Using free-view eye-tracking to study spoken language processing

Instructors: (LSA 363)

Michael K. Tanenhaus

(mktanenhaus@gmail.com)*

John C. Trueswell

Graduate Assistants:

Katie Carbary

Austin Frank

Natalie Klein

Course Website:

http://www.ircs.upenn.edu/igert/LSA363

Overview*

- **We review *some* of the "visual world" literature with an eye towards:**
  - ***highlighting how visual world paradigm can be used to address issues in:***
    - **phonetics**
    - **spoken word recognition**
    - **parsing,**
    - **reference resolution**
    - **production**
    - **experimental pragmatics and interactive conversation**
  - ***considering methodological issues that arise when using eye movements to examine spoken language processing***
  - ***providing students with some hands-on experience designing visual world experiments***

*** all bad puns should be attributed to JCT unless otherwise noted**

Lectures

Date	Topic and lecturer
7/5	intro, eye movements (JCT, MKT)
7/9	speech, words, pitch accents (MKT)
712	sentence processing, expectation (JCT)
7/16	referential domains, implicature (MKT)
7/19	parsing to learn and vice versa (JCT)
7/23	production, perspective-taking (JCT, MKT)
7/25	conversation, future directions (MKT, JCT)

Labs and Discussion

- Times to be arranged
  - Tuesday and Wednesday (We need three 2 hour blocks)*
  - Discussion and demos with head-mounted ISCAN, TOBII portable

- Labs:

  Lab 1: Referential ambiguity and point of disambiguation (T, W July 10-11th)
  Lab 2: Speaker eye gaze in verb learning and parsing (T, W July 17-18th)
  Lab 3: Common ground in spoken conversation (T, W July 24-25th)

* Each student attends one section each week (your assigned section)

Why use eye movements to study spoken language in a “Visual World”?

Why spoken language?

- Talking and listening are primary
  - All societies have spoken language
  - All children learn to converse, initially by talking about the world.
- Other forms of language are derivative
  - Most languages don’t have writing systems

Ignoring speech can encourage scientific balkanization, making it easier to ignore duality of patterning:

speech perception--> **word recognition**-->syntactic processing--> sentence interpretation--> **discourse** representation

Duration:

cap*/*cap*tain** (in strong position, vowel in ***cap is longer)

But information structure of discourse can affect duration.

Put the cap above the captain. Now put the cap/captain…

Put the CAPtain

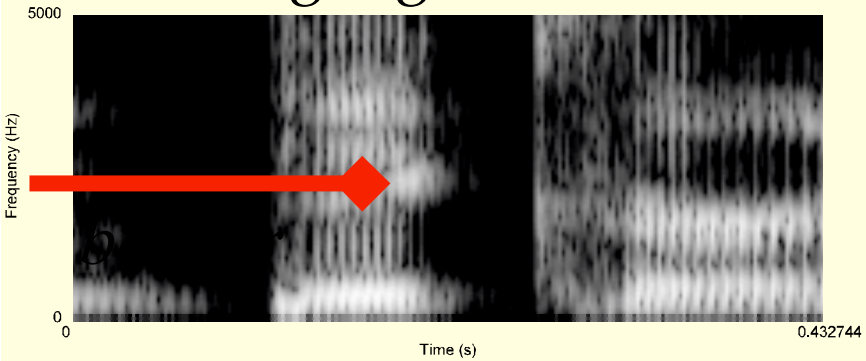
Moral: "high" and "low" level subsystems can share the same input data

Why use saccadic eye movements to study spoken language?

- Consider what spoken language is like from two perspectives
  - Language unfolding over time as a sequence of transient acoustic events.
  - Conversation as a joint activity between two or more people.

