Embodied communication: Hand gesture

Susan Wagner Cook
University of Iowa

Gesture is robust

- Babies gesture
- Blind people gesture
- “Armless” person gestures

Gesture

Why do we gesture?

- What’s the mechanism by which gestures are produced?

- What’s the function of gesture?

Why do we gesture?

- What's the mechanism by which gestures are produced?

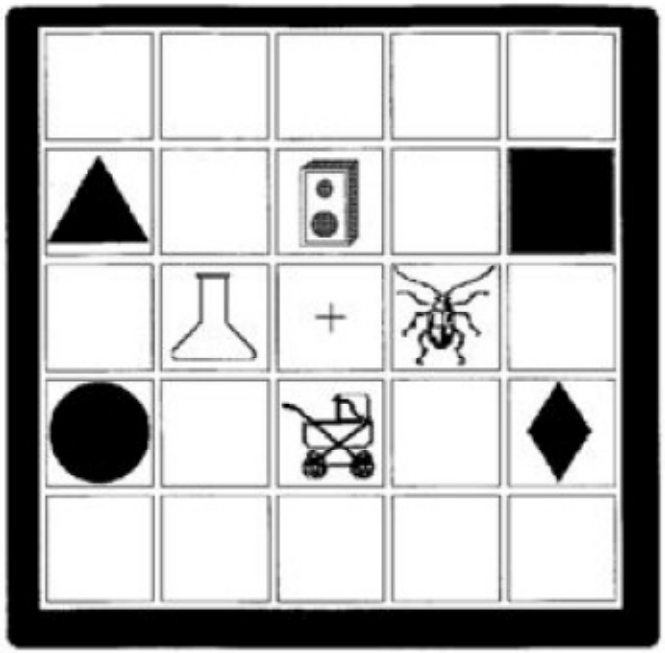
- *What's the function of gesture?*

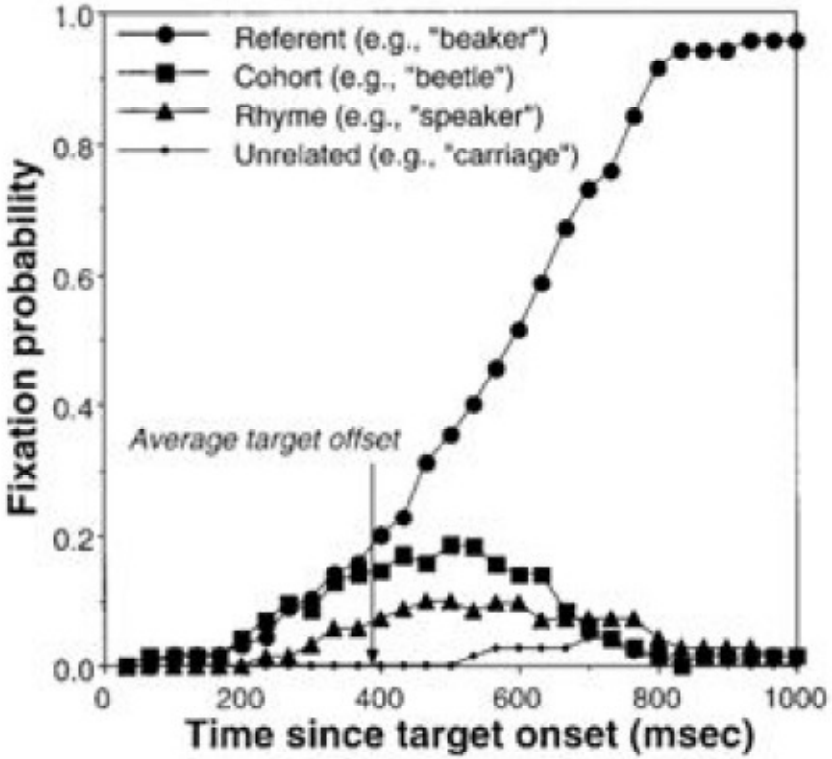
Words are reliable.

- We can talk on the phone.

Words are reliable.

- Listeners can quickly link incoming speech to referents in the world.





Speakers use both.

- Even in highly constrained situations, speakers produce gesture along with their speech

  – Pointing during picture naming (Bello, Capirci & Volterra, 2004)

Hand gesture reveals the embodiment of communication.

What is embodied cognition?

Wilson, 2002:

1.) Situated – involves perception and action

2.) Time pressured – real time interaction with the environment

3.) Off load onto environment

4.) Environment part of cognitive system

5.) Cognition is for action

6.) Off-line cognition is body based

Speech and gesture

- Produced by a single body
- Communicate a single idea
- Complementary encoding of information
  - Bodily versus abstract

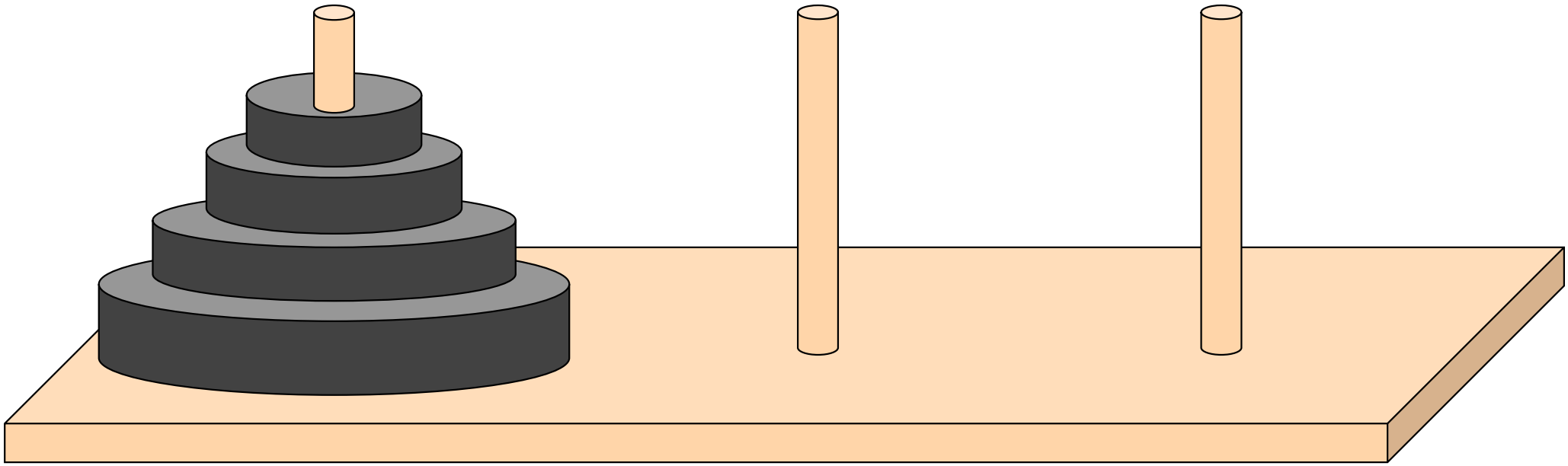
Do speakers use gesture to communicate perceptual-motor experience to listeners?

Tower of Hanoi

Goal: Move the tower from the left peg to the right peg.

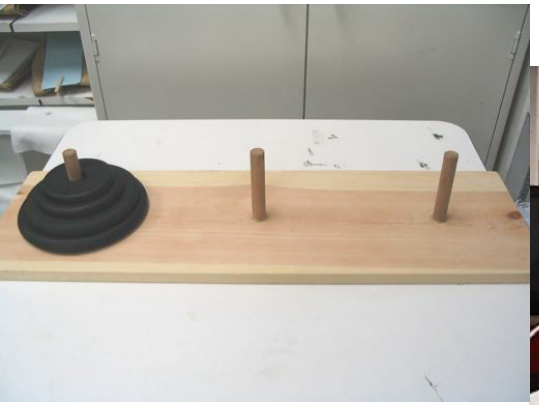
Rules: 1.) Only move one disk at a time

2.) Larger disks cannot be placed on top of smaller disks



Explanation Conditions

- **Real objects** - grasp and lift heavy objects
- **Computer** - click and drag virtual objects



Real objects

Speaker solves TOH according to condition.

Speaker explains TOH.

Listener solves TOH.

Computer

What do speakers say?

- *Where* to move each of the disks.

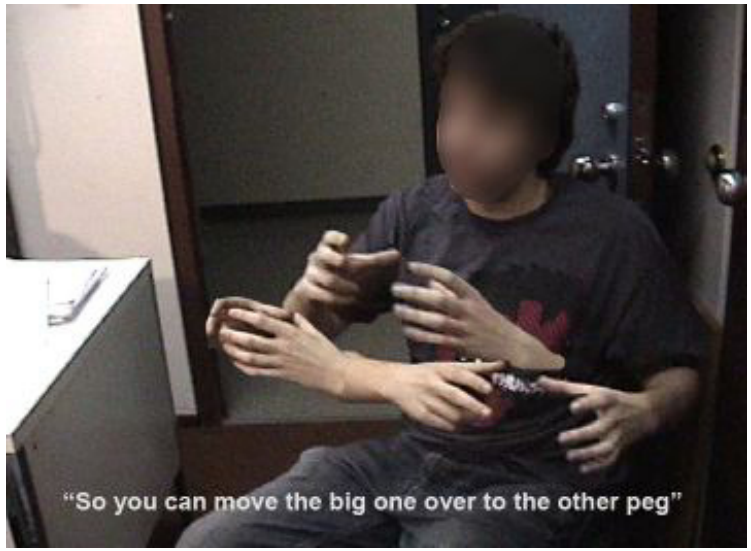
"Ok, so you have the four disks. They're from biggest to smallest, on the left peg and you have to get them all the way to the right. So if you move the top one, the smallest one, all the way over and then the one under it over so you have the two big ones and then littler, littler. And then you can put the smaller one on the middle one so that you can move the second biggest one over."

What do speakers say?

- *Where* to move each of the disks.

- Importantly, this verbal description does not change across conditions
  - Same distribution of word types and tokens
  - Listeners cannot reliably identify speakers' experience when given auditory information

What do speakers do?



- Depict movement of the disk
  - Trajectory
  - Hand shape
  - Number of hands

What about listeners?

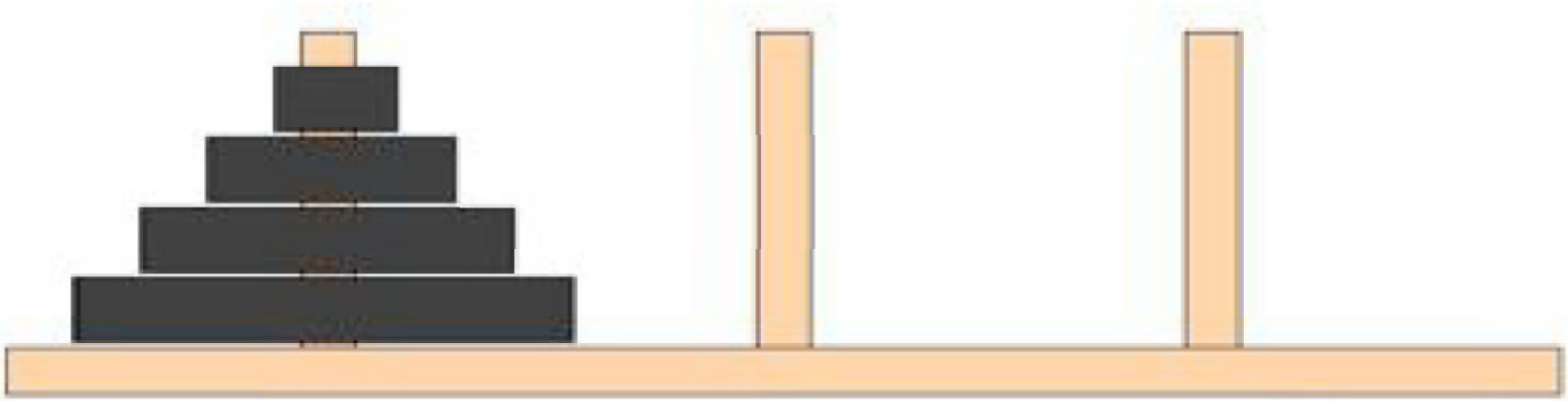
- Listeners were equally good at solving the task across conditions.
  - They knew <u>where</u> to move the disks.

