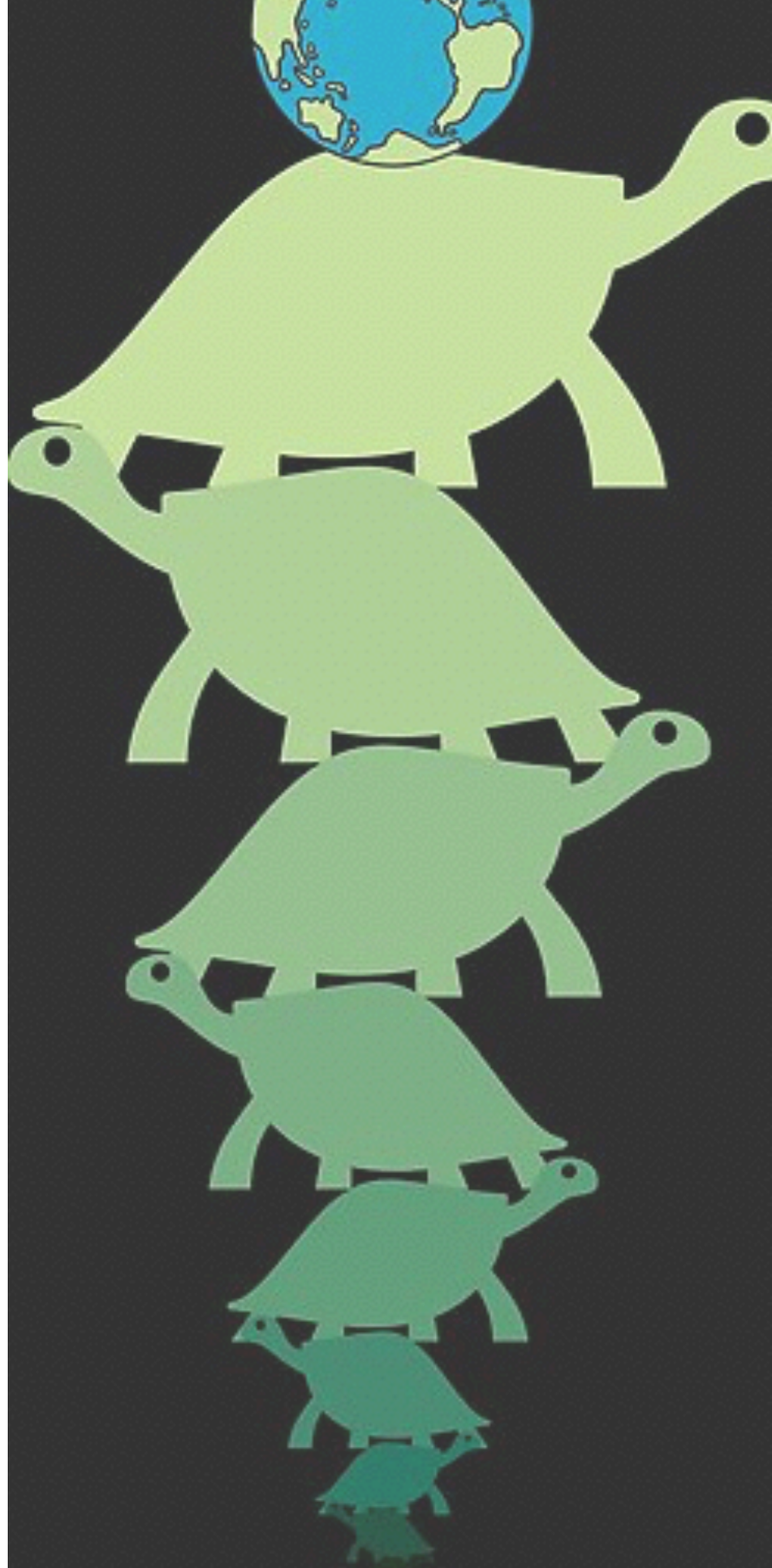
Ambiguity all the way down:

Inferring intentions from the acoustic signal

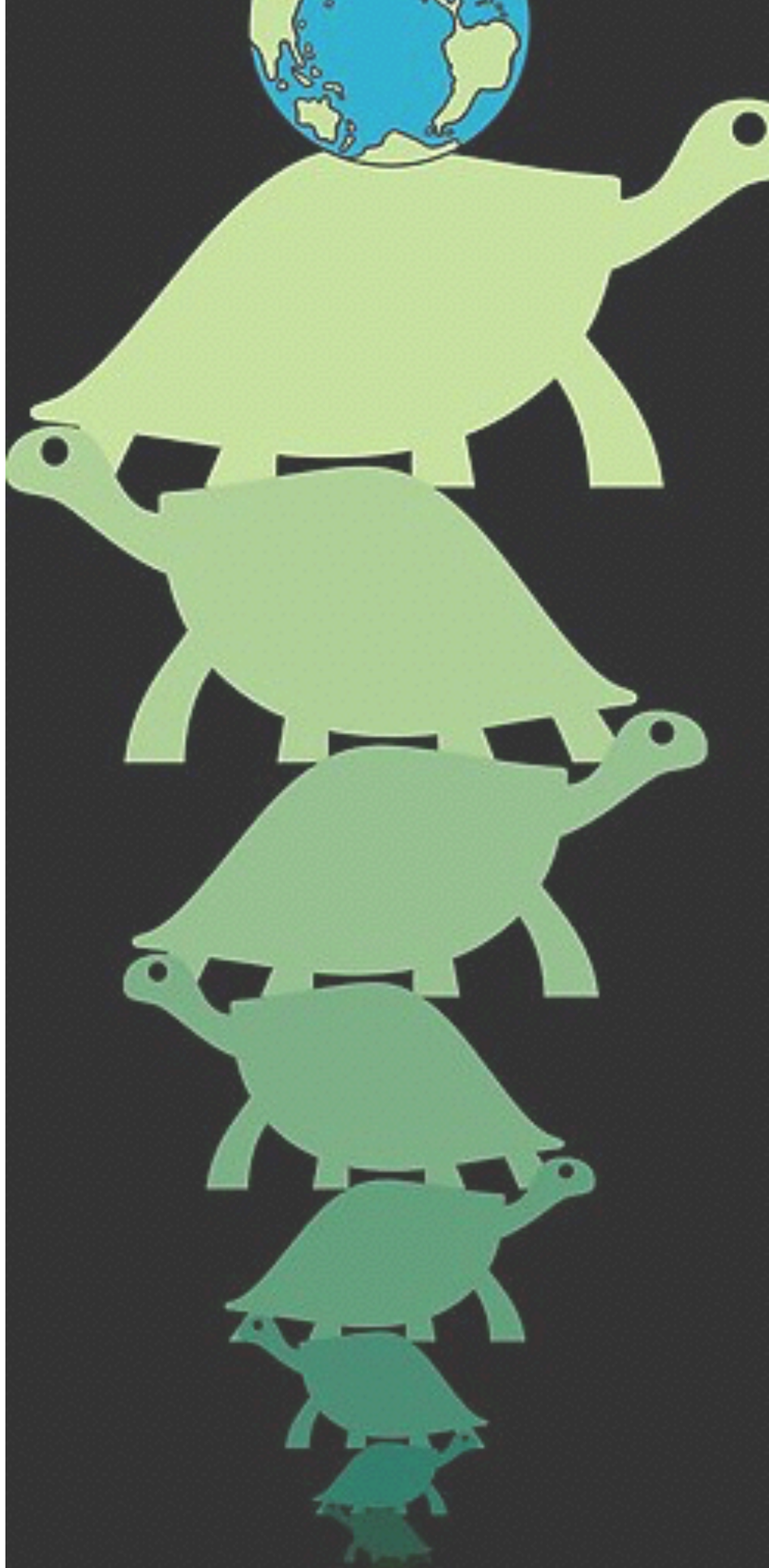
Chigusa Kurumada
University of Rochester
ckuruma2@ur.rochester.edu



Milliseconds

why do we care?

Chigusa Kurumada
University of Rochester
ckuruma2@ur.rochester.edu



German vs. Japanese

Rolf

Chigusa

Would you like some tea?

Thanks. I'm okay.

I'm not asking if you are okay or not.

Do you want tea? Or not.

… No, thanks.

message component

⇩

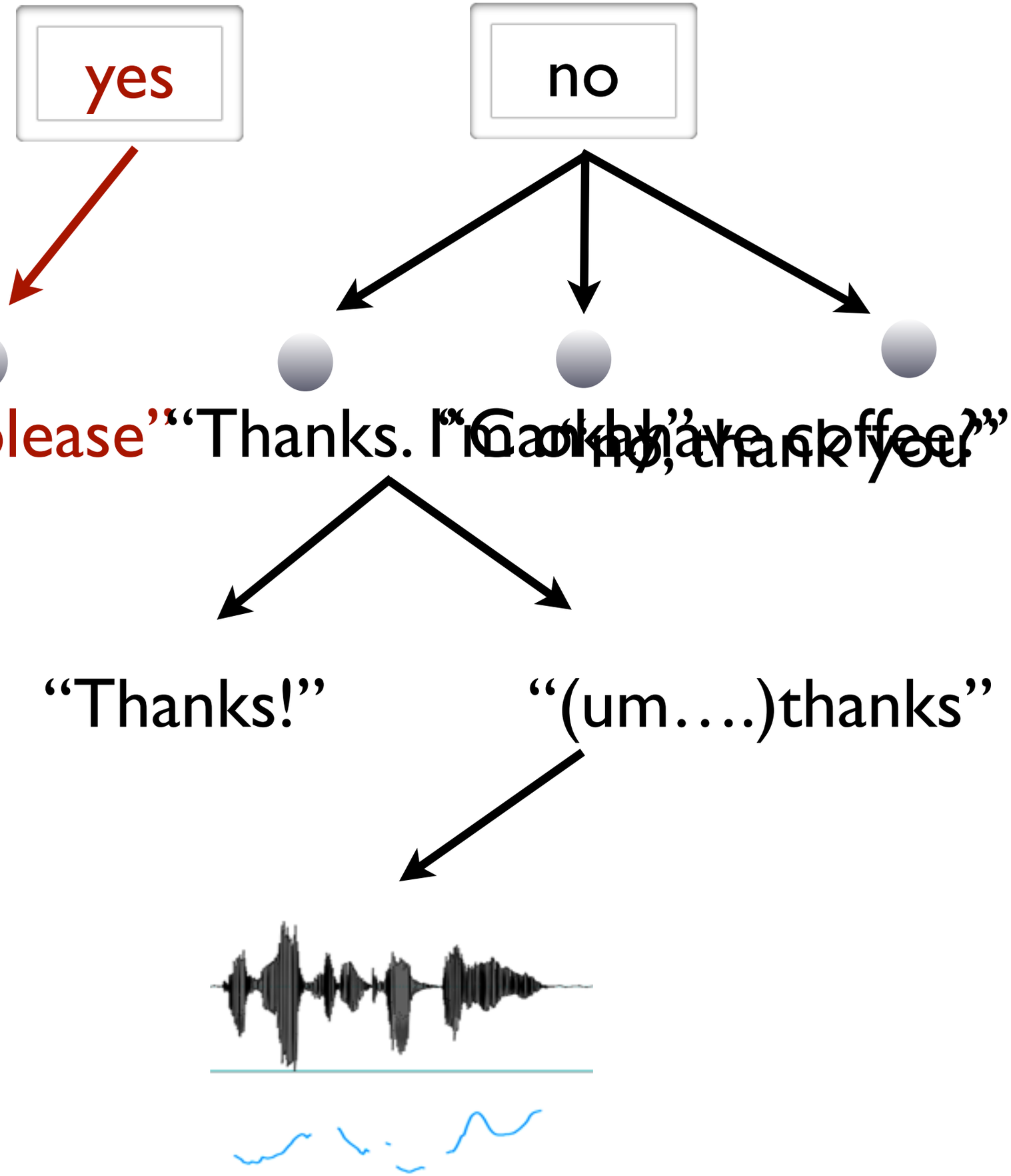
lexical grammatical component

⇩

phonological component

⇩

output system



Bock & Levelt (1994)

