The Visual-World Paradigm
Adult Language Processing

SS16 - (Embodied) Language Comprehension

Maria Staudte

(Embodied) Language Comprehension

- Language is "embodied" & "situated":

  - **Embodied:** Language is inseparably bound to our body and our physical interaction with the world

  - **Situated:** Language is typically used and understood in context

Embodiment

Acquisition

Adult Processing

Language is situated

- Spoken language and environment provide huge amounts of information simultaneously

- Processing needs to be fast!

- Using one to facilitate processing the other:

  - Visual information (non-verbal cues)

  - Visual (scene) information **Attention**

  - Linguistic information

Language is situated

- We see things when we listen and speak

- We often try to ignore the visual information — this requires cognitive control (cf. *automatic* processing)

- We often use the visual information — this requires cognitive control, capacity sharing and cross-modal integration
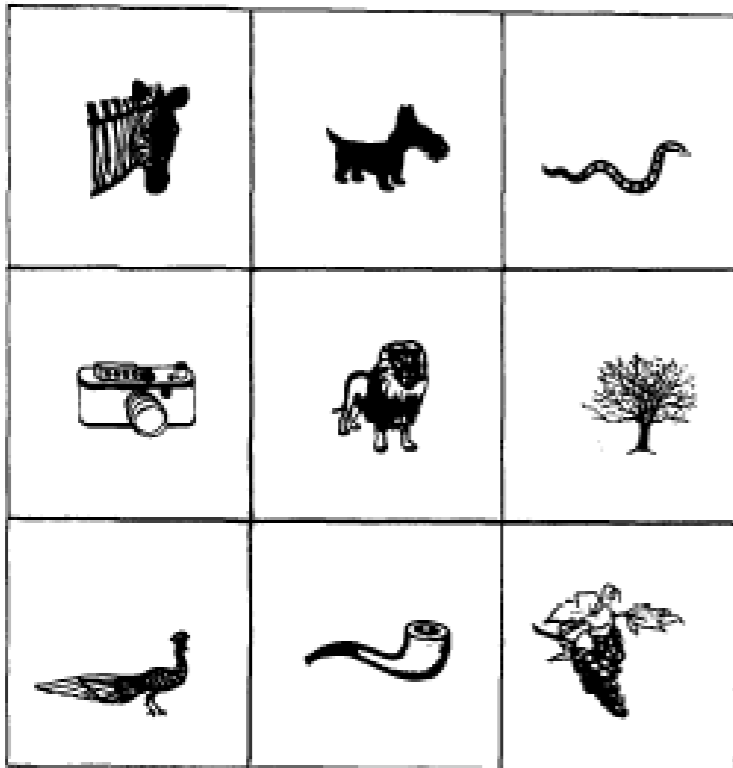
Cooper, 1974

FIG. 1. The slide accompanying Story 3.

While on a *photographic/safari* in *Africa*, I managed to get a number of breathtaking shots of the wild *terrain*. These included pictures of rugged mountains and *forests*, as well as muddy *streams/winding their way* through big *game/country*. One of my best shots though was ruined by my scatter-brained *dog/Scotty*. Just as I had slowly *wormed* my way on my *stomach*

Eyetracking

Use "visual attention" as index to underlying processes

Eye-tracking in scenes
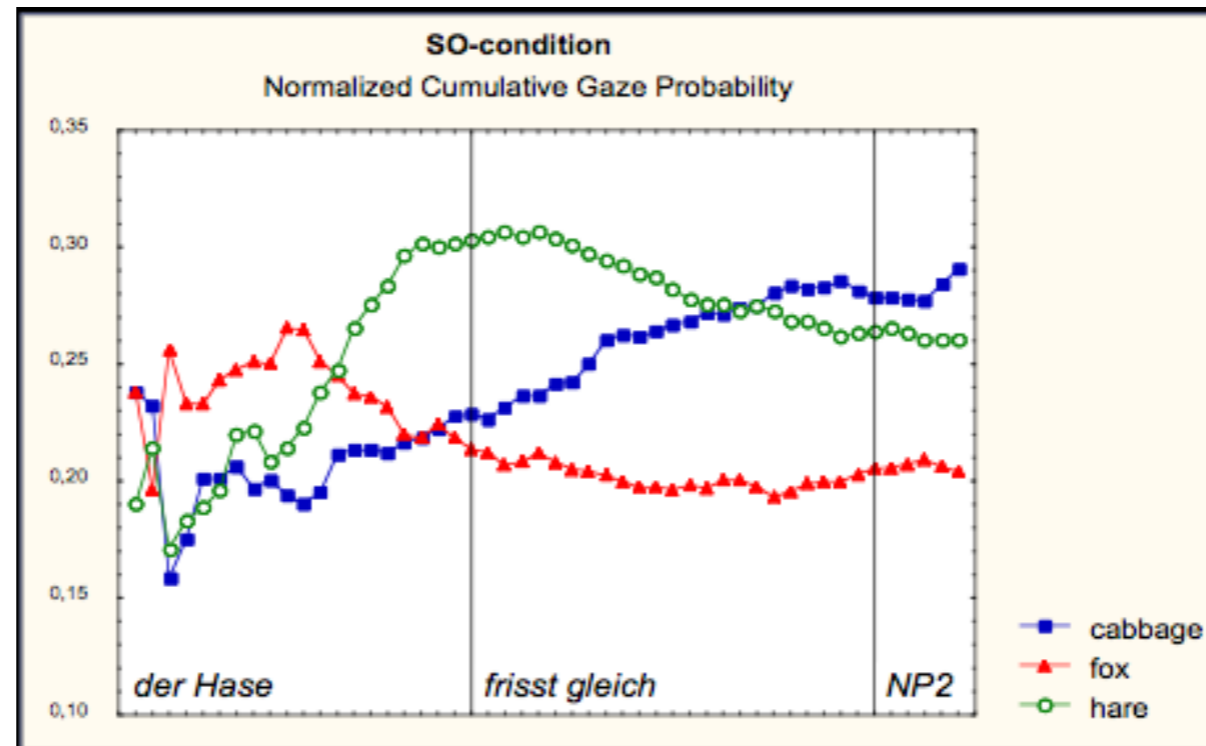
- Attention to objects in the scene is closely time-locked to comprehension

  - Makes it possible to use eye-tracking in scenes during utterance presentation to investigate spoken comprehension