- Were listeners affected by speakers' gestures?
  - Did they know <u>how</u> to move the disks?

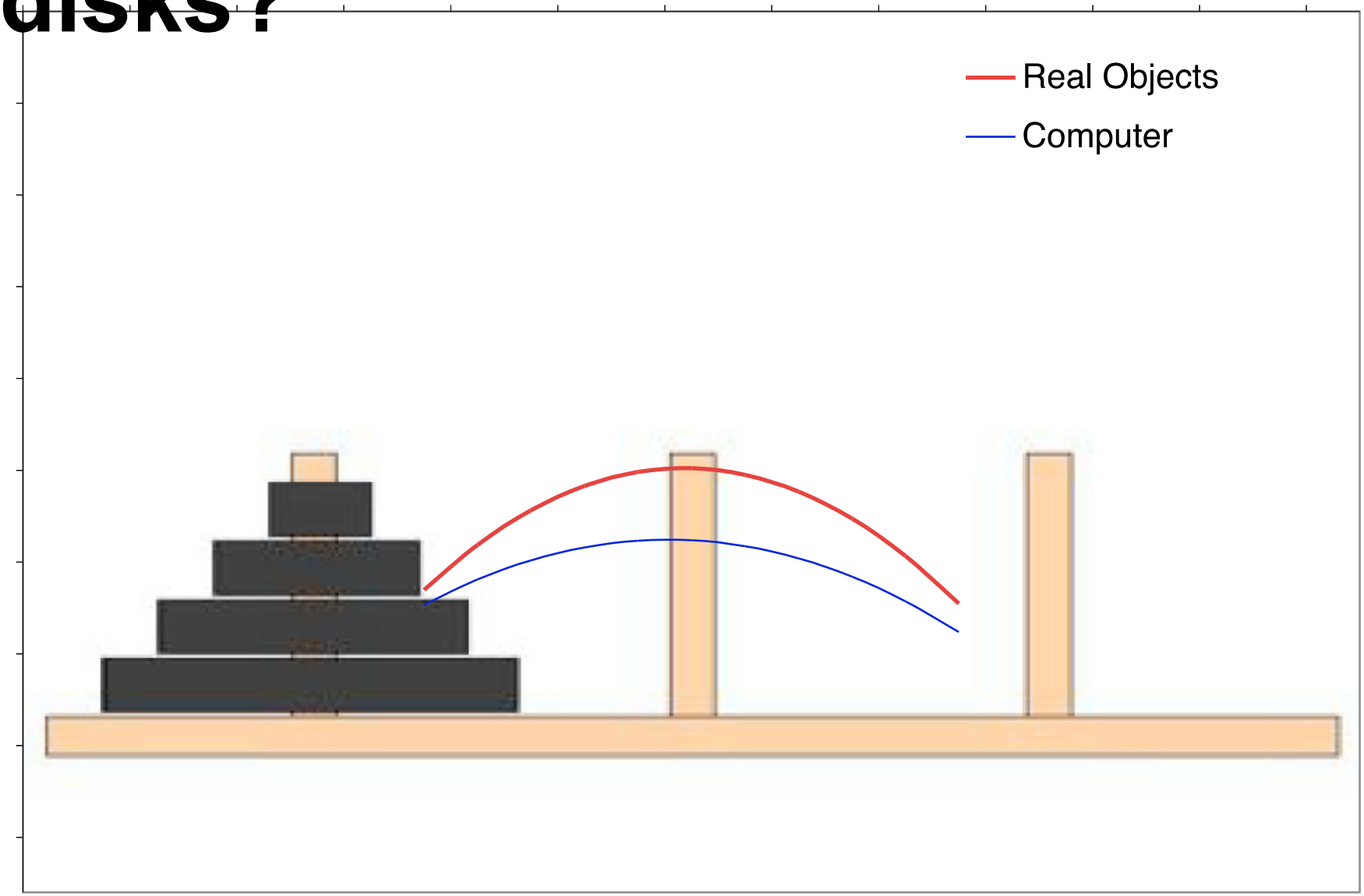
How do listeners move the disks?

Real Objects

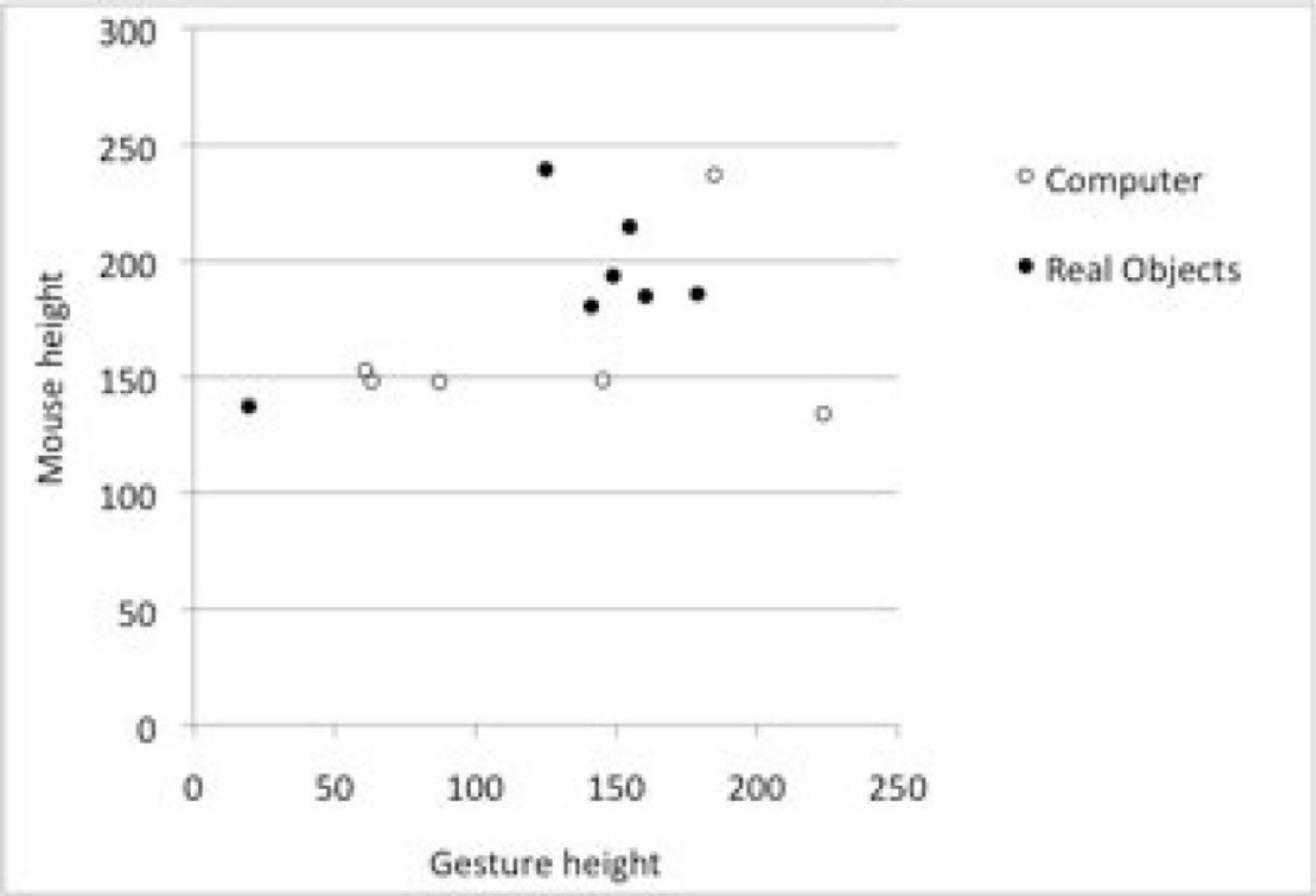
Computer



How do listeners move the disks?



Does this come from gesture?



Do speakers use gesture to communicate perceptual-motor experience to listeners?

Yes.

Speech and gesture

- Produced by a single body
- Communicate a single idea
- Complementary encoding of information
  - Bodily versus abstract

Speech and gesture

- Produced by a single body
- Communicate a single idea
- Complementary encoding of information
  - Bodily versus abstract
  - > Should be trade-off in information representation

Do speakers use gesture to incorporate the body into communication?

Study

- Speech and pointing gesture

-> If gesture is truly integrated with speech, it should influence speech at all levels.

