Visual Attention in Human-Robot Interaction

When robot gaze helps human listeners

Gaze in HRI

* Can robot gaze fulfill similar functions as human gaze?

  * Coordinating dialogue (e.g. Cassell et al. '99)

  * Information structure (e.g. Mutlu et al. '06)

  * Engagement (e.g. Sidner et al.'05, Breazeal et al.'06, Kuno et al.'07, Yamazaki et al.'08)

  * Visual reference, grounding

  * Signaling and checking understanding / joint attention

People do...

✤ ...look at an object they are about to mention

✤ (about 1 sec before mentioning)

✤ ...look at an object that someone else mentions

✤ (about 200-500 msec after mentioning)

✤ ...look at where the speaker looks

Should robots do...

✤ ...look at an object they are about to mention ?

   ✤ (about 1 sec before mentioning)

✤ ...look at an object that someone else mentions ?

   ✤ (about 200-500 msec after mentioning)

✤ ...look at where the speaker looks ?

Should robots do...

* ...look at an object they are about to mention ?
    * (about 1 sec before mentioning)

* Will people look at an object the robot mentions ?

    * (about 200-500 msec after mentioning)

* Will people look at where the robot looks?

Research Questions

* Is cognitively-motivated robot gaze interpreted as a visual reference? Is it beneficial for interaction?

    * Can it be used to quickly identify *intended referents*?

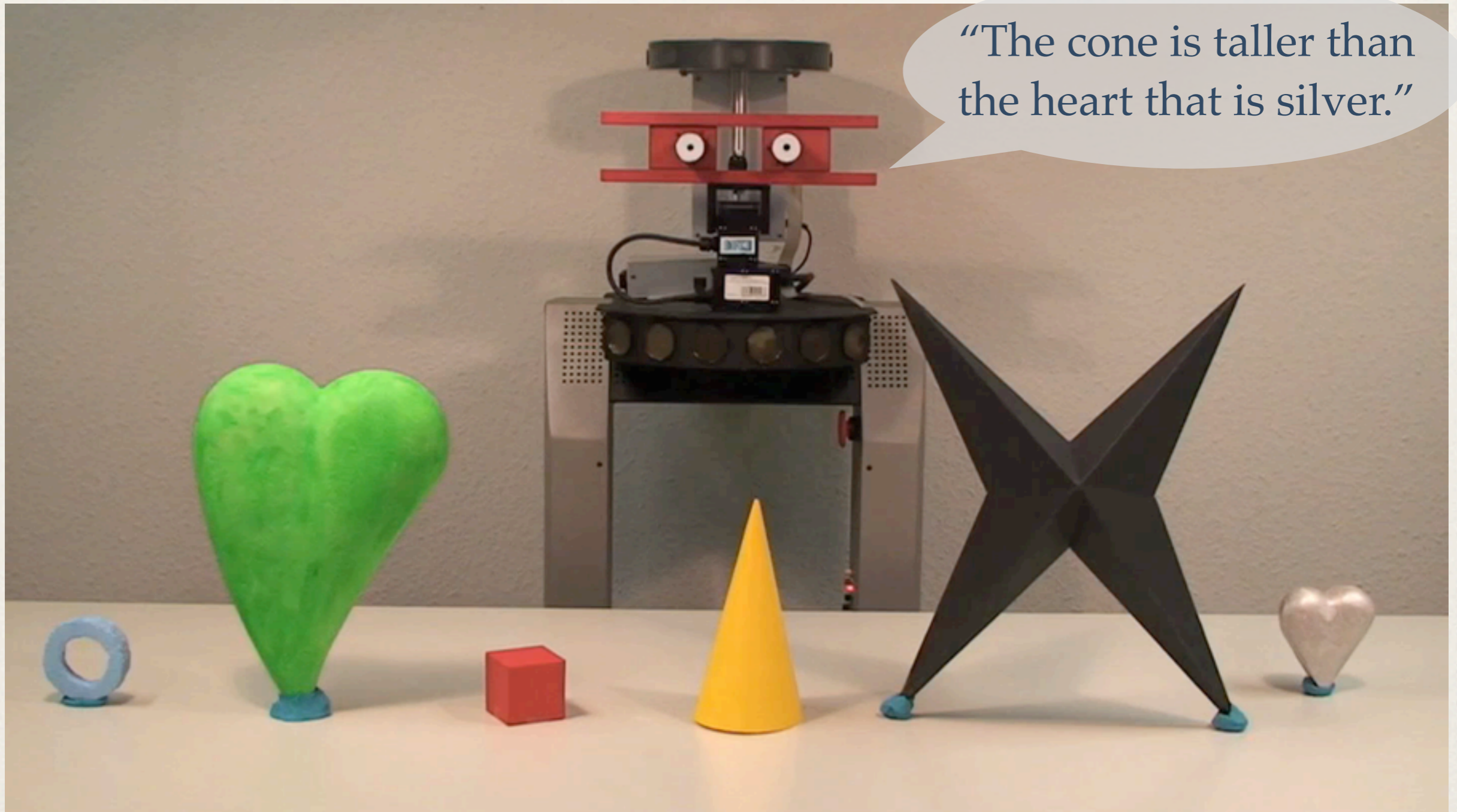* How are visual and linguistic references integrated?