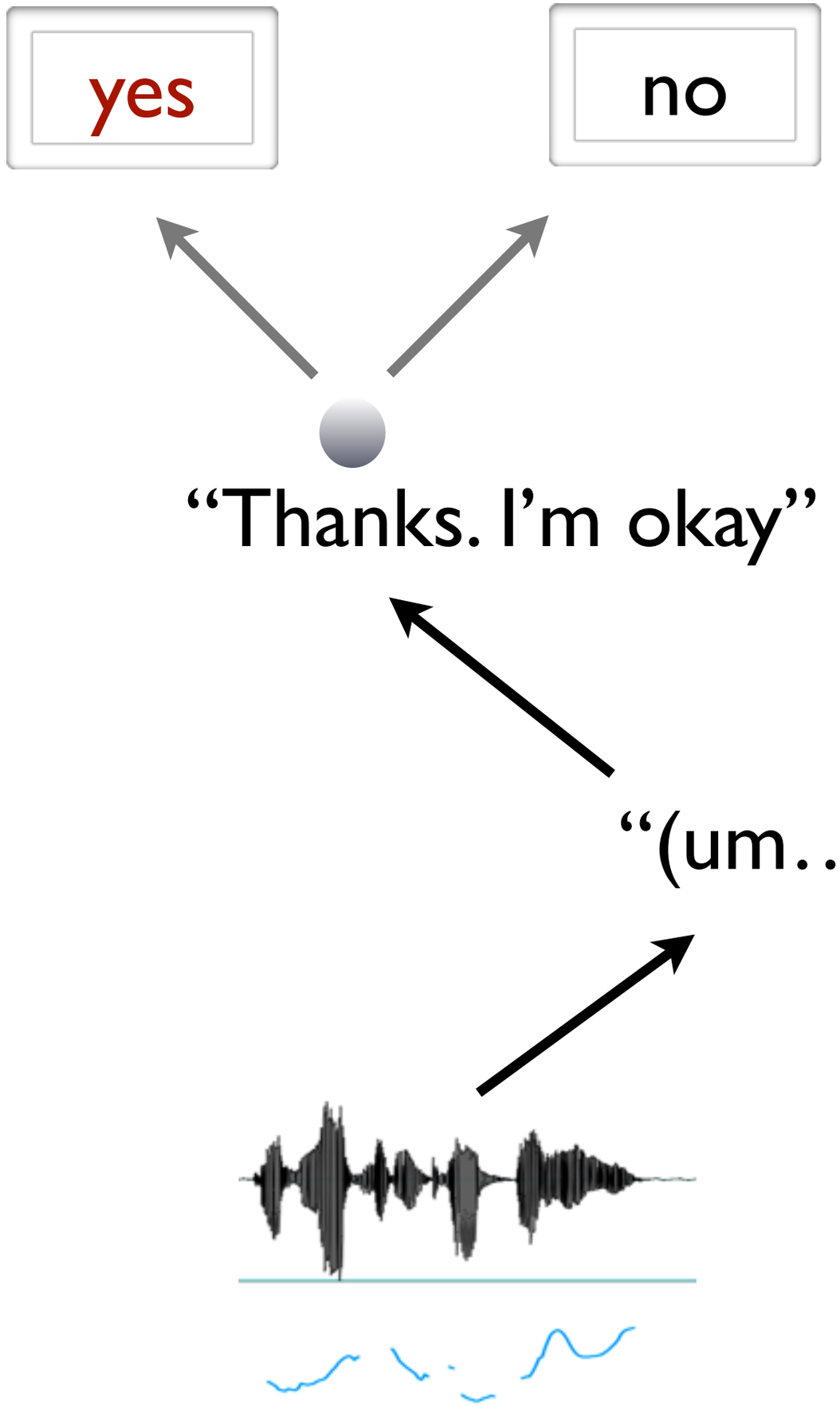
message component

lexical grammatical component

phonological component

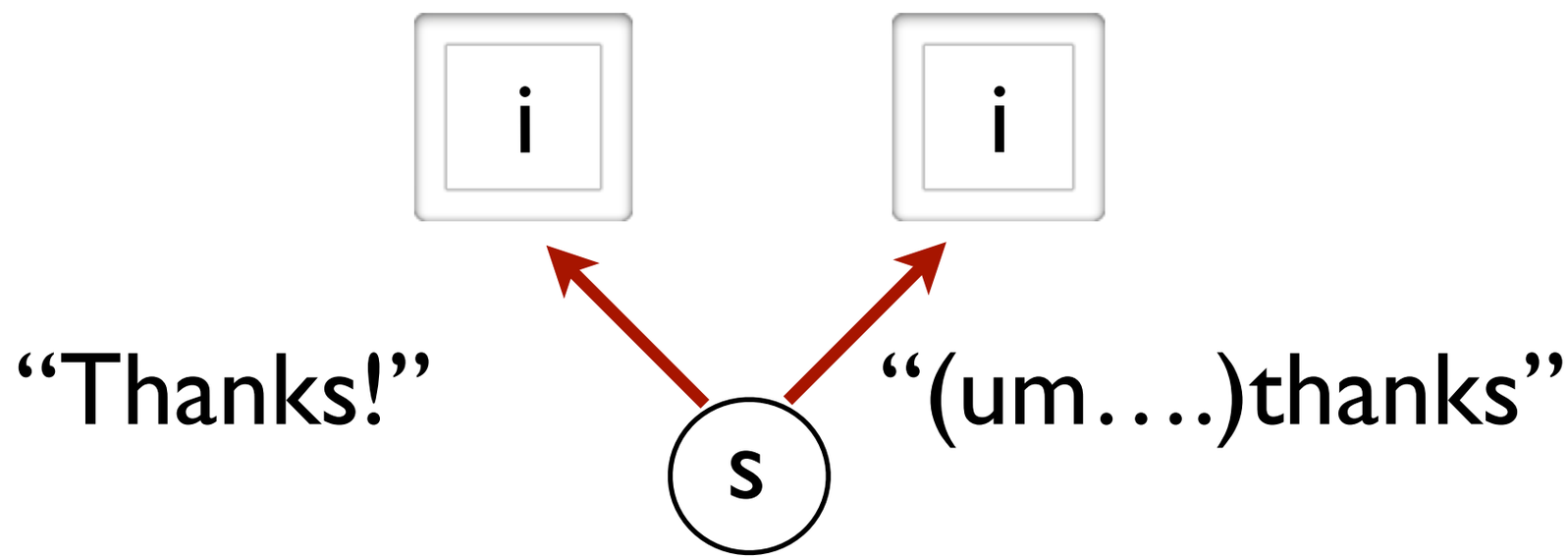
output system

e.g., Pierrehumbert & Hirschberg (1990)



Inferring intentions from the signal

- Listeners use the signal ("s") as evidence to inferentially arrive at the speaker's intention ("i")
- Ambiguity provides us with a lens through which to study the inference



Today

1. Ambiguity in language comprehension
2. Speaker-dependent inference over Question vs. Statement intonation contours
3. Real-time ambiguity resolution in understanding the speaker's communicative intentions
4. Discussion: What does the investigation of language comprehension tell us about ambiguity?

Variability and ambiguity

phoneme

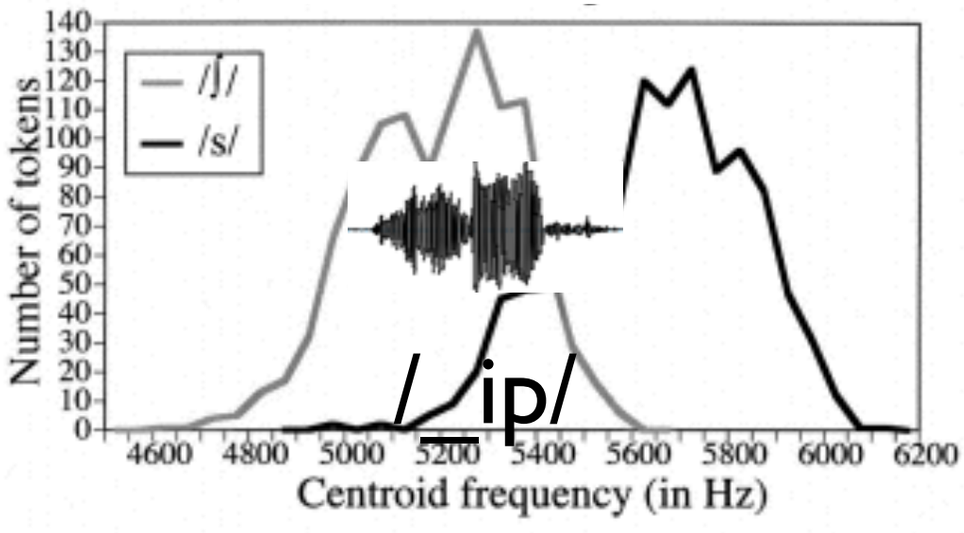
↑

acoustic signal

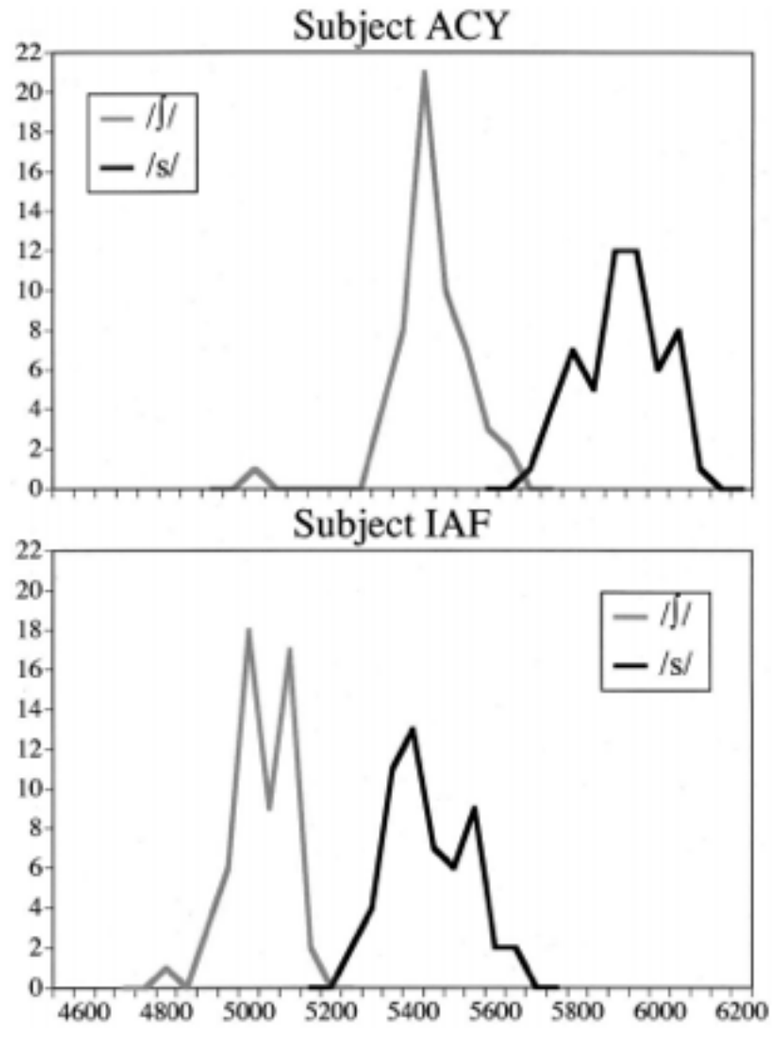
/sh/

/s/

20 speakers



individual speakers



Newman et al. 2001

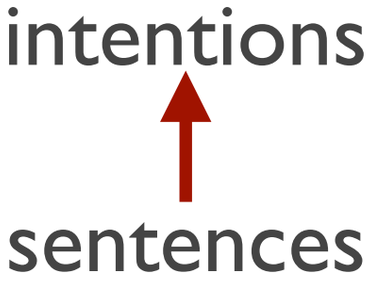
Variability and ambiguity

meanings

↑

words

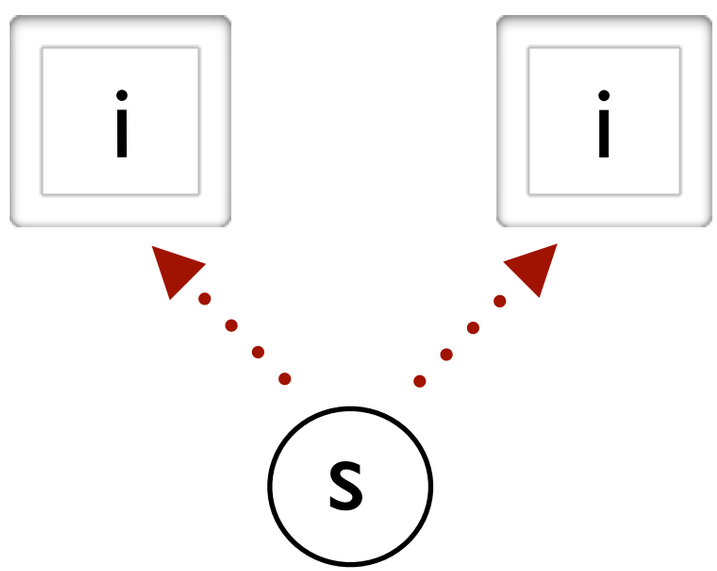
Variability and ambiguity



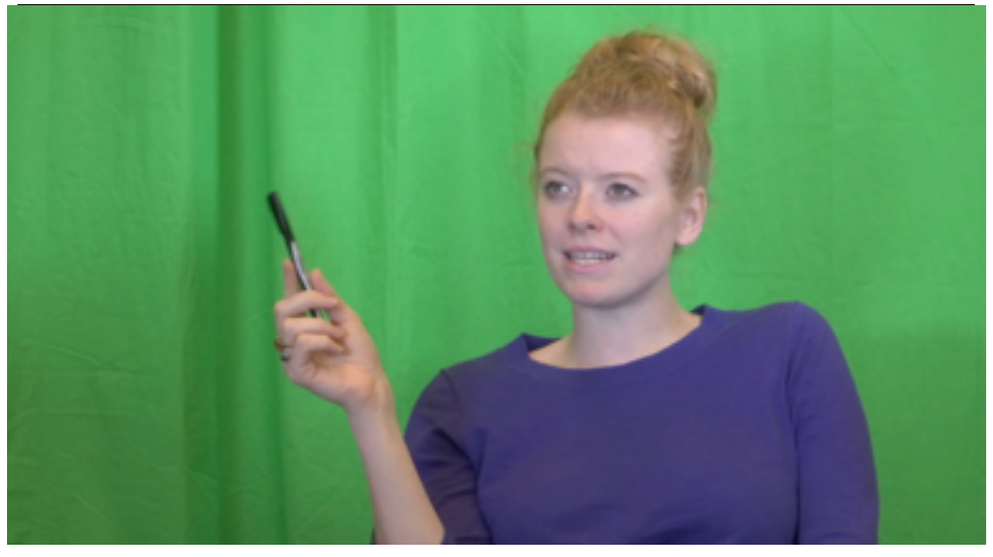
“I saw a boy with a binocular.”

Natural language is ambiguity-ridden

- The same physical input can support multiple hypotheses (e.g., sounds, words, intentions).
  - Different speakers use language differently
- The human brain is resolving the ambiguity at the rate of 2.5 words (4-6 syllables) / second

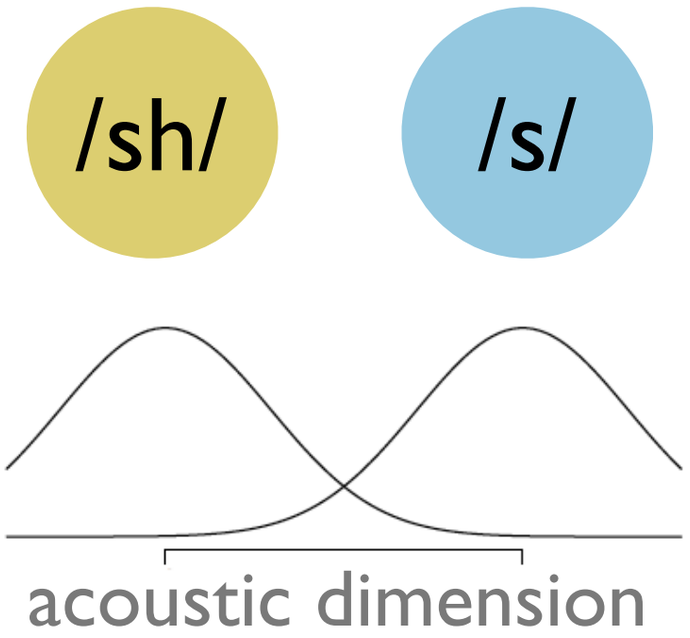


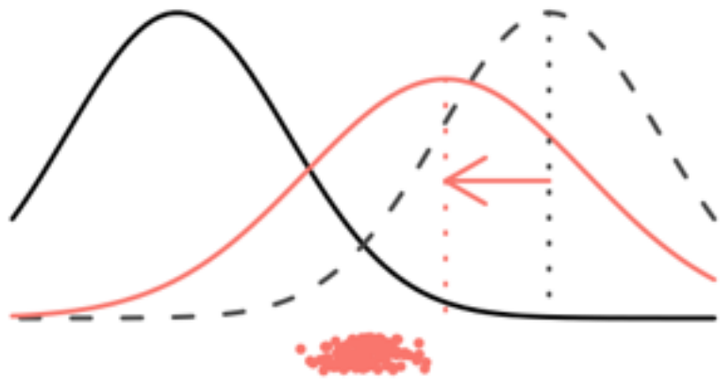
Speaker-dependent adaptation: sounds



(Liu & Jaeger 2018)