  - Permits us to examine use of scene information for comprehension

The Measure

- Tracker samples at 60/250/500/1000 Hz

  - New sample every 16,6/4/2/every millisecond

- x,y - coordinates per sample

- Fixations & saccades
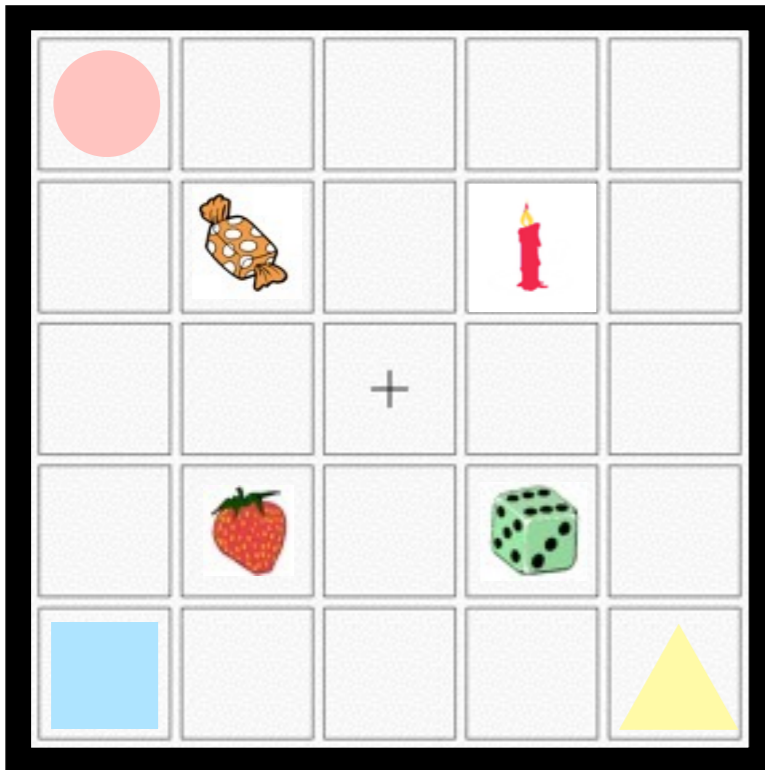
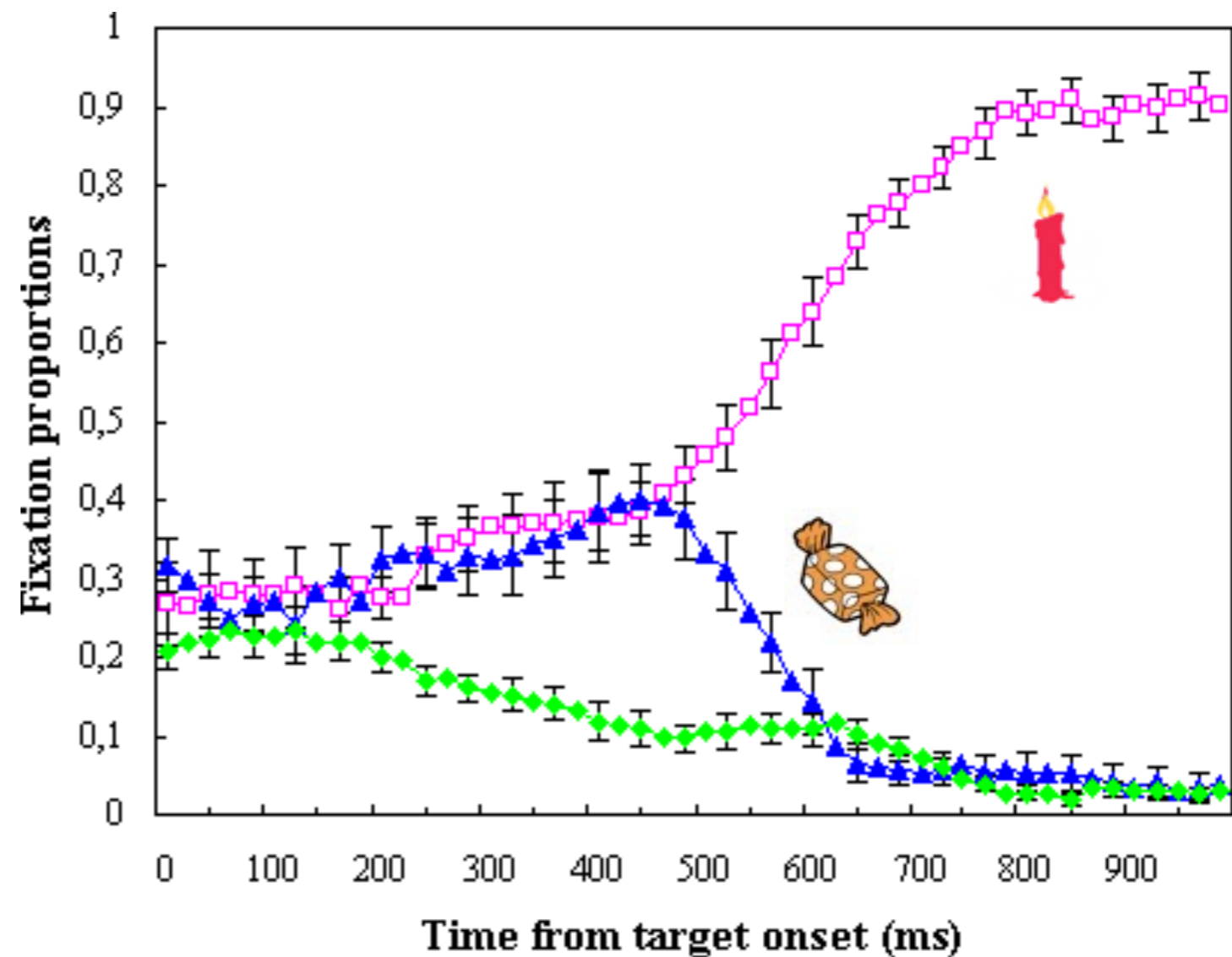 - min duration of ~80ms = fixation

The Measure
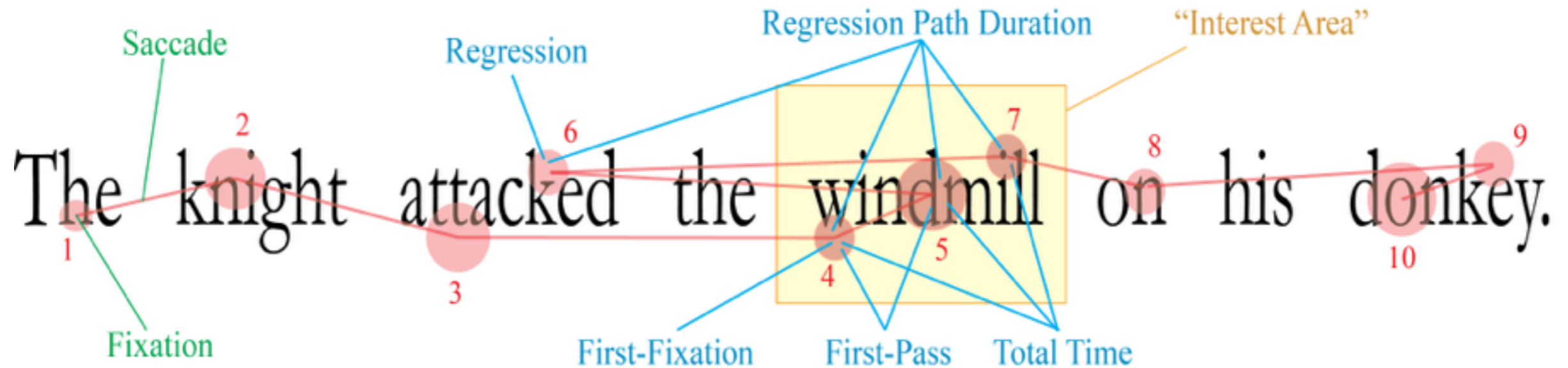
- Fixation proportion vs probability