Multimodal Reference
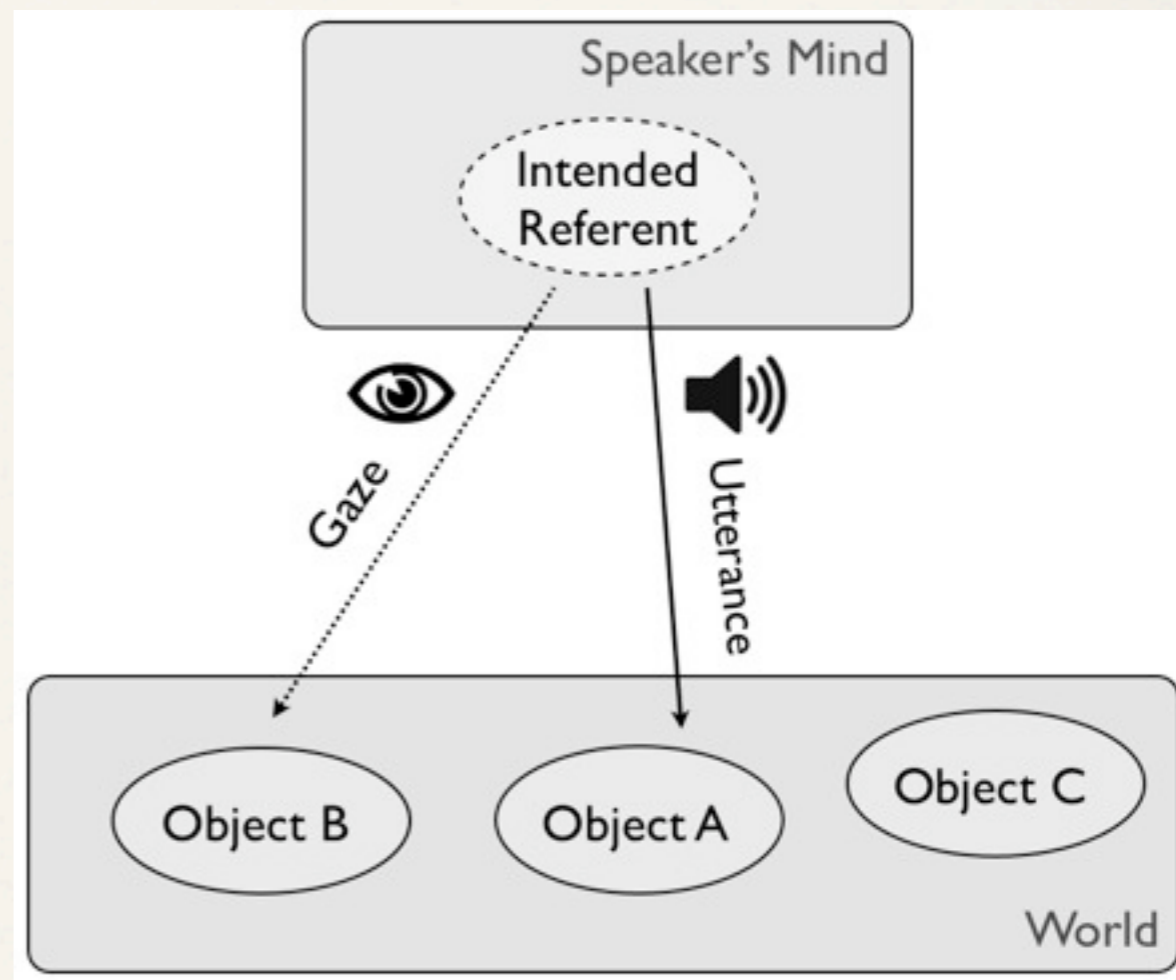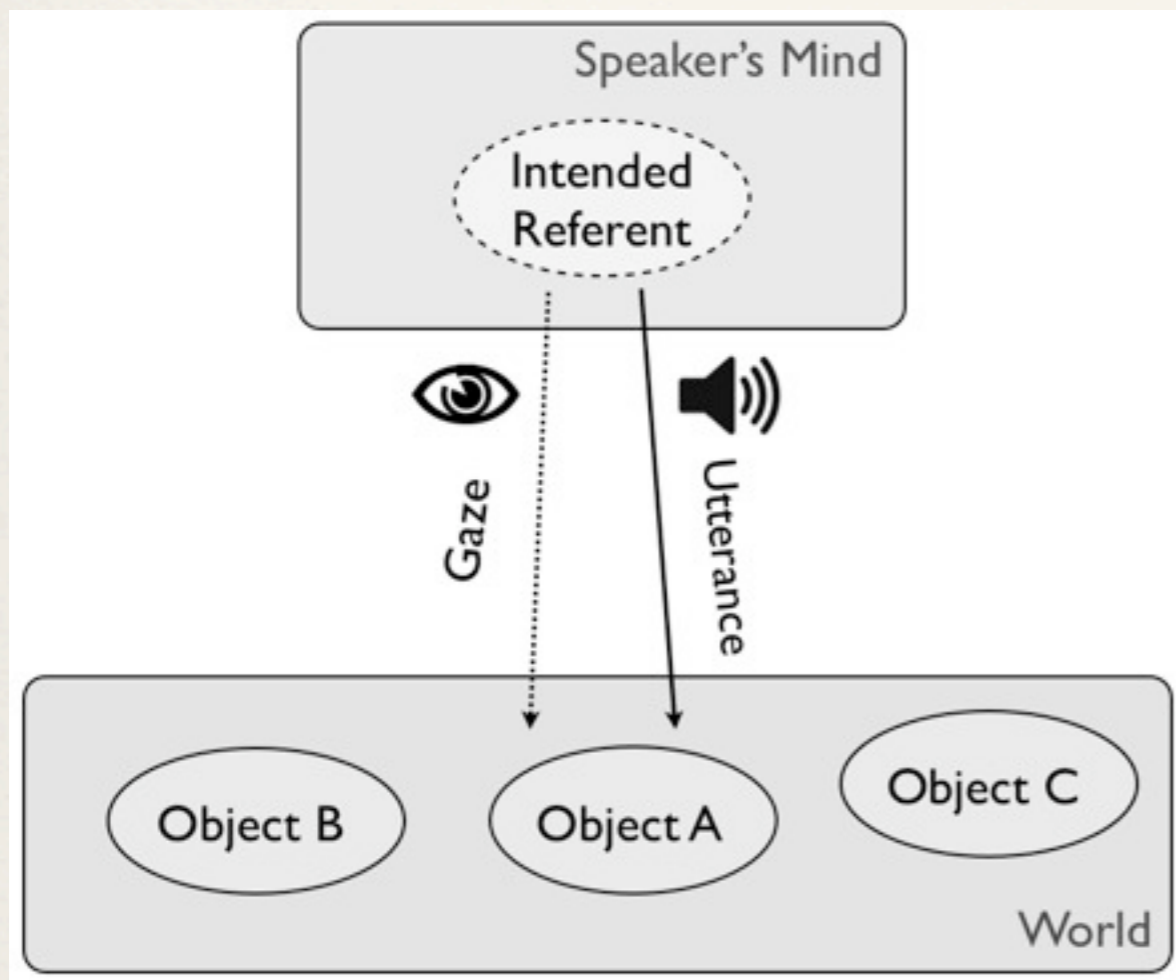
Experiment 1