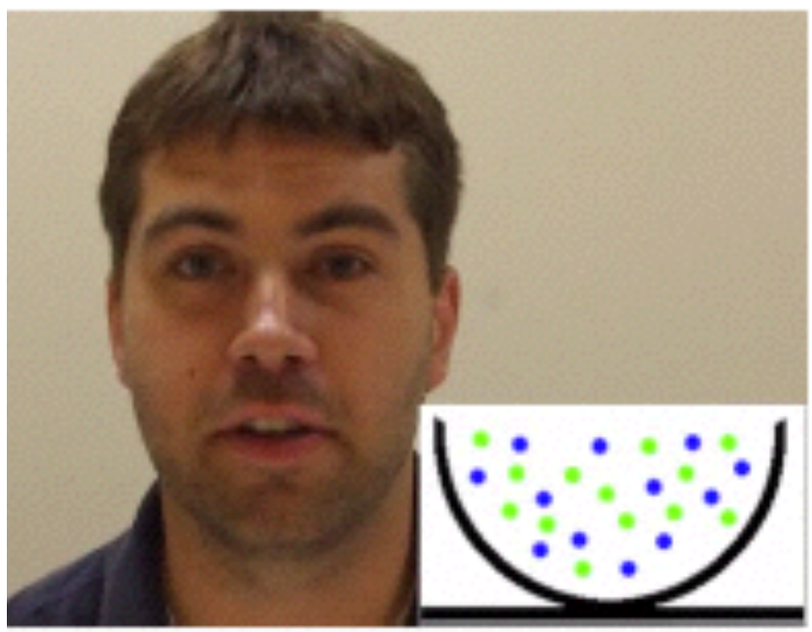
“medicine”





(e.g.,, Perceptual learning in phoneme categorization: Norris et al., 2003;2016; Vroomen et al.,2004, 2007; Kraljic & Samuel, 2007; Kleinschmidt & Jaeger, 2011; 2015)

Speaker-dependent adaptation: meaning

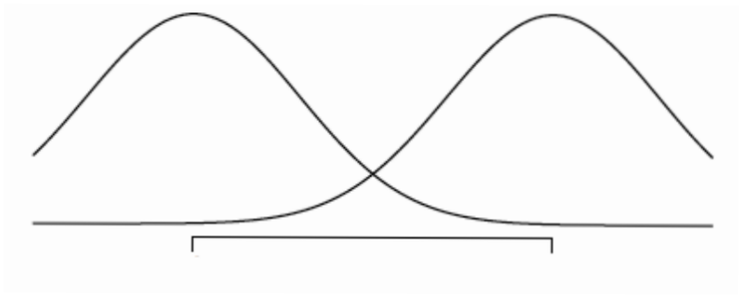


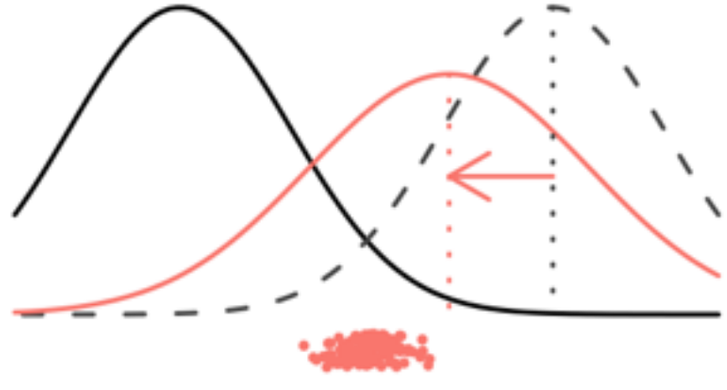
"many of the dots are blue"

NEXT

meaning1

meaning2

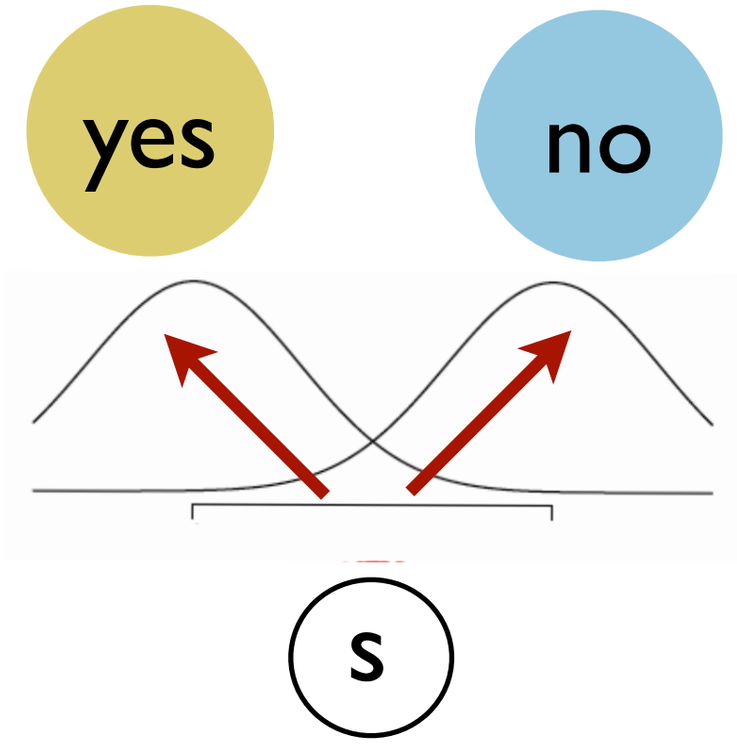




Yildirim, Degen, Tanenhaus, & Jaeger (2016)

Our hypothesis

- Mappings between intentions and acoustic realizations of speech can be similarly probabilistic
- Overtime listeners updating their assumptions about p(linguistic signal | intention, speaker)

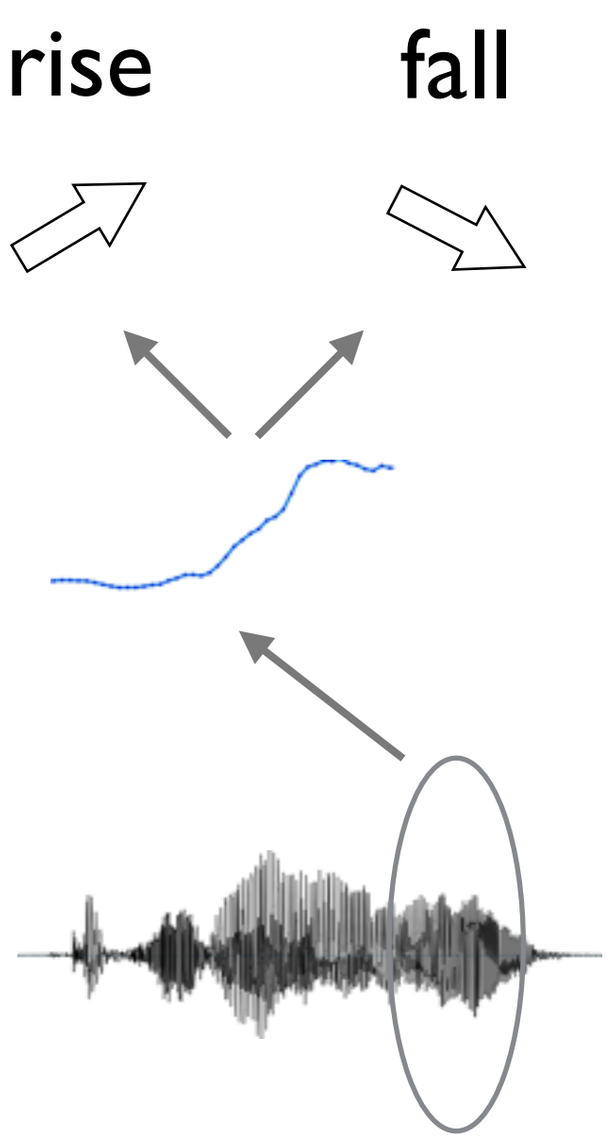


2

Speaker-dependent interpretations of English intonation contours

with Andrés Buxó-Lugo

Intonation interpretation



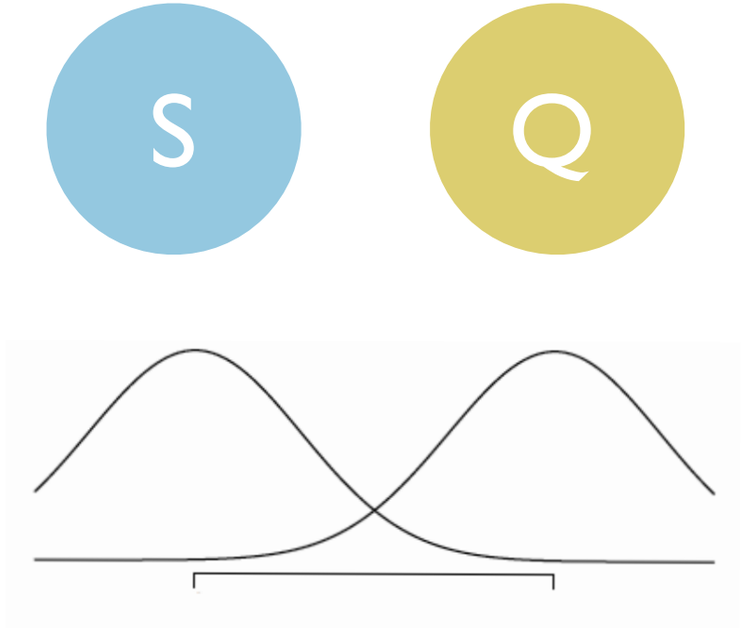
- Intonation is a powerful means to convey intentions.
  e.g., Rising and falling intonations are generally mapped onto questions and statements

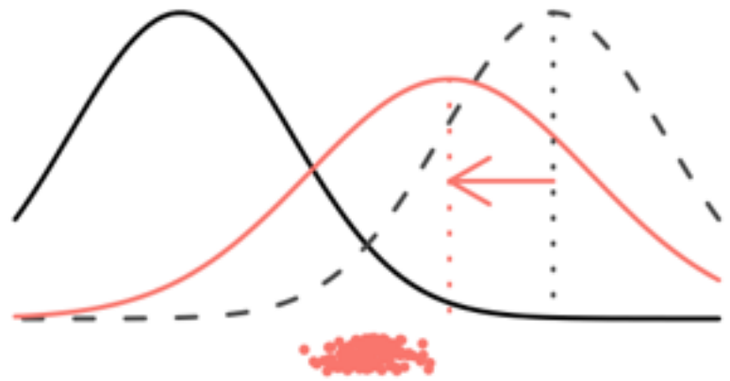
- Variability - ambiguity
  - e.g., American English vs. British English
    Adults vs. Children
    Animated vs. quiet speakers

(e.g., Bolinger, 1986; Breen et al,. 2012; Cutler, 1977; Dahan, 2015; Ladd, 1983; Watson, Gunlogson, & Tanenhaus, 2006; 2008; Ito & Speer, 2008; Pierrehumbert & Hirschberg, 1990)

Speaker-dependent adaptation

- Do listeners update their assumptions about p(intonation | intention = question, speaker) ?

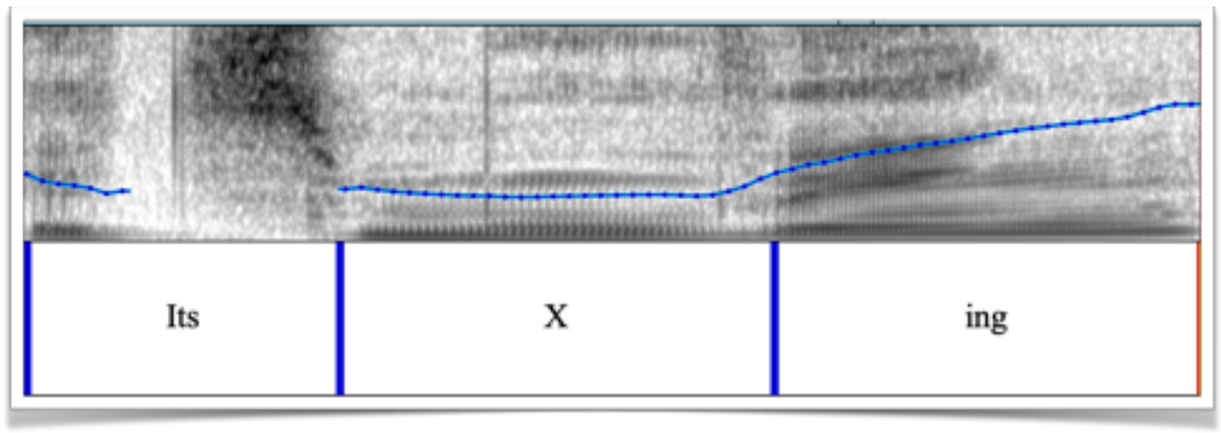




(Kurumada, Brown, & Tanenhaus, 2017; Buxó-Lugo & Kurumada, in preparation)