- Time to first fixation/saccade

- Fixation durations

- Inspection probability

Lexical access over time

"Pick up the *candle*"
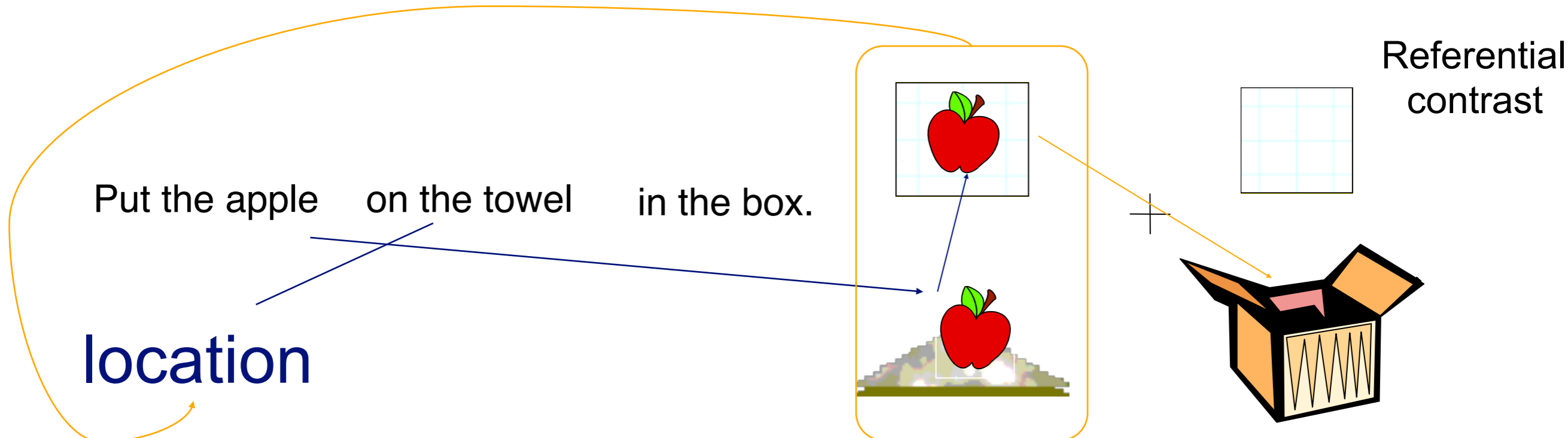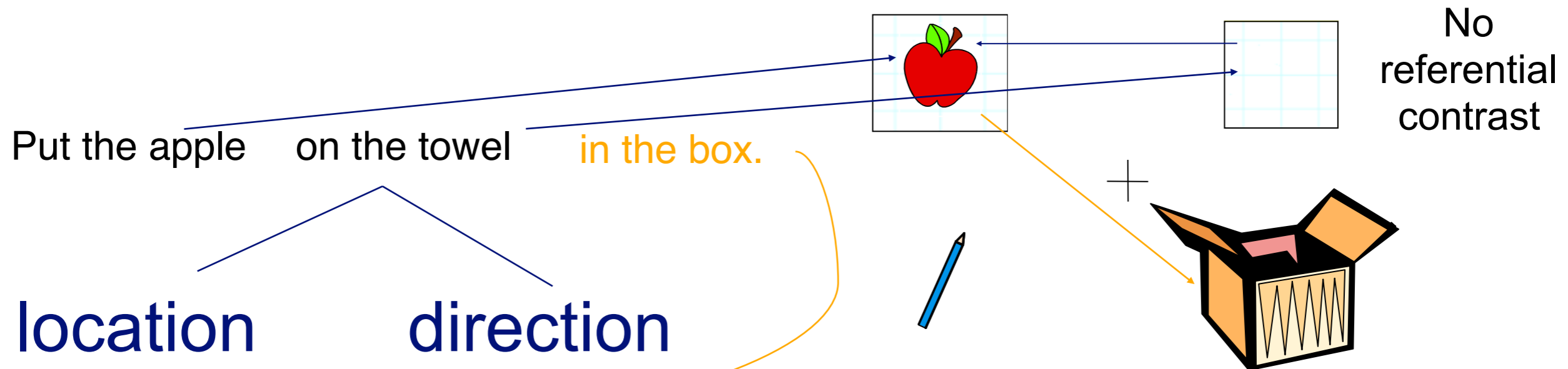
Attachment Ambiguity

(Tanenhaus et al. 2000)

No referential contrast

Put the apple   on the towel   in the box.

location   direction

Referential contrast

Put the apple   on the towel   in the box.

location

Incremental Semantic Interpretation

(Sedivy et al. 1999)

Two same-type objects
that differ in 1 property: size
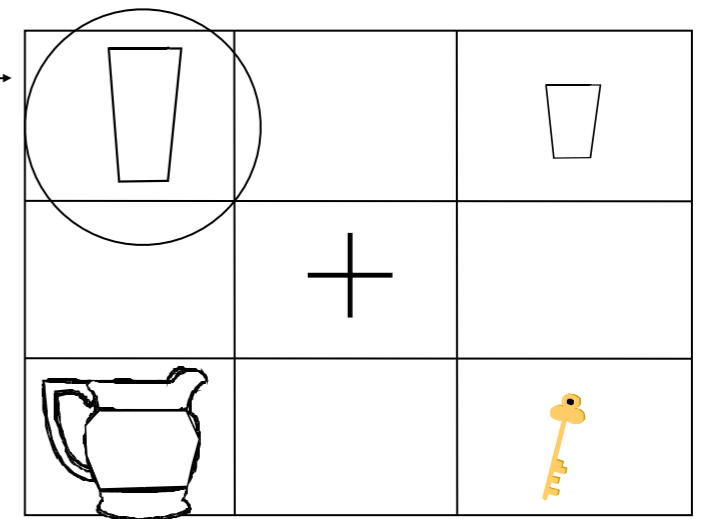
More visual referential ambiguity:

- Influence of visual contexts on

  - Mapping of reference to entities

  - Properties of objects (small, tall)
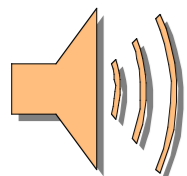
Pick up the tall glass and put it below the pitcher.

No contrastive objects
of the same type

- More rapid looks to the tall glass before hearing "glass" in the contrastive than non-contrastive condition
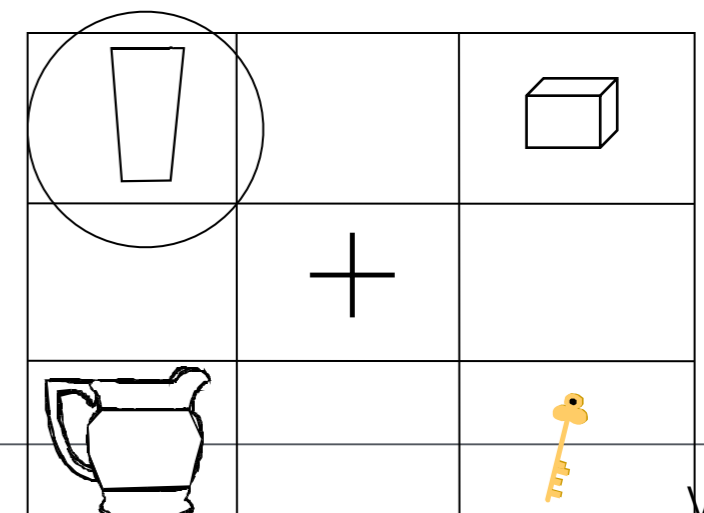
Predictive Listening

- Anticipation of objects (e.g. Altmann & Kamide 1999, 2007)

  - Semantic requirements of a noun to be used as argument for a verb

  - drink – object needs to be drinkable, liquid
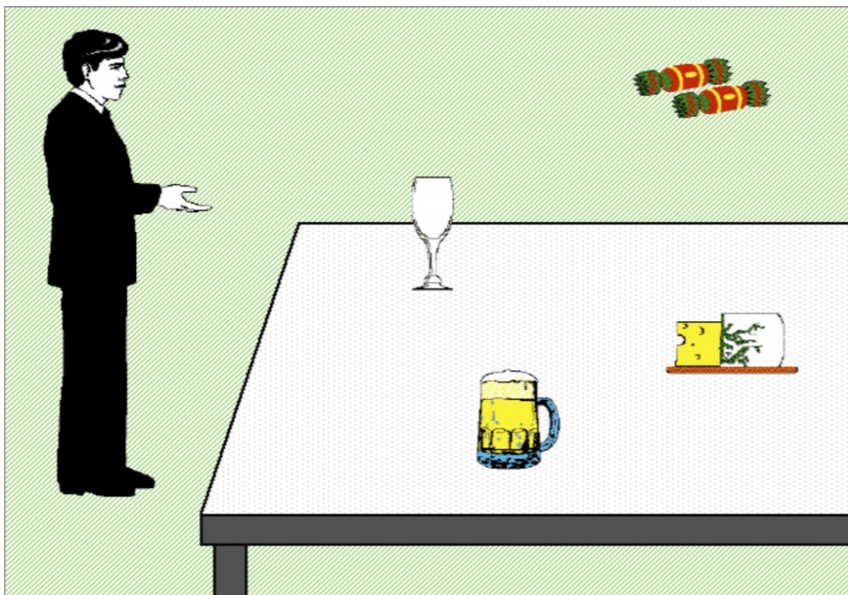
"The man will drink the wine"

Referential & Anticipatory Looking

- Can information associated with a 3-place verb be used to anticipate the second post-verbal argument — during hearing the first?