Objectives

* Do people follow robot gaze at all?

* If so, this could indicate:

  * An automatic response to robot gaze when it works as attention directing cue, or

  * A response based on the understanding that the robot also *sees*, i.e., has visual attention -> Does that entail *joint attention*?

* Does gaze as a referential cue affect comprehension of linguistic reference?

Example: true - congruent

Conditions

Statement	Gaze	Spoken Sentence:
		Gaze towards:
TRUE	congruent	The cone is taller than the heart that's silver. *\<cone\>* *\<silver heart\>*
TRUE	incongruent	The cone is taller than the heart that's silver. *\<cone\>* *\<green heart\>*
TRUE	neutral	The cone is taller than the heart that's silver. ----------- -----------------
FALSE	congruent	The cone is taller than the heart that's green. \<cone\> \<green heart\>
FALSE	incongruent	The cone is taller than the heart that's green. \<cone\> \<silver heart\>
FALSE	neutral	The cone is taller than the heart that's green. ----------- -----------------

Conditions

| Statement	Gaze	Spoken Sentence:
		Gaze towards:
TRUE	congruent	The cone is taller than the heart that's **silver**. *<cone>* *<silver heart>*
TRUE	incongruent	The cone is taller than the heart that's **silver**. *<cone>* *<green heart>*
TRUE	neutral	The cone is taller than the heart that's **silver**. ----------- -----------------
FALSE	congruent	The cone is taller than the heart that's green. <cone> <green heart>
FALSE	incongruent	The cone is taller than the heart that's green. <cone> <silver heart>
FALSE	neutral	The cone is taller than the heart that's green. ----------- -----------------

Conditions

Statement	Gaze	Spoken Sentence:
		Gaze towards:
TRUE	congruent	The cone is taller than the heart that's silver. *<cone>*　　　　　　　　　*<silver heart>*
TRUE	incongruent	The cone is taller than the heart that's silver. *<cone>*　　　　　　　　　*<green heart>*
TRUE	neutral	The cone is taller than the heart that's silver. ------------　　　　　　　　------------------
FALSE	congruent	The cone is taller than the heart that's **green**. <cone>　　　　　　　　<green heart>
FALSE	incongruent	The cone is taller than the heart that's **green**. <cone>　　　　　　　　<silver heart>
FALSE	neutral	The cone is taller than the heart that's **green**. ------------　　　　　　　　------------------

Conditions

| Statement	Gaze	Spoken Sentence:
		Gaze towards:
TRUE	**congruent**	The cone is taller than the heart that's **silver**. *<cone>* **<silver heart>**
TRUE	incongruent	The cone is taller than the heart that's silver. *<cone>* *<green heart>*
TRUE	neutral	The cone is taller than the heart that's silver. ----------- ------------------
FALSE	**congruent**	The cone is taller than the heart that's **green**. <cone> **<green heart>**
FALSE	incongruent	The cone is taller than the heart that's green. <cone> <silver heart>
FALSE	neutral	The cone is taller than the heart that's green. ----------- ------------------

Conditions

Statement	Gaze	Spoken Sentence: Gaze towards:
TRUE	congruent	The cone is taller than the heart that's silver. *\<cone\>* *\<silver heart\>*
TRUE	**incongruent**	The cone is taller than the heart that's **silver**. *\<cone\>* **\<green heart\>**
TRUE	neutral	The cone is taller than the heart that's silver. ----------- -----------------
FALSE	congruent	The cone is taller than the heart that's green. \<cone\> \<green heart\>
FALSE	**incongruent**	The cone is taller than the heart that's **green**. \<cone\> **\<silver heart\>**
FALSE	neutral	The cone is taller than the heart that's green. ----------- -----------------

