 years working on Google Search, it's that people's curiosity is endless. We see billions of searches every day, and 15 percent of those queries are ones we haven't seen before--so we've built ways to return results for queries we can't anticipate.

When people like you or I come to Search, we aren't always quite sure about the best way to formulate a query. We might not know the right words to use, or how to spell something, because often times, we come to Search looking to learn--we don't necessarily have the knowledge to begin with.

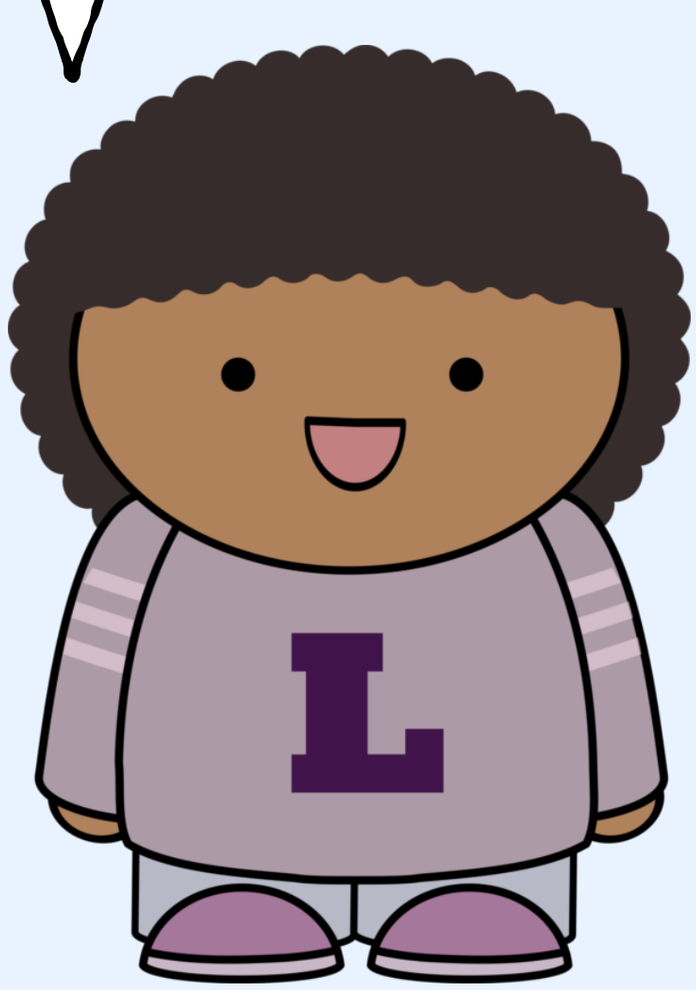
At its core, Search is about understanding language. It's our job to figure out what you're searching for and surface helpful information from the web, no matter how you spell or combine the words in your query. While we've continued to improve our language understanding capabilities over the years, we sometimes still don't quite get it right, particularly with complex or conversational queries. In fact, that's one of the reasons why people often use "keyword-ese," typing strings of words that they think we'll understand, but aren't actually how they'd naturally ask a question.