- "The woman will spread/slide the butter…

1. Inanimate goal ("on the bread")

2. Animate goal ("to the man")

(Kamide, Altmann & Haywood, 2003)

Referential & Anticipatory Looking

1. The woman will spread the butter on the bread
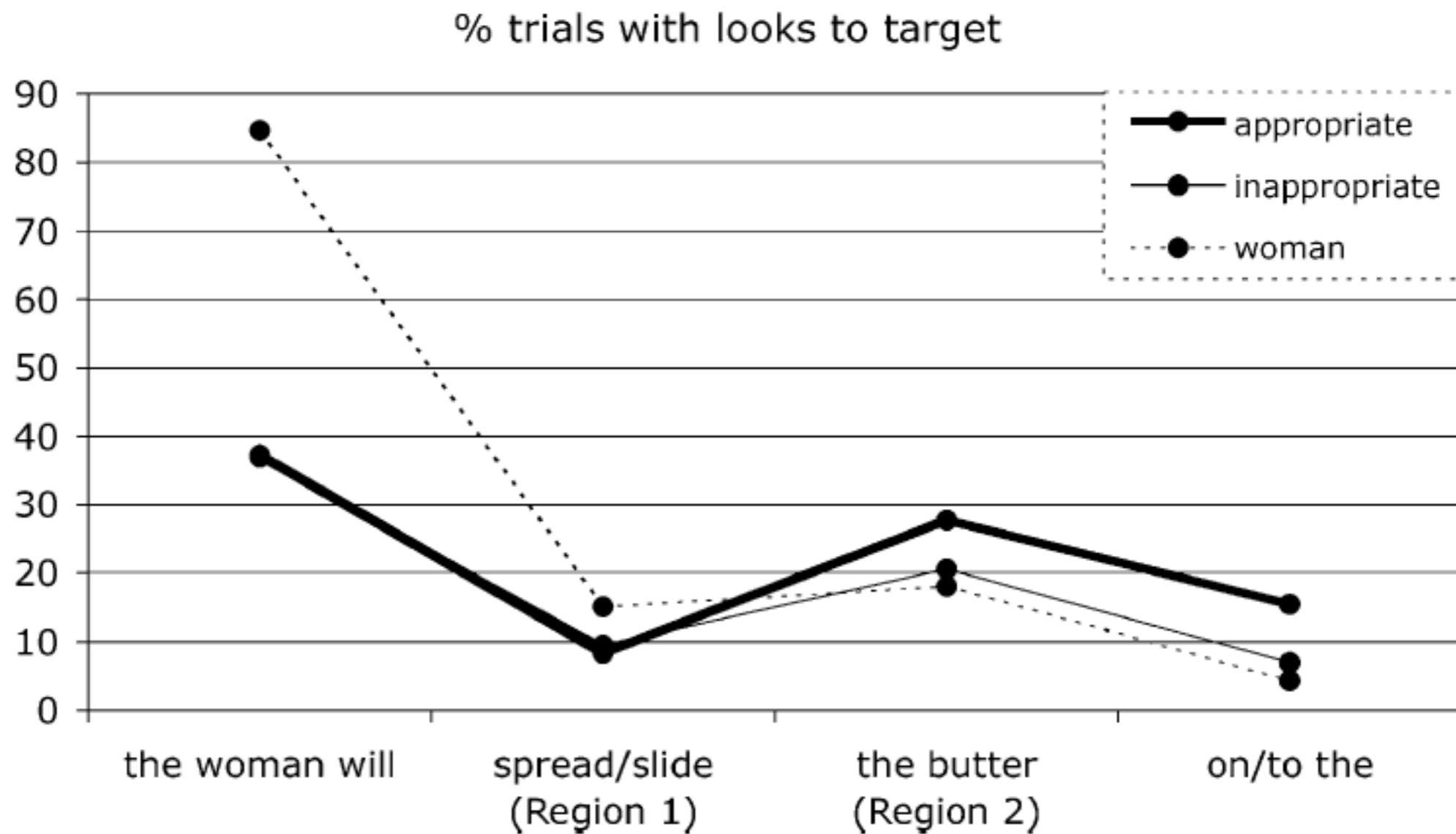
Referential & Anticipatory Looking

2. The woman will slide the butter to the man

Referential & Anticipatory Looking

% trials with looks to target

Asymetric distribution

- Looks to the **man**

  - Slide condition: 24,3%

  - Spread condition: 16,8%

- Looks to the **bread**

  - Slide condition: 24,3%

  - Spread condition: 27,4%

Bread could be a theme, too!

Composed predictions?

- Are the observed effects driven by lexical information associated with the verb?

- Are the effects due to the combination of distinct sources of information?

➡ Can information about the **agent** be combined with the **selectional restrictions of the verb** to predict the theme?

Composed predictions?

1. The man will ride the motorbike

Composed predictions?

2. The girl will ride the carousel

Conditions

- Main conditions:

  - The **man** will ride the motorbike

  - The **girl** will ride the carousel

- To assess combinatory effects (vs agent effects):

  - The man will **taste** the beer

 - The girl will taste the sweets

Hypotheses

- Compare looks towards the motorbike after

  - The man will **taste**

  - The man will **ride**

- Only agent-based information -> No difference

- Compare looks towards the motorbike after

  - The **man** will ride

  - The **girl** will ride

- Only verb-based information -> No difference

Hypotheses

- Compare looks towards the motorbike after

  - The **man** will taste

  - The **girl** will taste

- Agent-effect information -> man > girl

- Compare looks towards the motorbike after

  - The girl will **ride**

  - The girl will **taste**

- Verb-effect -> ride > taste

Composed predictions

% trials with looks to motorbike
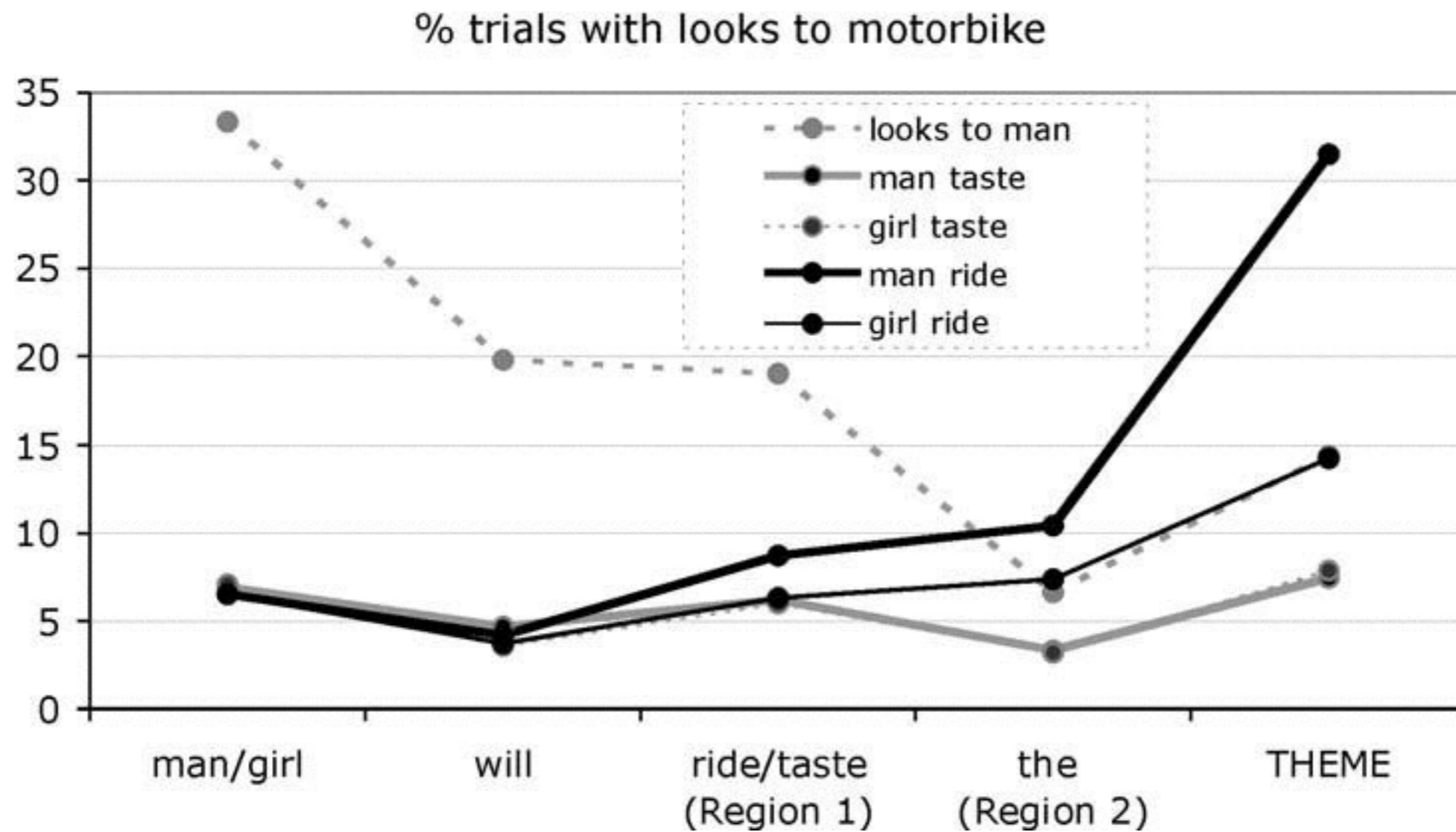
Composed predictions