False
Congruent vs. Incongruent
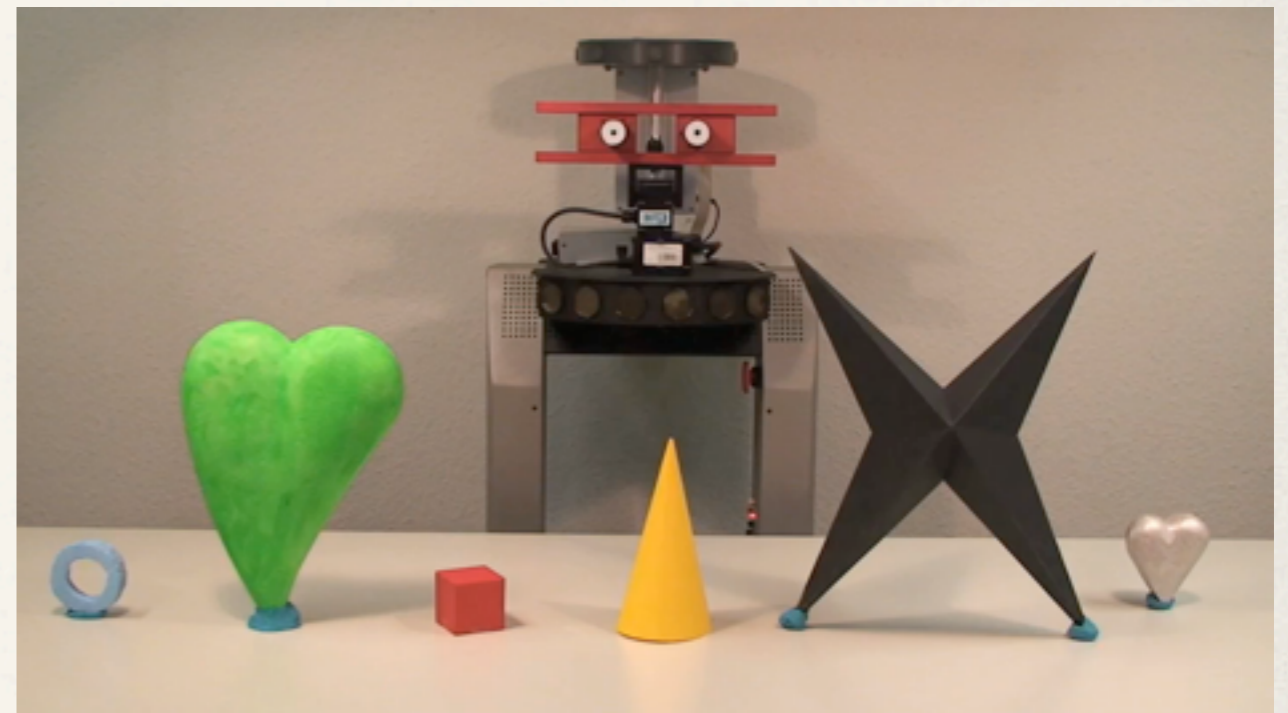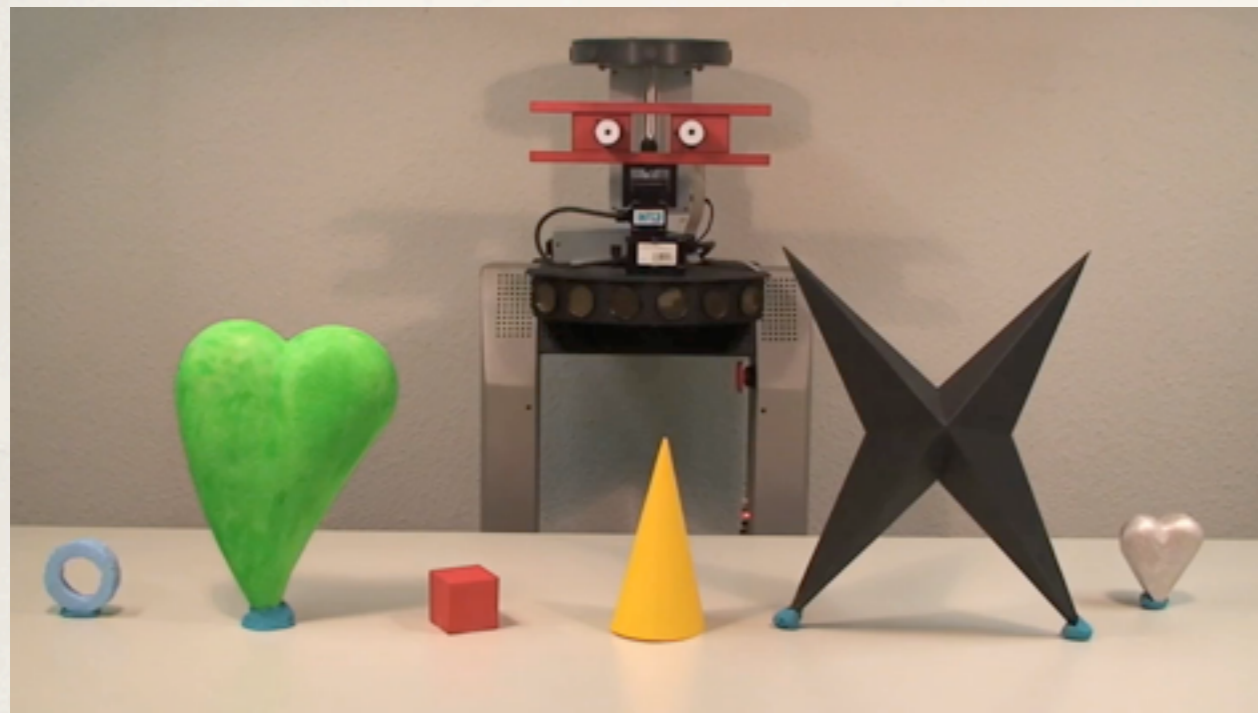
Design Details

- ✤ Task: Validate the robot's statement by pressing a button (correct/false) as fast and as accurate as possible

- ✤ 48 subjects

- ✤ 6 conditions
  - ✤ 24 items + 48 fillers = 72 trials

- ✤ Logistic Regression (inspection probability)