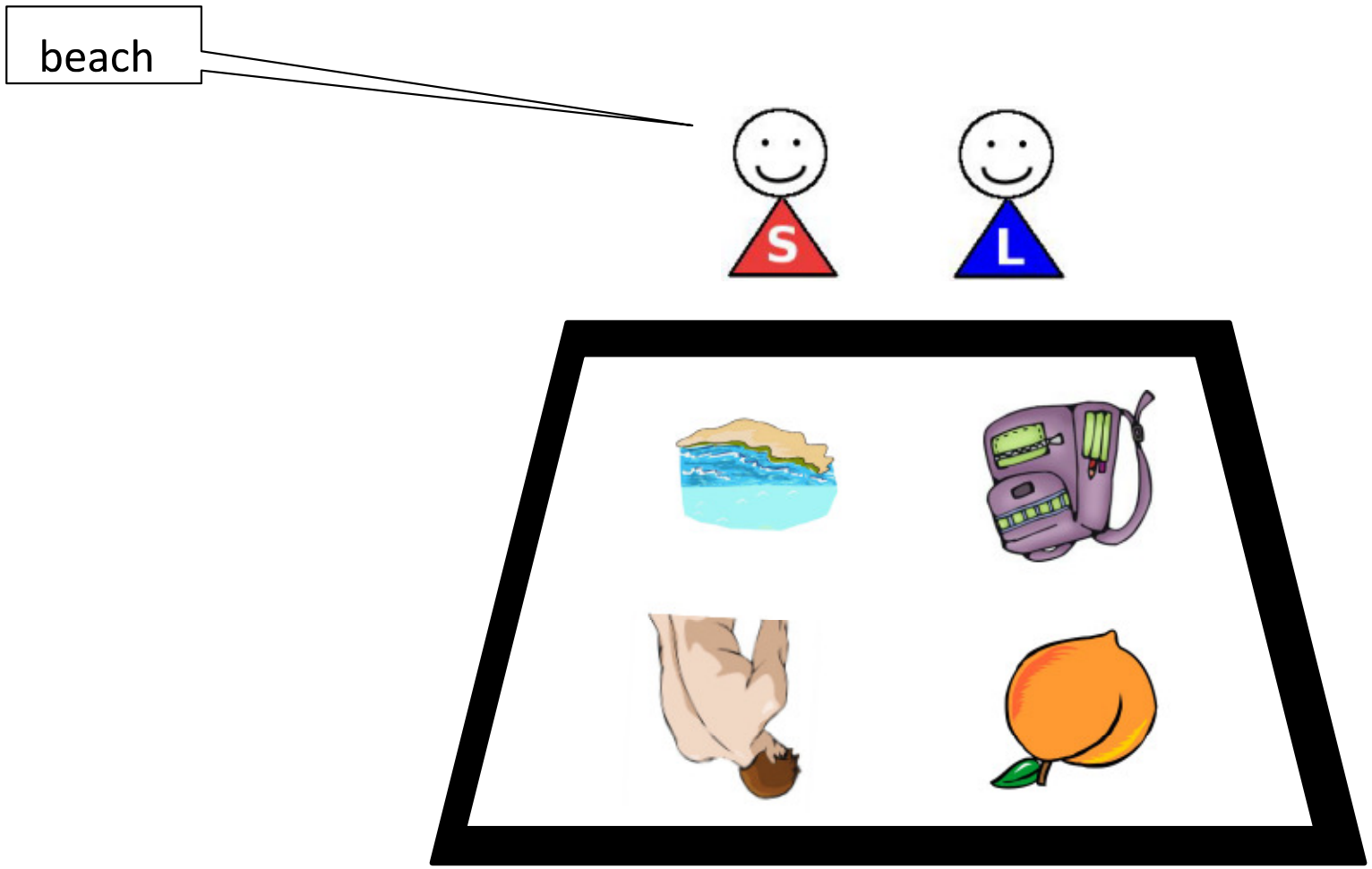
Lieberman, 1963

- Showed speech is more informative when produced in a less informative context

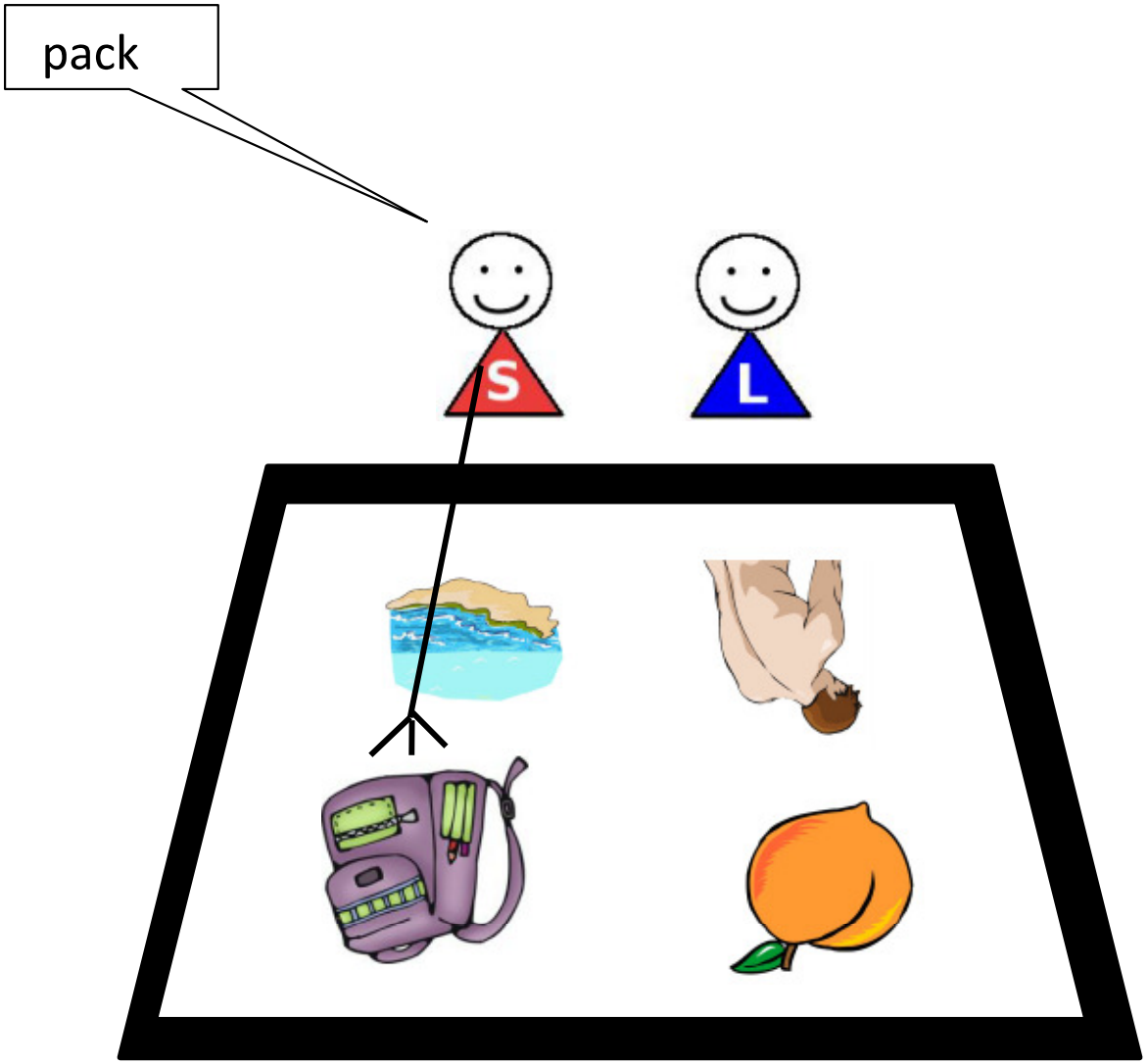
  “Stitch in time saves ____.

  “The party is at ____.

Method



Method



Manipulation

- Speaker points or does not point
  - Instructions over headphones:

  "Point at and name the picture in the upper right."

  "Name the picture in the upper right"

Method

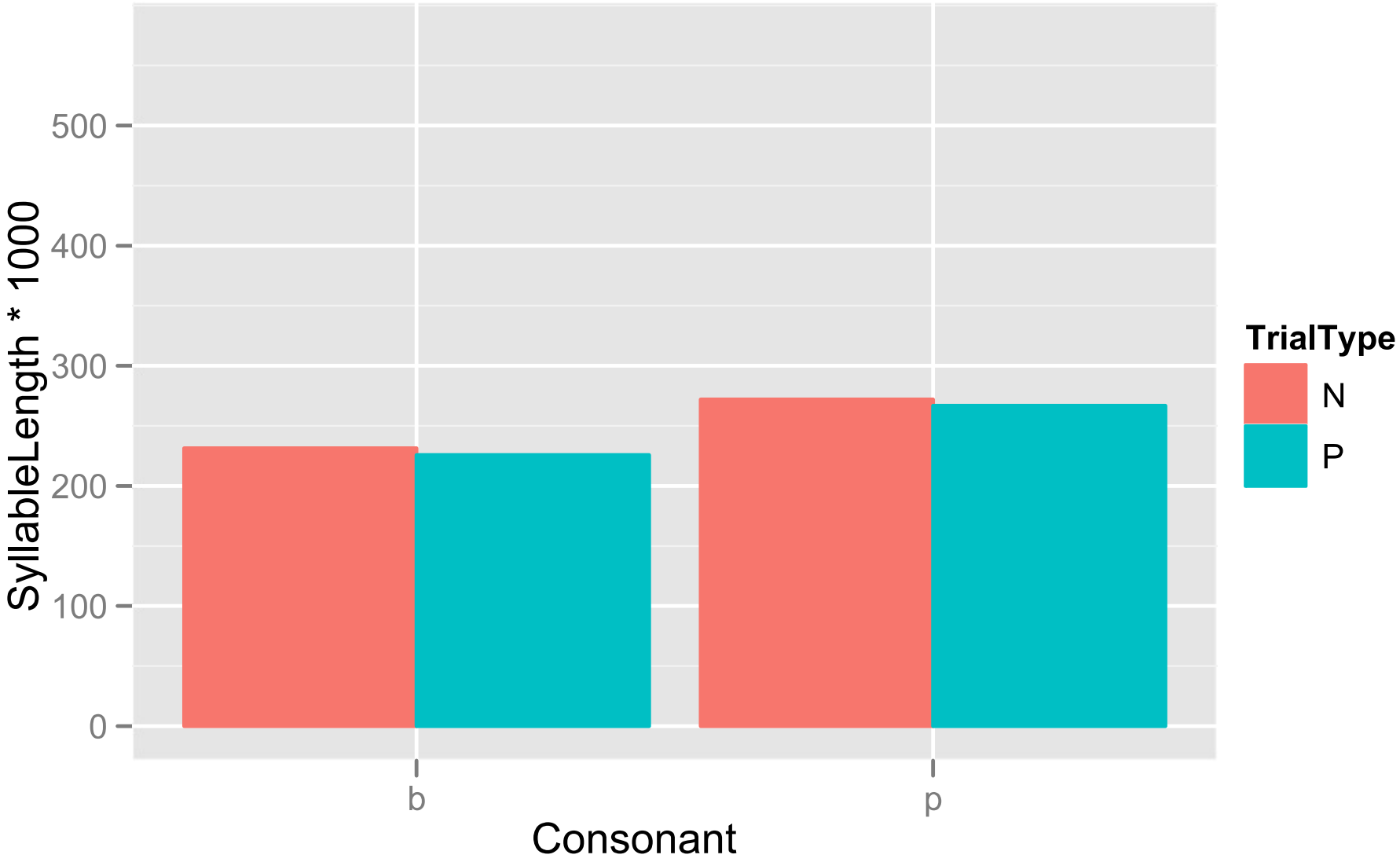
- Speaker: single word production task

- Listener: 2-back memory task with to-be-remembered information communicated from a partner

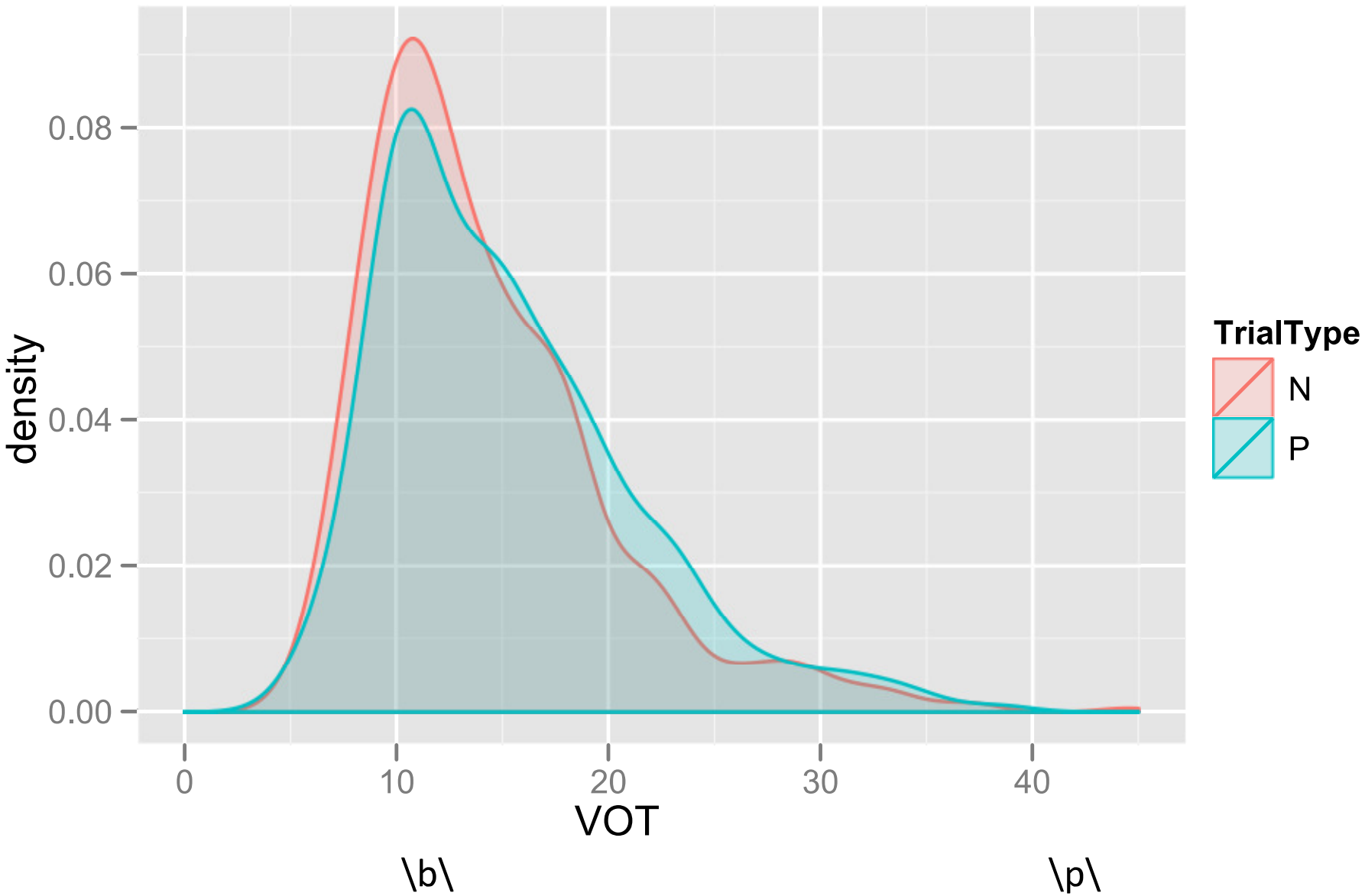
Results

- Analyze the acoustic properties of the speech
- Syllable length and voice onset time (VOT)