- Combinatory effects found

- (Almost) no simple agent or verb effects

  - More looks to motorbike after "girl will ride" than after "girl will taste"

Anticipatory Looks

- **Prediction:**

  - projection "forward in time"

  - Processor predicts a post-verbal argument / linguistic input

  - Projected structure evaluated against visual/mental context

- **Integration:**

  - current evidence

  - Thematic organisation of objects according to linguistic context

What's the hen, egg…?

- "The man will ride ... "

➡ Does the linguistic structure trigger a predictive process?

➡ Does the visual context suggest a plausible argument?

"Coordinated Interplay Account"

Visual

Scene

Auditory

Utterance

Listener
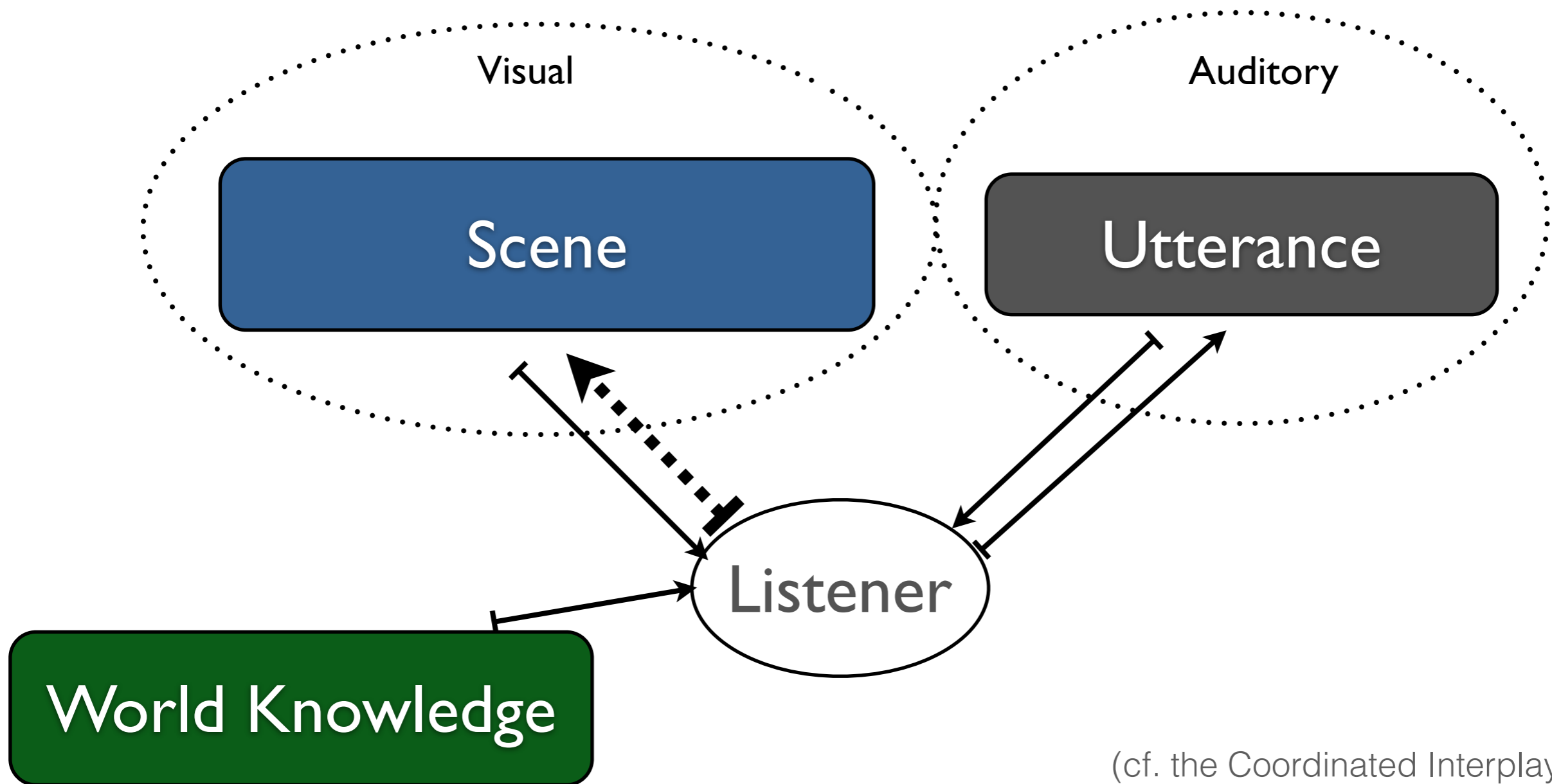
World Knowledge

(cf. the Coordinated Interplay Account
by Knoeferle & Crocker, 2006)

Questions

- What takes **precedence**, world knowledge or scene information?

- What exactly is **predicted**?

World knowledge vs scene info

(Knoeferle & Crocker, 2006)