Study 1-a: Production

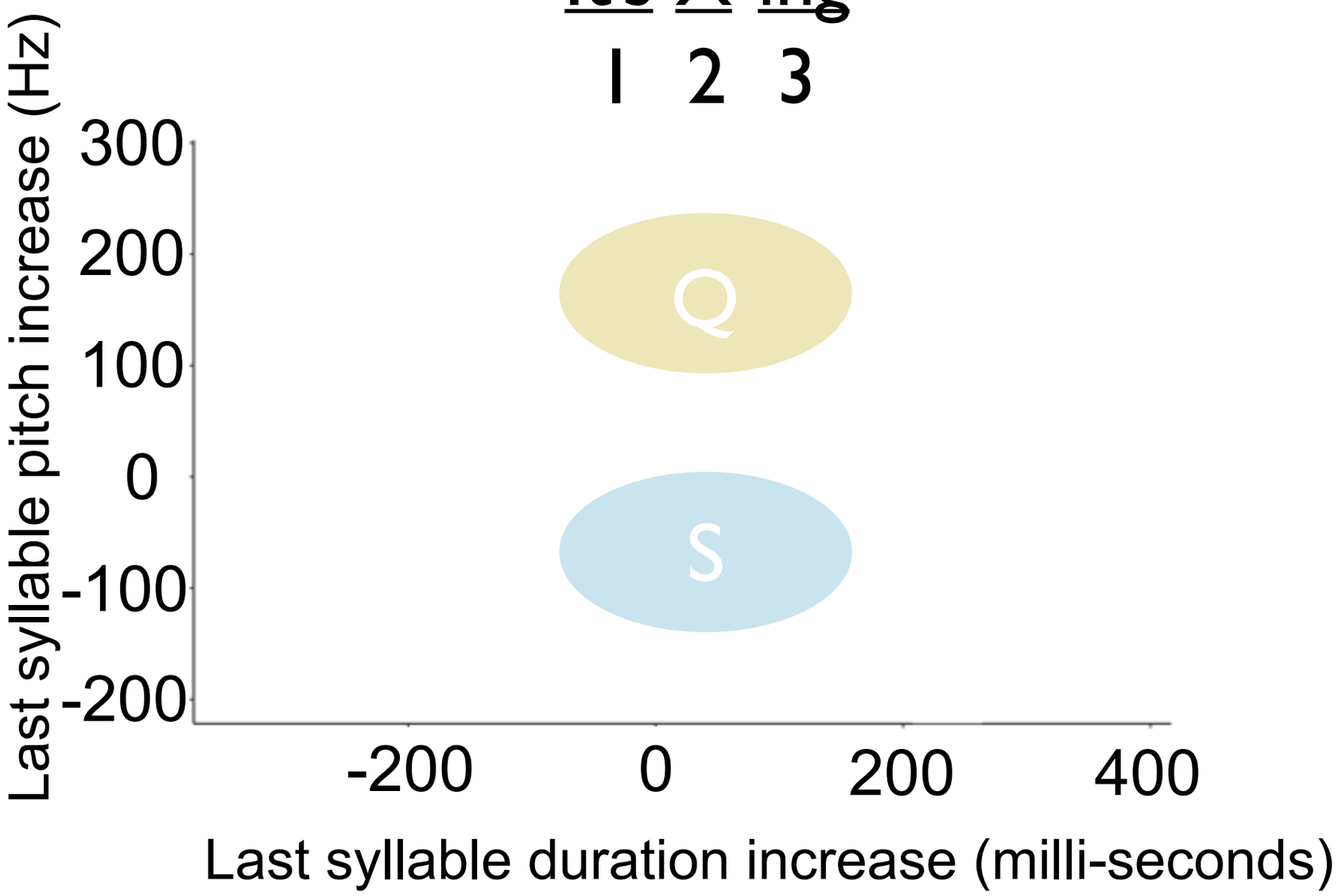
- "It's X-ing" (e.g., It's raining, It's raining?)



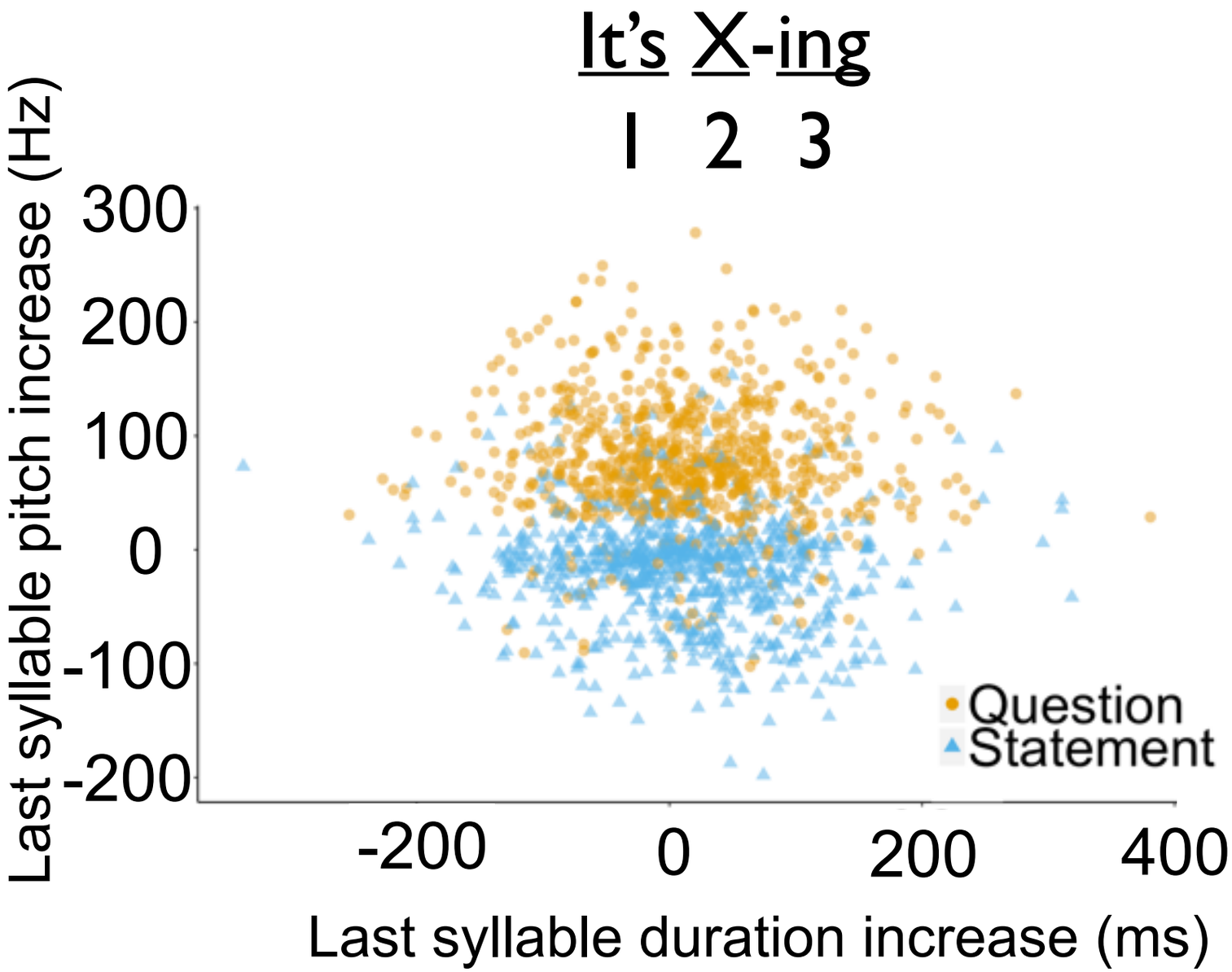
- Q: How much variability is there in the input?
  - 33 subjects
  - 24 questions and 24 statements

(Buxó-Lugo & Kurumada, in preparation)

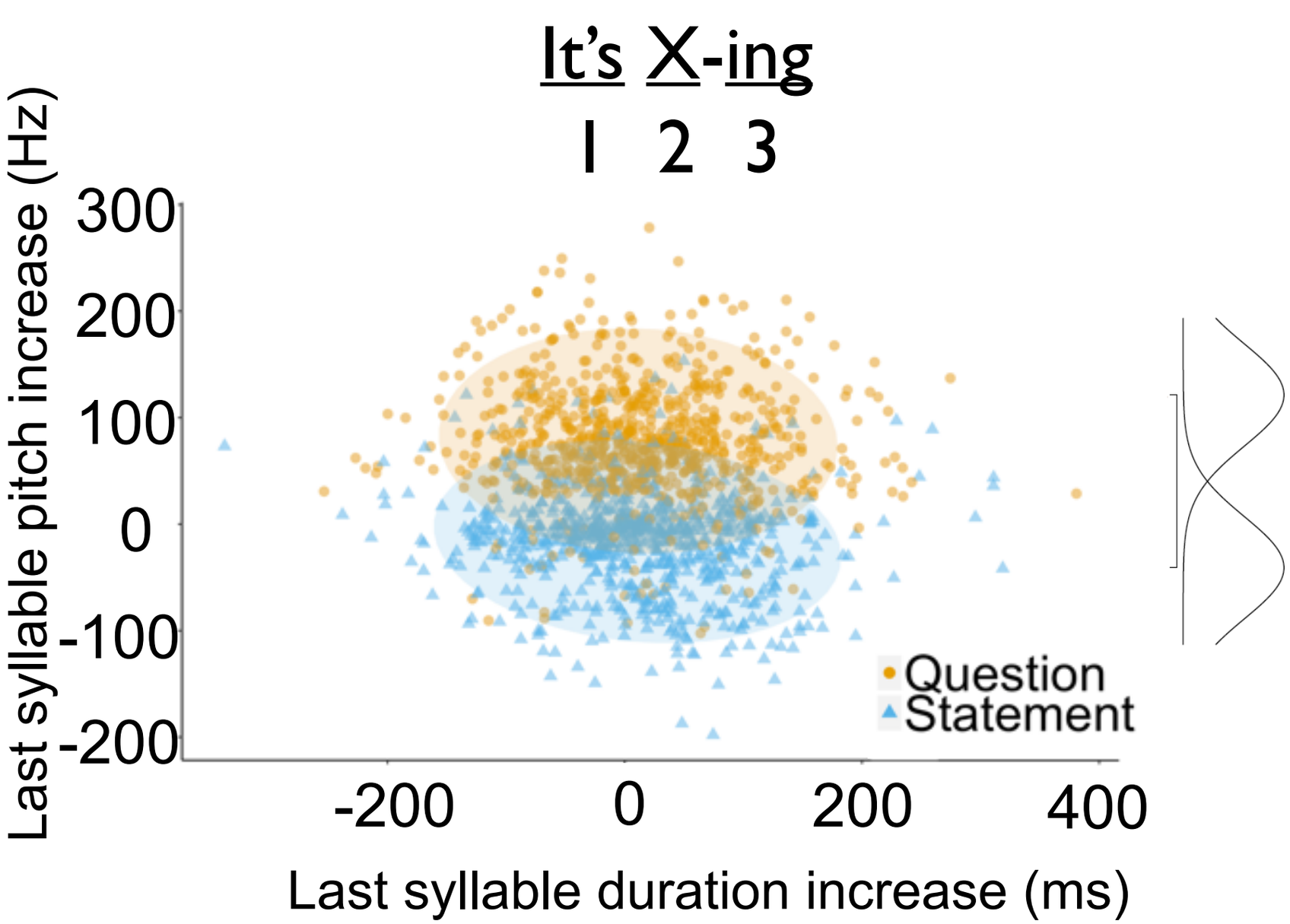
Study 1-a: Results



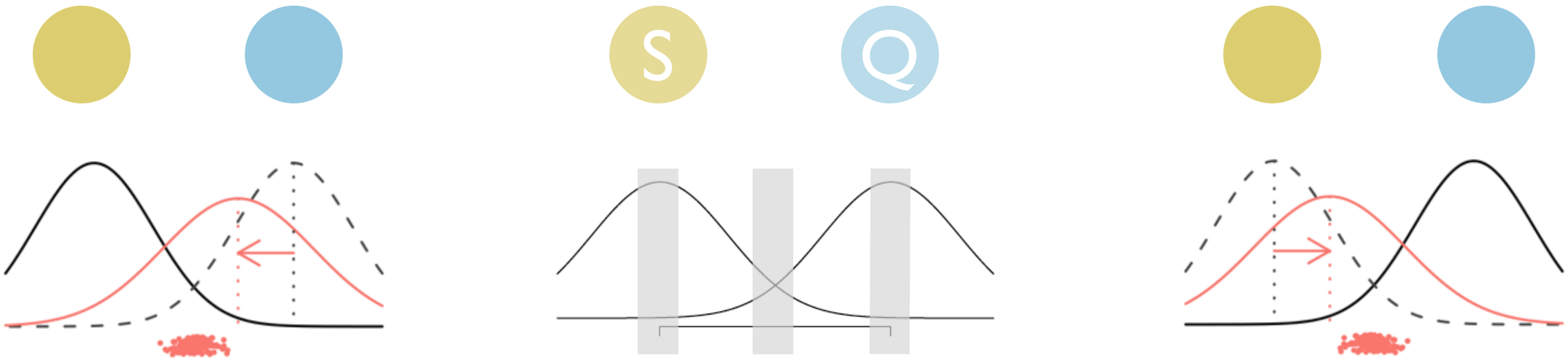
Study 1-a: Results



Study 1-a: Results



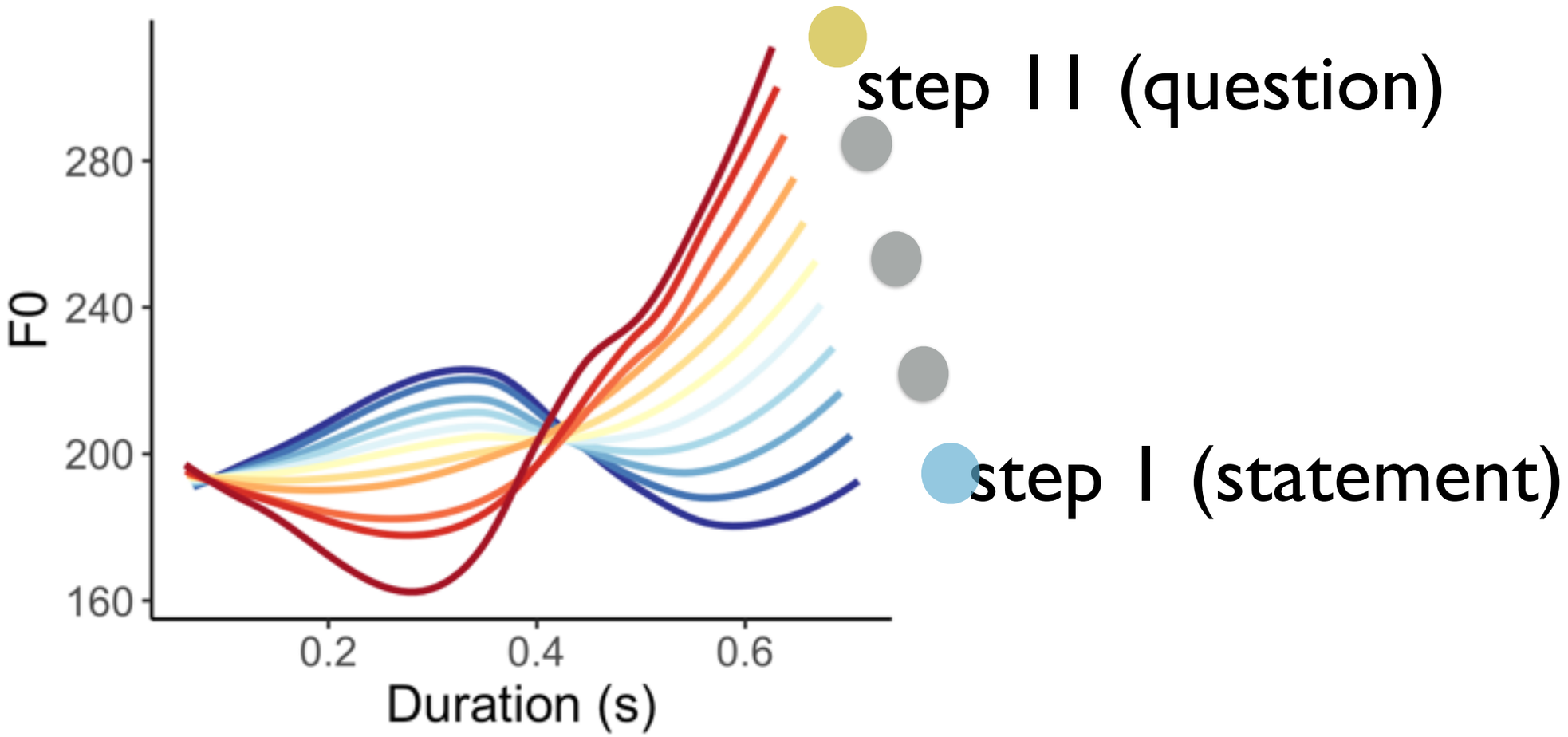
Adaptation to speaker's intonations?