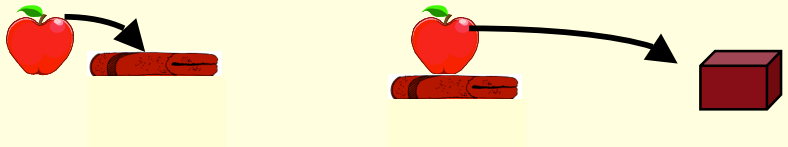
The unfolding signal

1. Language unfolds over time.



beetle, beacon, beak, beep…

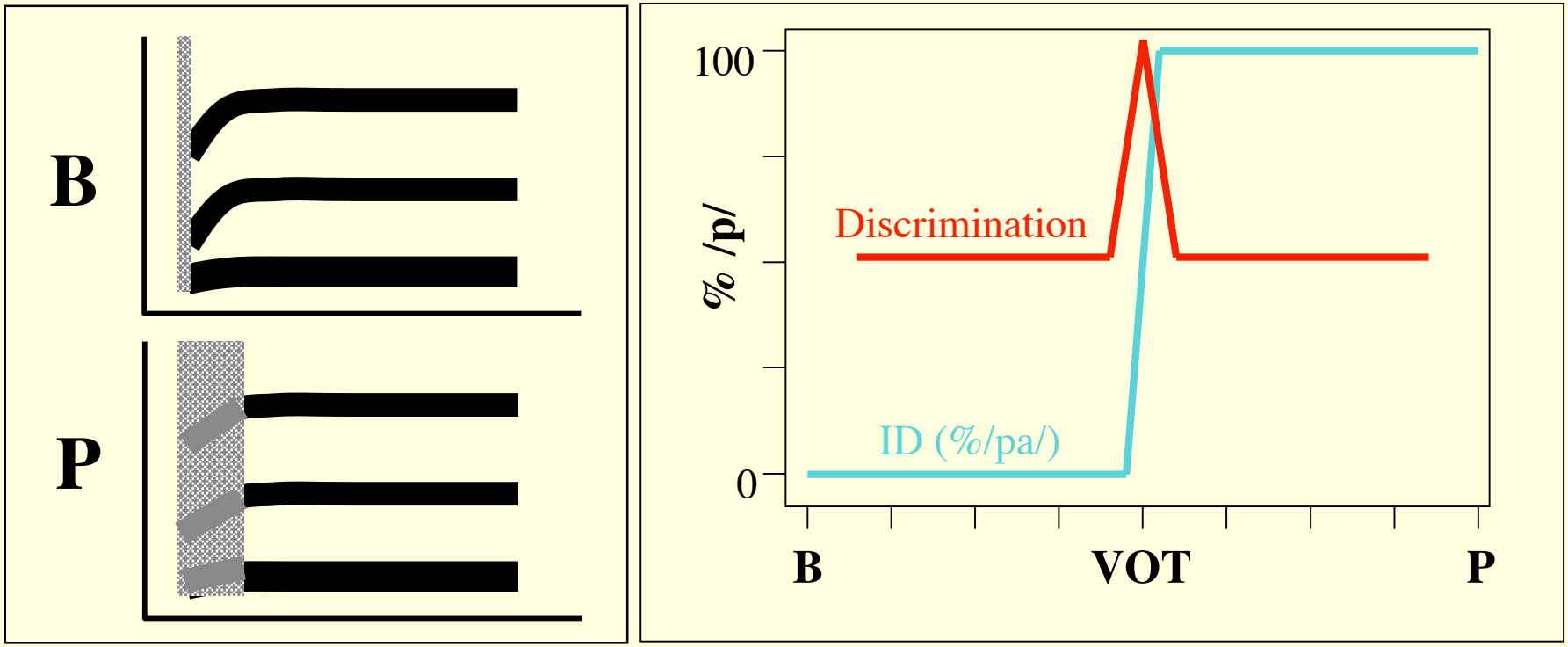
Put the apple on the towel…



2. Processing is closely time-locked to the unfolding utterance.

3. Requires monitoring moment-by-moment comprehension with careful control of the signal.

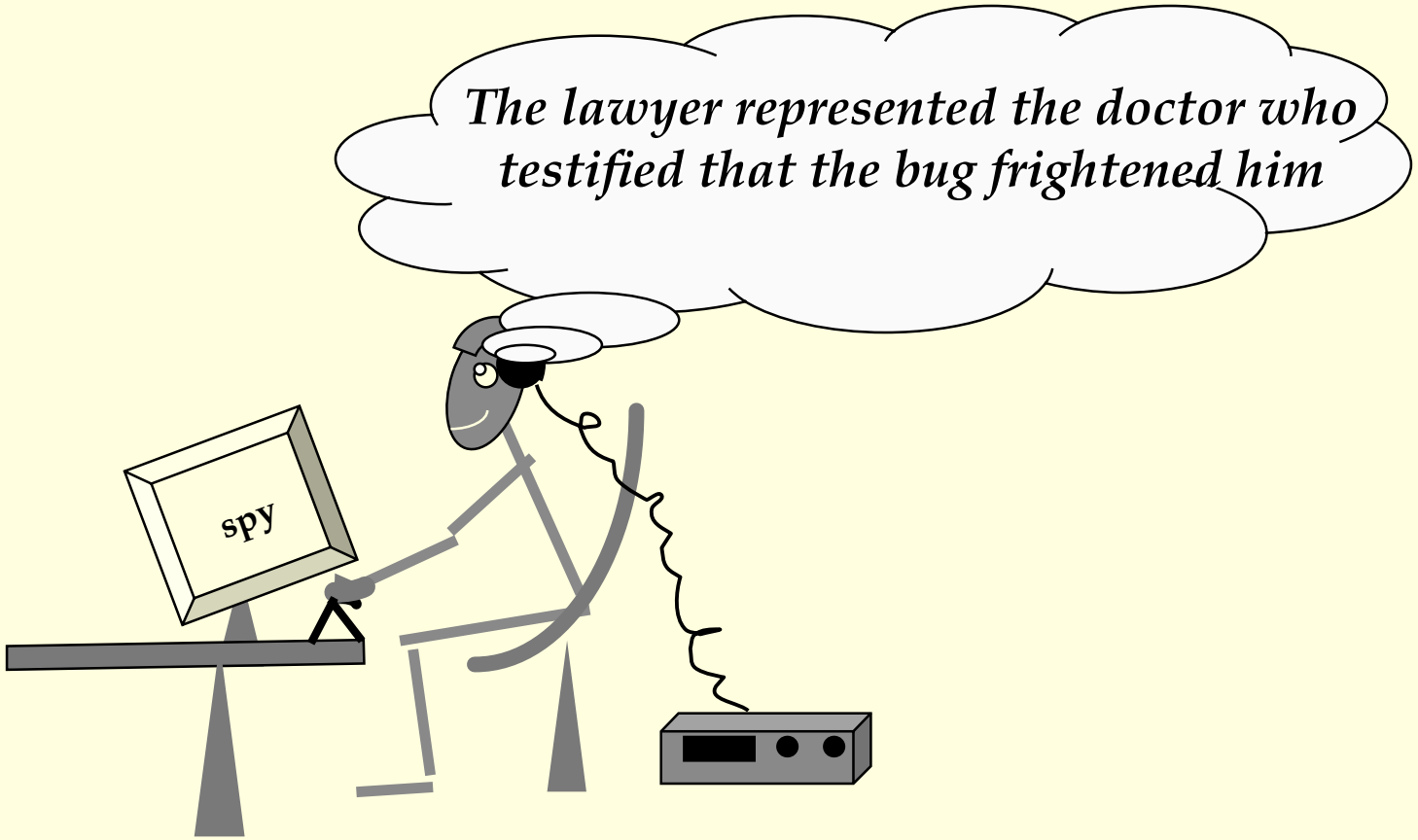
Example 1: Categorical Perception



Example 2: Sentence processing

Paradigms have required de-contextualized language

e.g., cross-modal priming

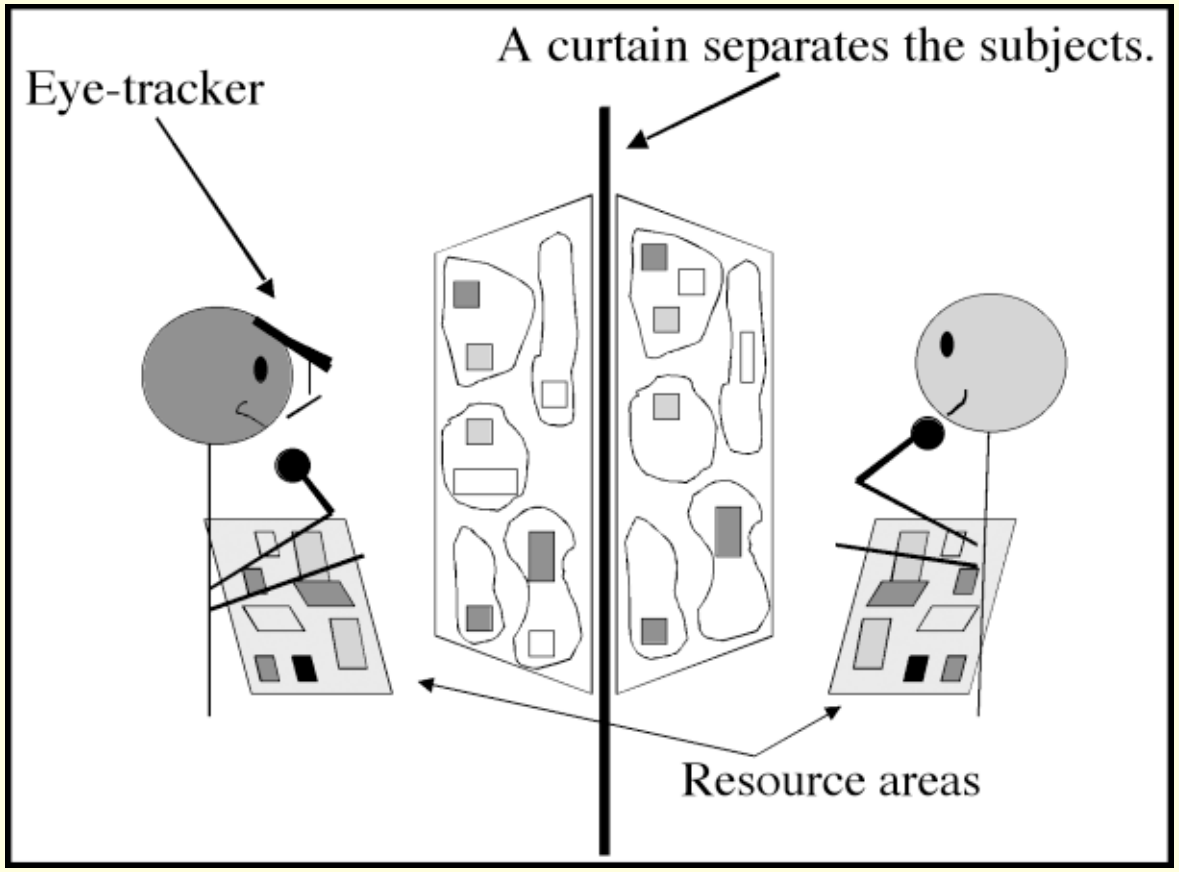


Interactive conversation

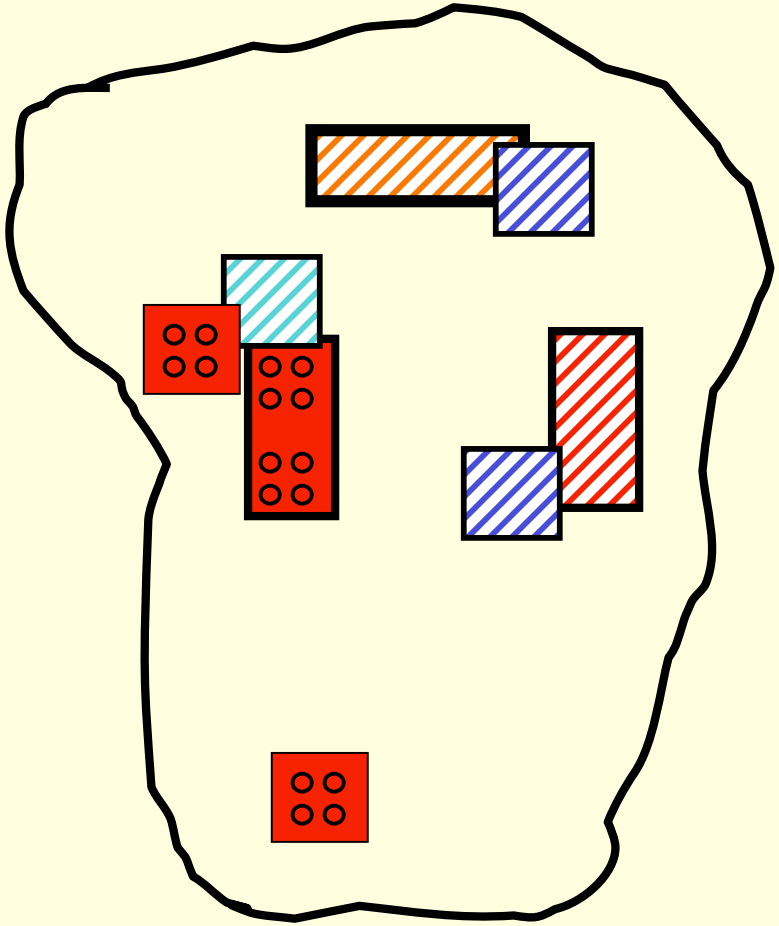
Gibberish?

1 *ok, ok I got it* **ele…**ok
2 alright, *hold on*, I got another easy piece
1 *I got a* well wait I got a green piece RIGHT above that
2 above this piece?
1 well not exactly right above it 
2 **it can't be above it**
1 it's to the…it' doesn't wanna fit in with the cardboard
2 it's to the right, right?
1 yup
2 **w- how? *where***
1 ***it's*** kinda line up with the two holes
2 line 'em right next to each other?
1 yeah, vertically
2 vertically, meaning?
1 **up and down**
2 **up and down**

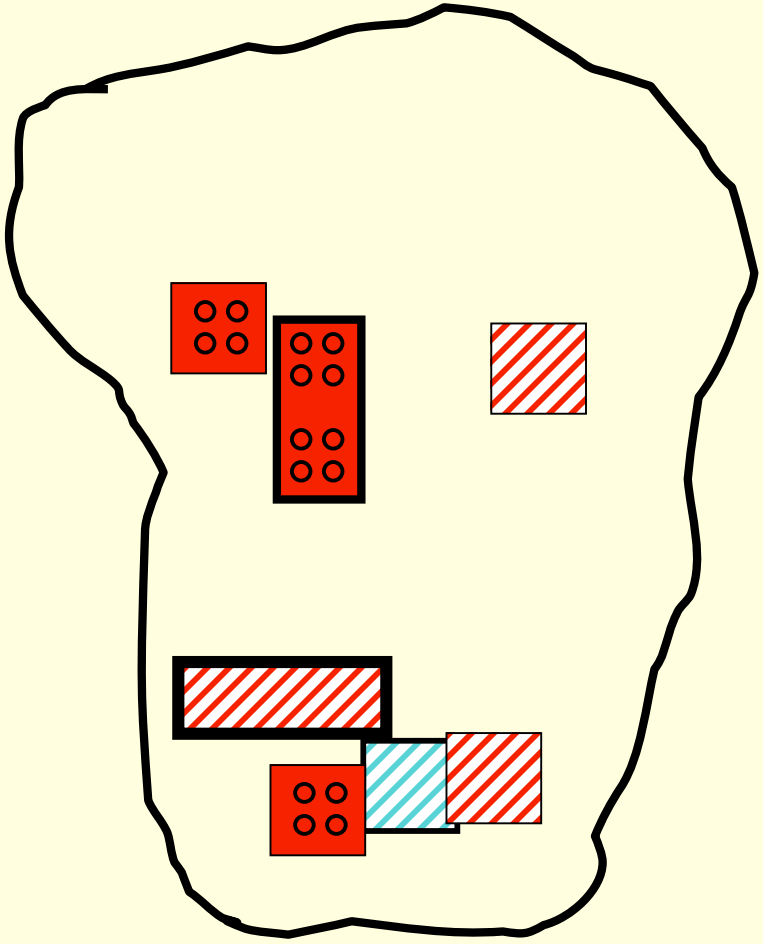
Brown-Schmidt & Tanenhaus (in press, *Cognitive Science*)



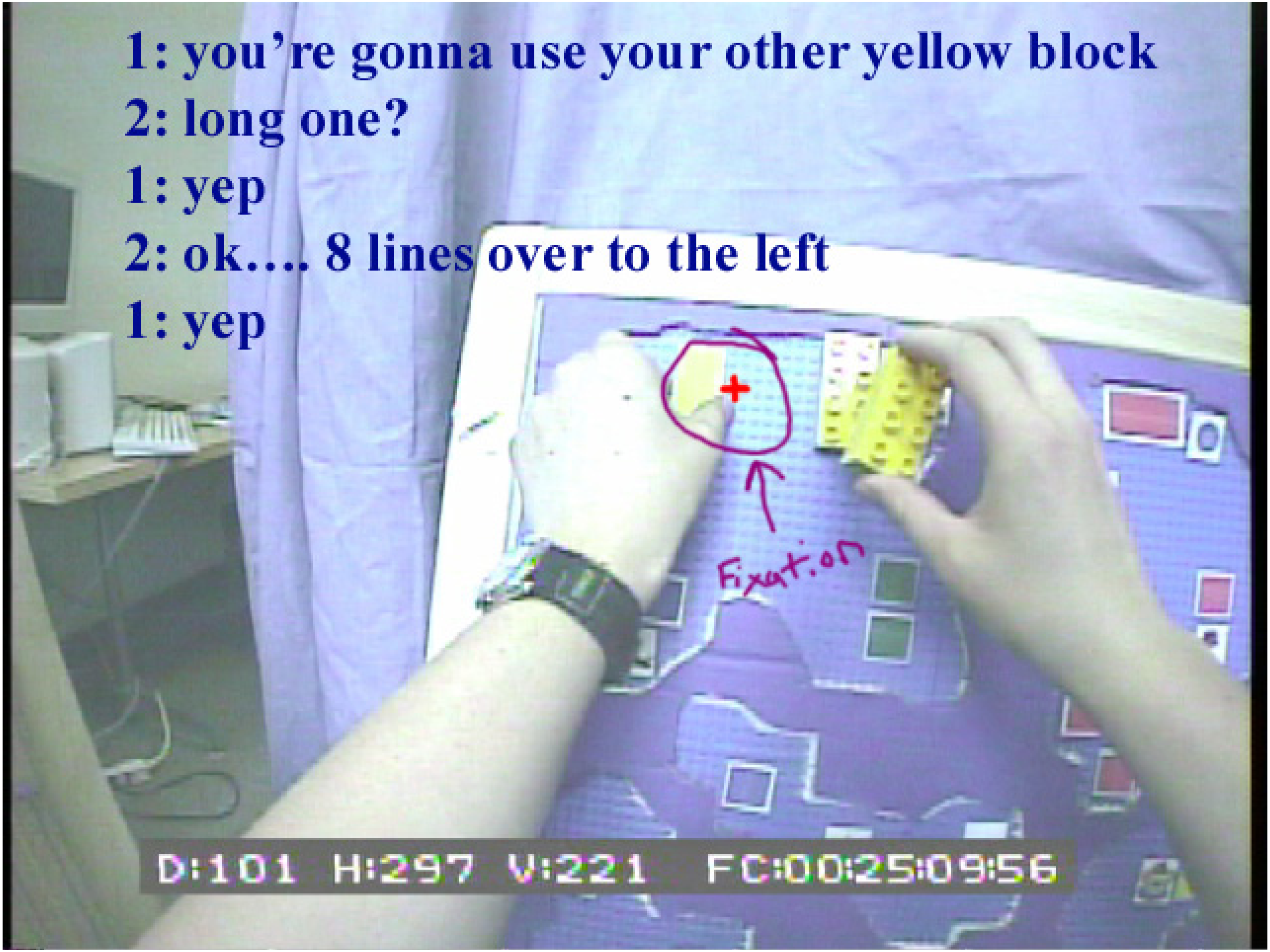
Participants replace stickers / with blocks
This is mid-way through the task:



Subject 1's Board



Subject 2's Board


1: you're gonna use your other yellow block
2: long one?
1: yep
2: ok…. 8 lines over to the left
1: yep
Fixation
D:101 H:297 V:221 FC:00:25:09:56

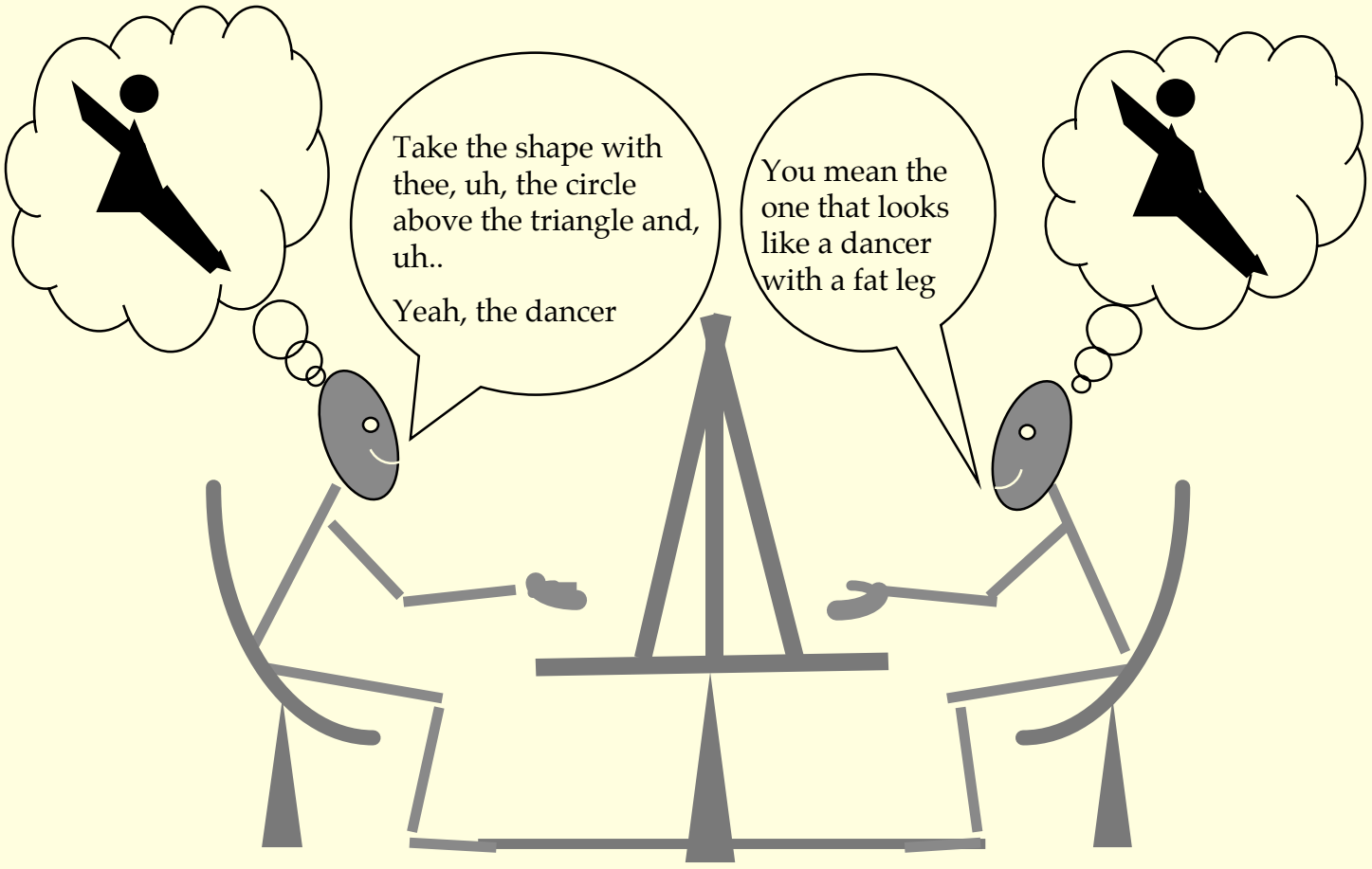
Interactive conversation

Gibberish?

1 *ok, ok I got it* **ele…**ok
2 alright, *hold on*, I got another easy piece
1 *I got a* well wait I got a green piece RIGHT above that
2 above this piece?
1 well not exactly right above it 
2 **it can't be above it**
1 it's to the…it' doesn't wanna fit in with the cardboard
2 it's to the right, right?
1 yup
2 **w- how? *where***
1 ***it's*** kinda line up with the two holes
2 line 'em right next to each other?
1 yeah, vertically
2 vertically, meaning?
1 **up and down**
2 **up and down**

Spoken Language in a Different Tradition

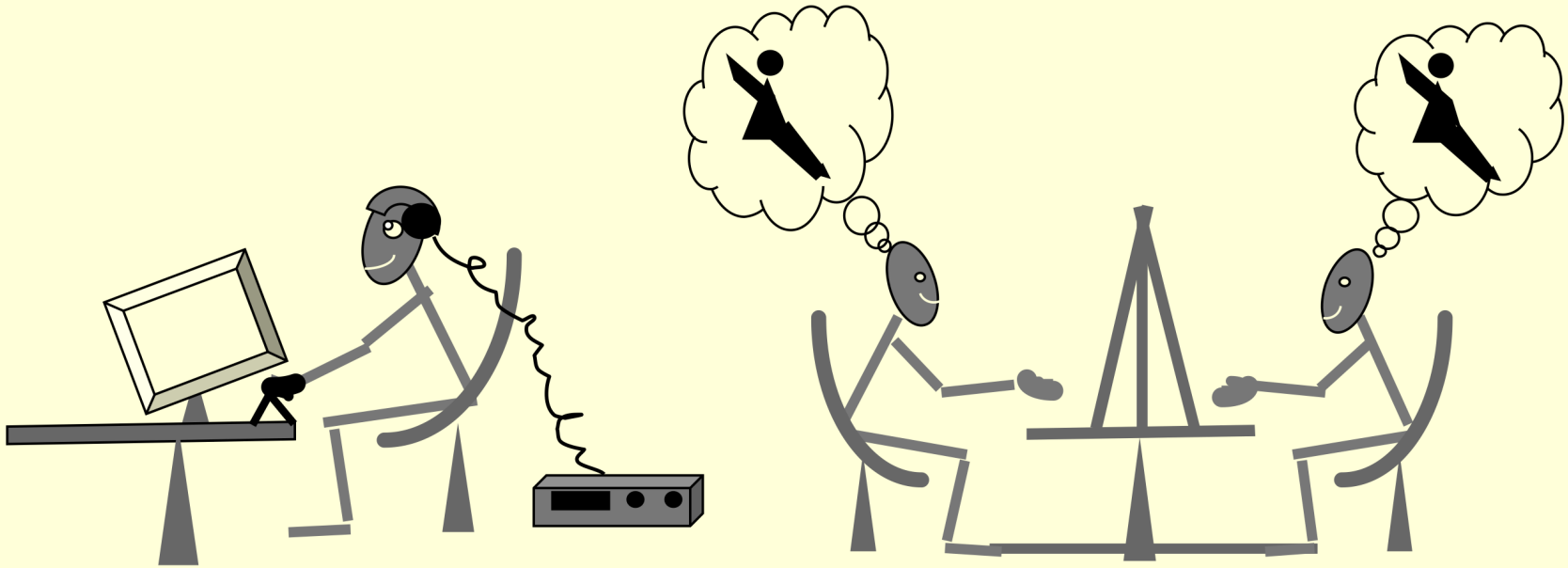
Prototypical experiment requires participants to interact in goal-driven task



Reasons for separate traditions

Methodological:

It has been difficult to study language in “natural” contexts with precision necessary to examine time-course.



Why eye movements?

- ballistic measure
- can be used with continuous speech
- natural response measure
- low threshold response
- subject is unaware
- does not require meta-linguistic judgment
- closely time-locked to speech
- plausible linking hypothesis:

> Probability of eye movement at time (t) is a function of activation of possible alternatives plus some delay for programming and execution (~200ms)

Eye movements allow us to study spoken language in rich contexts with precision necessary to examine time-course.

Is this a good thing?

No: (confuses language with non-linguistic stuff)

We should be studying how we construct/generate linguistic representations. In visual world, we introduce task-specific strategies, etc.

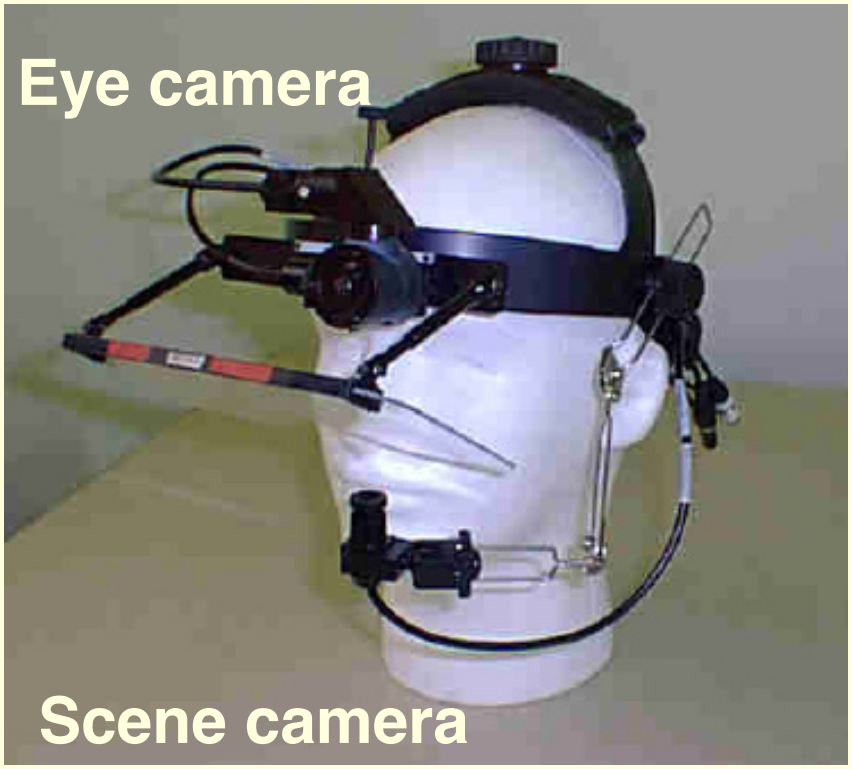
Yes:

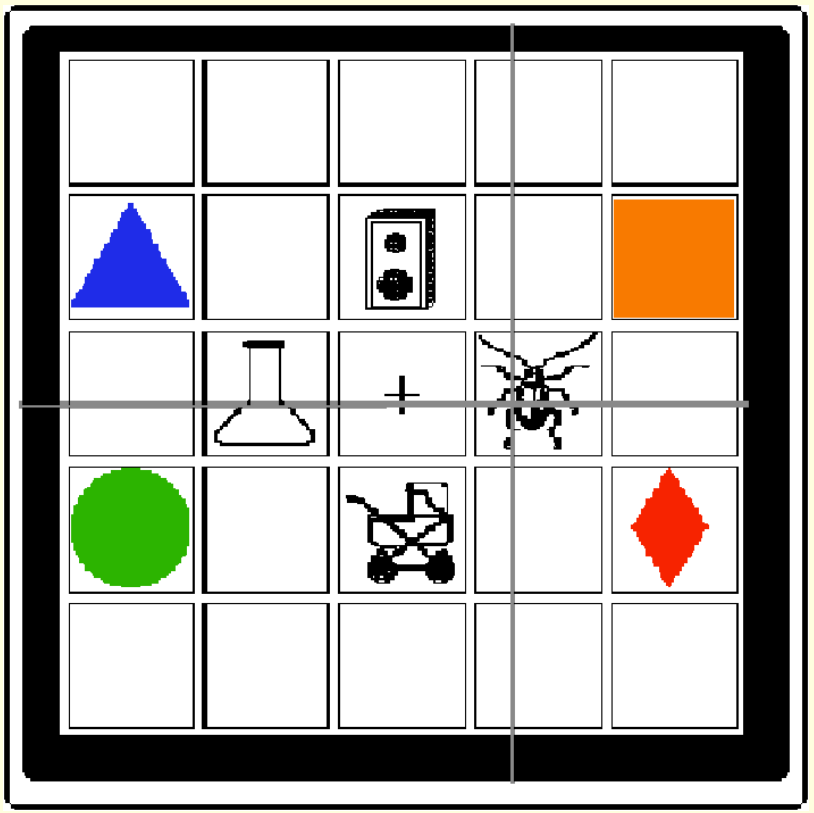
There is always a context, it always matters; understanding how the system works in a rich environment can be more informative about basic principles than studying the system in an impoverished environment (analogy with vision; bottom-up or top-down).