Theme = Pilot

Agent = Wizard/Detective

Stereotypical actions:
- Wizard & jinx
- Detective & spy

Depicted actions:
- Wizard & spy
- Detective & serve food

World knowledge vs scene info

Theme = Pilot

Agent = Wizard/Detective

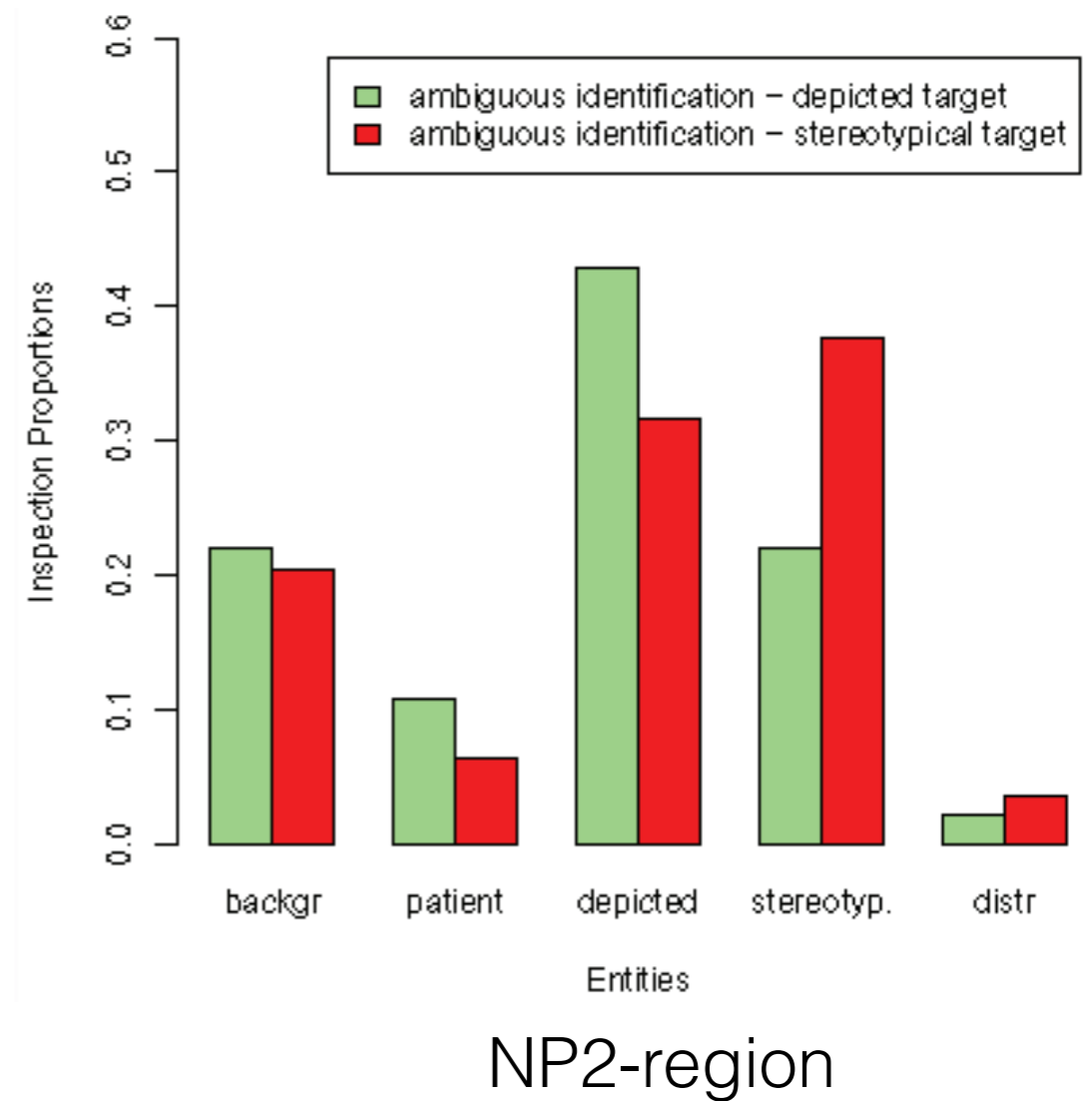Stereotypical actions:
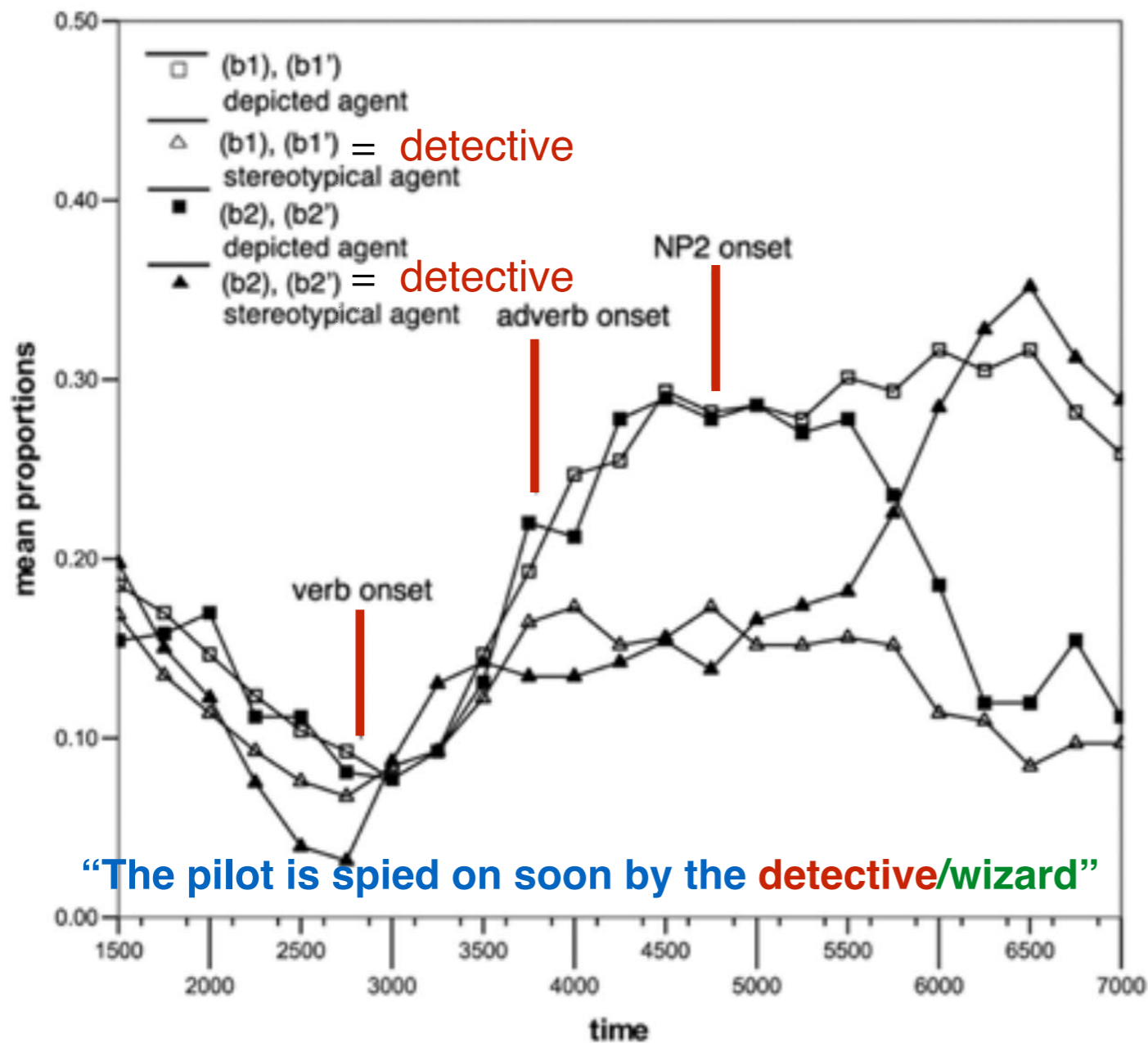- Wizard & jinx
- Detective & spy

Depicted actions:
- Wizard & spy
- Detective & serve food

"Den Piloten verzaubert / verköstigt / bespitzelt gleich…"
   "The pilot is jinxed / served / spied on soon by…"

Precedence & Revision

"The pilot is spied on soon by the **detective**/**wizard**"