- prediction: depending on the patterns of production by a given speaker, ambiguous tokens receive opposing interpretations

Study 1-b: Stimuli

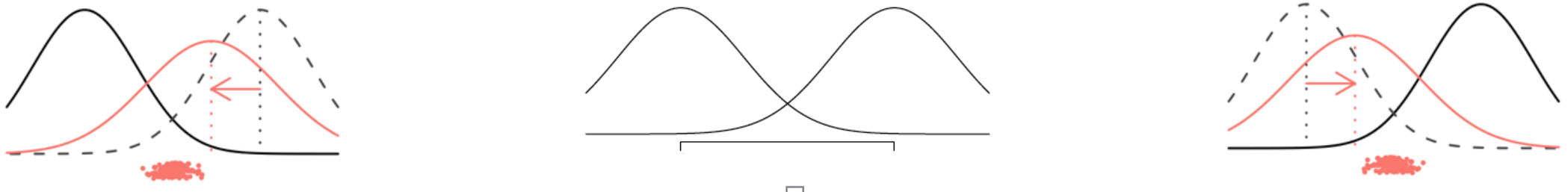
"It's moving"



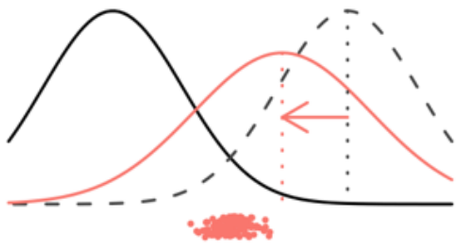
Study 1-b: Design (n=180)

Pre-exposure (24 trials)
"It's cooking" sampled from Steps 1-11
"Is this a question or a statement?"

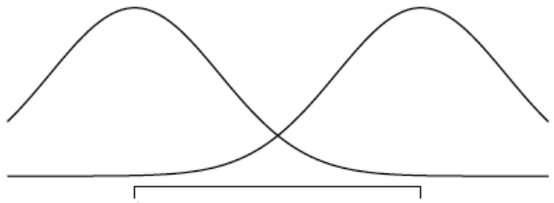
Exposure (30 trials) with feedback
3 between subject conditions



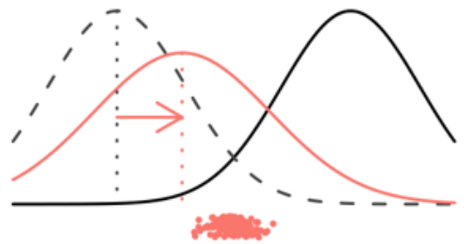
Post-exposure (24 trials) : identical to the pre-exposure phase



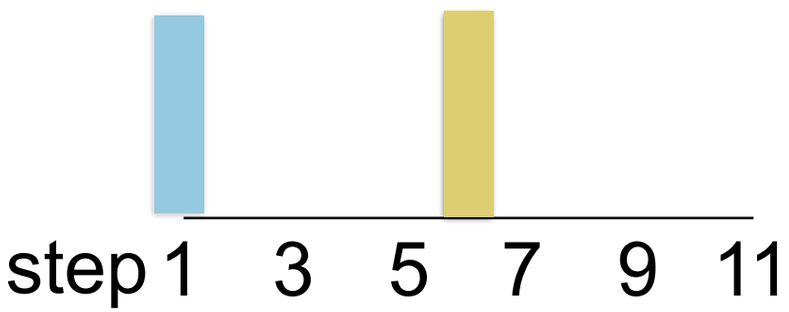
Question-biasing

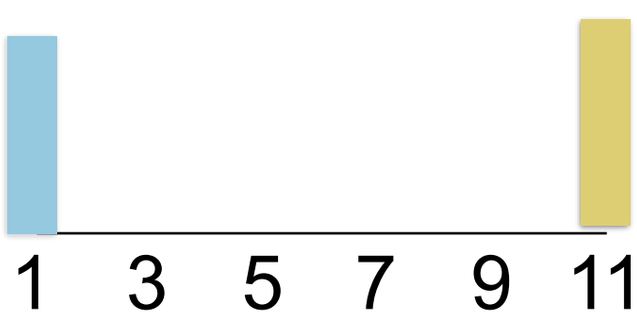


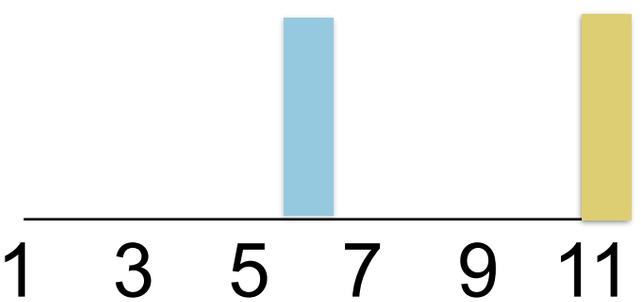
Non-ambiguous

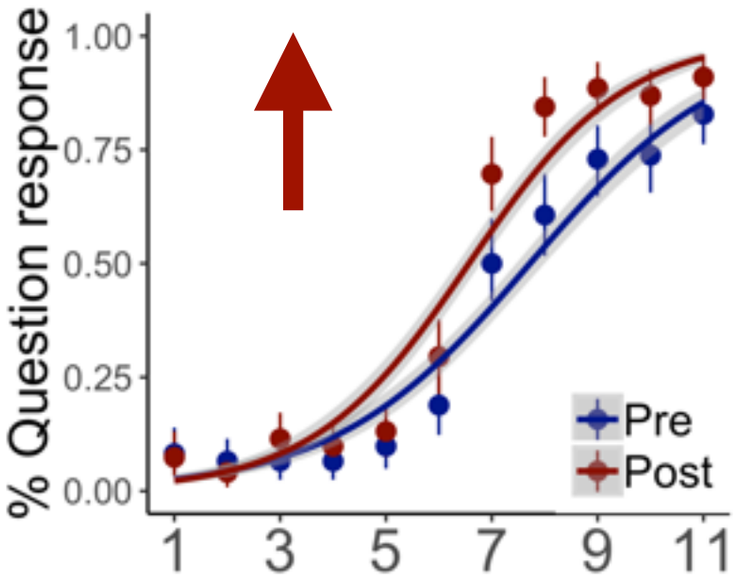


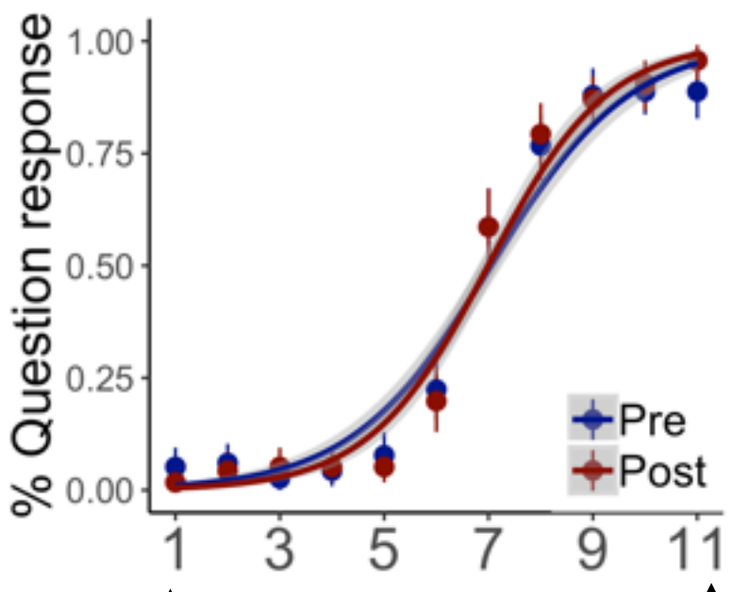
Statement-biasing

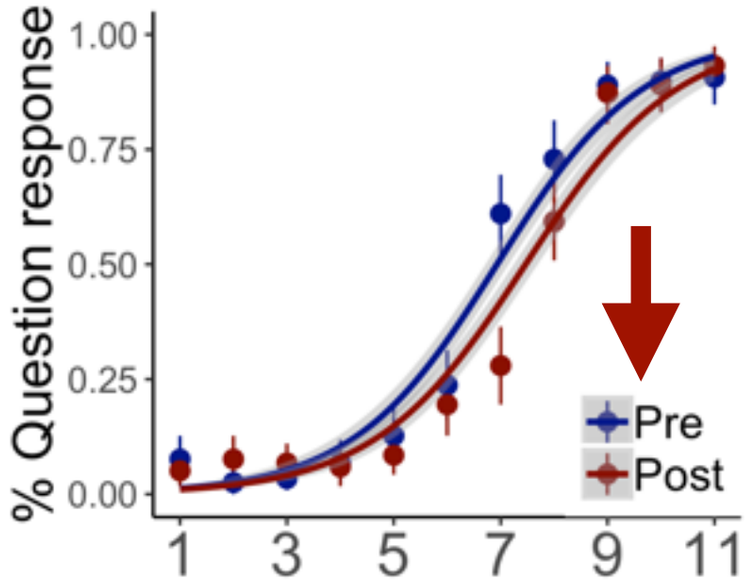








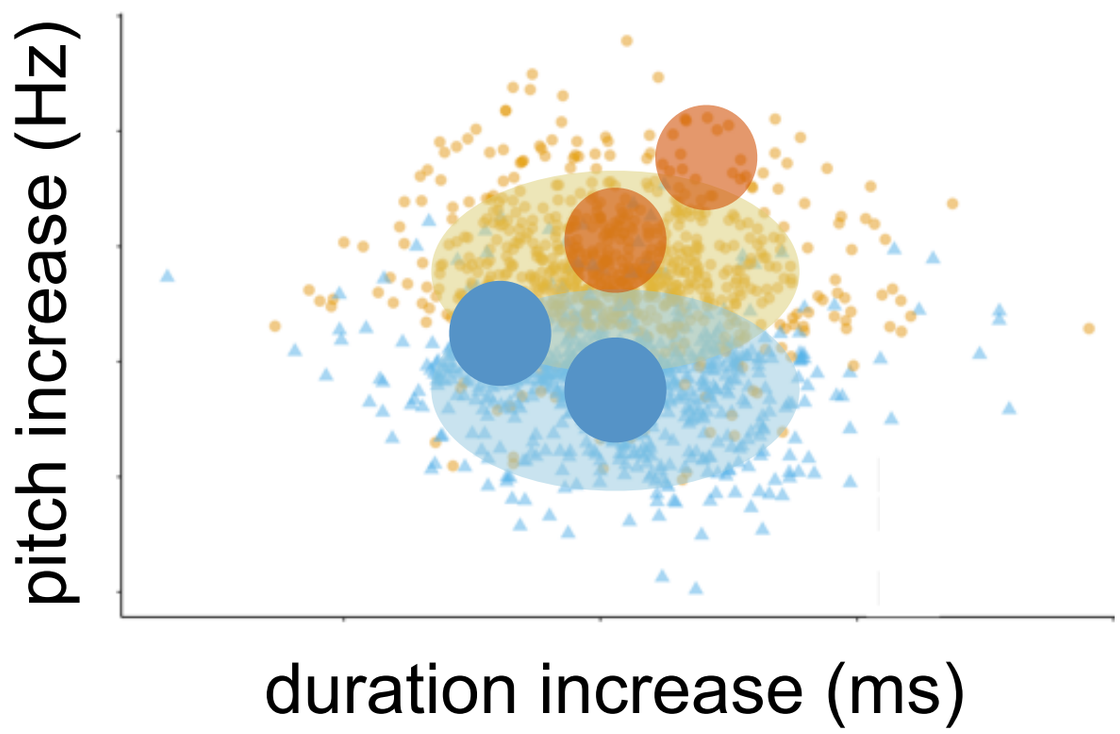




most statement-like

most question-like

Study 1: Summary



- Listeners rapidly (after 30 tokens of exposure) update their assumptions about p(intonation | intention, speaker) to better resolve ambiguity.

Speaker-dependency: Questions

e.g.,

- How much input necessary?
- Can we track language uses of multiple speakers simultaneously?
- Can you apply this logic to an author/book/literary genre?
- Does this an explicit modulation of judgment patterns? Or does our real-time inference process get modulated?

3 Real-time ambiguity resolution in pragmatic inferences

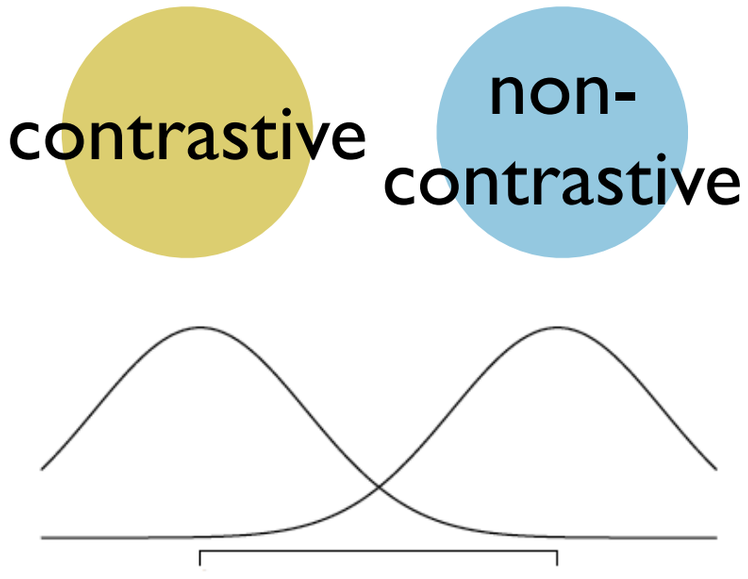
with Sadie Dix et al.