- ✤ Linear Mixed-Effects Models (response time)
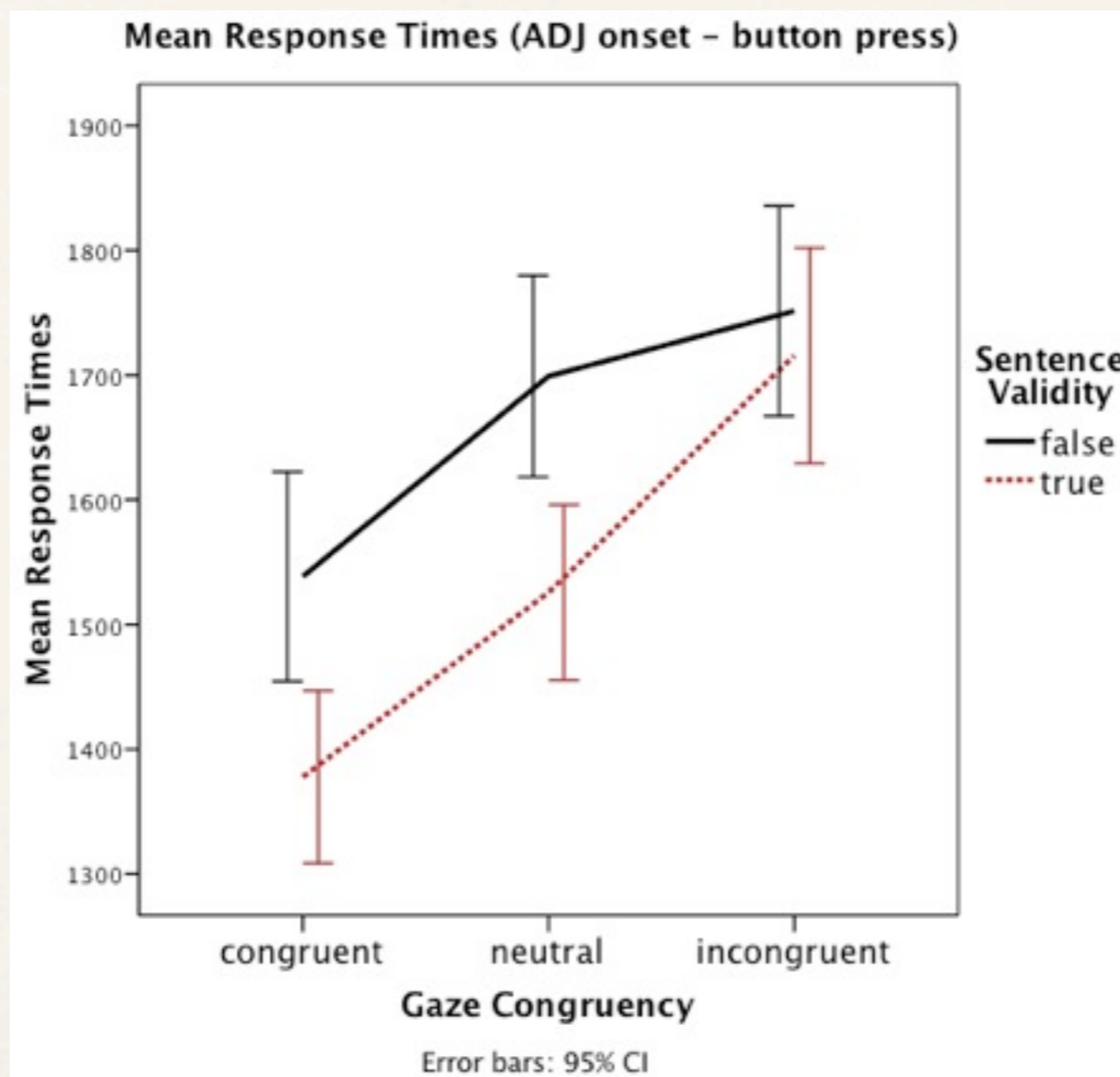
Eye-Movements