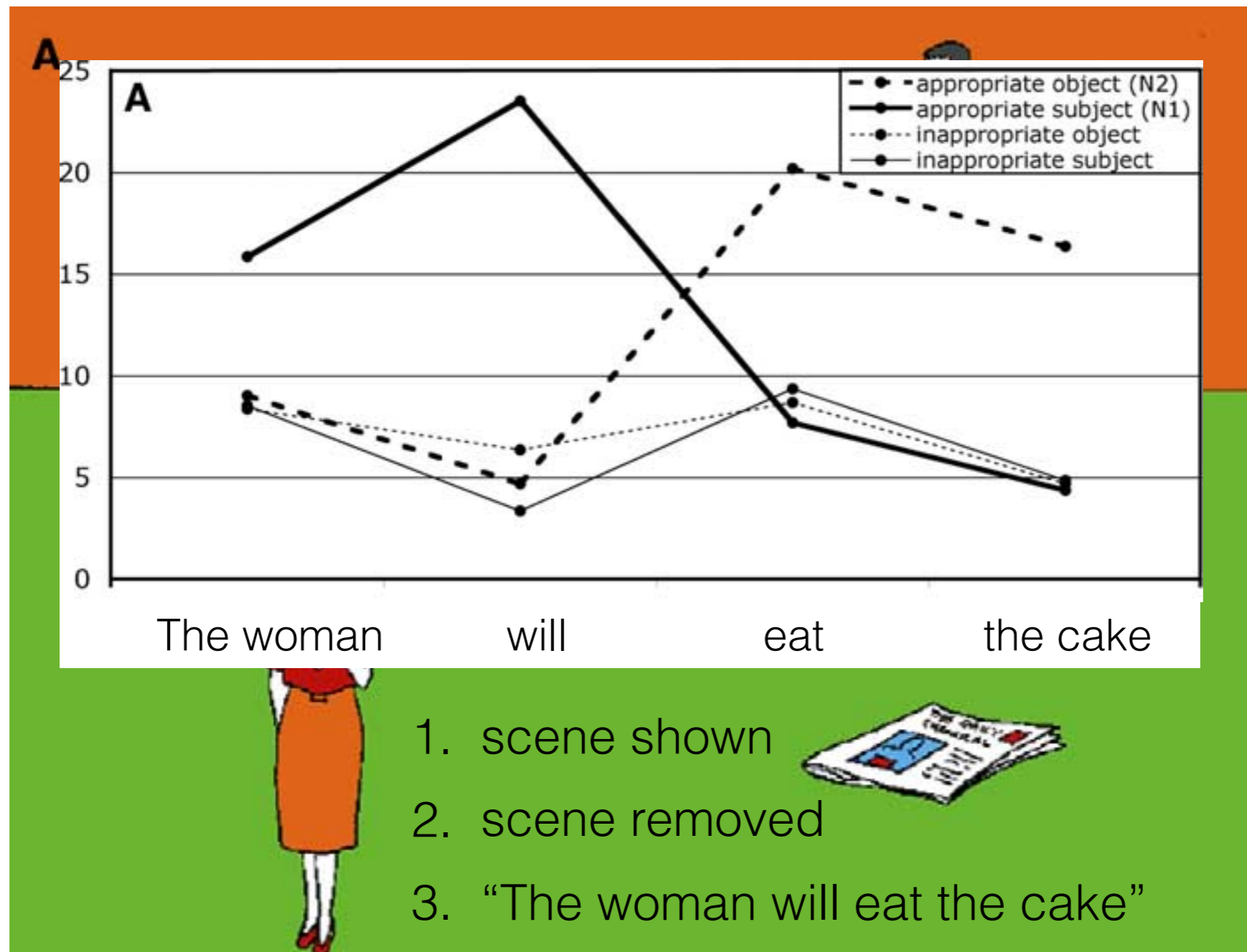NP2-region

Concurrent & previous scene information

- Concurrent scene information extremely dominant!

- Eyes move to relevant objects, automatically? why?

- Relevant objects and/or relevant locations?

- Scene information changes, objects move/disappear

Blank-Screen Paradigm
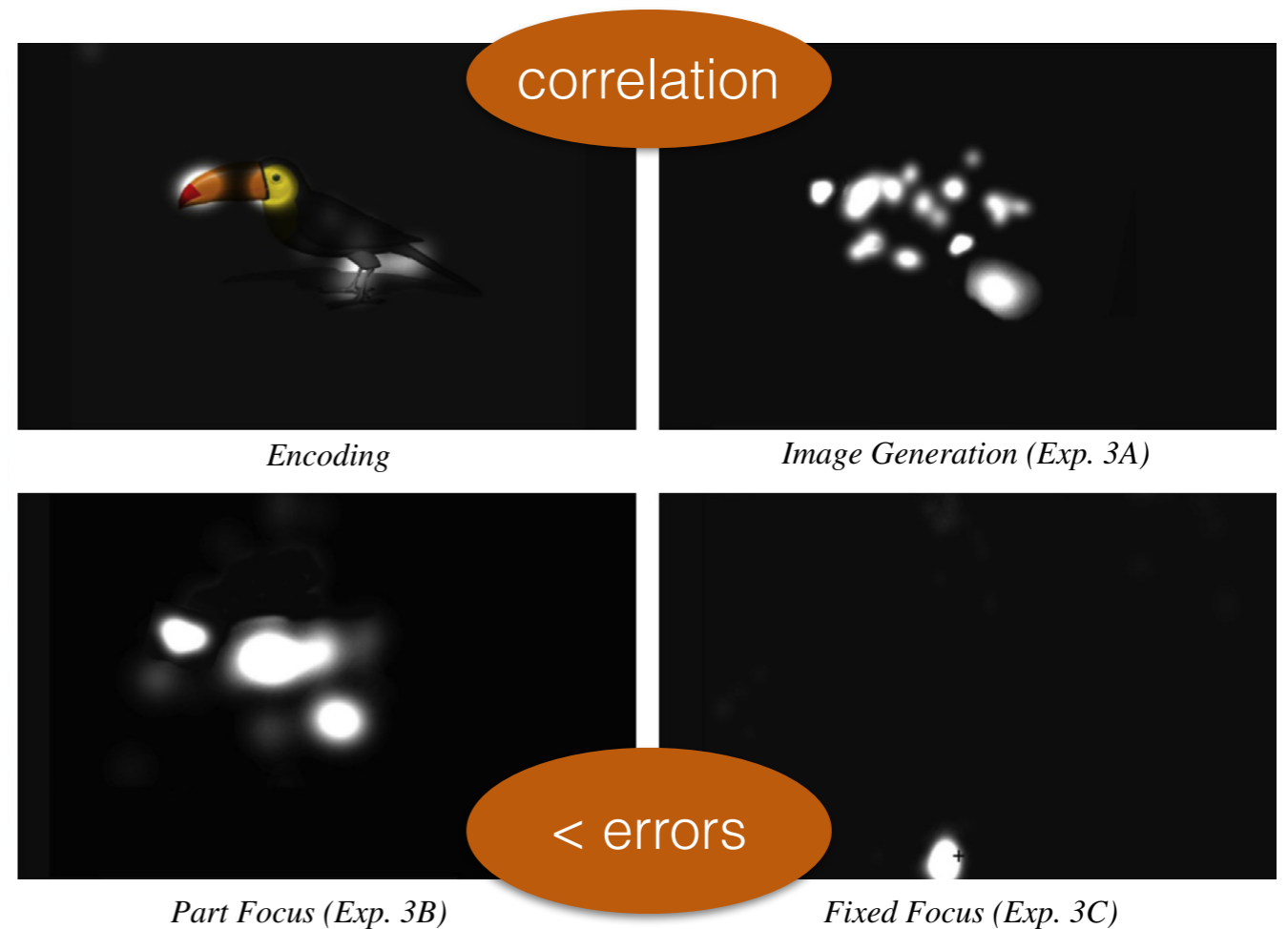

(Altmann, 2004)

Why such eye-movements?

- People look at relevant things in a scene

- People look at locations where relevant things **have been**

- Not for perceiving visual information

- For what?

Why such eye-movements?

- Hebbian learning / processing

  - Objects are located somewhere

  - Featural and spatial information is associated

  - One activates the other

- Looking at a location might activate what was at that location!

Why such eye-movements?

(Laeng et al., 2014)

Encoding

Image Generation (Exp. 3A)

correlation

Part Focus (Exp. 3B)

Fixed Focus (Exp. 3C)

< errors

a) Imagery
b) Free recall "What was the color of the beak of the toucan bird?
c) Fixed view recall "…"

More changes…

- Scene info + utterance

  - Eye-movements to utterance-related objects in scene

- Scene changes/disappears

  - Eye-movements to where these objects where in the scene

- What if spoken utterance proposes changes to scene?