We can manipulate action/goals/language/display

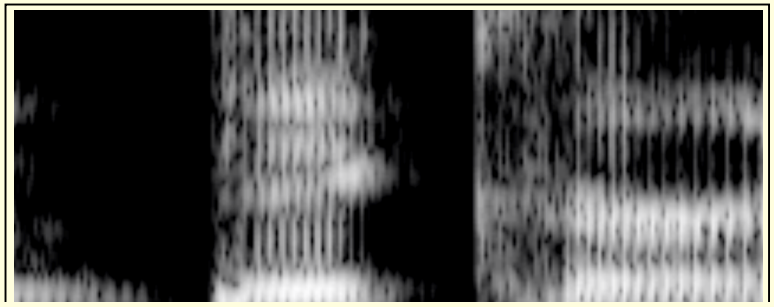
Cards Video

Allopenna, Magnuson & Tanenhaus (1998)

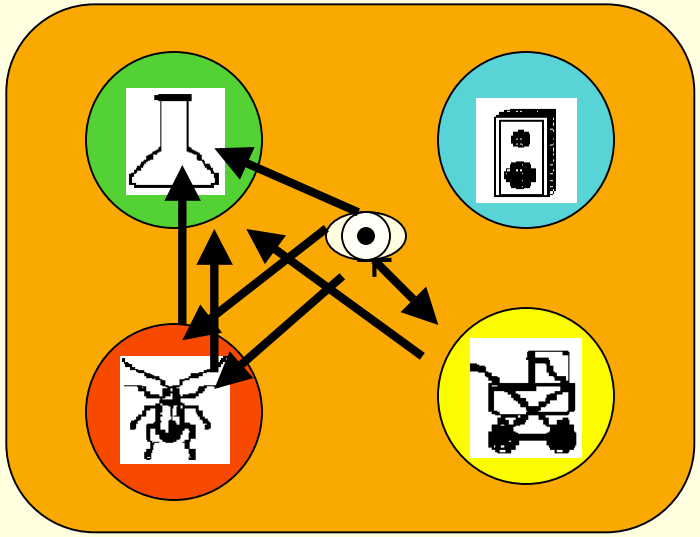


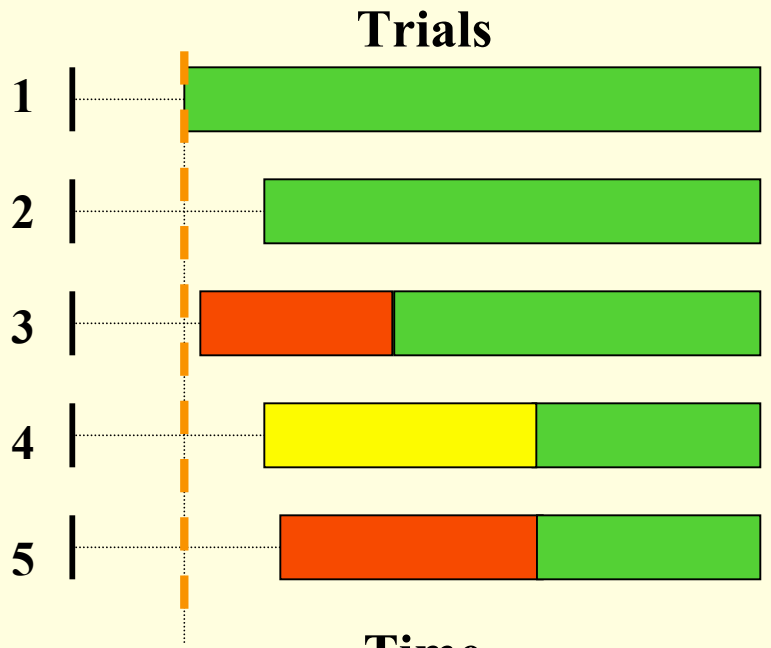


Pick up the beaker



Fixation Proportions over Time

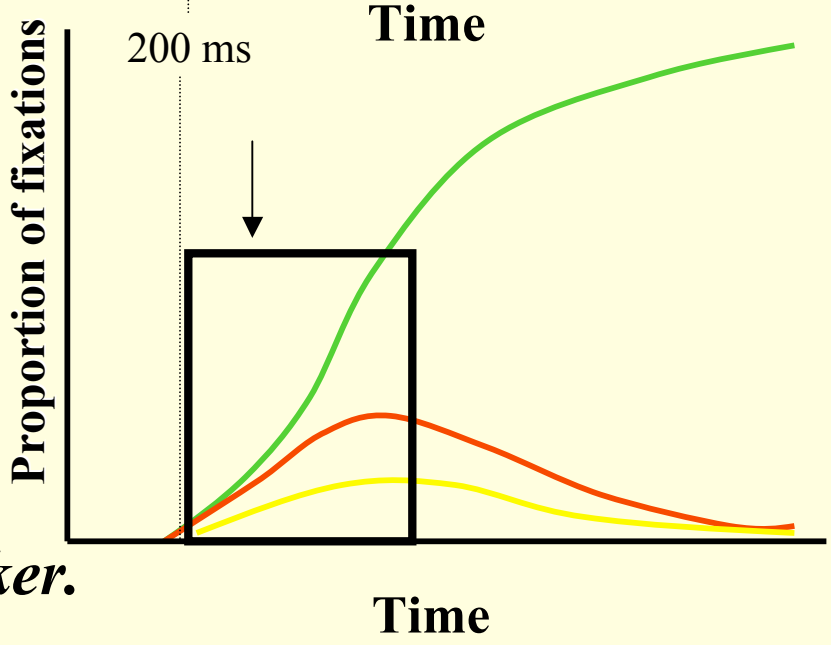




Target = beaker

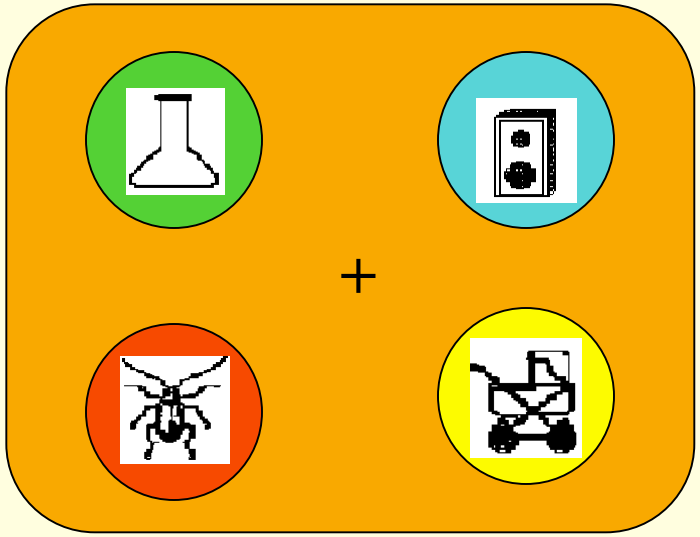
Cohort = beetle

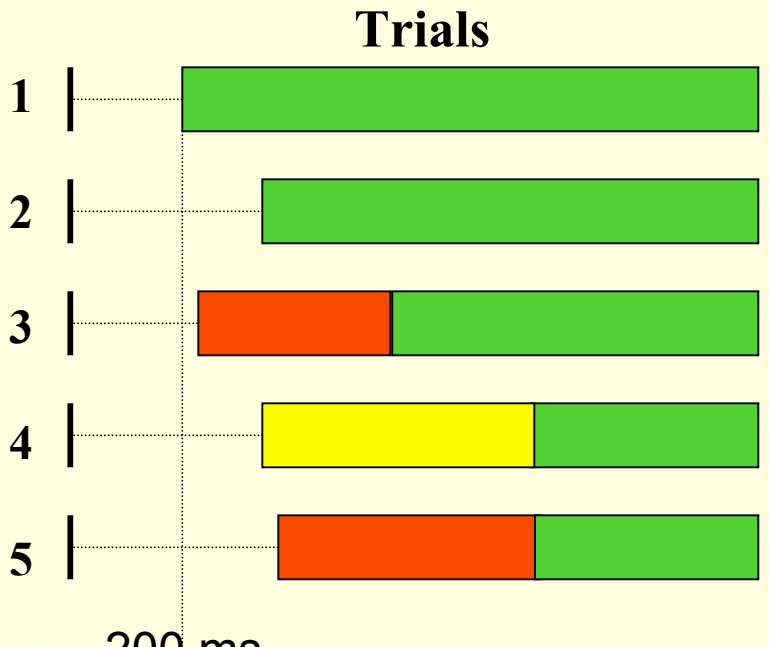
Unrelated = carriage

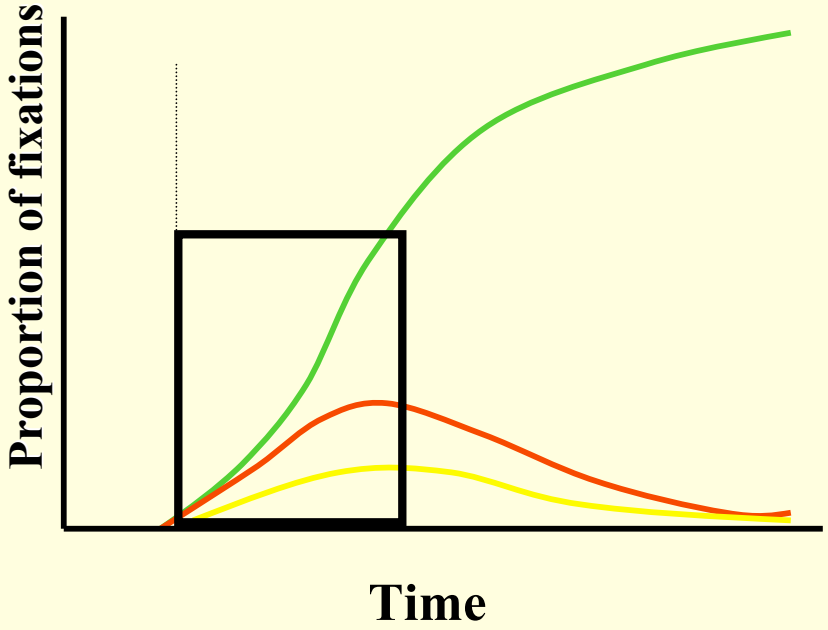


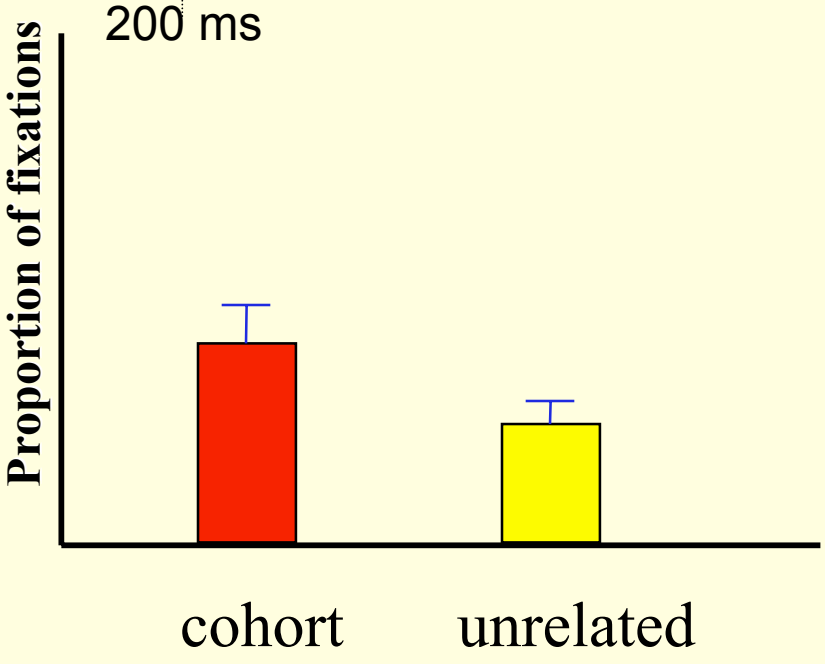
Look at the cross. Click on the beaker.