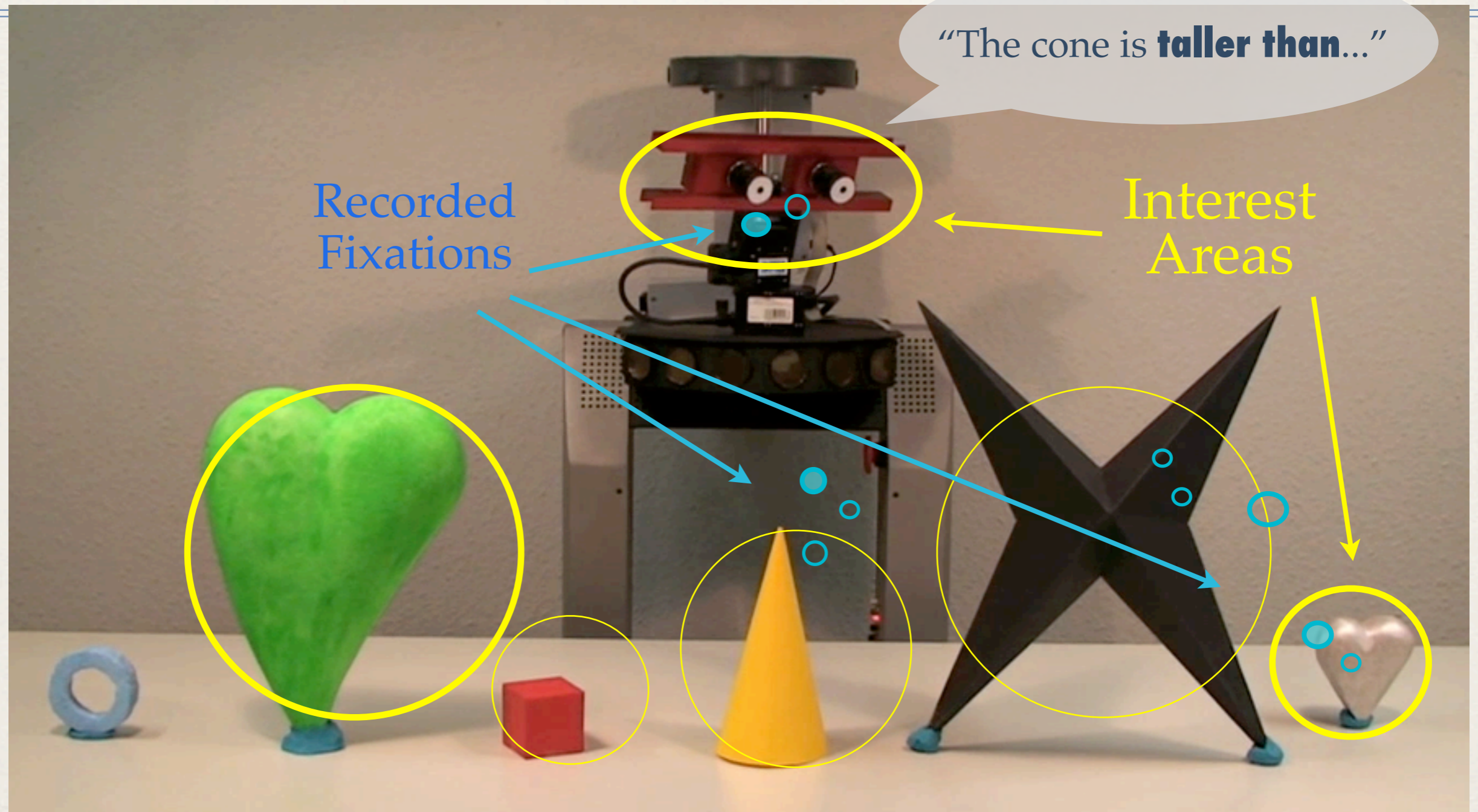
Results: Eye Movements
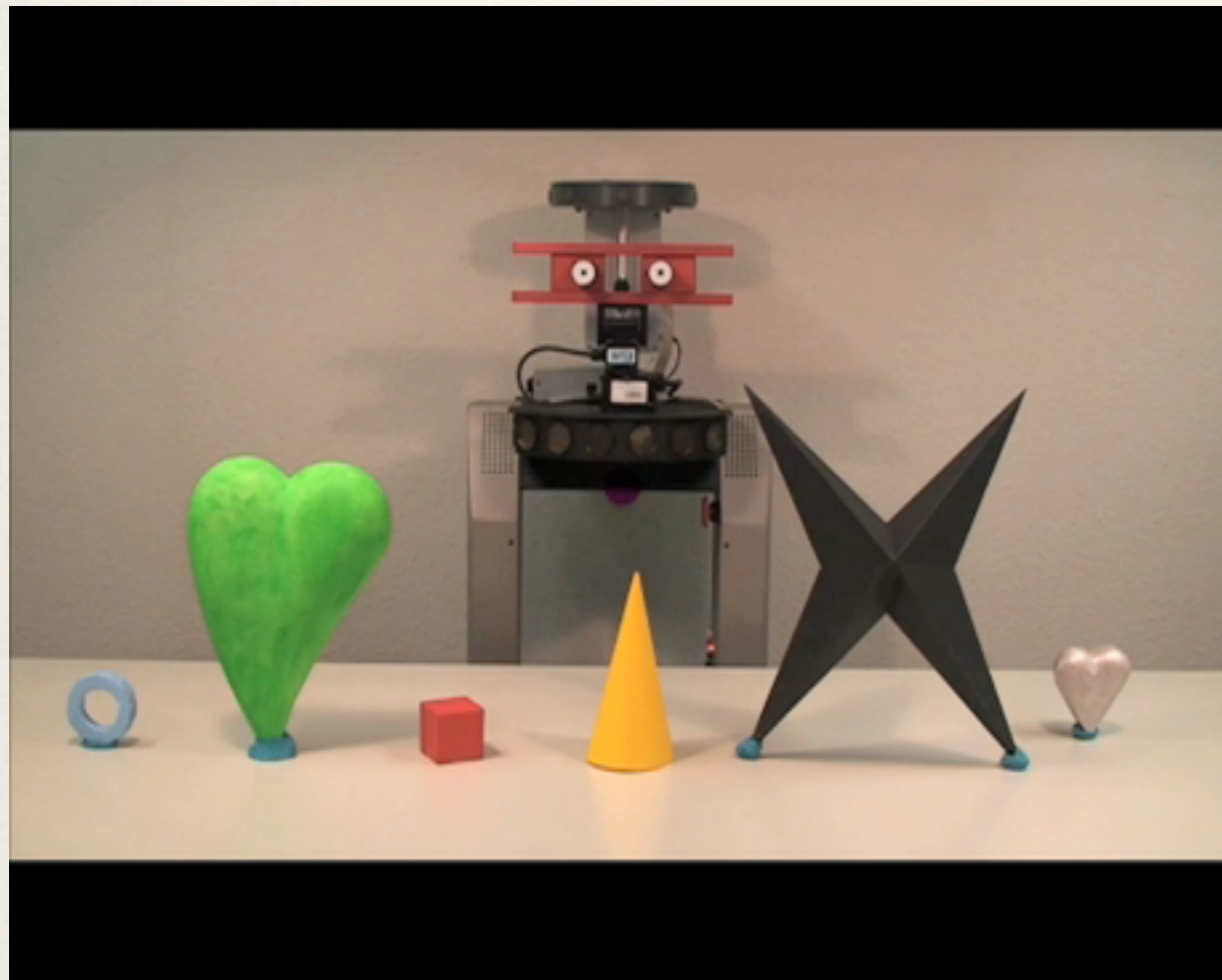