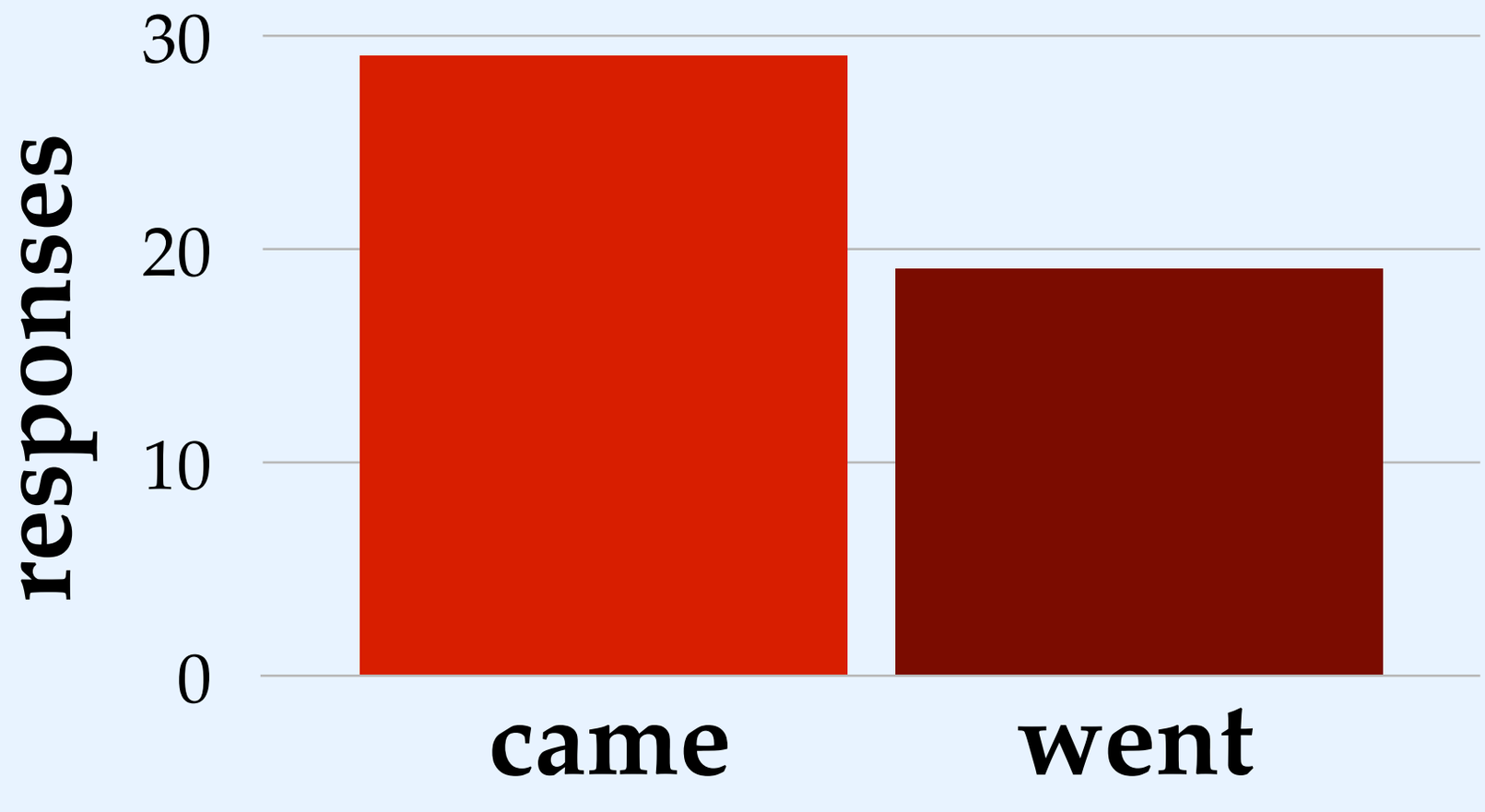
With the latest advancements from our research team in the science of language understanding--made possible by machine learning--we're making a significant improvement to how we understand queries, representing the biggest leap forward in the past five years, and one of the biggest leaps forward in the history of Search.

Applying BERT models to Search

Last year, we introduced and open-sourced a neural network-based technique for natural language processing (NLP) pre-training called Bidirectional Encoder Representations from Transformers, or as we call it--BERT, for short. This technology enables anyone to train their own state-of-the-art question answering system.

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!





Masked Language Modeling

Enter text with one or more "[MASK]" tokens and BERT will generate the most likely token to substitute for each "[MASK]".

Sentence:

Last weekend Jo [MASK] 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!

Mask 1 Predictions:

25.2% **came**
21.6% **went**
15.8% **drove**
13.5% **flew**
6.9% **traveled**

Masked Language Modeling

Enter text with one or more "[MASK]" tokens and BERT will generate the most likely token to substitute for each "[MASK]".

Sentence:

Last weekend Joe [MASK] 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!

Mask 1 Predictions:

28.3% **went**
18.5% **flew**
17.1% **came**
14.0% **drove**
8.8% **traveled**

Why does modeling perspective matter?

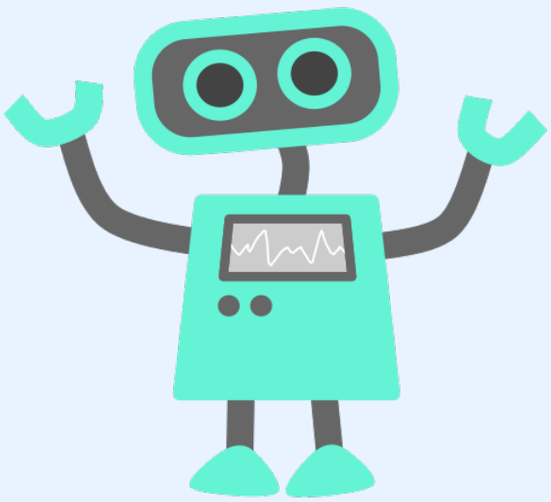
Commonsense / contextual reasoning from text