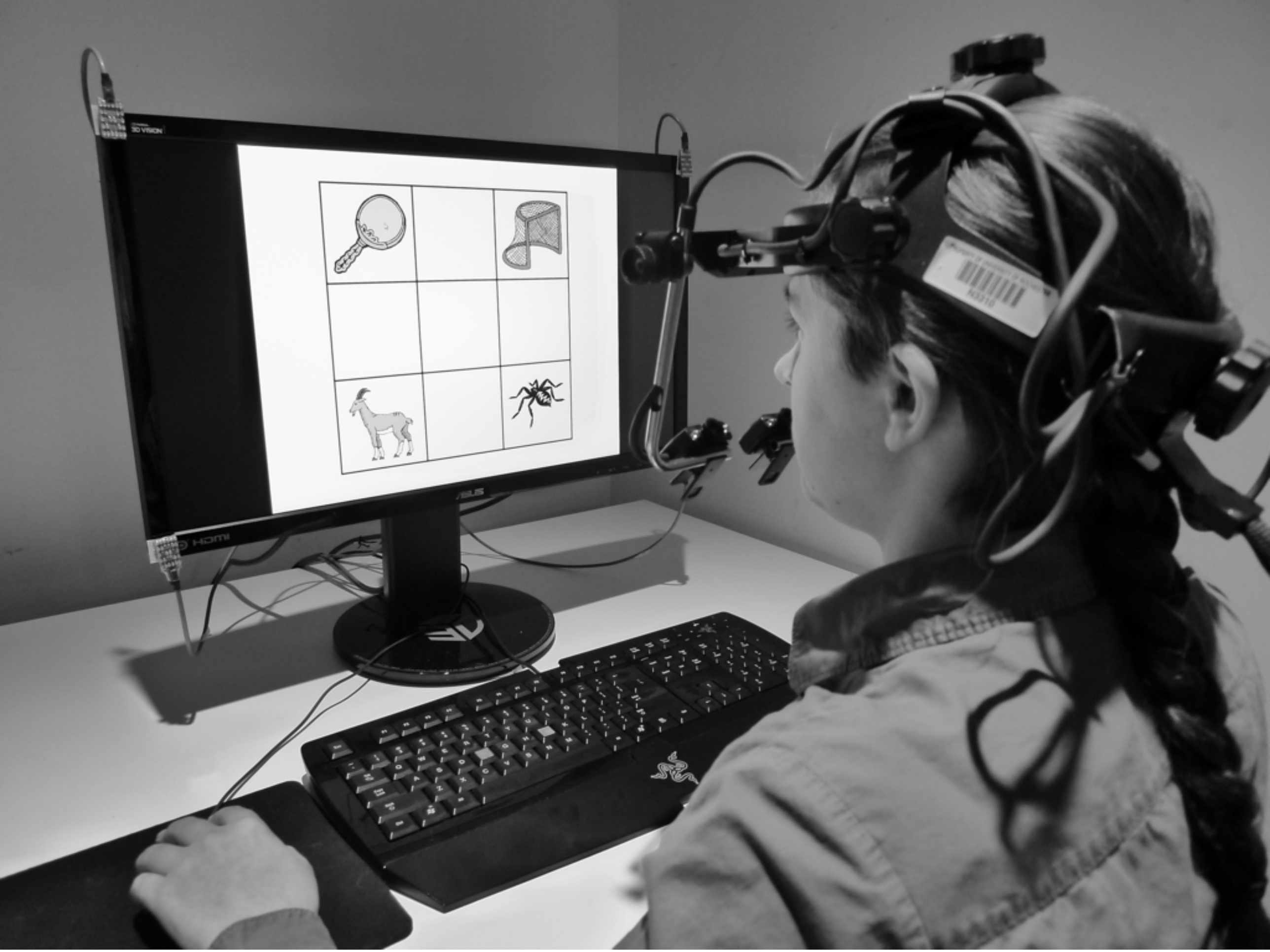
Inferences based on adjectives

- "Can you pass me the large cup?"
- Ambiguity between two intentions

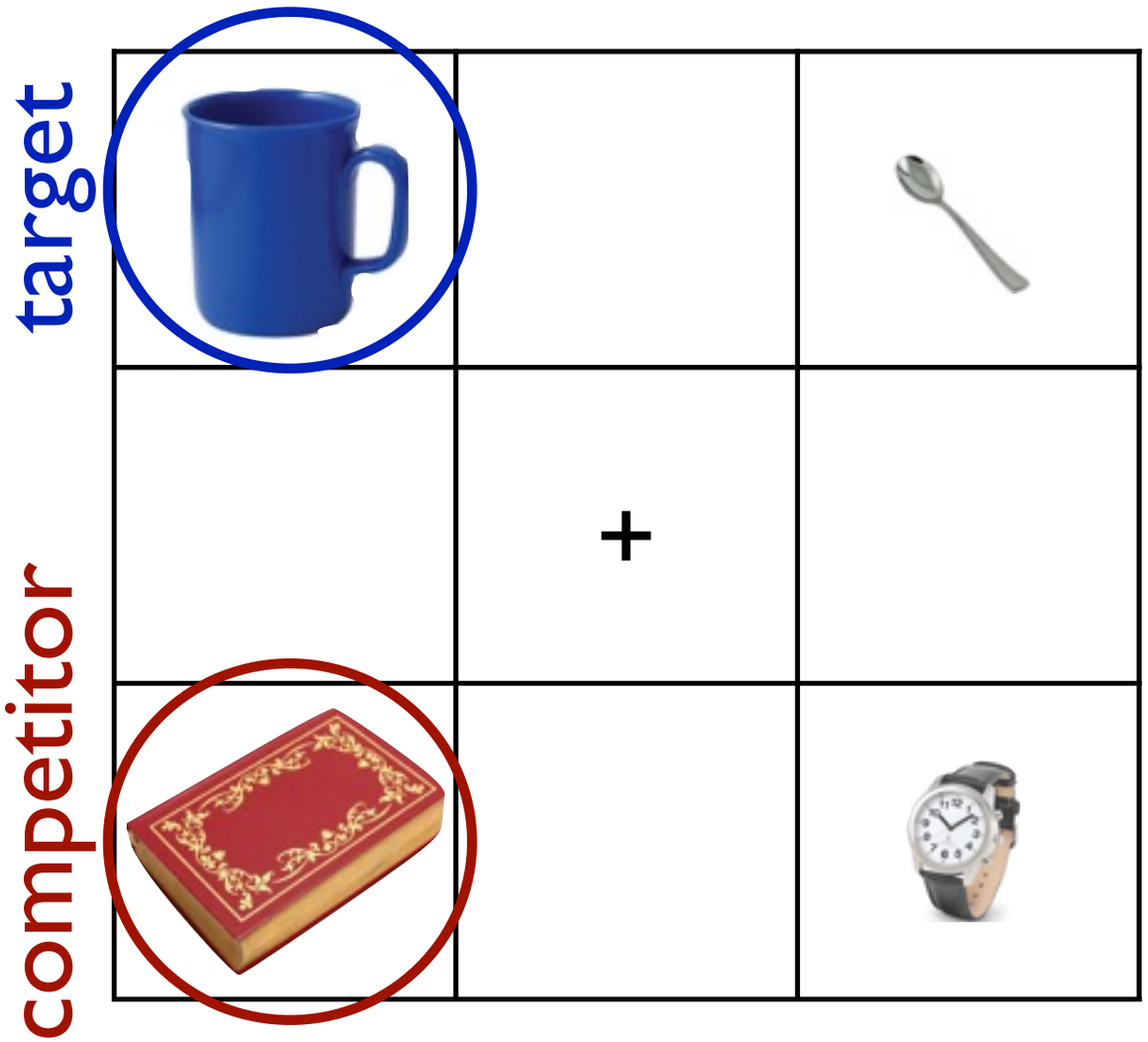
  1) "large" with respect to a standard

  2) "larger" in contrast to a contextual alternative





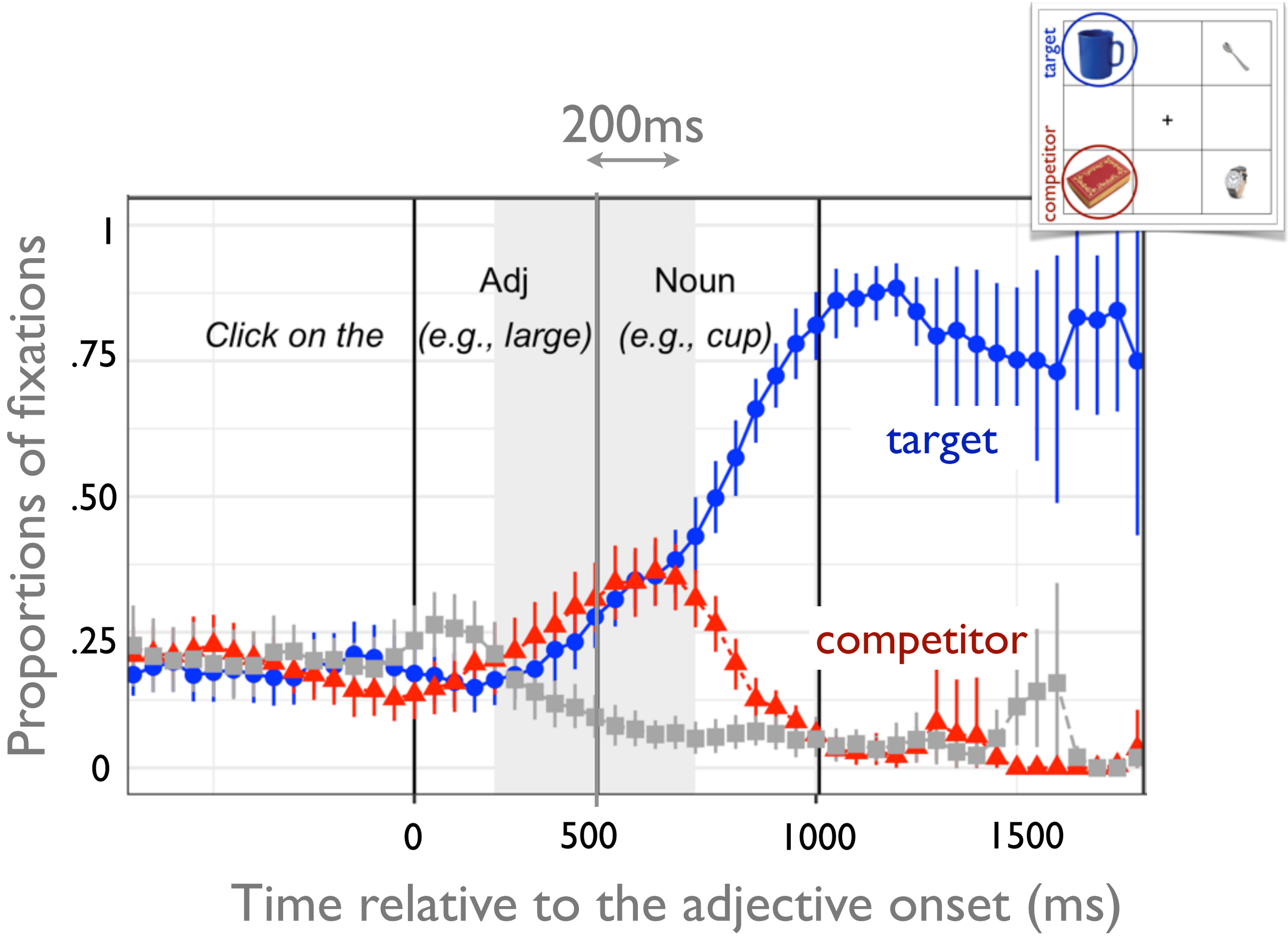
Study 2-a: Eye-tracking experiment



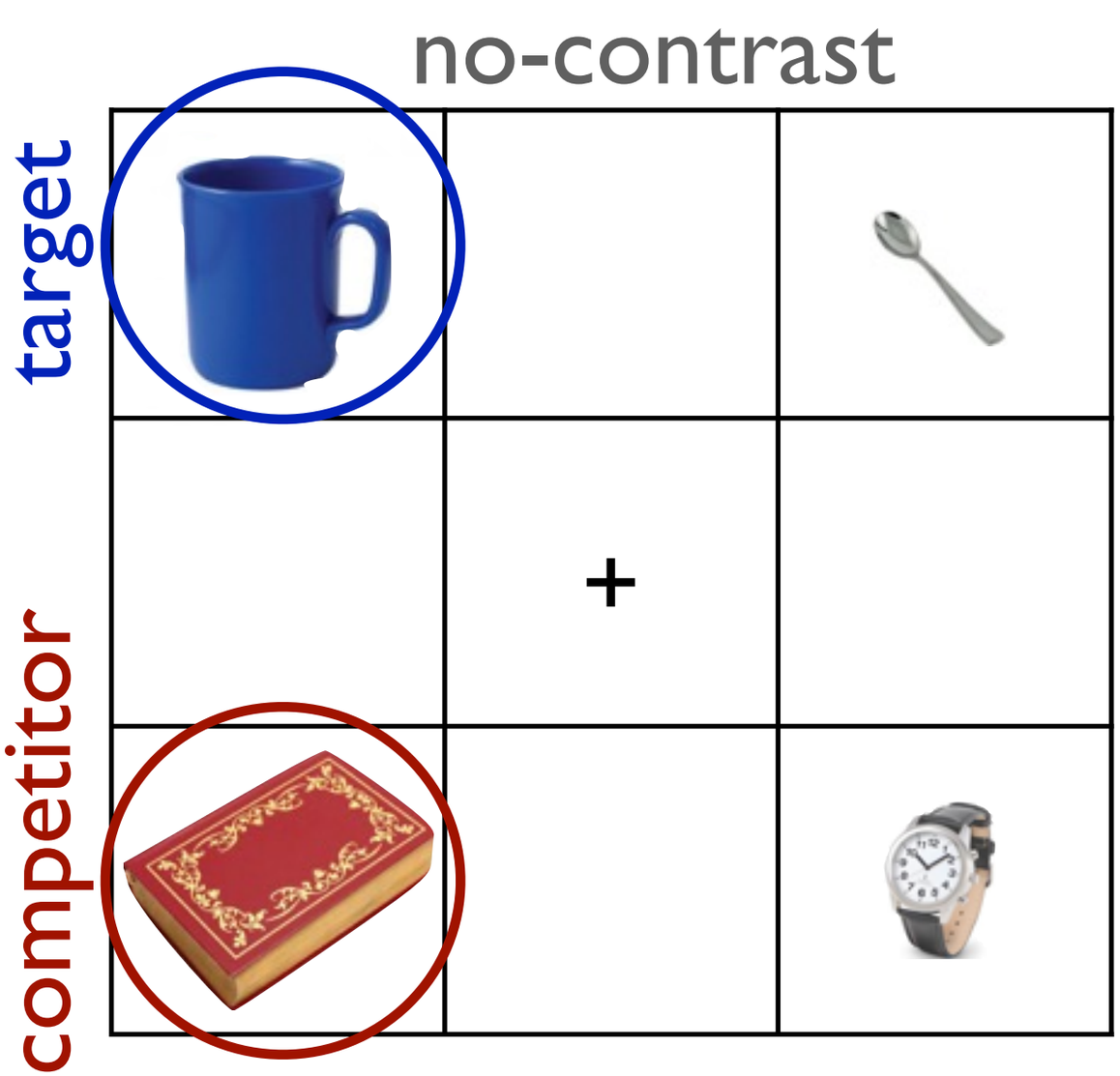
- 2 large objects and 2 small objects
- target = mentioned
- competitor = compatible with the said adjective