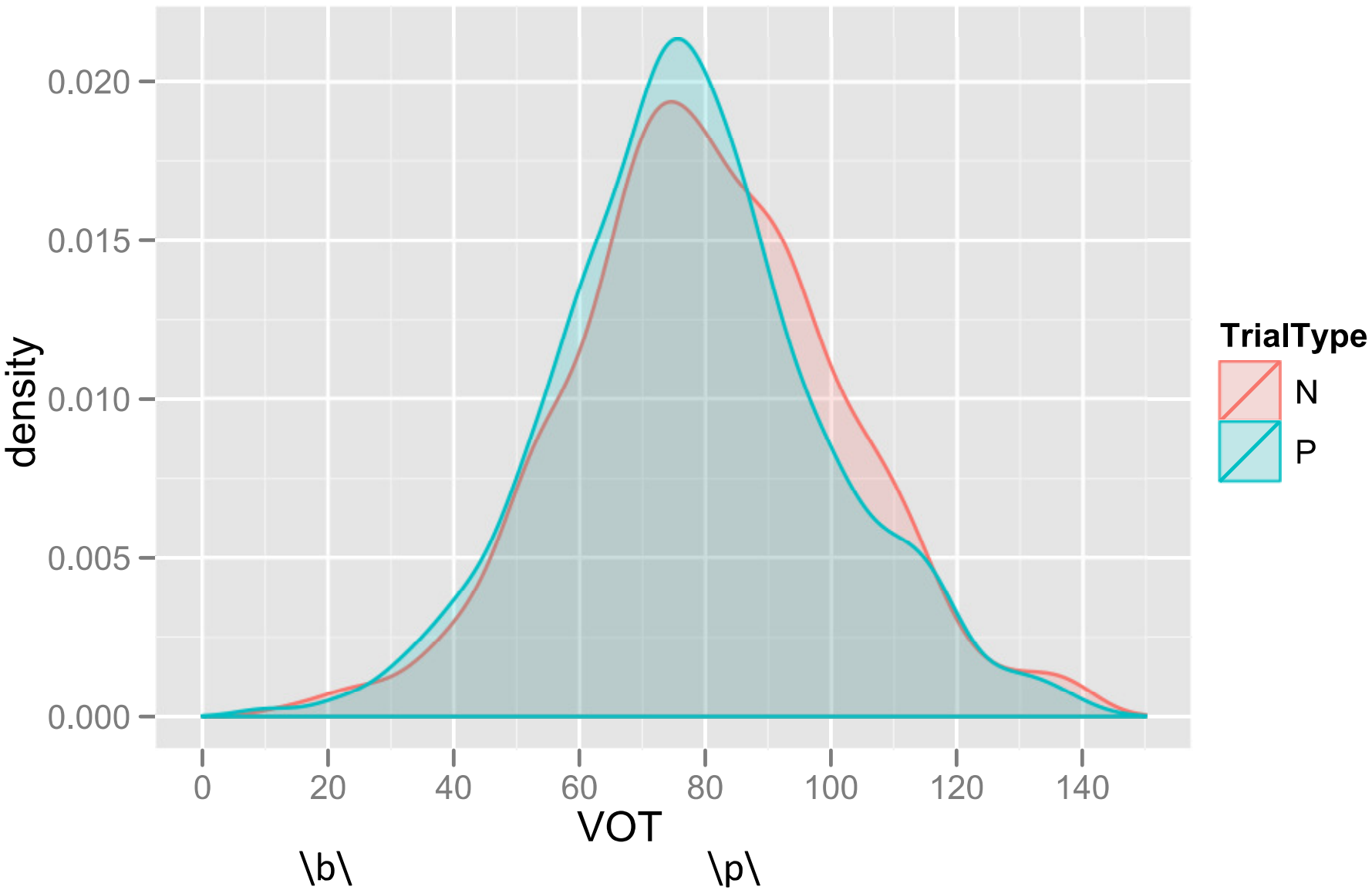
Results – syllable length



Results – VOT \b\



Results – VOT - \p\



Information is distributed!

- Speakers produce less informative speech when they produce accompanying pointing gesture

Speech and gesture

- Produced by a single body
- Communicate a single idea
- Complementary encoding of information
  - Bodily versus abstract
  - > Should be trade-off in information representation

Do speakers use gesture to incorporate the body into communication?

Yes.

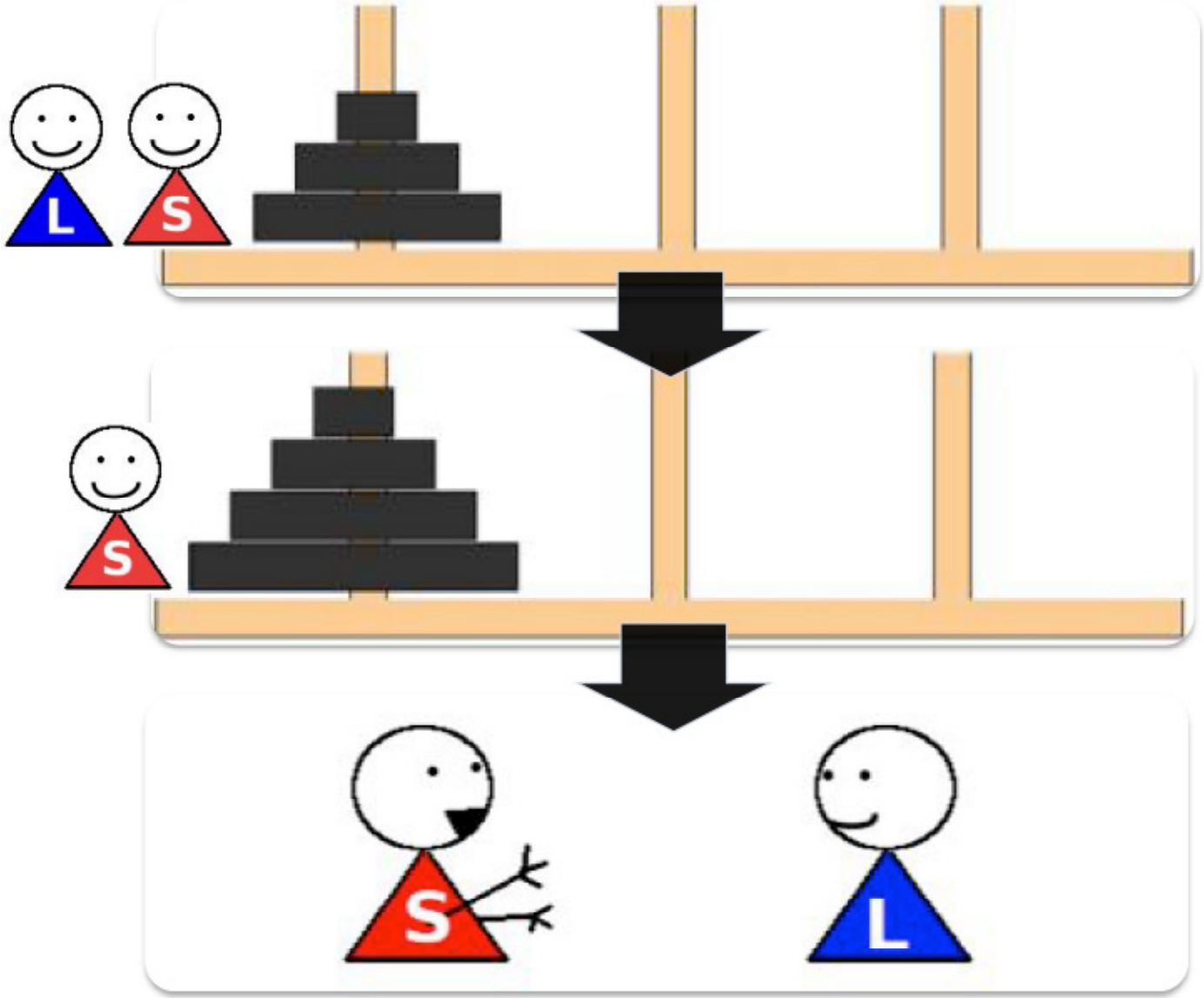
Speech and gesture

- Produced by a single body
- Communicate a single idea
- Complementary encoding of information
- Flexible encoding?
  - > Information in gesture should be adjusted according to communicative context

Are speakers gesturing for their listeners?

Gesture pragmatics

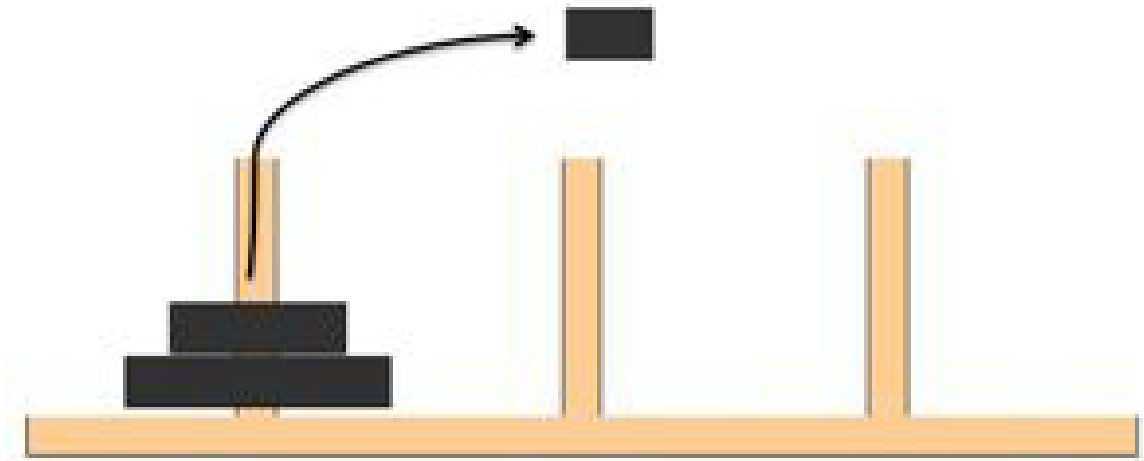
- Gesture should be less informative when it is produced to a more knowledgeable listener.