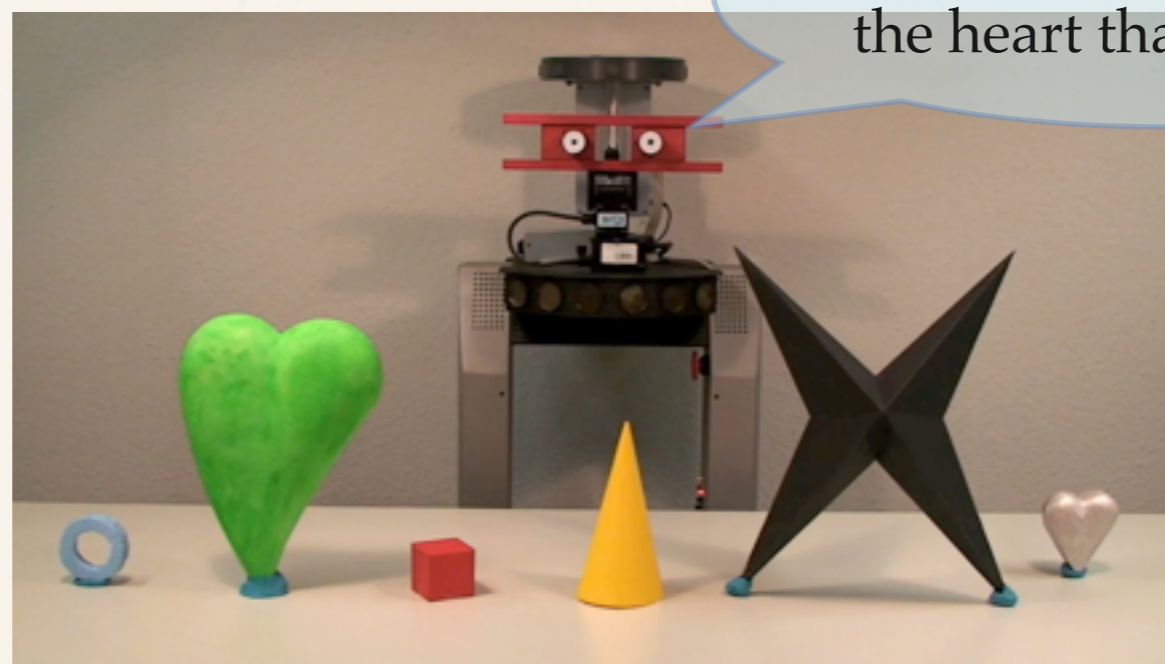True - congruent
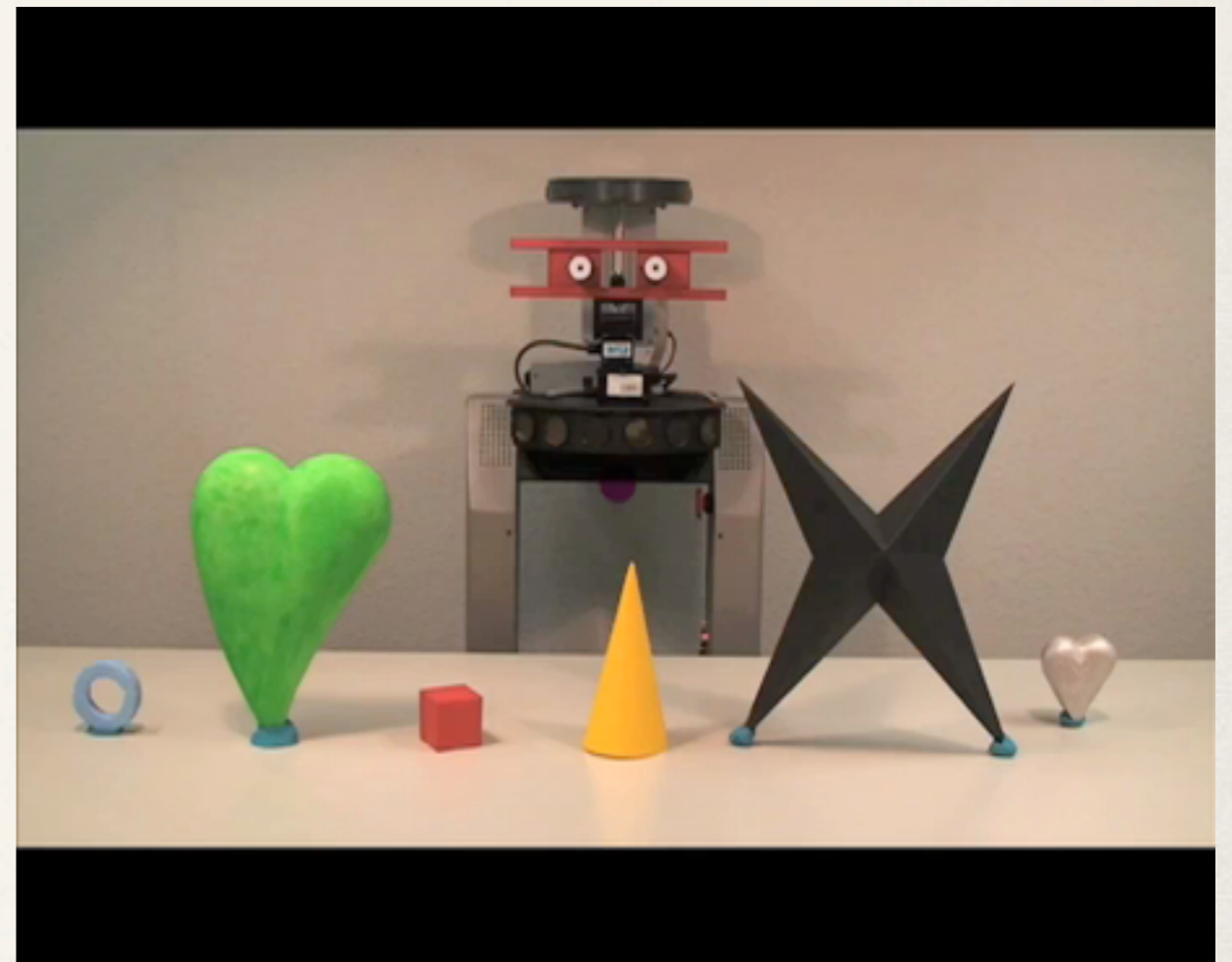
True - incongruent