Fixation Proportions summed over an interval

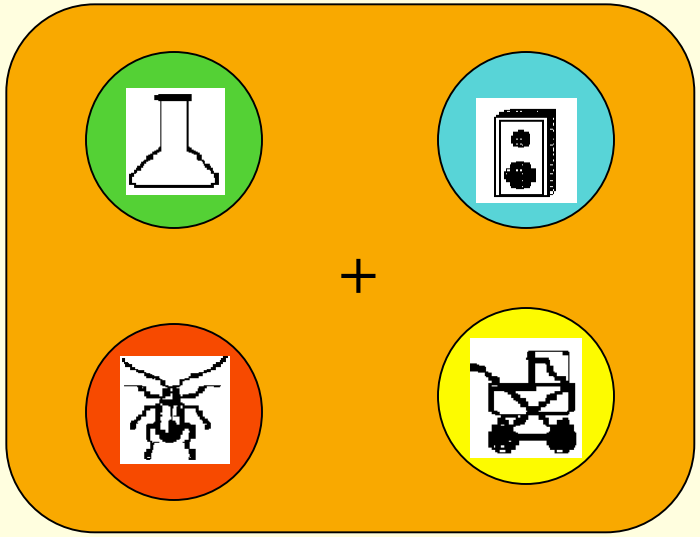


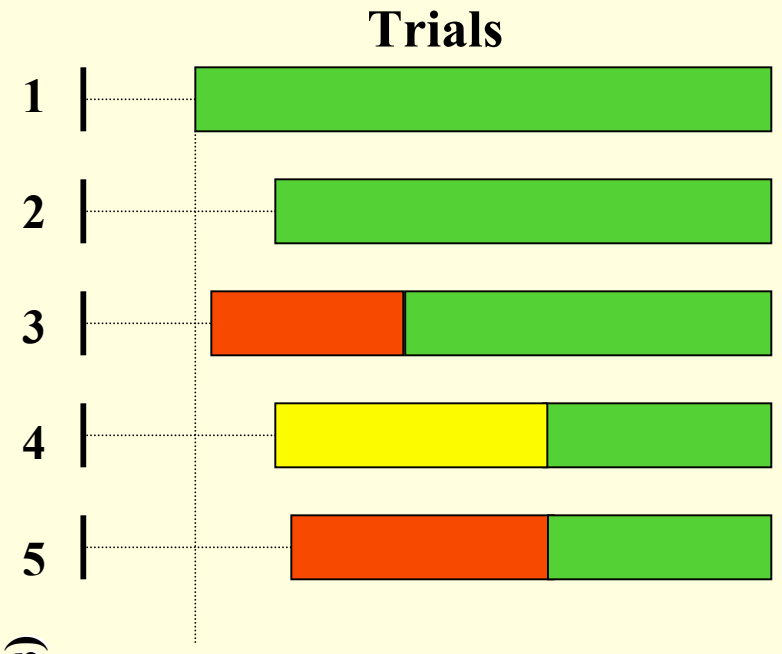


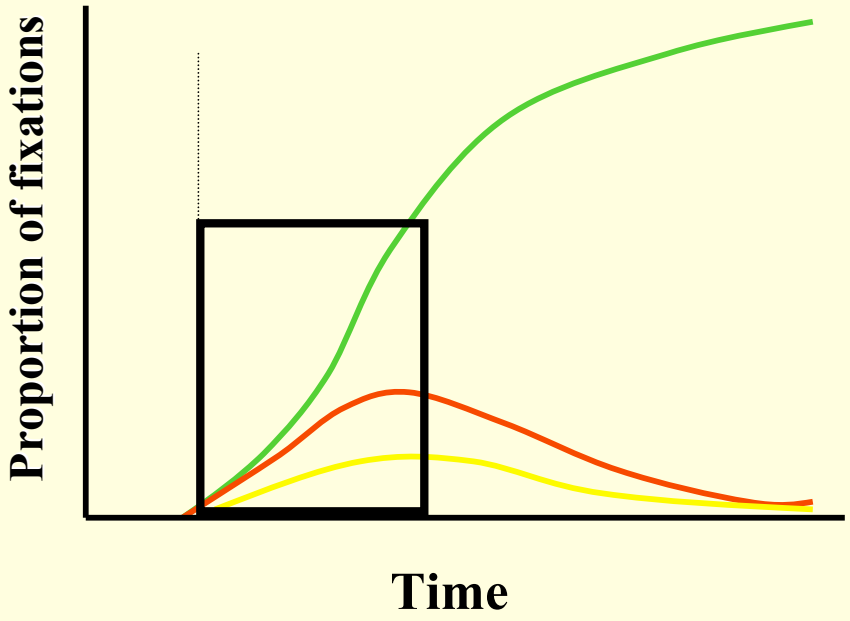


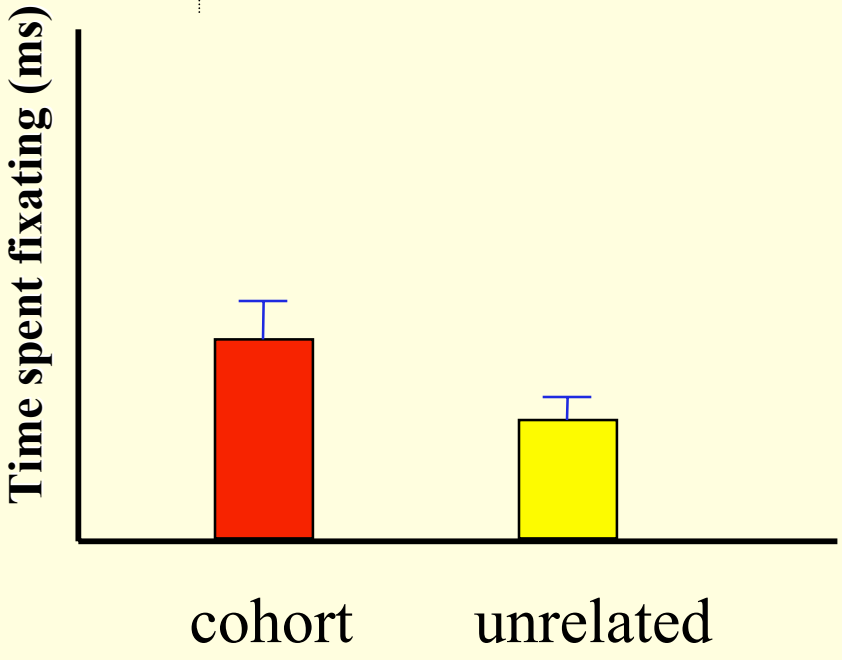


Fixation time over an interval

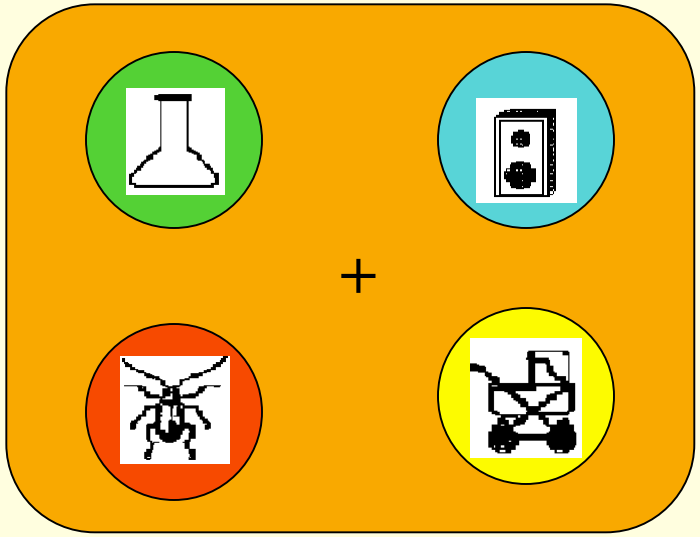


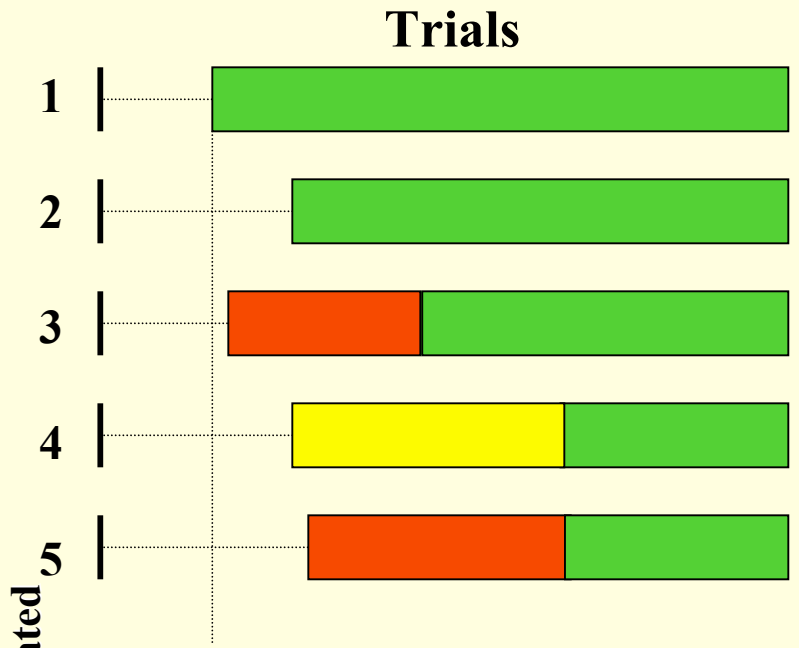


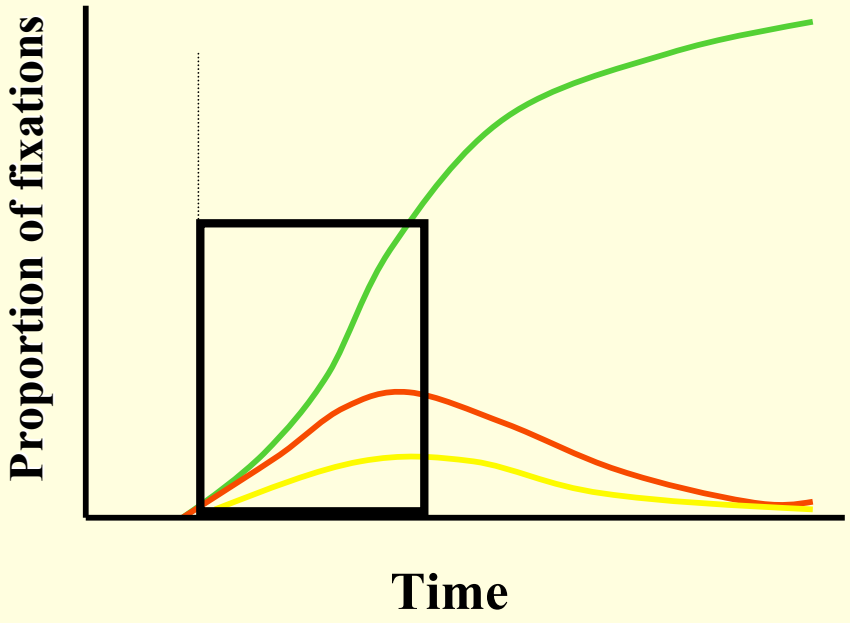


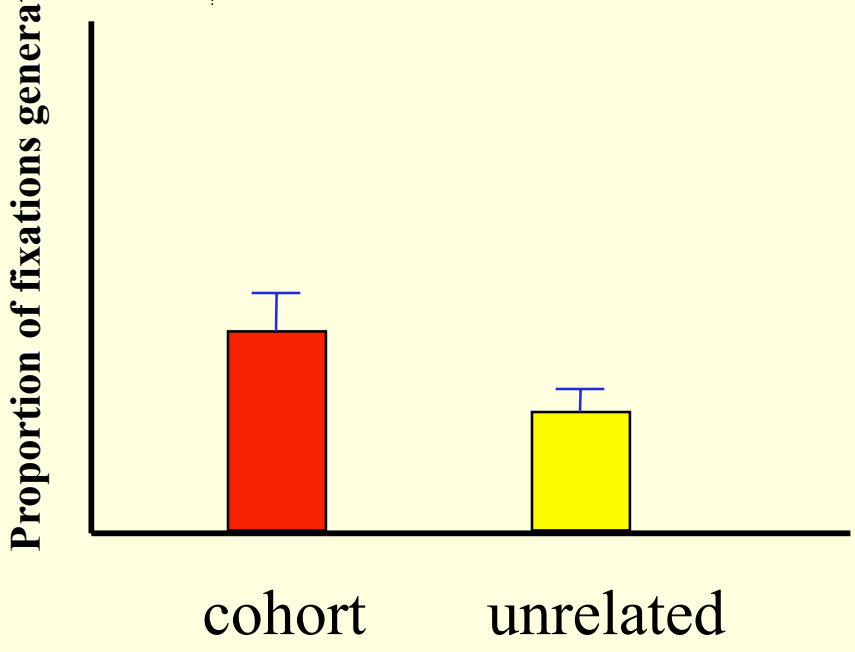


Fixation time over an interval

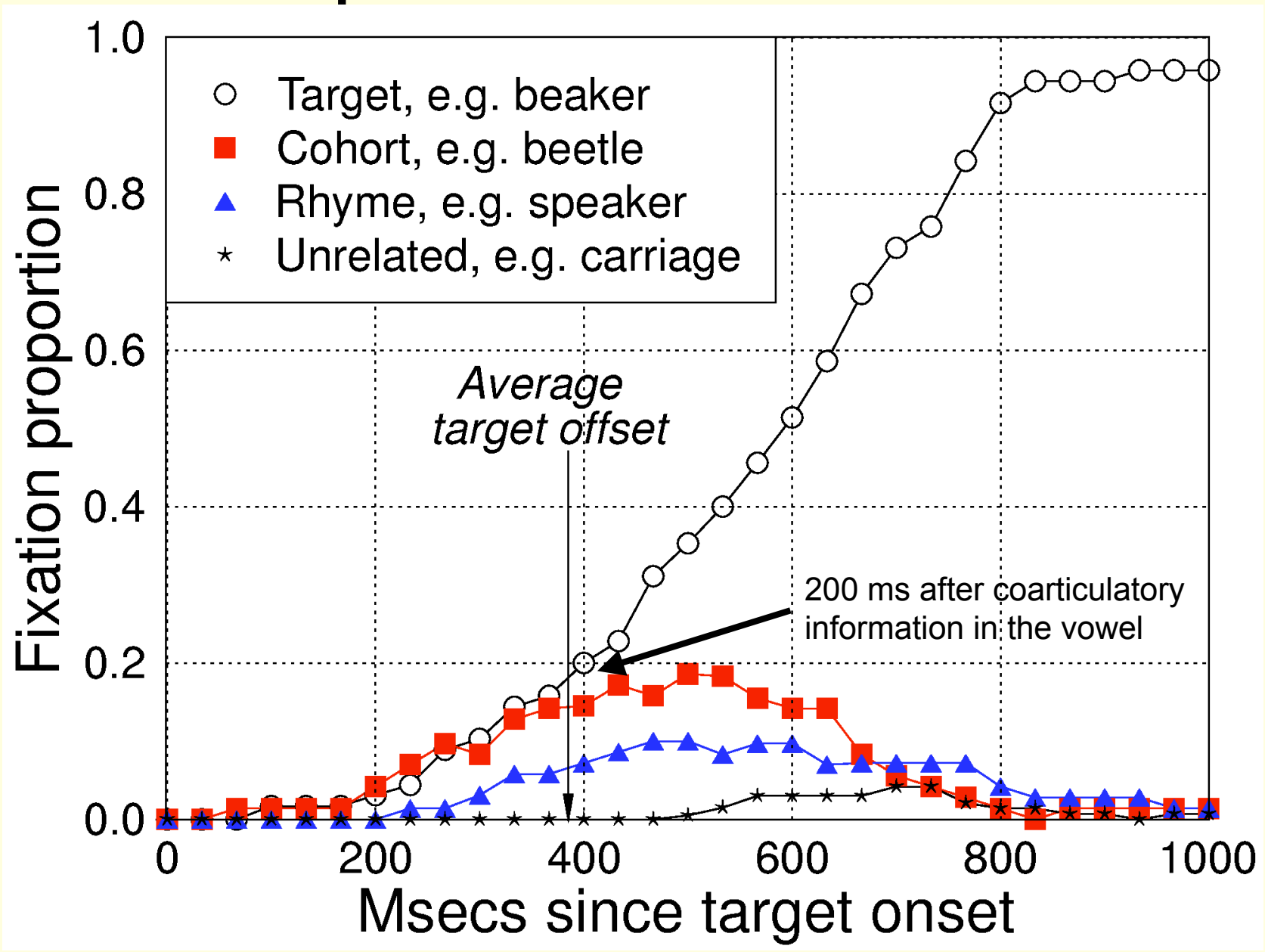




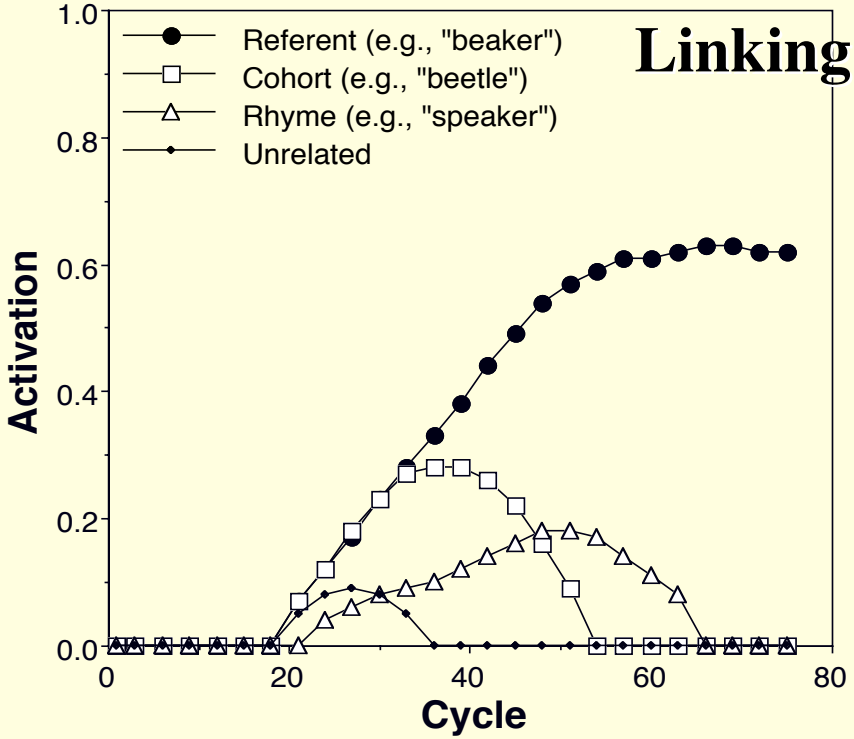




Allopenna et al. results



Linking hypothesis