L
S
S
S
L

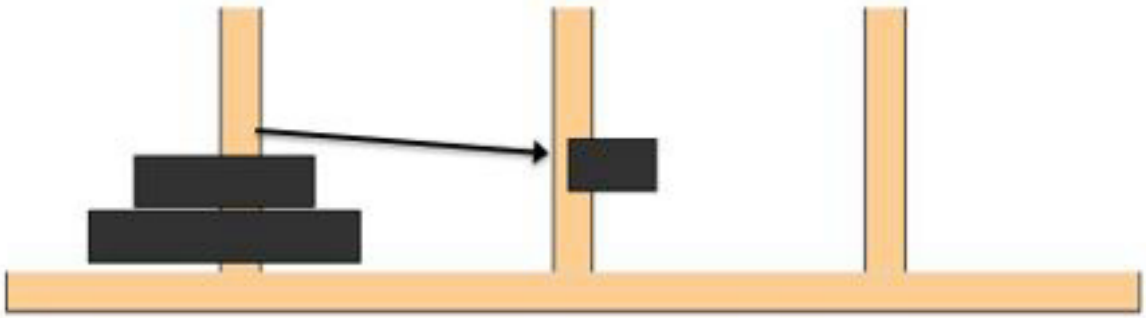
Common vs. Uncommon Ground

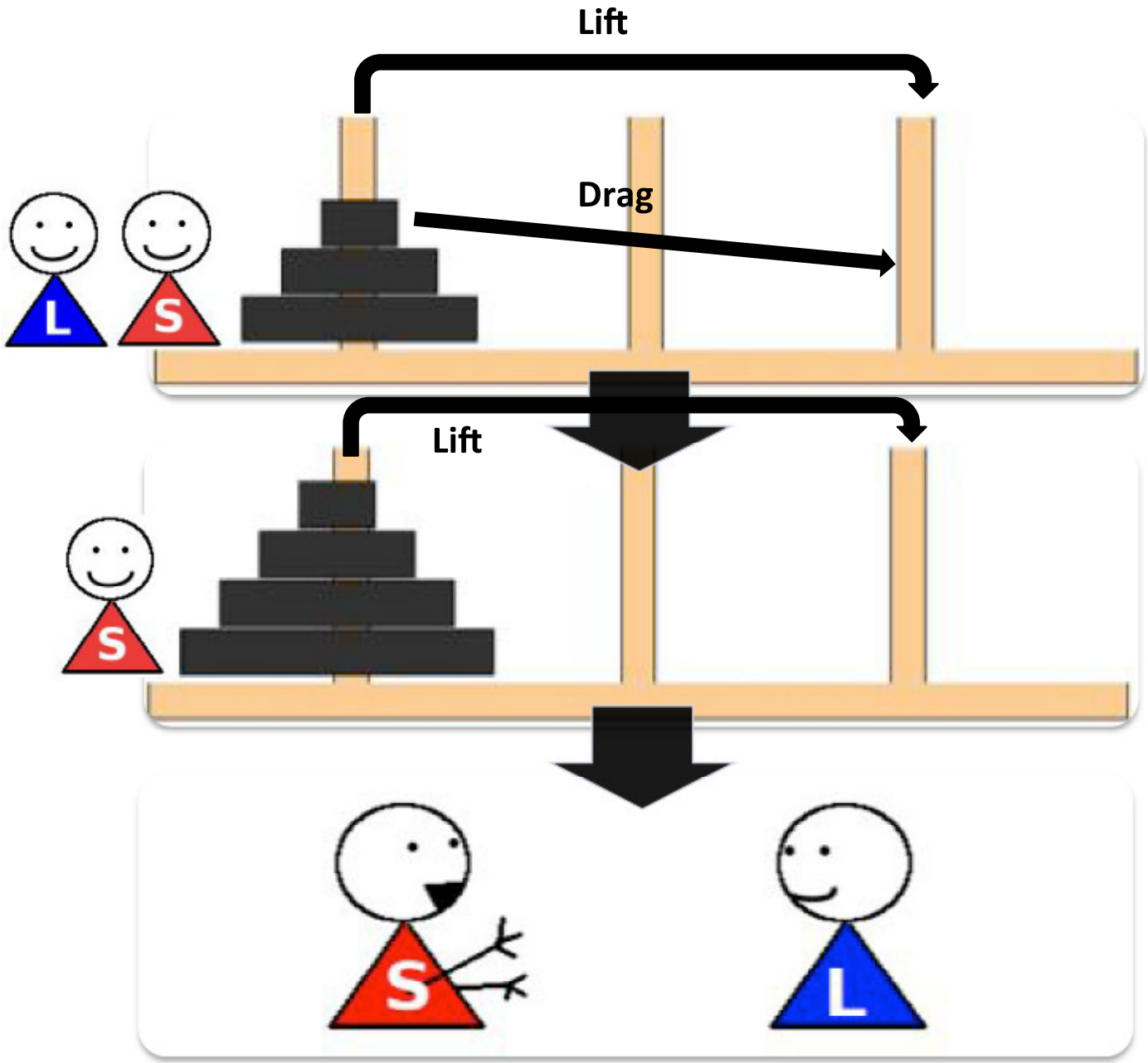
- Manner is manipulated

LIFT



DRAG





Common Ground Condition

Uncommon Ground Condition

Results: Speech

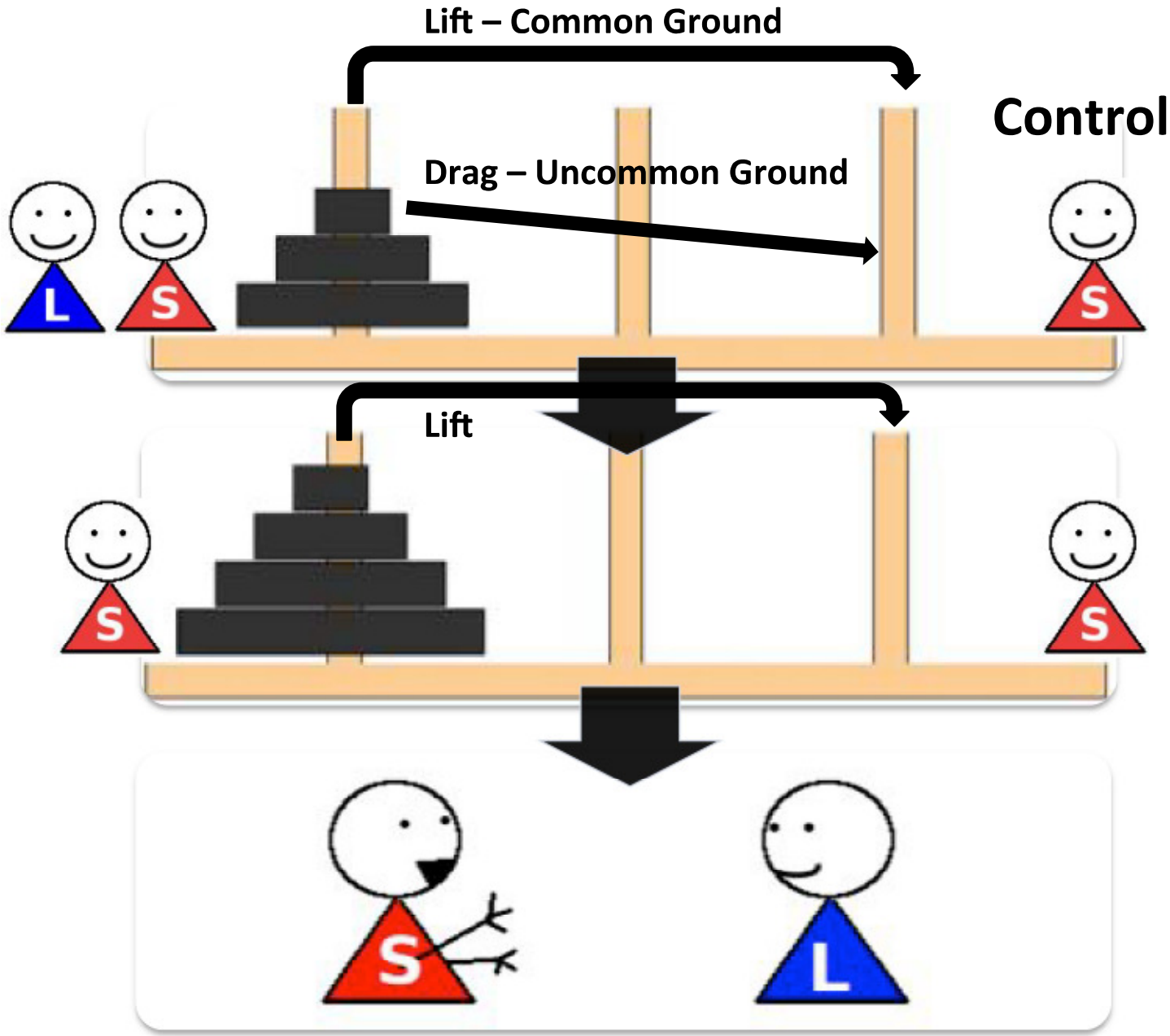
- No explicit mention of manner change
  - No speaker ever used the word “lift”
- No differences
  - Word count
  - Types of words
  - Token frequency

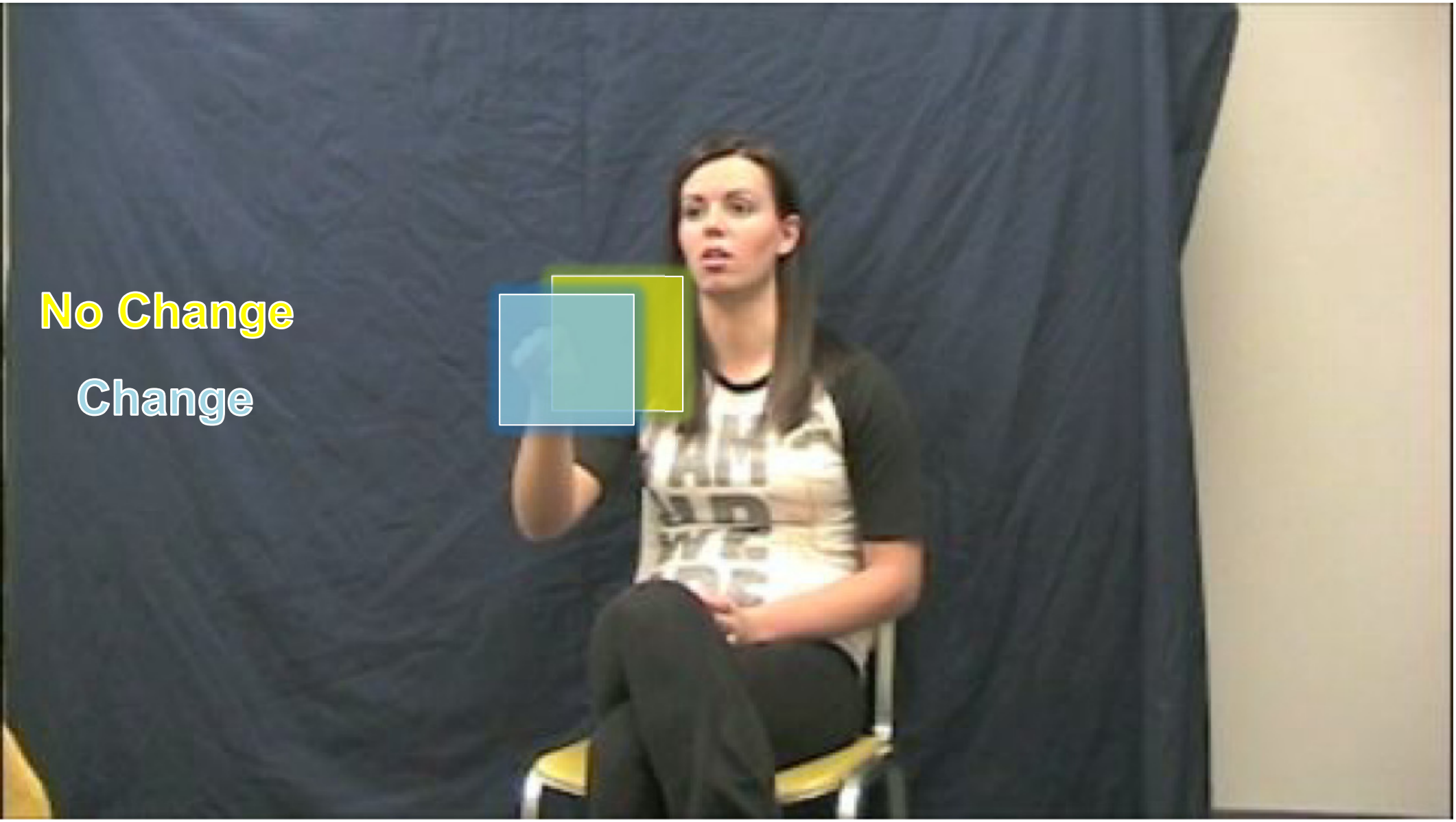
Results: Gesture

- Gesture rate – comparable for gest/word and gest/sec

- Gesture space
  - No differences in size of gesture – length or width of gestures produced
  - Gesture height


Uncommon Ground
Common Ground


Lift – Common Ground
Control
Drag – Uncommon Ground
L
S
S
Lift
S
S
S
L


No Change
Change

Pragmatics

- Speakers modulate their gestures when they know their listener has misleading or no information
  - Raise them up higher
  - More informative for listeners!

Are speakers gesturing for their listeners?

Yes.

- Speakers adjust the form of their gesture for their listener

- Do speakers adjust the content of their gestures for their listener?

Do speakers adjust the content of gesture?