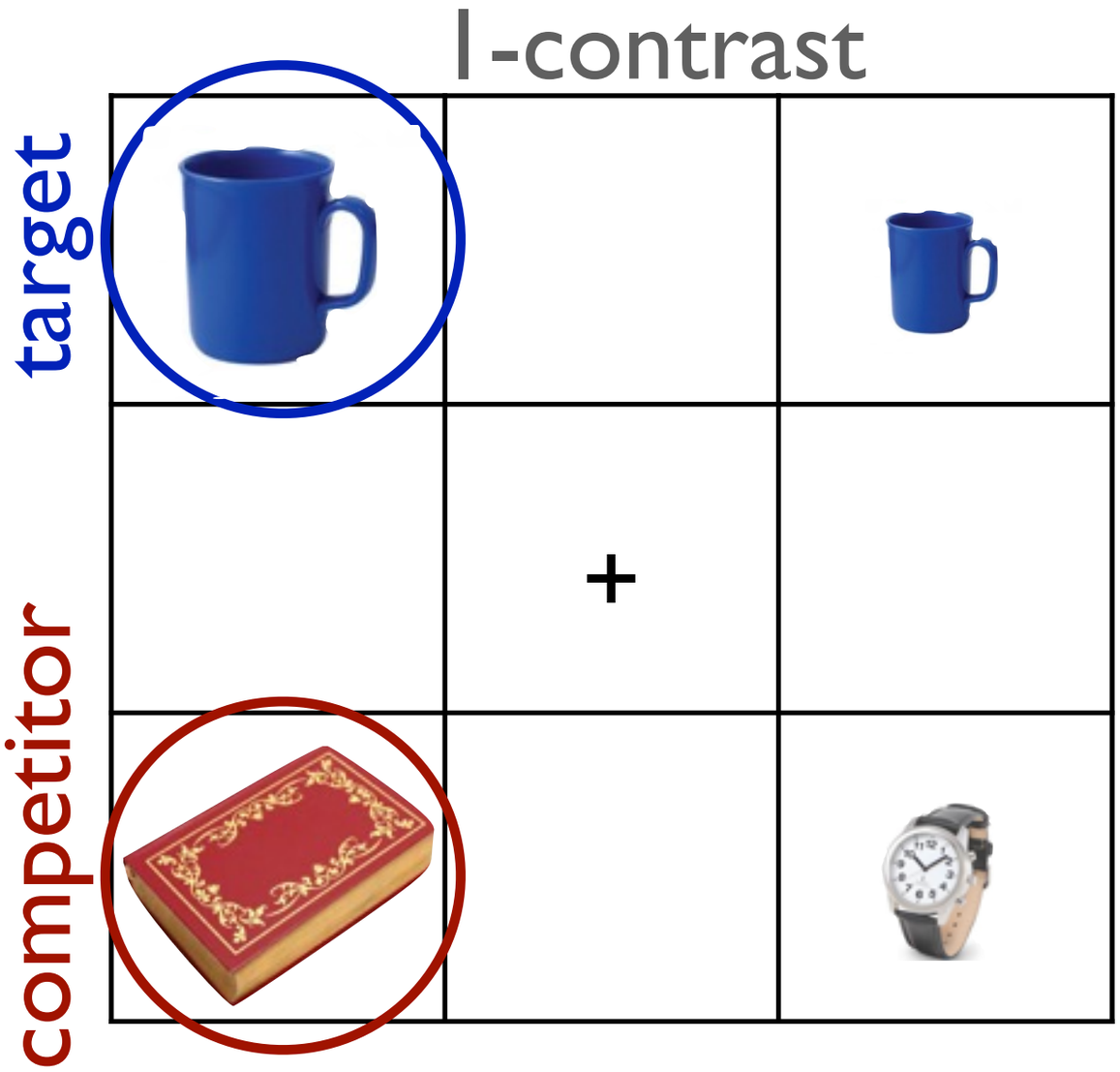
"Click on the large cup"

Sedivy et al., (1999); Grodner & Sedivy (2011)


200ms
target
competitor
Click on the
Adj
(e.g., large)
Noun
(e.g., cup)
target
competitor
Proportions of fixations
1
.75
.50
.25
0
0
500
1000
1500
Time relative to the adjective onset (ms)

Study 2-a: Eye-tracking experiment



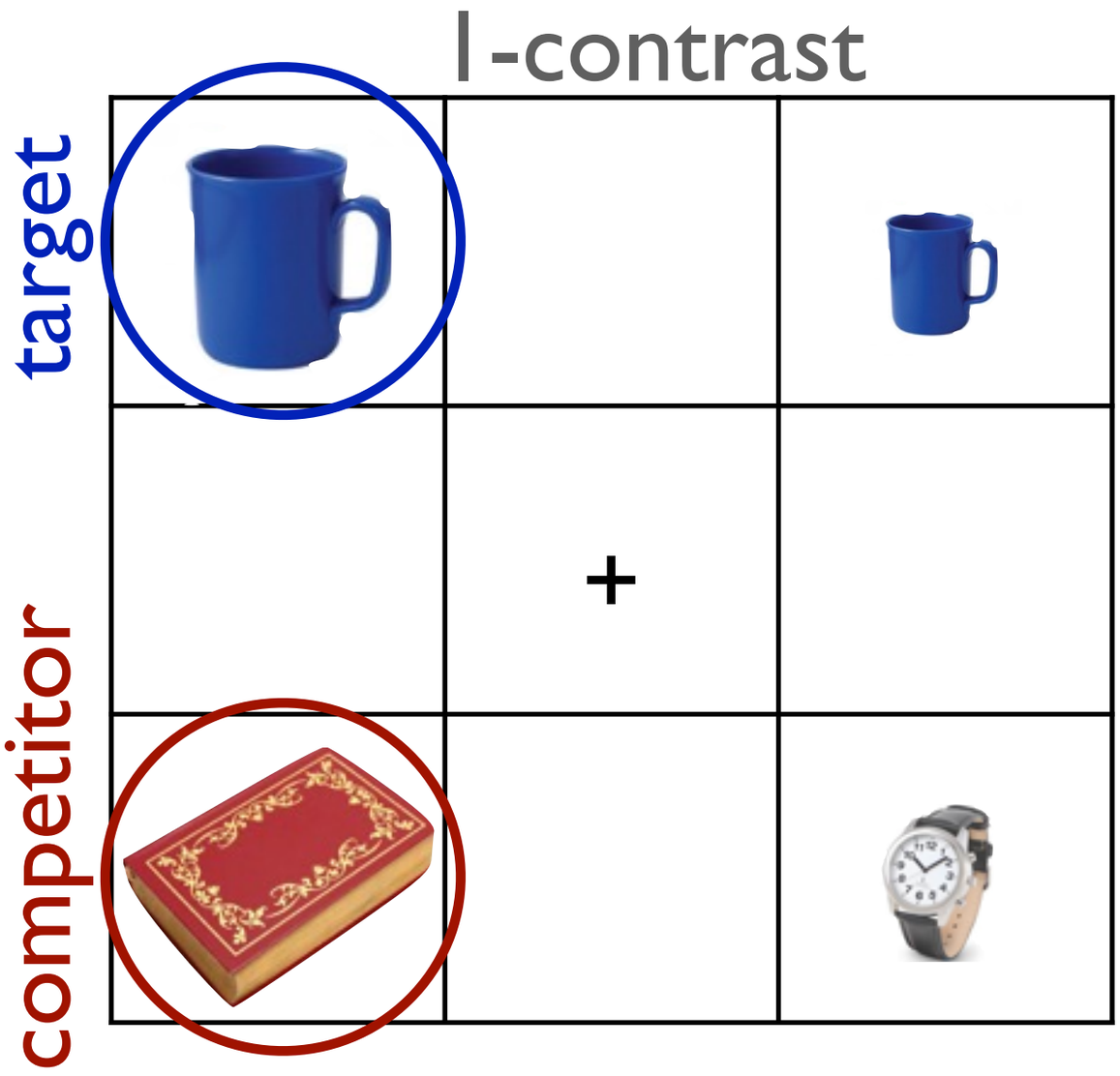


"Click on the large cup"

Sedivy et al., (1999); Grodner & Sedivy (2011)

Study 2-a: Eye-tracking experiment

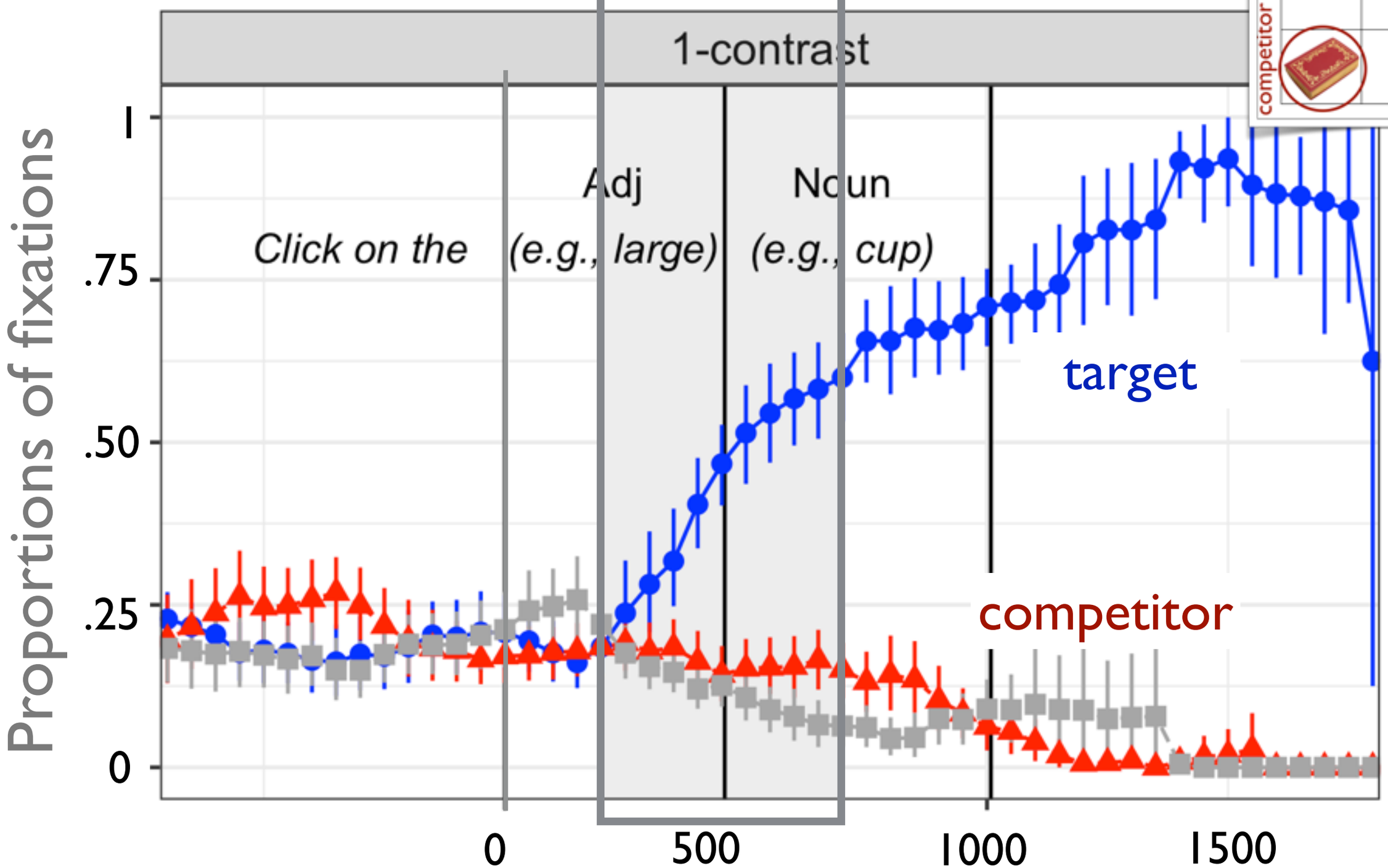
- 1 contrast set
- the adjective is more likely to convey the contrastive interpretation
- "large" can trigger fixations to the target



"Click on the large cup"

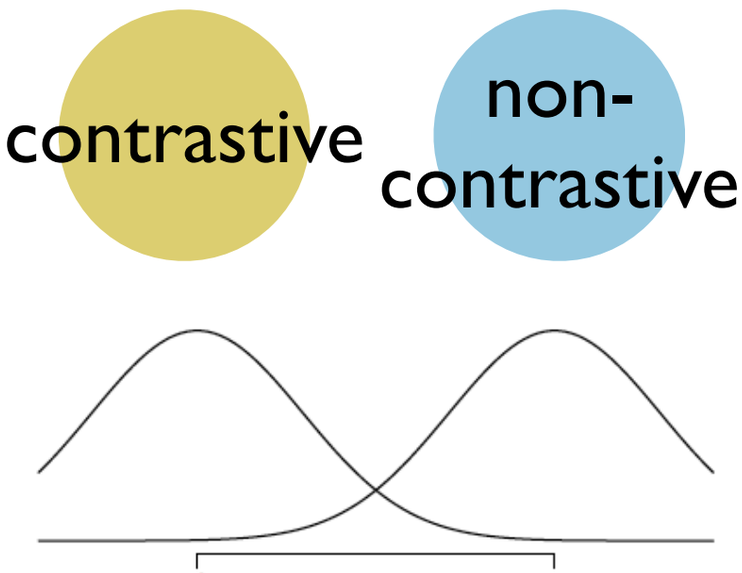
Sedivy et al., (1999); Grodner & Sedivy (2011)

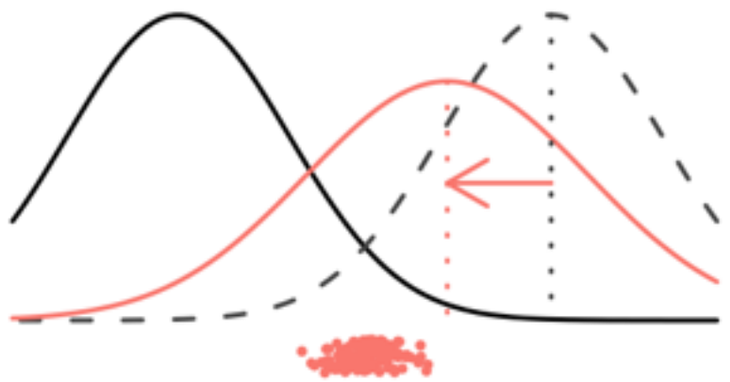
contrastive interpretation



Study 2-a: Summary

- Listeners derive the contrastive interpretation when there is a unique contrast set (i.e., 1-contrast condition)
- Does updated expectations change the time-course of real-time language comprehension?