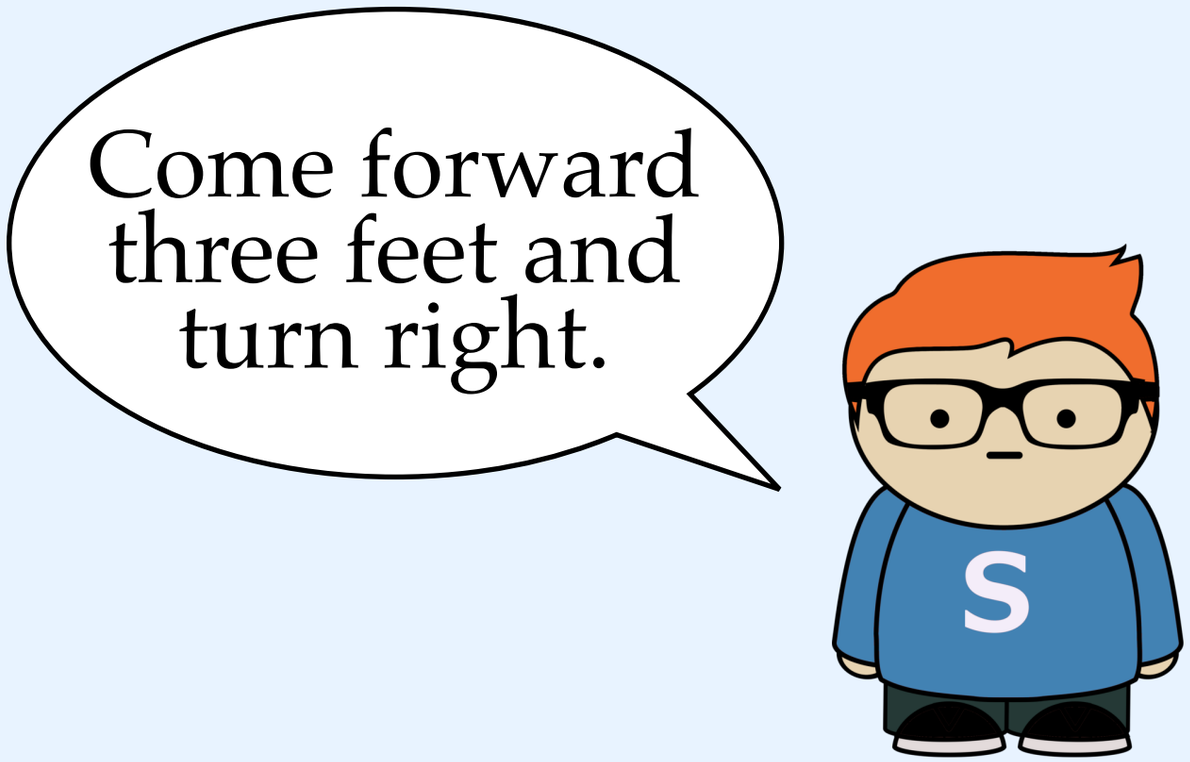
When I arrived at the cafe, it was empty.
*By the time Amanda came, ***MASK*** was really busy.*

Stance/sentiment detection

The professor, a jerk who nobody likes, assigned too much reading.

Human-robot interaction





Future Work

Bayesian modeling

- Conversational adaptation
- Social knowledge

Probing neural network understanding

- Perspective identification in stance detection
- Exemplar effects in proper name representations

Low-resource languages

- ASR for **very** low-resource languages
- Multilingual MT for related low-resource languages

Acknowledgements

Tessa Patapoutian
UMass Amherst

Brian Dillon
UMass Amherst

Rajesh Bhatt
UMass Amherst

Daniel Altshuler
Hampshire College

Mohit Iyyer
UMass Amherst

What do we know about perspective?

- Speakers sometimes fail to consider the listener's point-of-view (Epley et al. 2004, Lin & Epley 2010)
- Shifting to someone else's perspective is costly (Keysar & Barr 2003, Brown-Schmidt 2009, Köder et al. 2015, Long et al. 2018)
- The speaker is the most common perspective holder (Roberts 2012, Harris 2012).