- Examined mother-child communication about safety

- Stakes are high when mothers think a situation is unsafe that their child thinks is safe (disparity)

Do speakers adjust the content of gesture?

- As a function of what speaker thinks listener needs to know

1	2	3	4
very safe	kind of safe	kind of unsafe	very unsafe

Safety Rationales

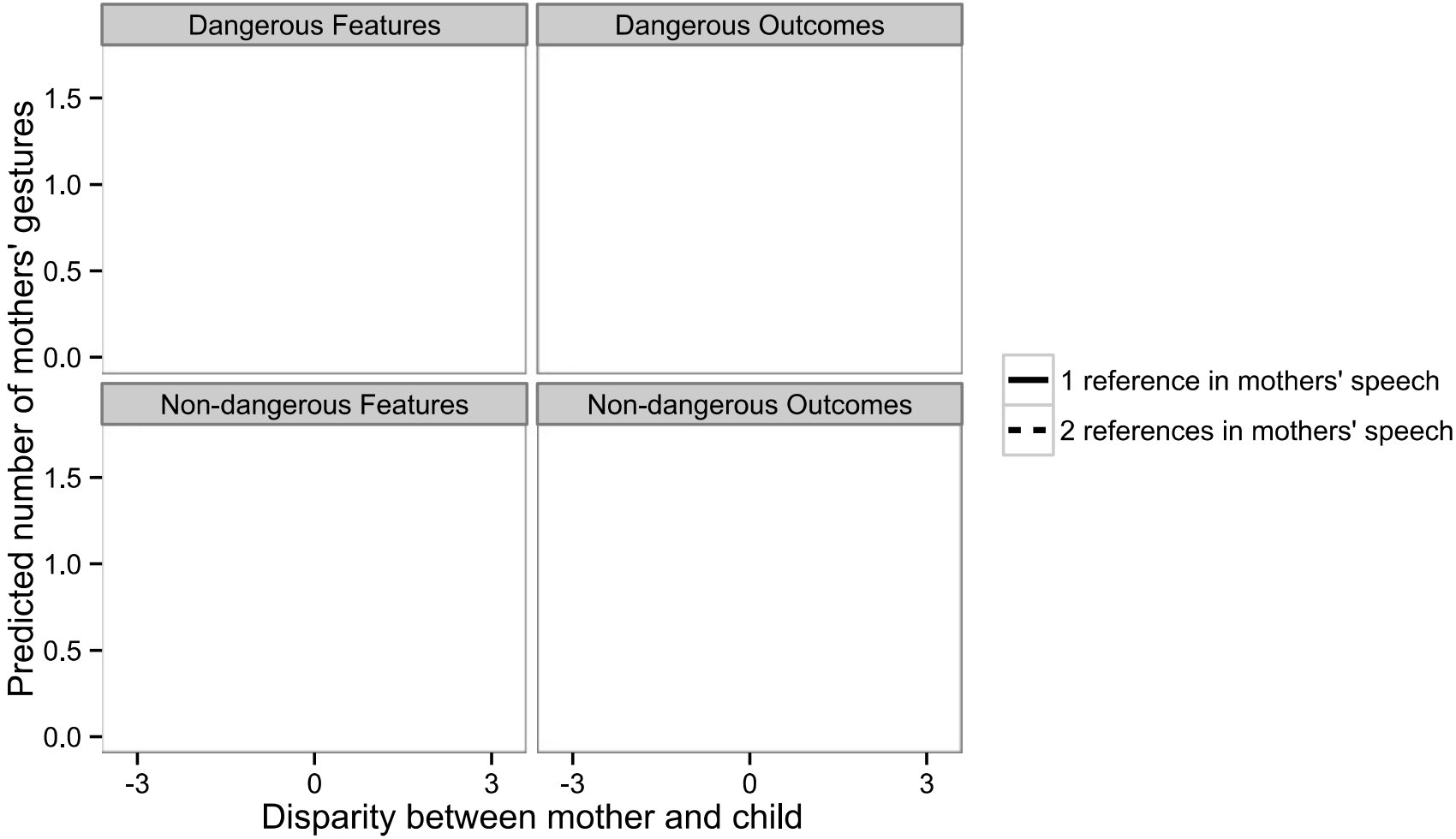
- References to **features** – directly observable
  - Dangerous – *He's not wearing a helmet.*
  - Non-dangerous – *That hill is not very steep.*
- References to potential **outcomes** – could occur
  - Dangerous – *He could get hit by a car.*
  - Non-dangerous – *He can easily turn before the street.*

What does the listener need?

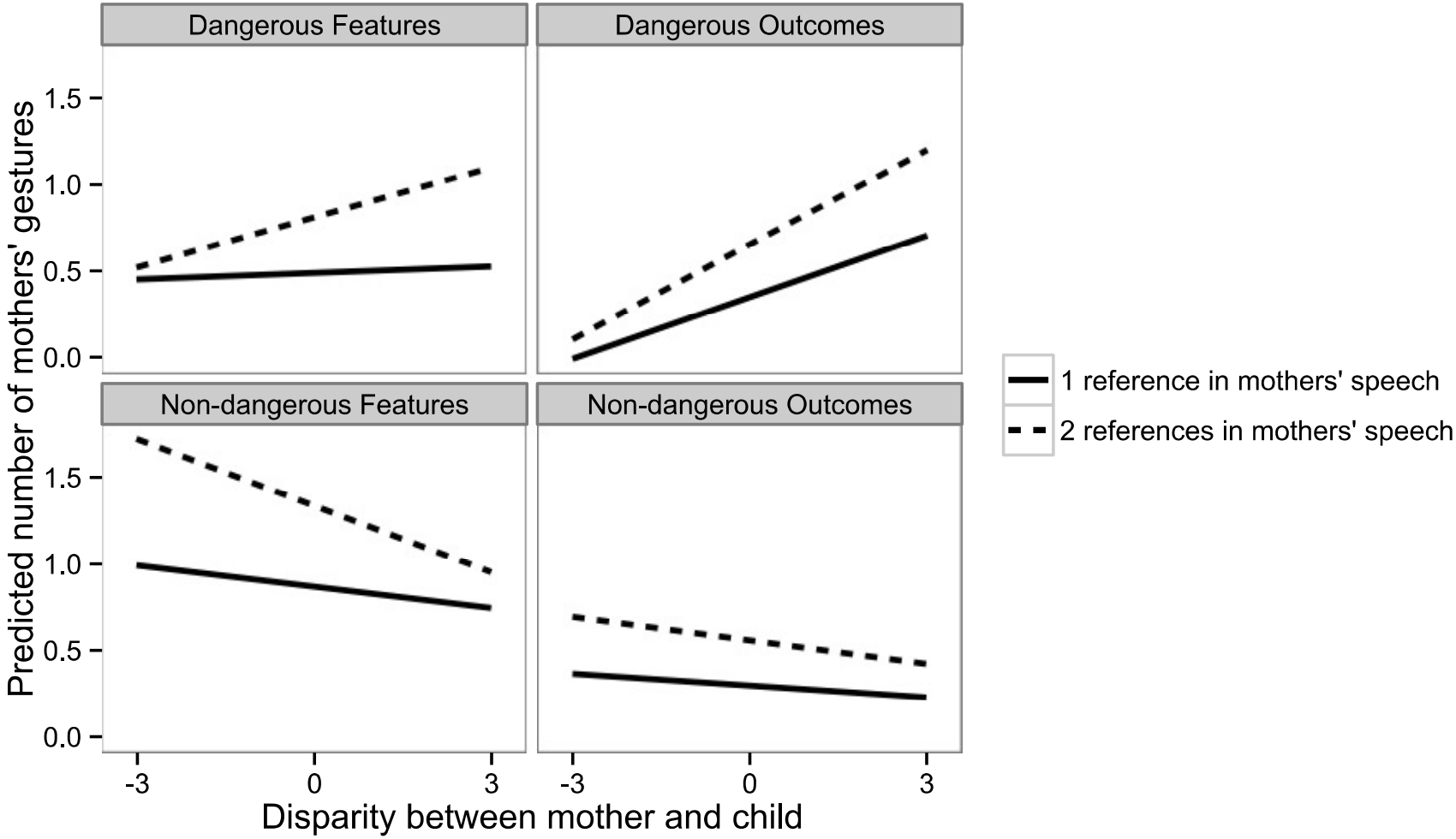
- Mom thinks dangerous, child thinks safe
  - > Child needs to know about danger

- Mom thinks safe, child thinks dangerous
  - > Child needs to know about safety

Mothers' gesture production as a function of disparity



Mothers' gesture production as a function of disparity



Information is highlighted.

- Mothers use gesture to highlight relevant information for their children.

- Child’s perspective informs mothers’ gesture

Speech and gesture

- Produced by a single body
- Communicate a single idea
- Complementary encoding of information
- Flexibly deployed

What about listeners?

Are listeners sensitive to *gesture*?

Gesture, or ???