The "represented" scene
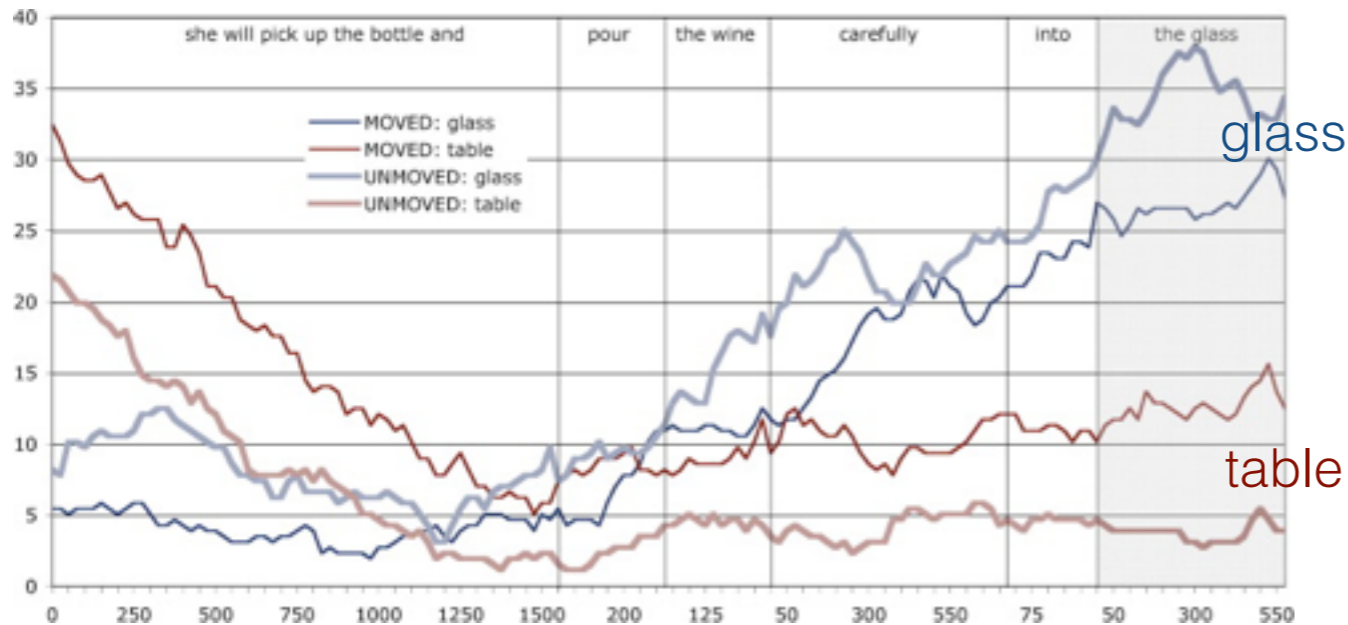

a. The woman will put the glass on the table

b. The woman is too lazy to put the glass on the table

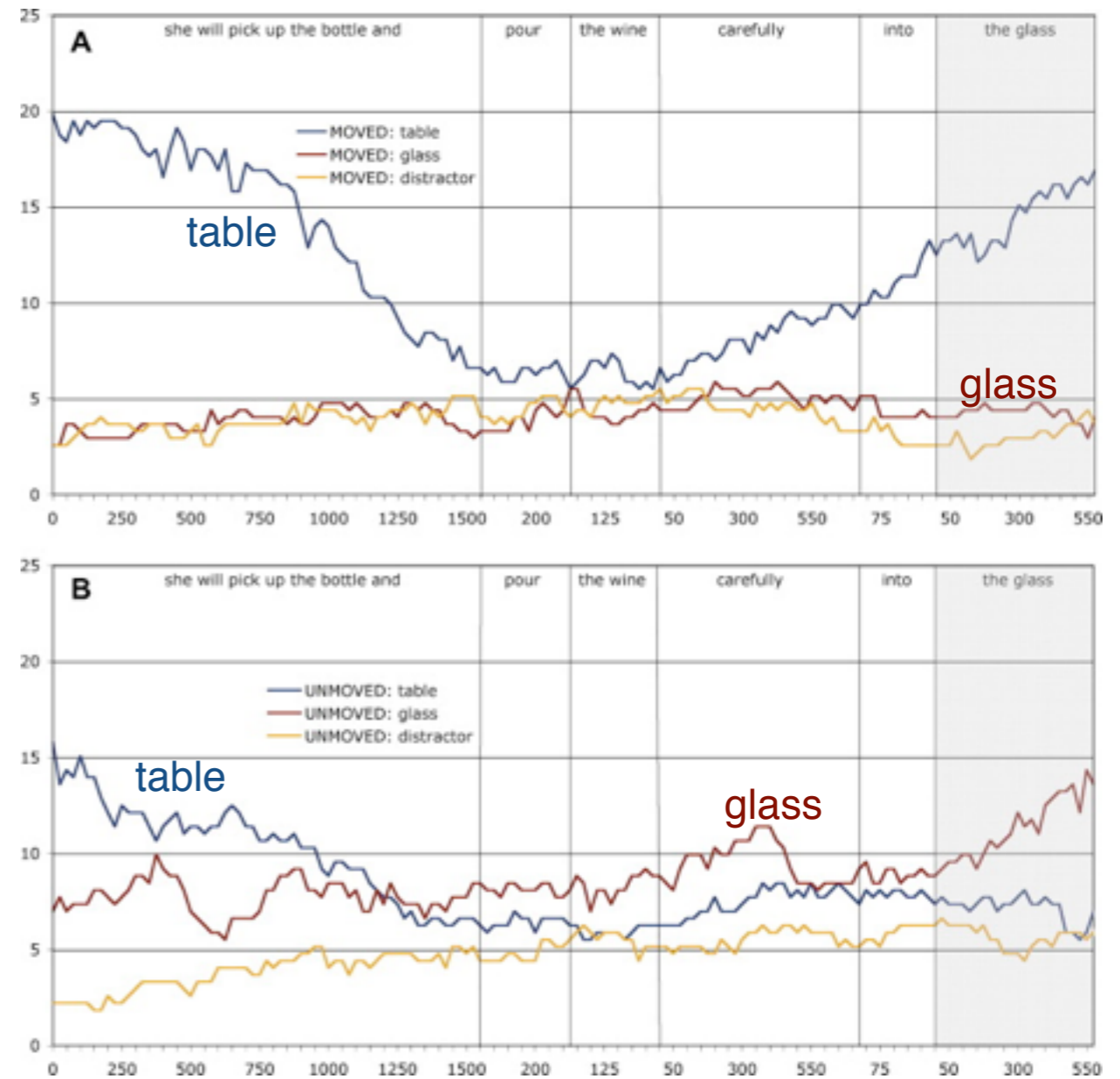She will pick up the bottle, and pour the wine carefully into the glass

(Altmann & Kamide, 2009)

The "represented" scene

Summary

- Eye-movements

  - are a result of (incremental) information integration

    - Lexical access

    - Syntactic ambiguity resolution

    - Interpretation of contrastive features in reference resolution

  - index predicted contents

  - help (?) memory retrieval

  - index online up-to-date **representations** of the context

Outlook

- Next time: "Visual World Paradigm II"

  - What takes precedence ✓

  - What is predicted?

References

- Altmann, G.T.M. (2004). Language-mediated eye movements in the absence of a visual world: the 'blank screen paradigm. Cognition 93, B79-B87.

- Altmann, G.T.M. and Kamide, Y. (2009) Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. Cognition 111 (1), 55-71.

- Cooper, R. (1974). The control of eye fixation by the meaning of spoken language. Cognitive Psychology 6, 84-107.

- Kamide, Y., Altmann, G.T.M., and Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. JML 49, 133-156.

- Knoeferle, P. & Crocker, M.W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. Cognitive Science 30, 481-529.

- Laeng, B., Bloem, I., D'Ascenzo, S., and Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. Cognition 131, 263-283.

- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. Science 268 (5217), 1632-1634.