- Activation converted to probabilities using the Luce (1959) choice rule

$$S_i = e^{ka_i}$$

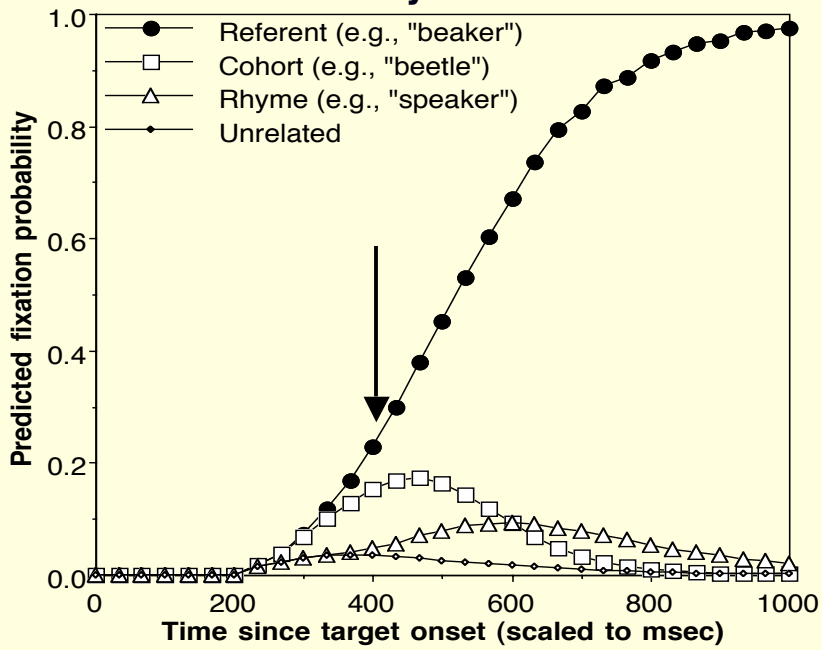
S: response strength for each item
a: activation
k: free parameter, determines amount of separation between activation levels (set to 7)

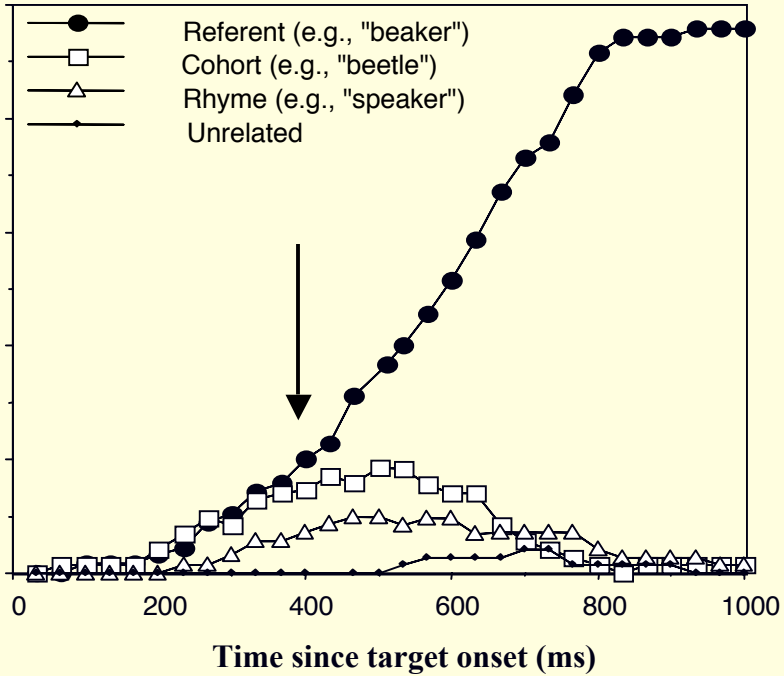
$$L_i = S_i / \sum S_i$$

Choice rule assumes each alternative is equally probable given no information; when initial instruction is *look at the cross* or *look at picture X*, we scale the response probabilities to be proportional to the amount of activation at each time step:

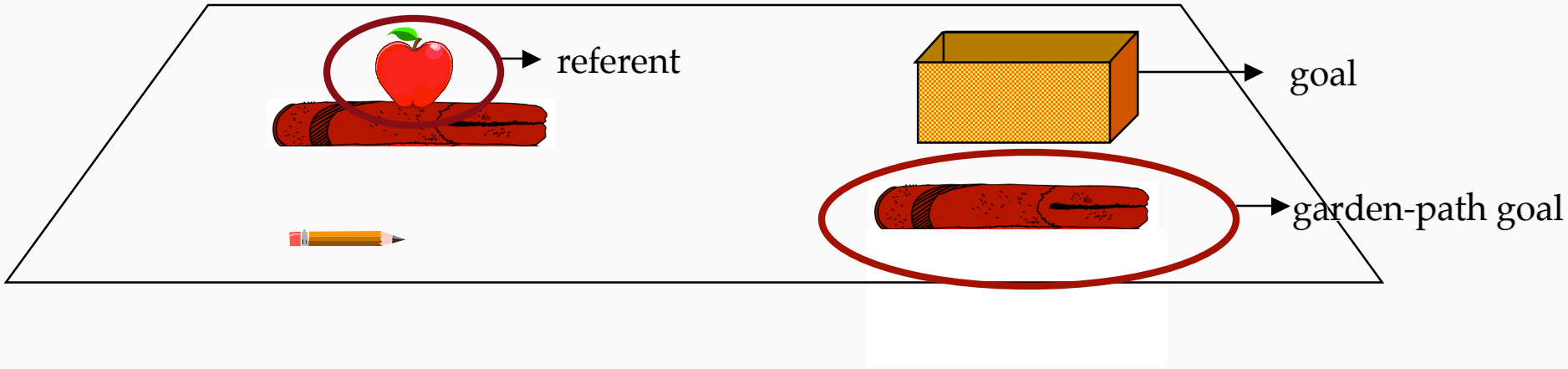
$$d_i = \text{max\_act}_i / \text{max\_act\_overall}$$

$$R_i = d_i L_i$$



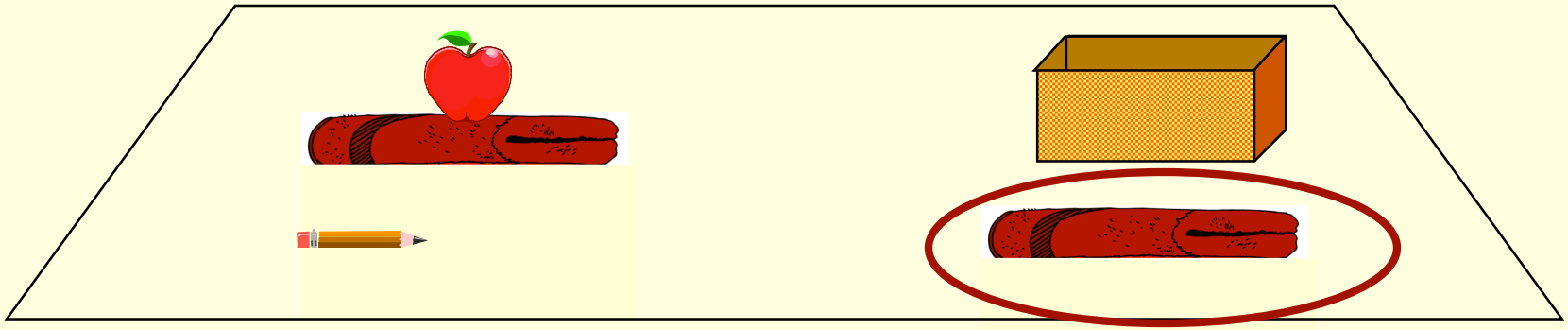


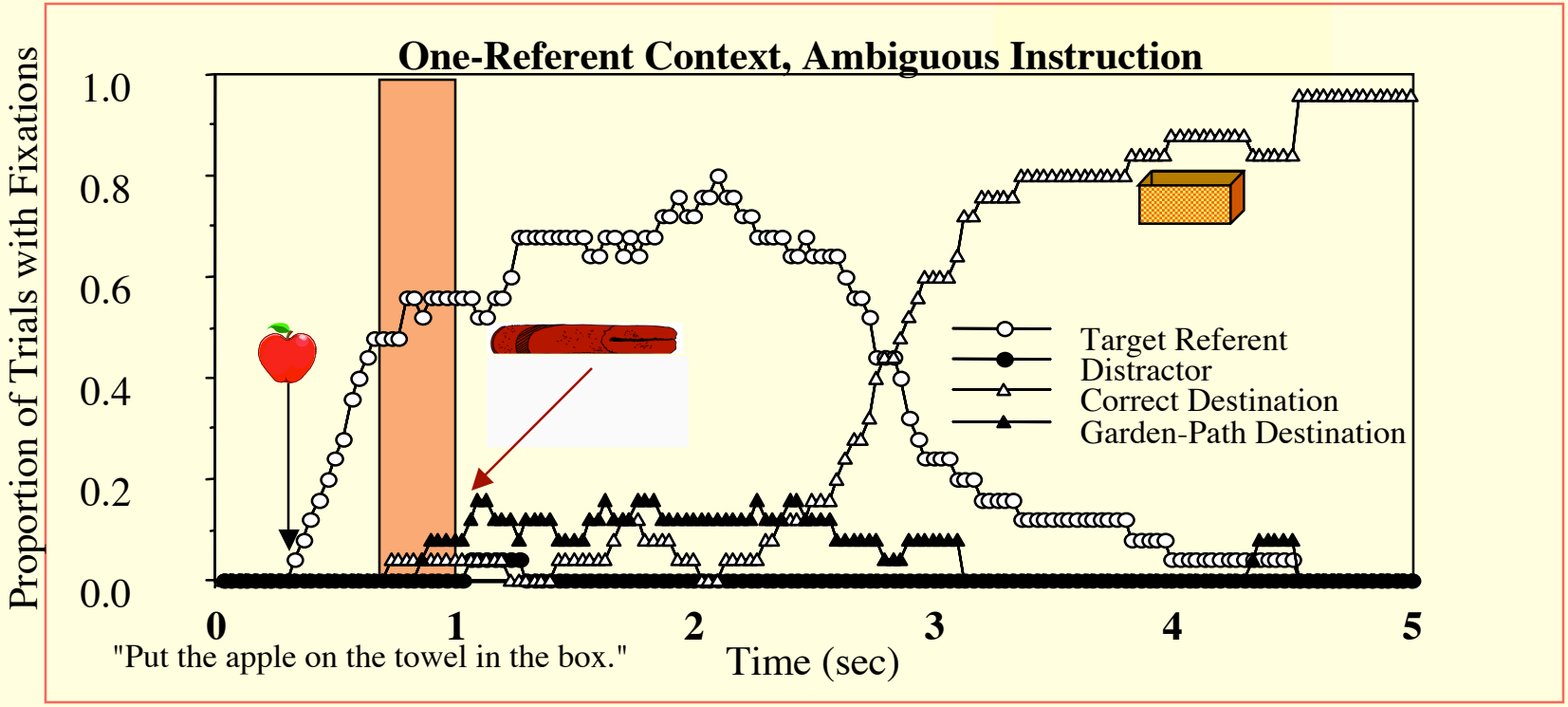
Put the apple on the towel in the box



Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy (1995) ***Science,***
Spivey, et al. (2002) ***Cognitive Psychology***

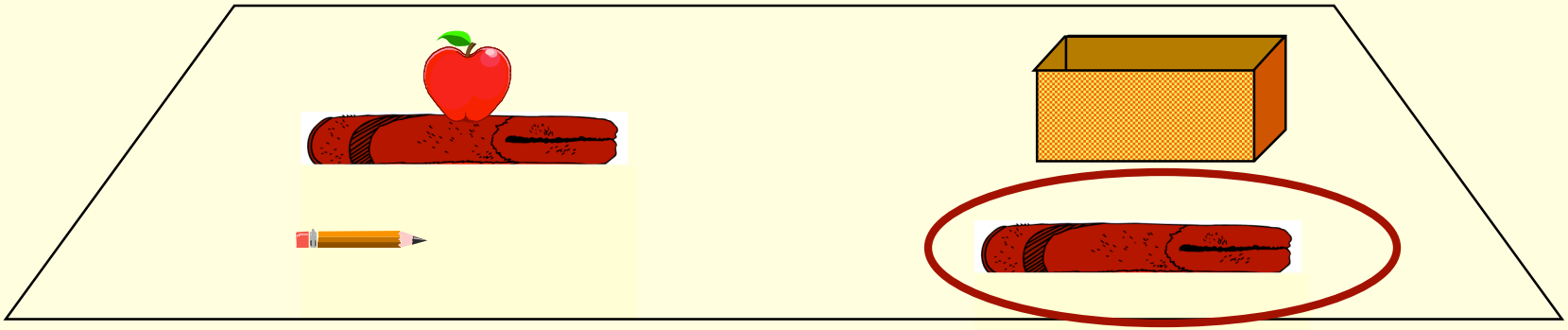
Put the apple on the towel in the box.

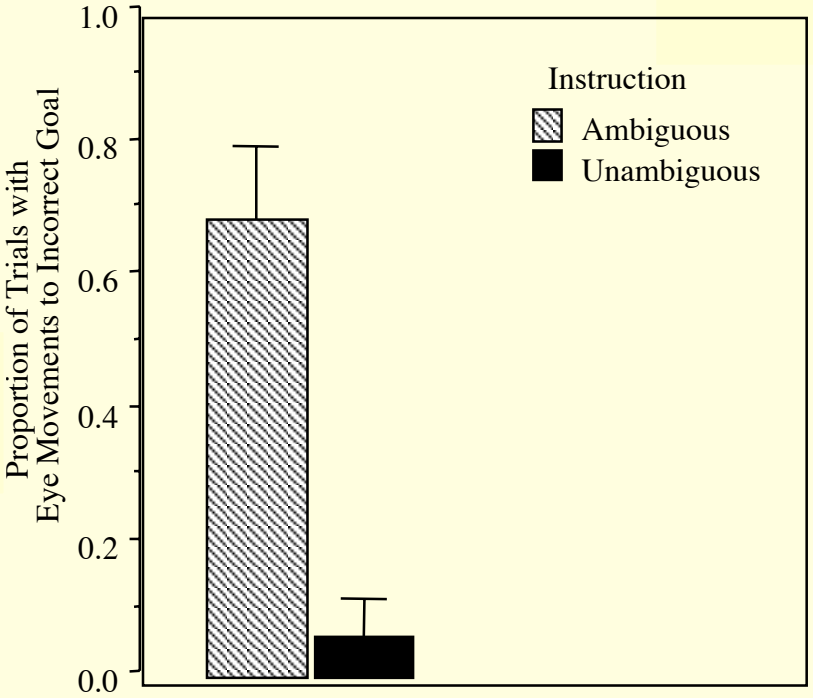




Put the apple on the towel in the box.