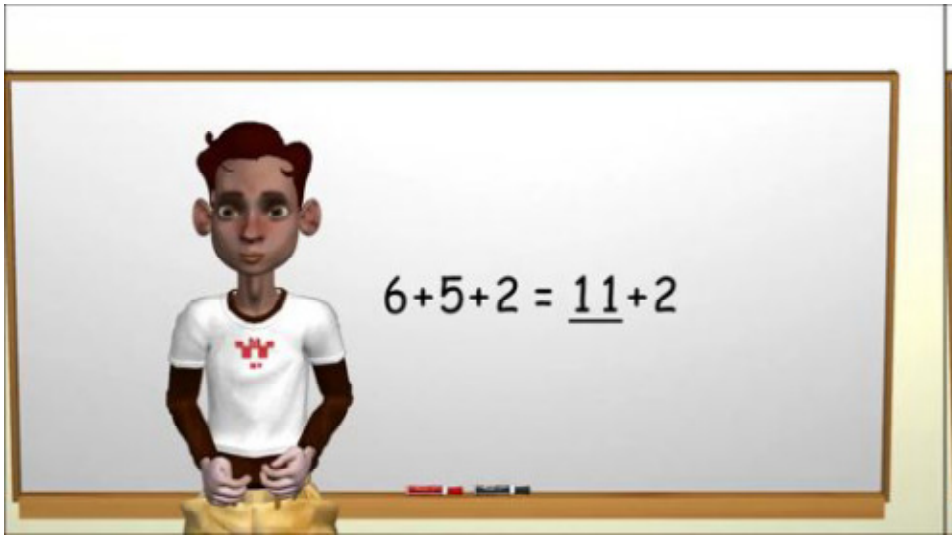
- Eye gaze
- Prosody
- Timing
- …

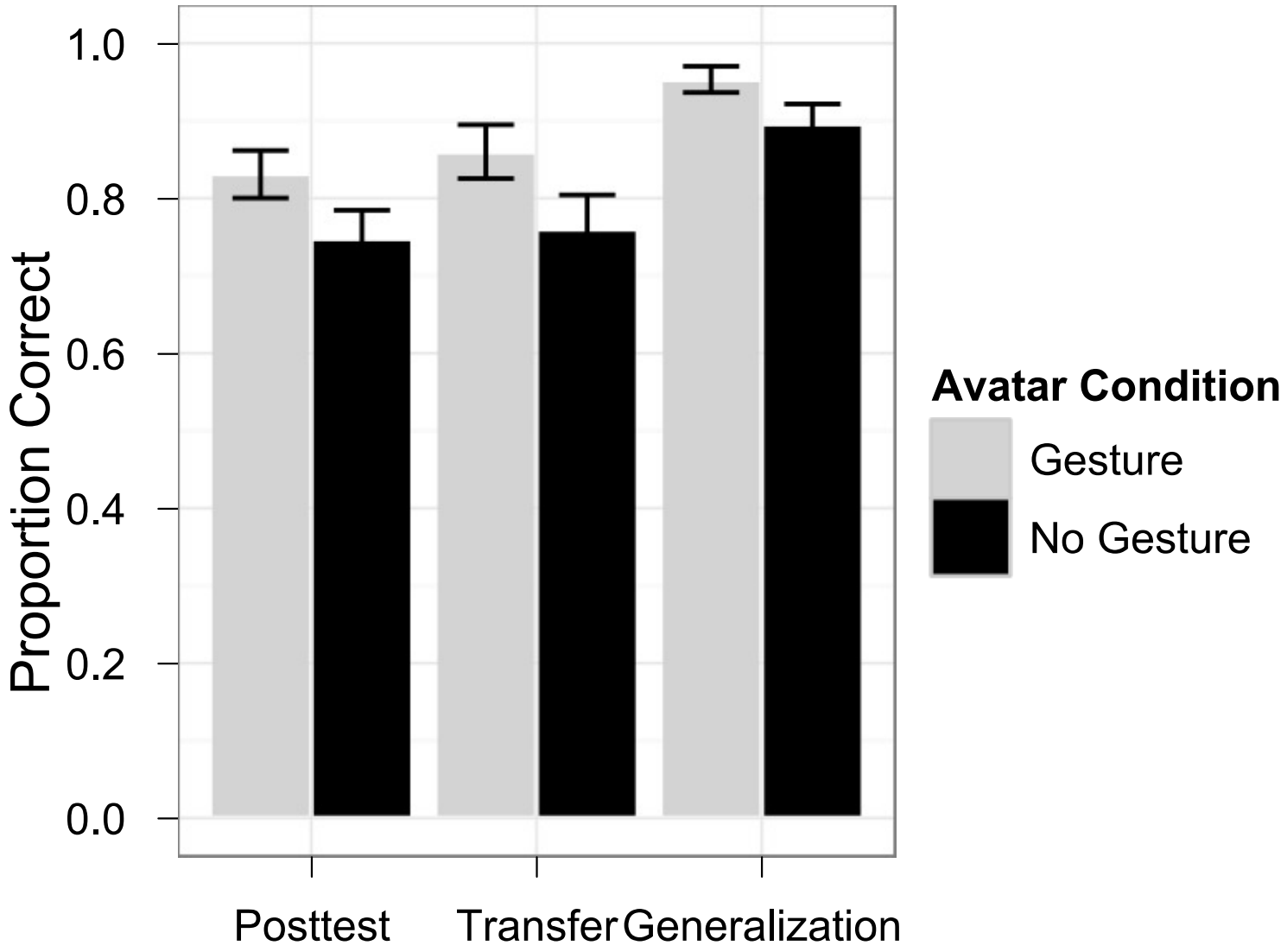
Avatar Study

- Use a gesturing avatar to control for other nonverbal behaviors

  (eye gaze, posture, …)

Avatar Study




1.0
0.8
0.6
0.4
0.2
0.0
Proportion Correct
Posttest
Transfer
Generalization
Avatar Condition
Gesture
No Gesture

Avatar Study

- Reveals that ***gesture*** seems to be important, at least for learning

- Hands are important!

What are listeners sensitive to when they are sensitive to gesture?

- Hands and bodies
- Temporal coordination

Synchrony versus bodies.

- Need synchronous non-bodily stimulus

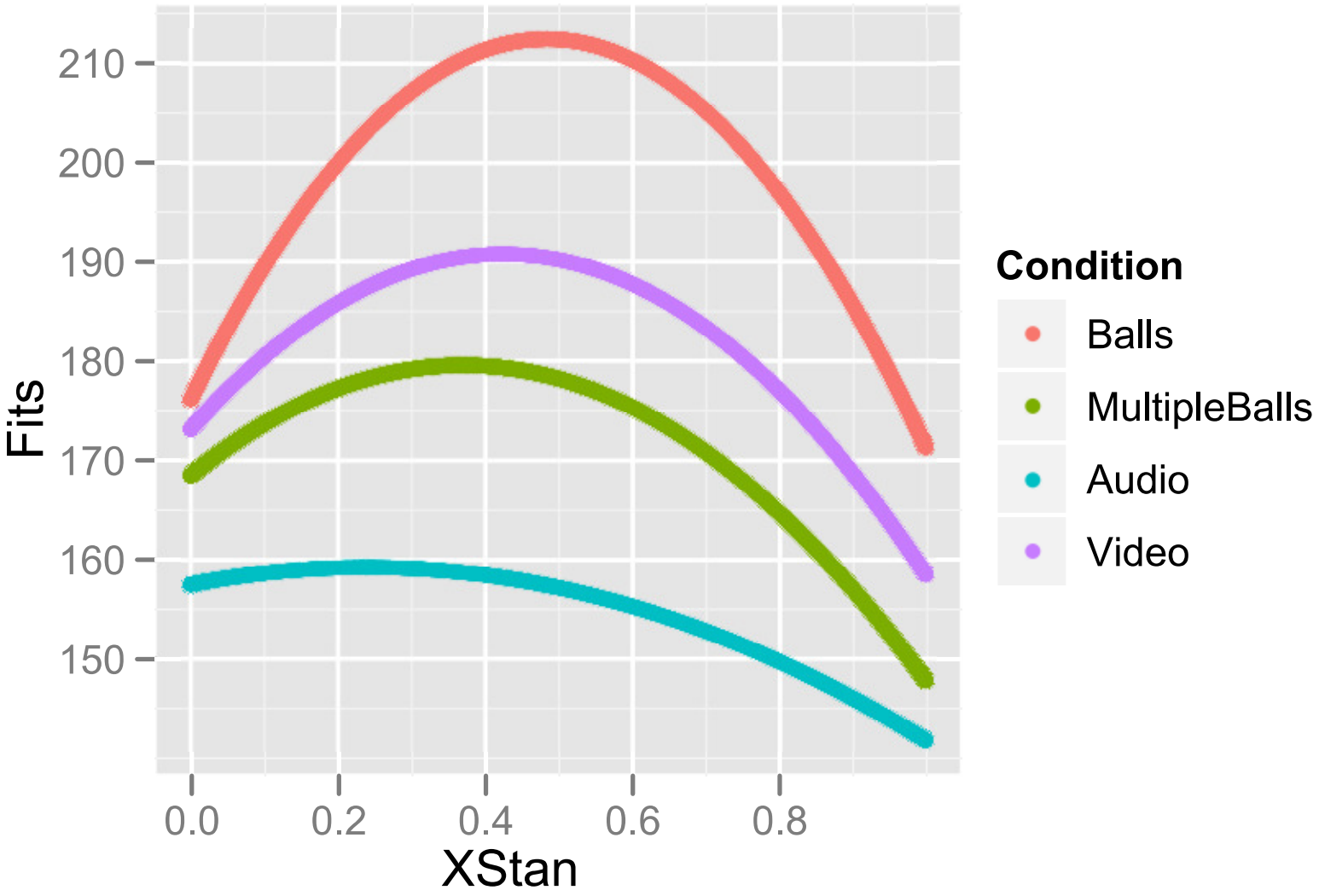
-> Balls that depict hand trajectory

Ball Perception

- Perception study
  - Audio
  - Visual + Audio (Speech + Bouncing Balls)
  - Visual + Audio (Speech + Three Pairs of Bouncing Balls)
  - Video

[redacted]

Real Objects Explanation



Real Objects Explanation

- Differences by video
  - Greatest curvature in the ball condition, then multiple balls, less in the video, curvature is not reliable in the audio

Implications

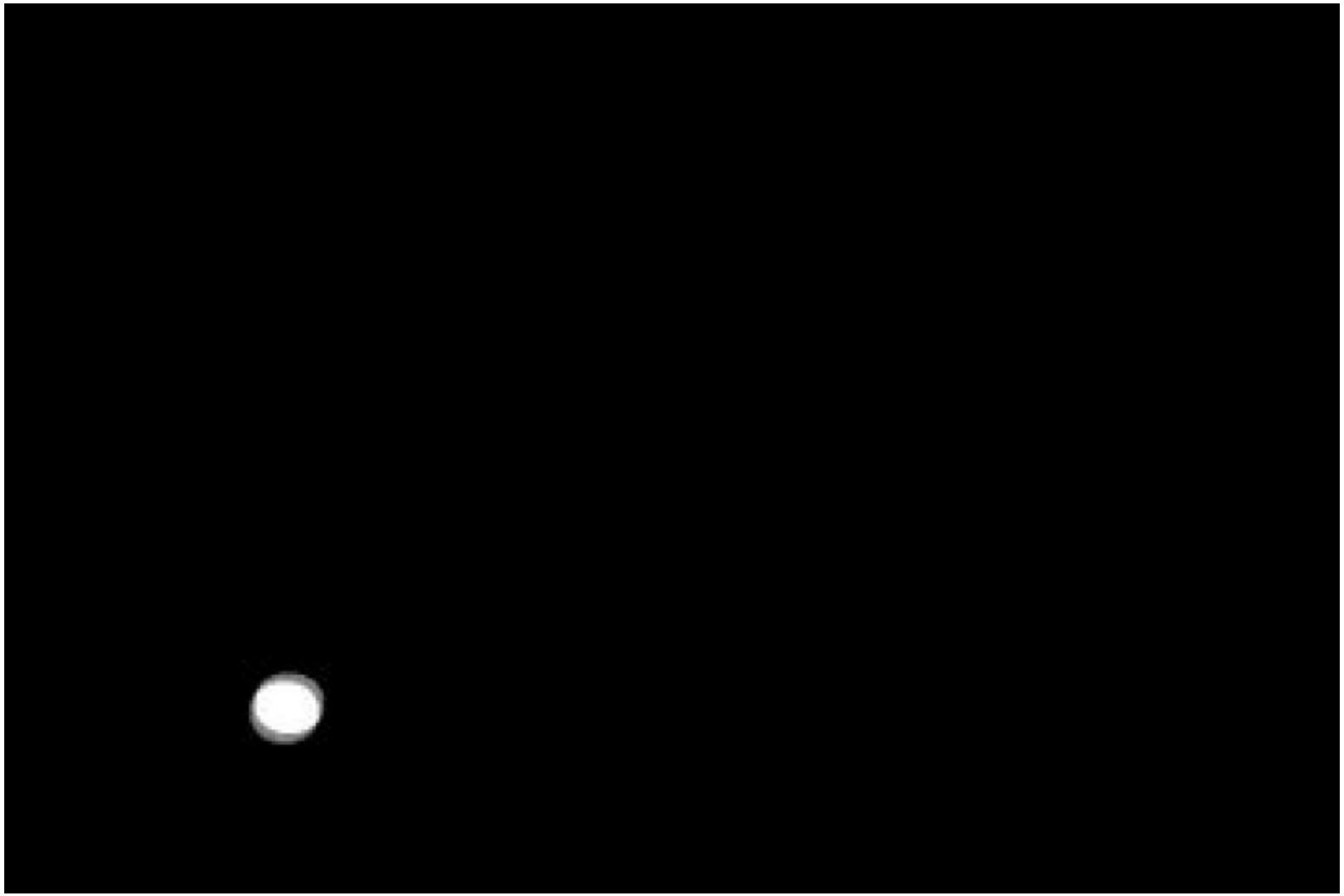
- Suggests people integrate synchronous visual information with speech, even motion that is not necessarily speaker-produced
  - Subjects reported ignoring the balls