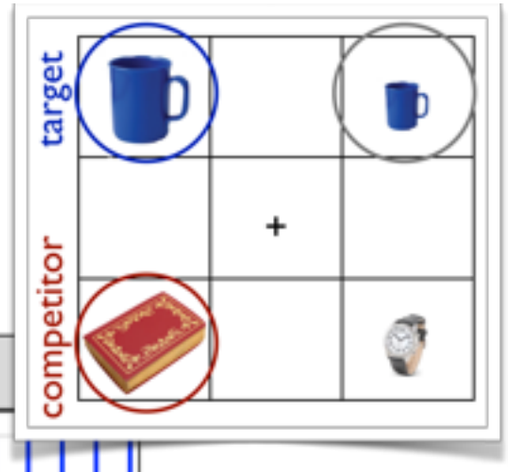
Study 2-b: Manipulations

- Instruction: *This speaker has a communicative impairment, which can cause linguistic problems.*
  i.e., this speaker may not use language in an expected manner

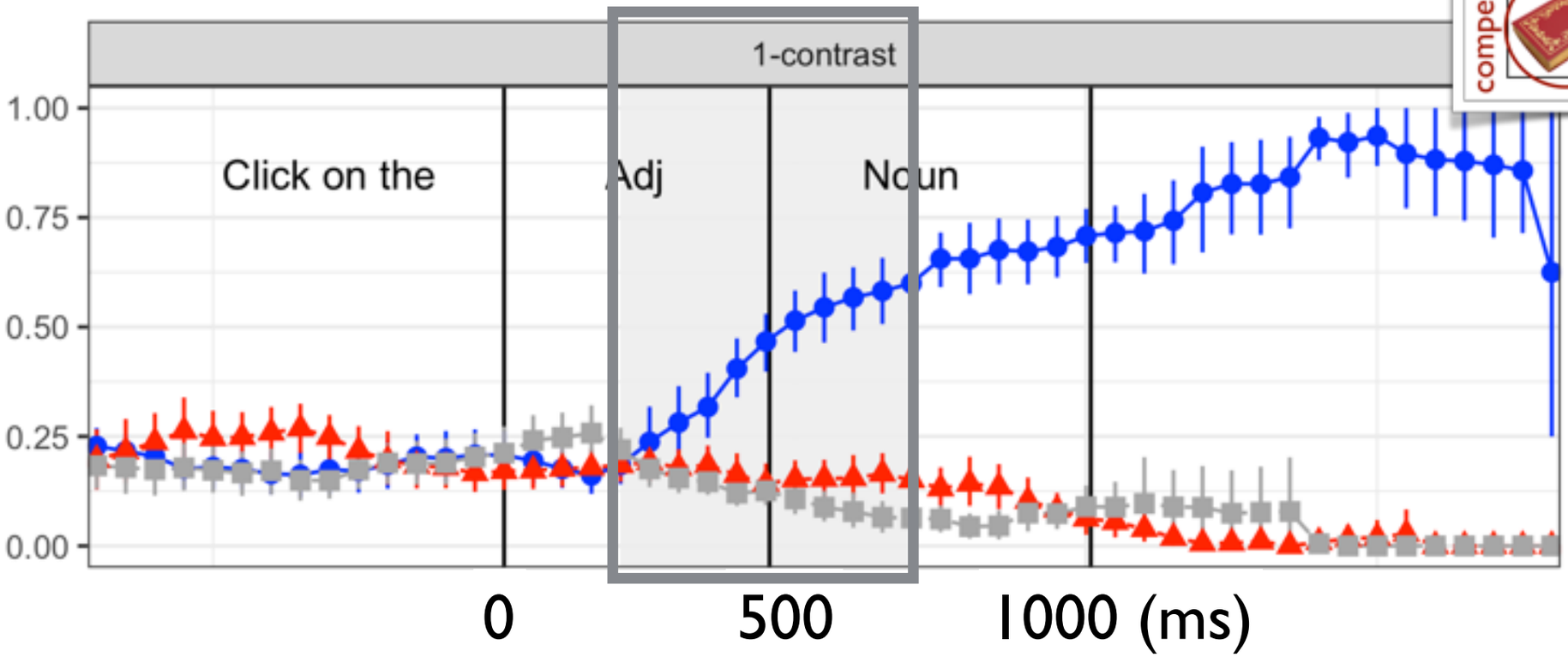
- Redundant (over-informative) adjective uses (e.g., "the large yellow banana" when there is only one banana = adjective non-contrastive)

- Prediction: If listeners process the signal based on the updated expectation, they will be less likely to make the contrastive inference.

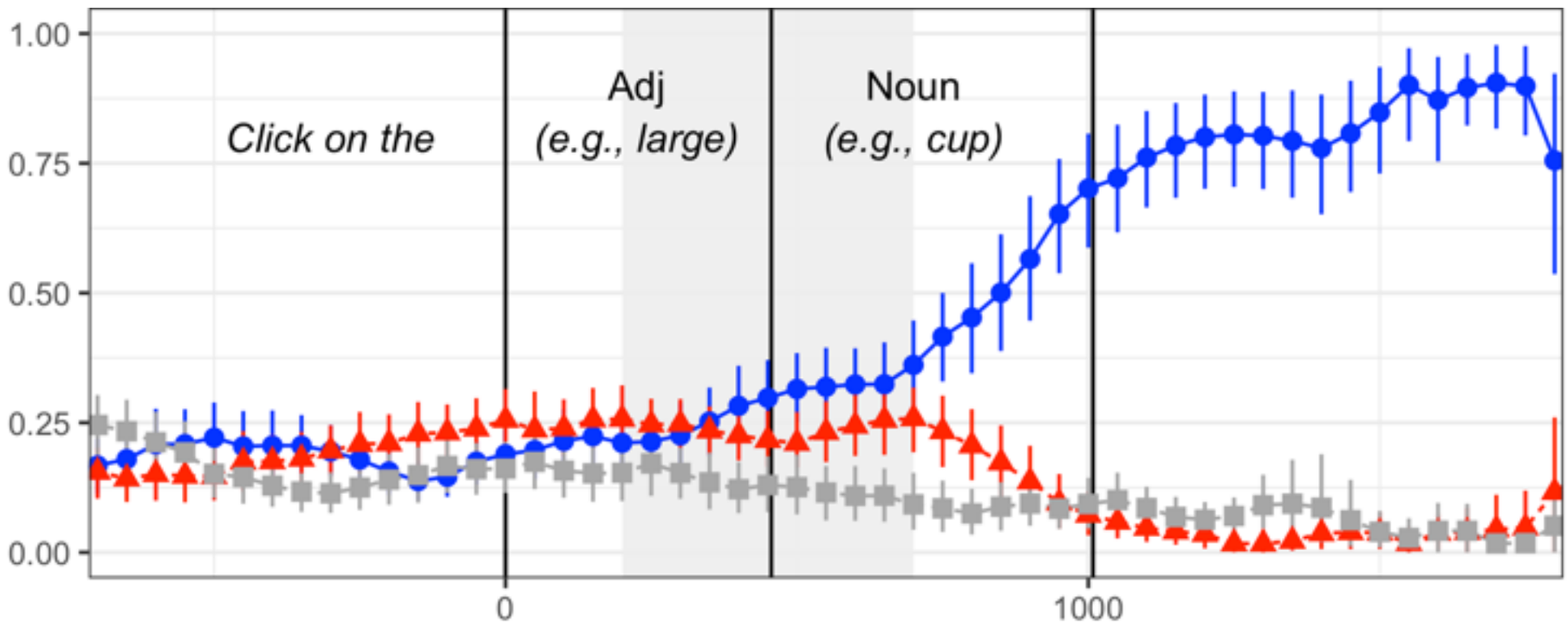
Experiment 2-b: Results



Study 2-a (ordinary speaker)



Study 2-b (redundant speaker)



Study 2: Summary

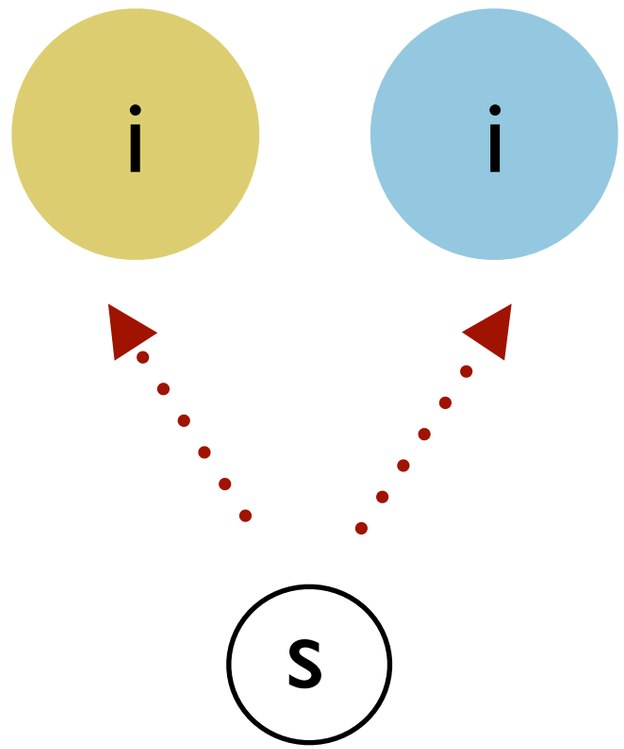
- Listeners update their expectations about the likelihood with which the speaker uses adjectives to convey the "contrastive" interpretation (e.g., the larger of the two)
- Adaptation of expectations about the speaker's language use modulates the amount of ambiguity listeners experience on a milli-second by milli-second basis.

4

Discussion

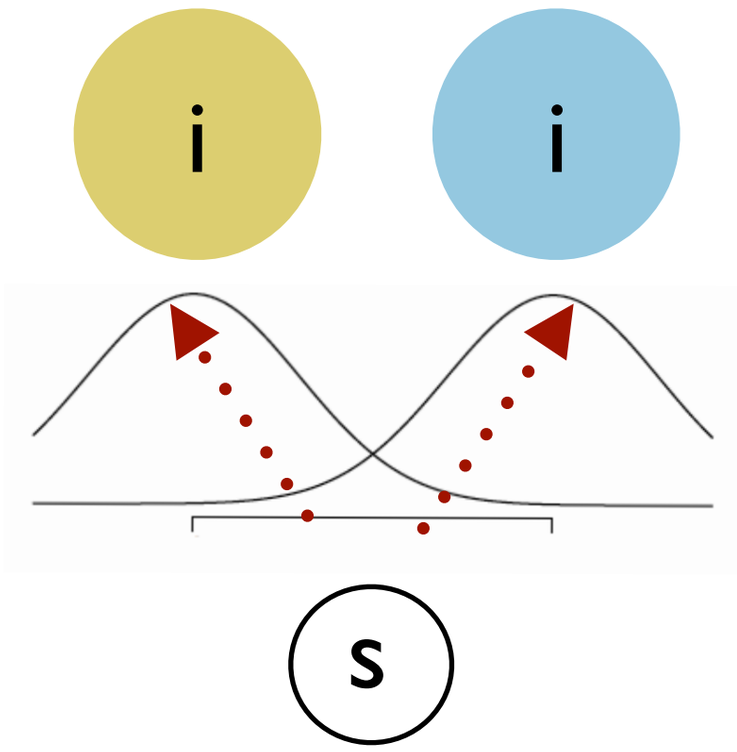
Ambiguity all the way down

- Inherent ambiguity in signal-intention mappings
- Even when listeners are not consciously experiencing any problem, the brain is consistently resolving the ambiguity.



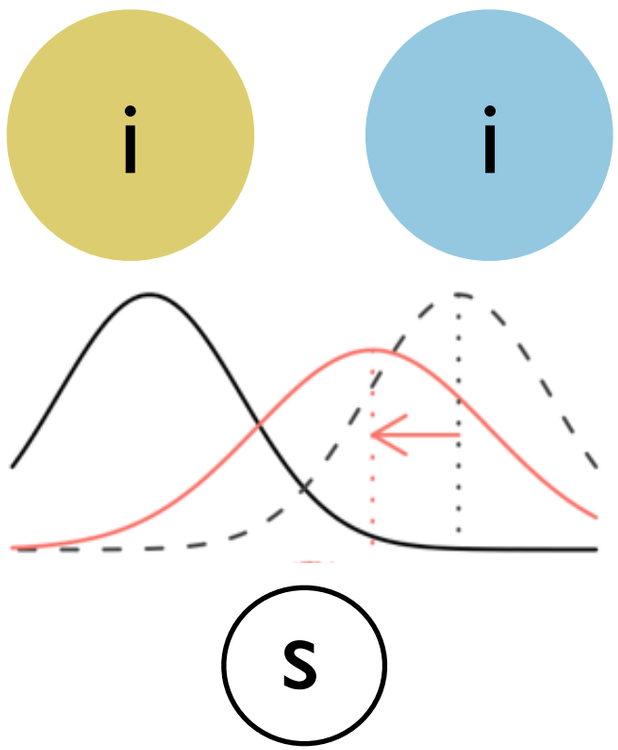
What allows us to resolve ambiguity?

- Fast and accurate ambiguity resolution relies on 1) an underlying model of possible signal-intention mappings and 2) flexible fine-tuning of expectations according to recent experiences in context.



Where does ambiguity exist?

- Is ambiguity a property of the linguistic signal? Or does ambiguity emerge in the process/act of perceiving and interpreting the language?
- Roles of expectations and experiences



Thank you!

For discussions: Andres Buxó-Lugo, T. Florian Jaeger, Xin Xie, HLP lab
For R/Praat scripting: Meredith Brown, Dave Kleinschmidt
For testing and annotation: Sadie Dix, Bethany Gardner, Cameron Morgan, Rebecca Lawrence, Anaclare Sullivan, Sherwin Nourani, Manasvi Chaturvedi, Nicole Vieyto

http://kinderlab.bcs.rochester.edu/