Put the apple that's on the towel in the box.





Other measures/terms

Contingent analyses:

response contingent:

speech (McMurray et al., 2002, *Cognition*)

reference (Runner et al., 2003, *Cognition*)

look contingent:

time to target from picture X after point in the signal

(Dahan & Gaskell, in press, *JML*)

Point of disambiguation (POD):

Eberhard et al., 1995, *JPR*,

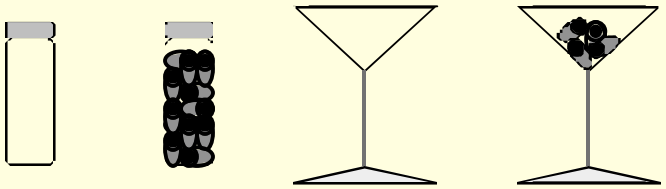
Anticipatory eye movements:

Altmann & Kamide, 1999, *Cognition*

The boy ate the cake

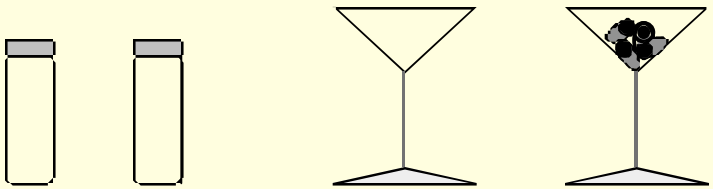
Point of Disambiguation

Late Match



Pick up the empty martini glass ...

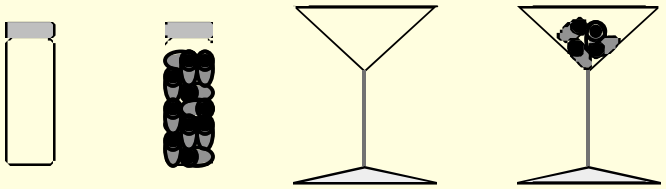
Early Match

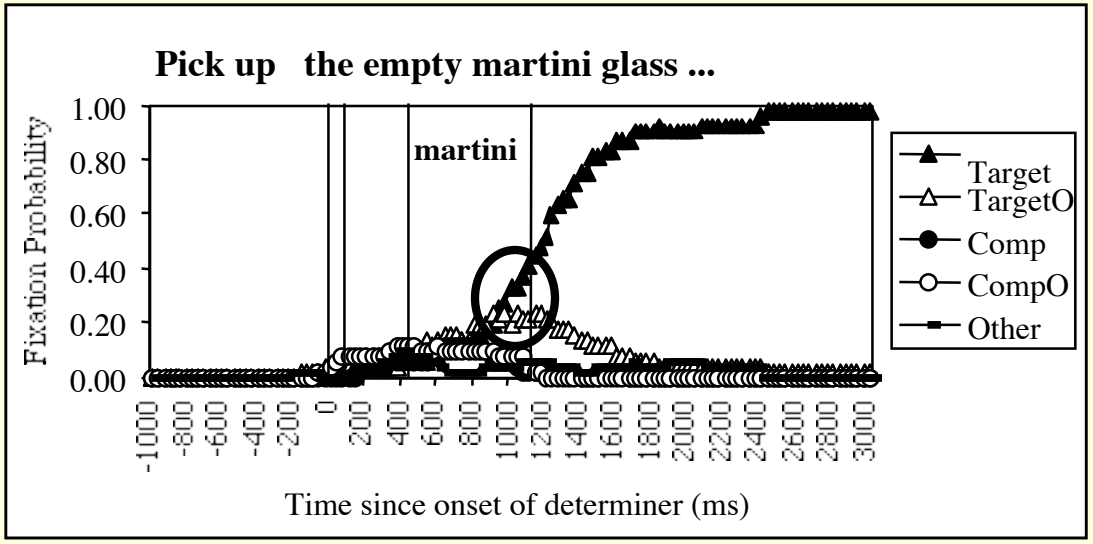


Pick up the empty martini glass ...*

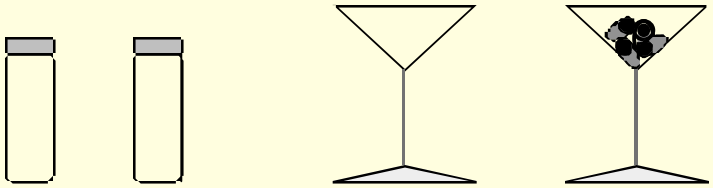
* Assumes listeners will immediately interpret prenominal adjectives such as *empty* contrastively (i.e., there should be one empty X and one not empty X). See Sedivy et al. (1999) *Cognition* for supporting evidence.

Late Match

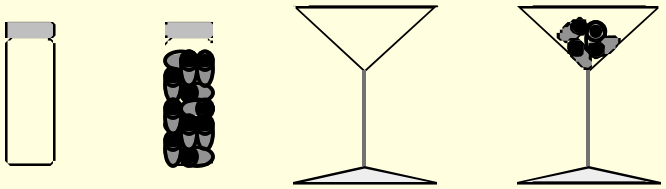


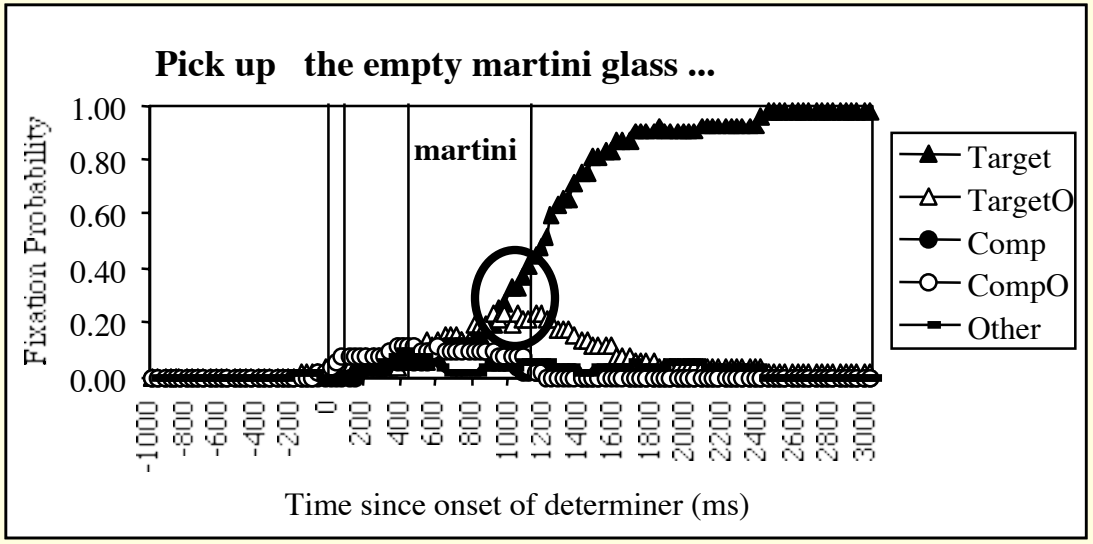


Early Match

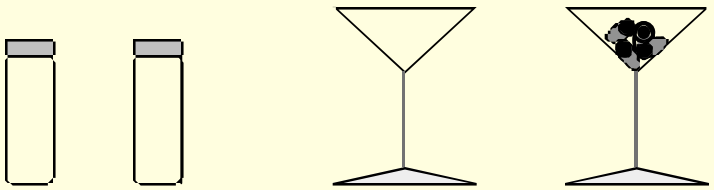


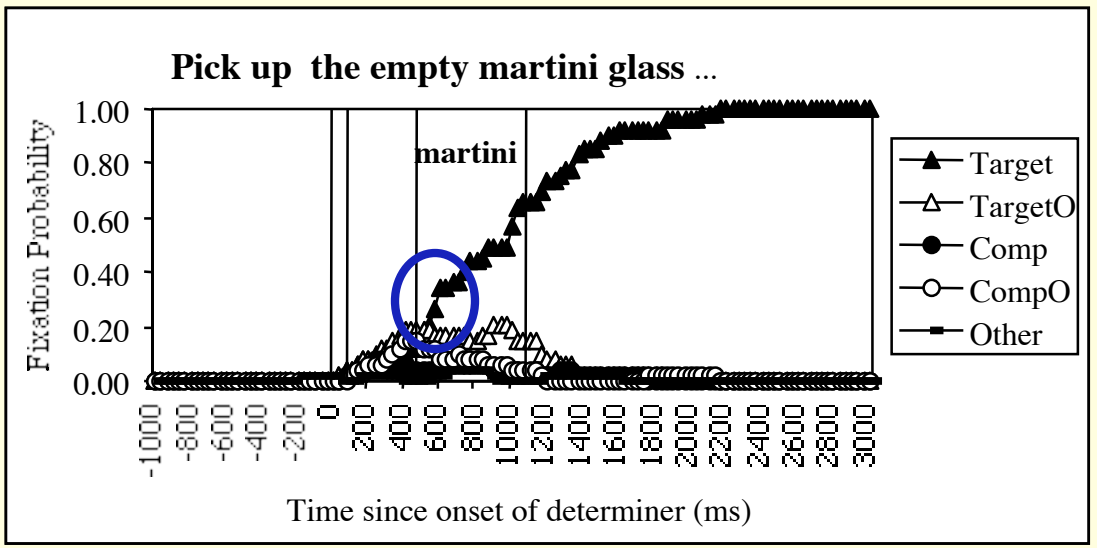
Late Match



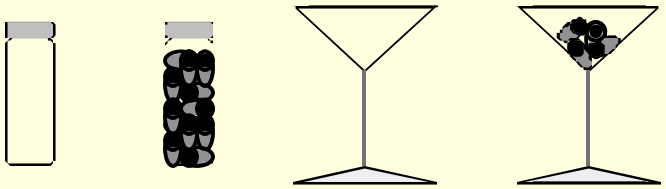


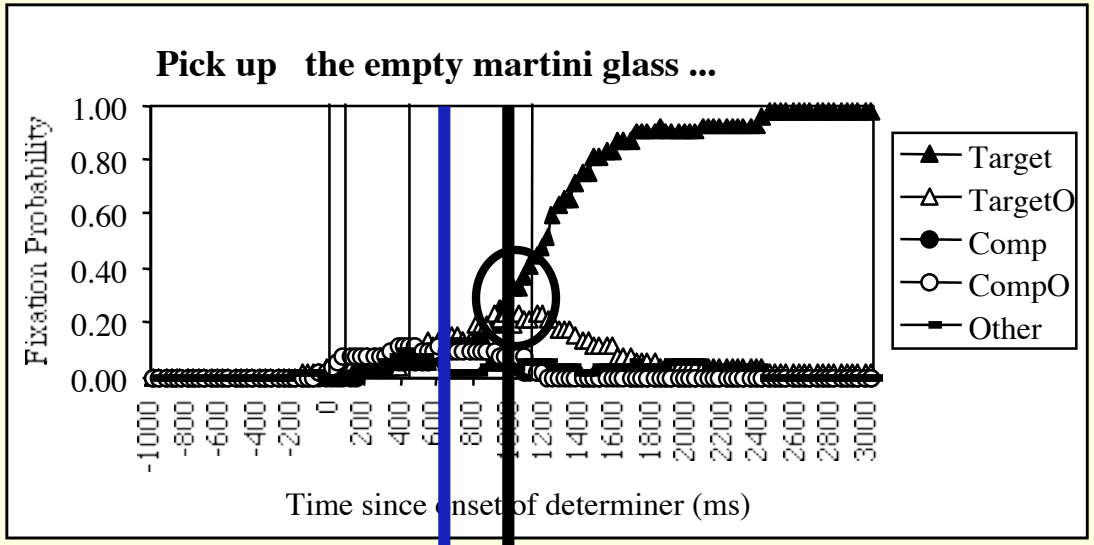
Early Match





Late Match





Early Match