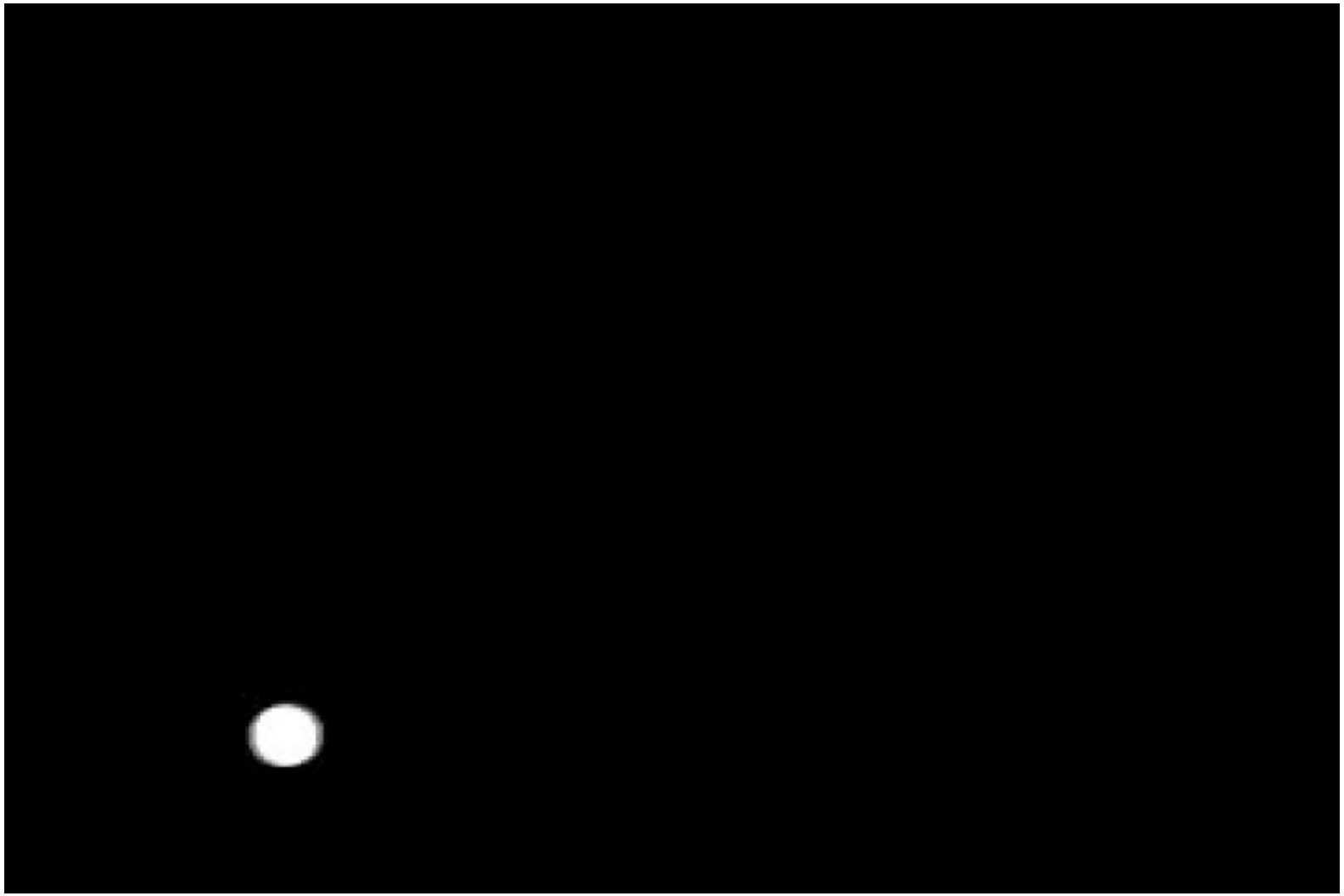
Asynchronous visual information

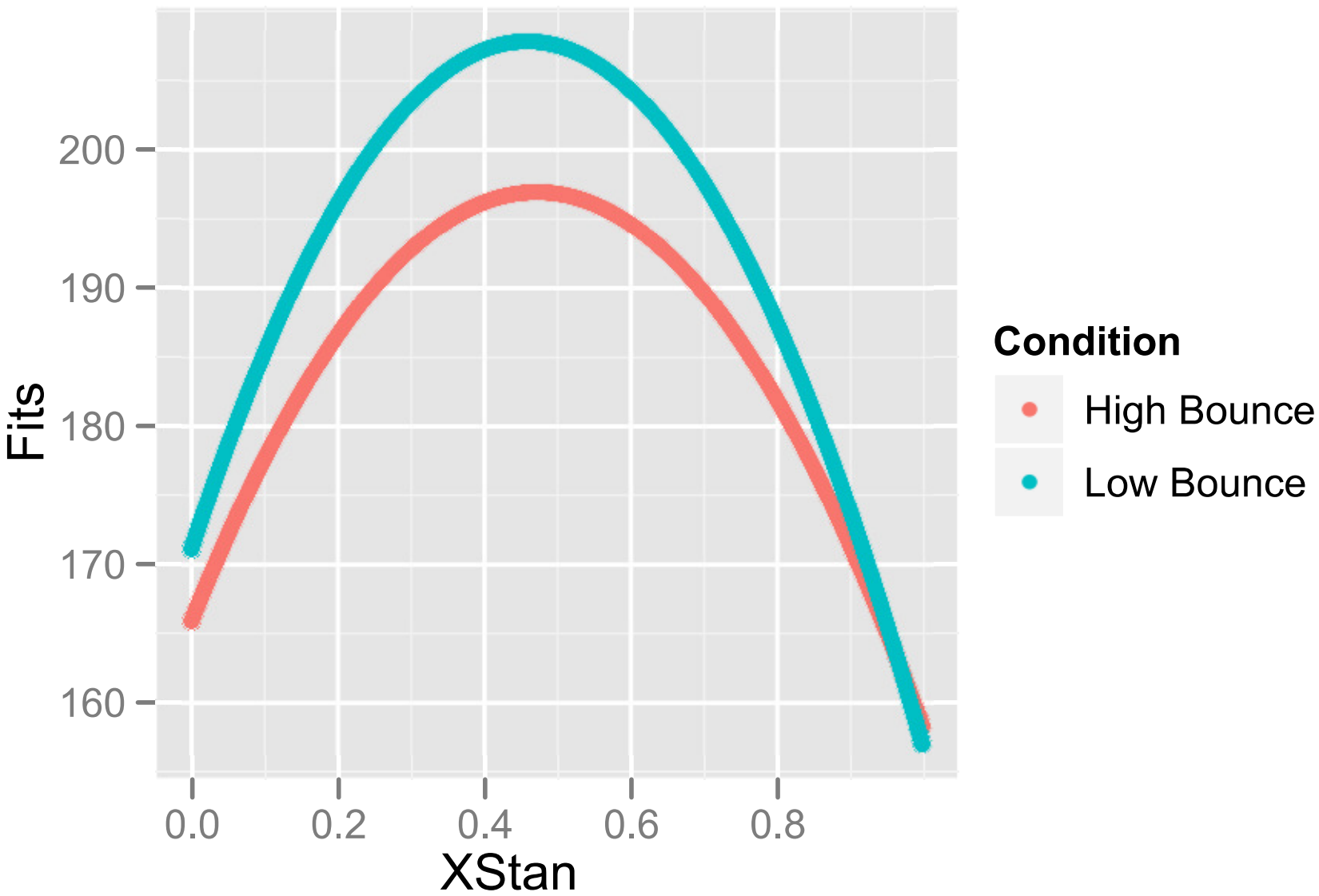
- What if listeners see a reliable visual signal that is not synchronous?

Bouncing Study

- Listeners watch a ball bouncing continuously during the audio

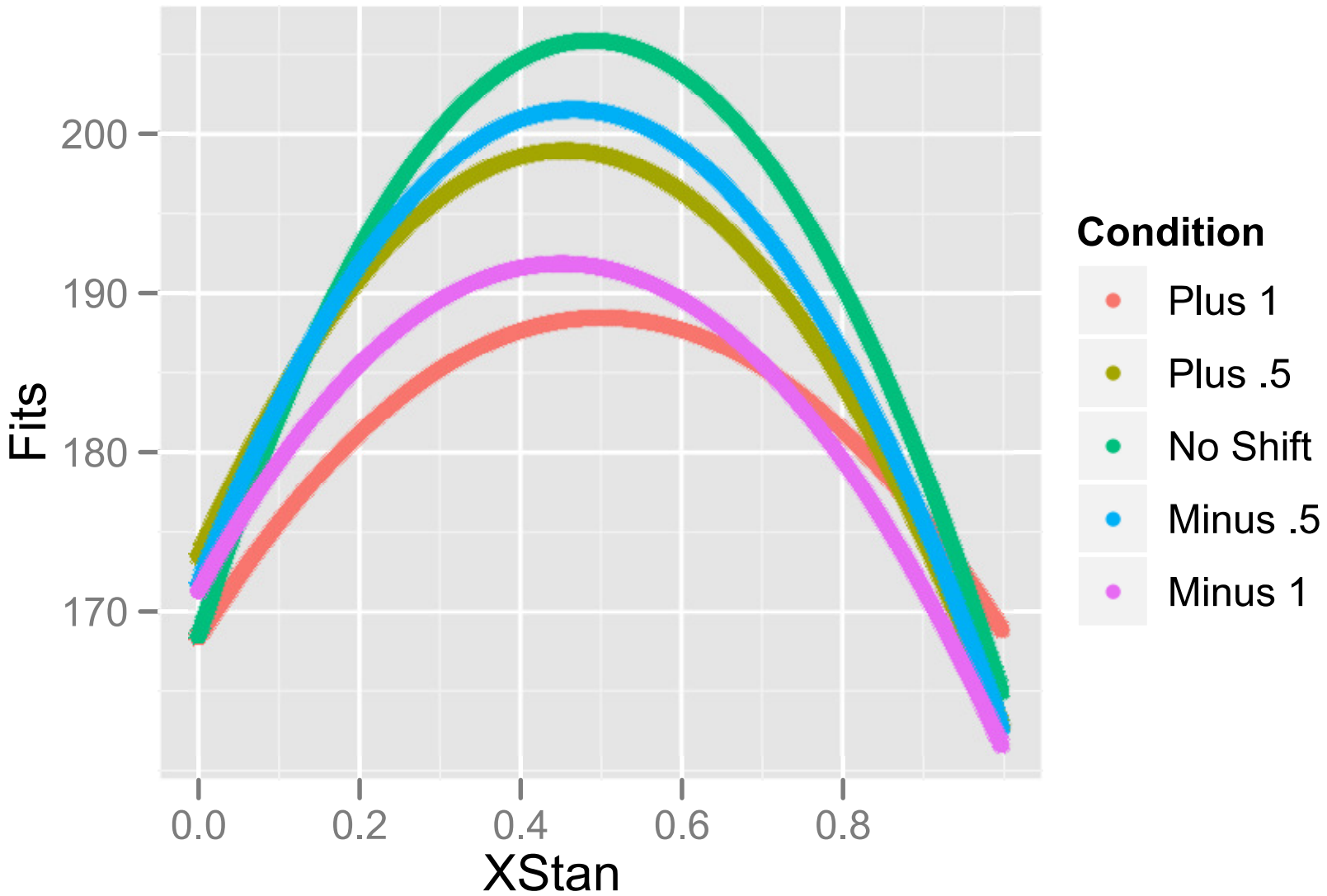





Fits
200
190
180
170
160
0.0
0.2
0.4
0.6
0.8
XStan
Condition
High Bounce
Low Bounce

Shifted video study

- Blurred face with audio and video are out of sync by .5 and 1 sec
  - -1 sec
  - -.5 sec
  - No shift
  - +.5 sec
  - +1 sec


Fits
200
190
180
170
0.0
0.2
0.4
0.6
0.8
XStan
Condition
Plus 1
Plus .5
No Shift
Minus .5
Minus 1

Shifted video study

- Already less integration with .5 second shift

- Less fall-off with visual ahead of auditory

Listeners are particularly sensitive to synchrony.

- Synchrony promotes integration across channels/modalities.
- Visual can anticipate auditory but not vice versa.

Speech and gesture

- Produced by a single body
- Communicate a single idea
- Complementary encoding of information
- Flexibly deployed

- Synchronously perceived

What is embodied cognition?

Wilson, 2002:

1.) Situated – involves perception and action

2.) Time pressured – real time interaction with the environment

3.) Off load onto environment

4.) Environment part of cognitive system

5.) Cognition is for action

6.) Off-line cognition is body based

Conclusions

- Gesture has broad impact on speakers and listeners, both in the moment in which it is produced, and over time.

- Communication is truly embodied.

Vielen Dank!

Funded by

NSF Grant 1217137

IES Grant R305A130016

DeLTA Center

Thanks to

Caitie Hilliard

Andrew Allyn

LDG

Emily Mudryj
Kelsey Walsworth
Stephanie Packard
Tana Truelove
Caitlin Polz
Lindsey Miller
Sieon Kim

Thanks to