Perspectival RSA model

Production study: $p(u \mid w) = \Sigma_a\, p(u,a \mid w)$

Comprehension study: $p(w \mid u) = \Sigma_a\, p(w,a \mid u)$

$$L_1(w, a \mid u) \propto S_0(u, a \mid w)\, p(w)$$

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:

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

Pragmatic listener:

$$
\begin{aligned}
L_1(w, a \mid u) &\propto p(u \mid w,a)\, p(w,a) \\
&\propto p(u, w, a) \\
&\propto p(u, a \mid w)\, p(w) \\
&\propto S_0(u, a \mid w)\, p(w)
\end{aligned}
$$

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

Lexical semantics

Two analyses of *go*:

- Its lexical semantics are not perspectival, but it is in pragmatic competition with *come* (Sudo 2018)

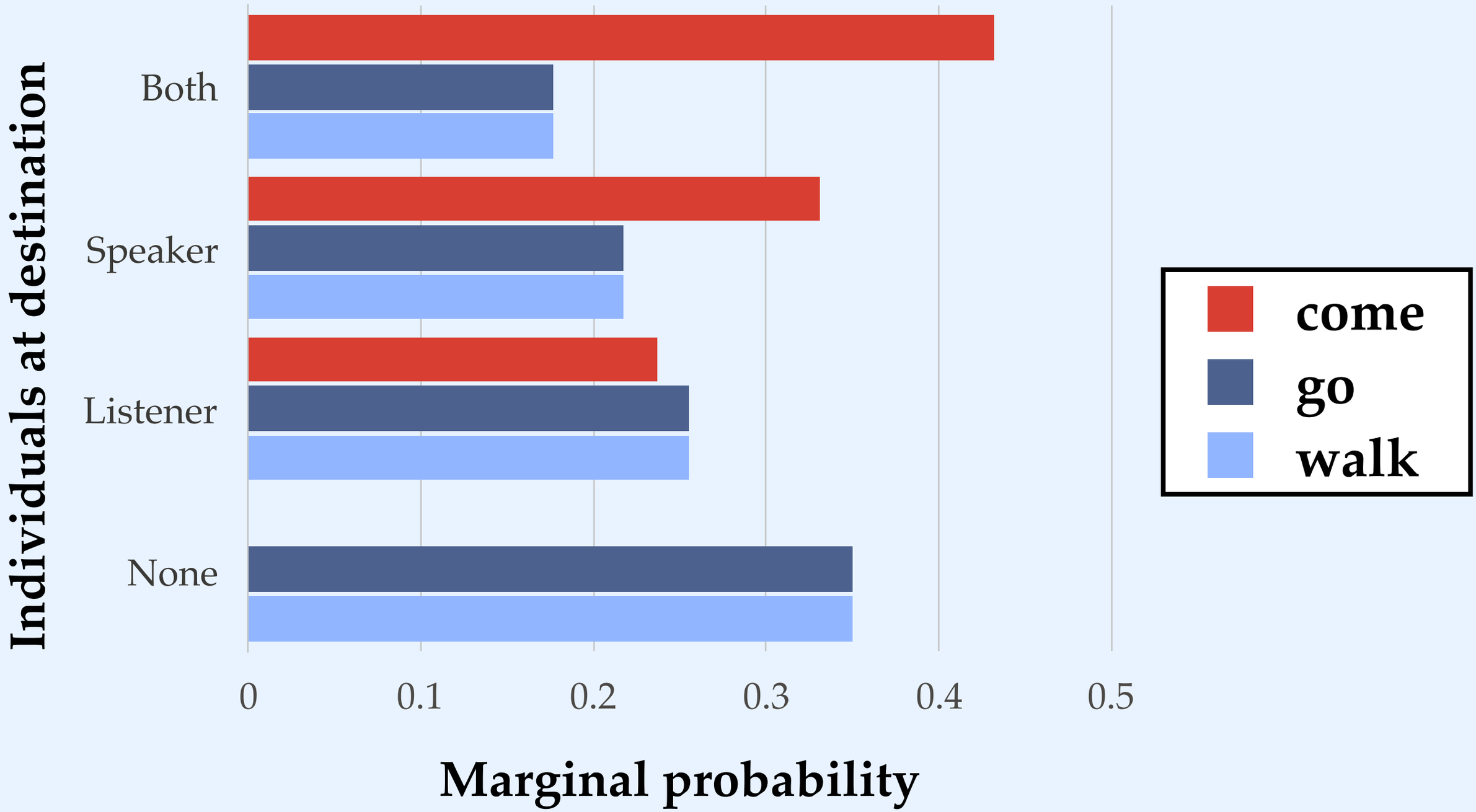
  prag-go: 1 if the subject is in motion.

- Its lexical semantics are anti-perspectival (Oshima 2008)

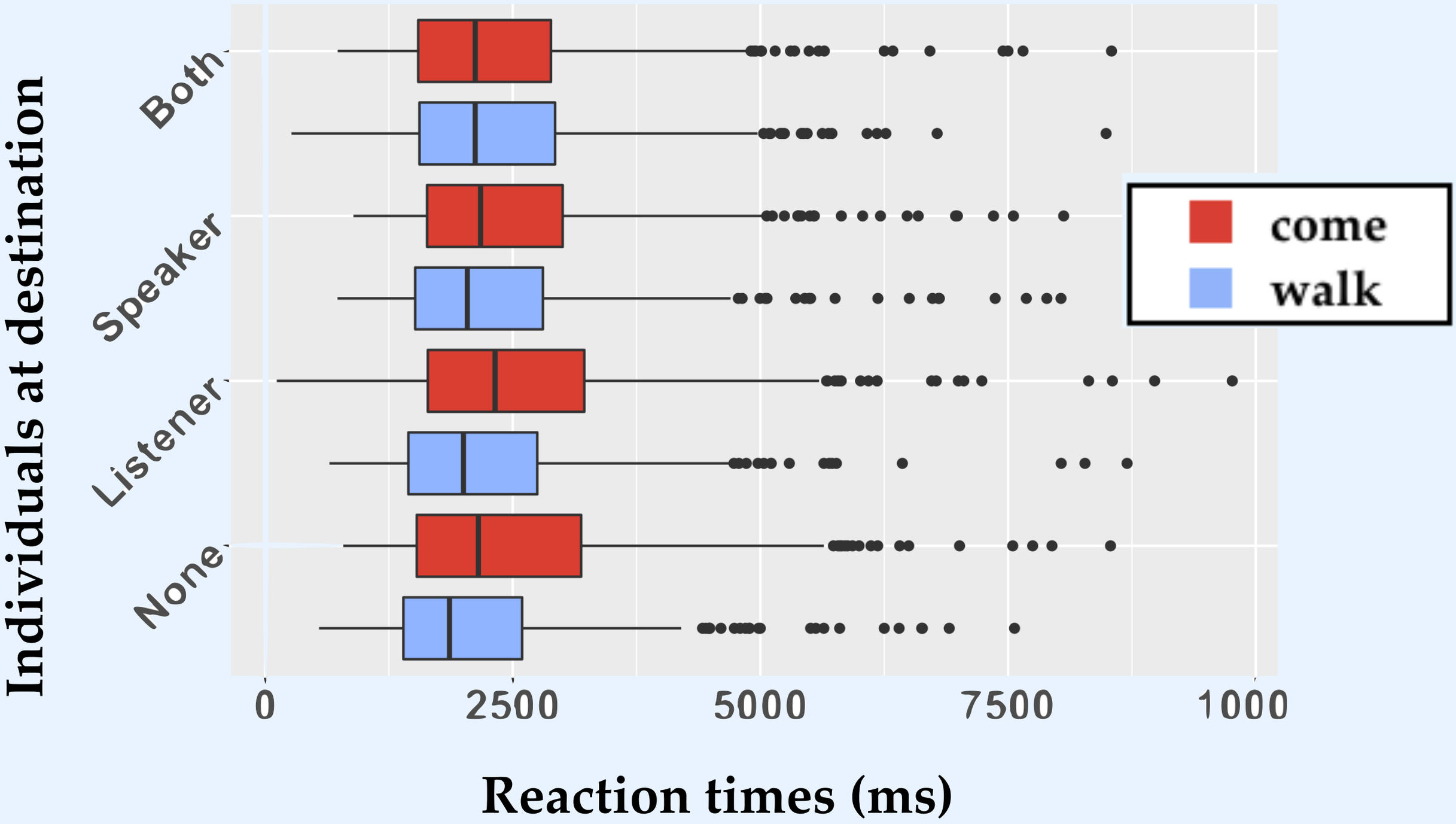
  lex-go: 1 if the subject is in motion and the perspective holder is not at the destination of motion.

Pragmatic Listener: marginal probabilities

p(w | *Thelma is* ___ *to Northampton*) with cost=0.5 & *prag-go*:



Reaction time results



Reaction time results

Condition	*walk* Cosineau-corrected mean (95% CI)	*come* Cosineau-corrected mean (95% CI)	Difference
both	2385 (+/- 87.5)	2366 (+/- 87.0)	-19
speaker	2261 (+/- 85.0)	2627 (+/- 105.2)	339
listener	2129 (+/- 79.5)	2537 (+/- 102.9)	408
none	2331 (+/- 92.5)	2469 (+/-91.3)	138