Results: Response Time

Results

* Gaze and utterance mediated eye-movements

    * People follow robot gaze to objects & look at mentioned objects

    * More frequent looks to object that was looked at compared to shape competitor that was not looked at

* When robot gazed and mentioned the same object (compared to different objects), people were faster to validate the utterance.

    * Match/mismatch of people hypotheses with actual utterance?

* Congruent gaze facilitates comprehension, incongruent gaze disrupts comprehension!

Experiment 2

Objectives

* Congruent Gaze facilitated comprehension, incongruent gaze disrupted comprehension!

* Is this effect a result of

    * *Intention Recognition*: Gaze reflects referential intentions

    * *Attention Direction* : Gaze directs visual attention - maybe even automatically

* Do people believe that robot gaze signals an intention of the robot to mention it? What do they believe when gaze and utterance are incongruent?

    * How to assess peoples beliefs (indirectly)?

False - Neutral Gaze

(Predicted) Correction:
—| The cone is <u>shorter</u> than the heart that's green.
—| The cone is taller than the heart that's <u>silver</u>.

False - Congruent Gaze
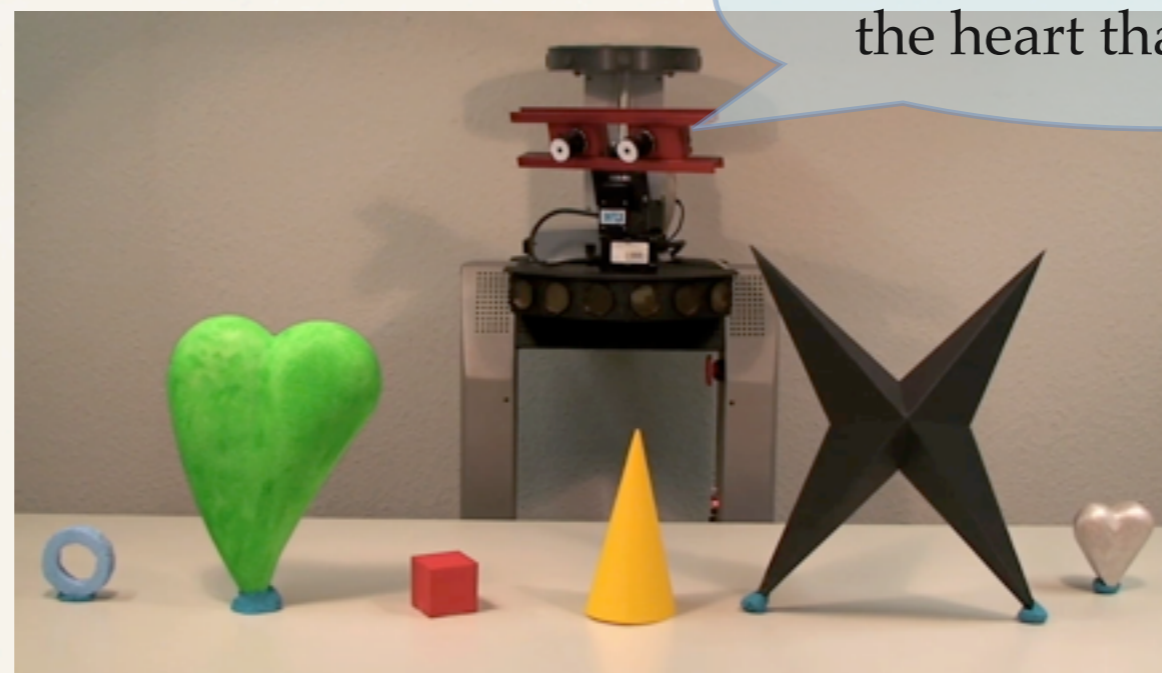
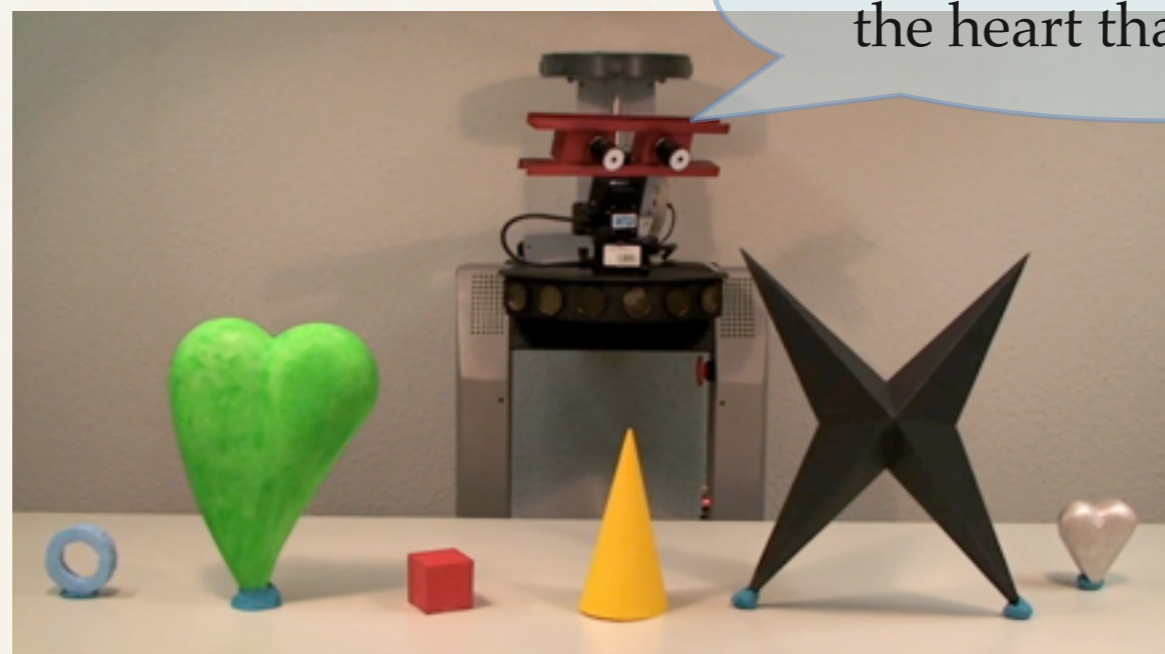"The cone is taller than the heart that's green."

(Predicted) Correction:
— **The cone is <u>shorter</u> than the heart that's green.**
— The cone is taller than the heart that's <u>silver</u>.

False - Incongruent Gaze

"The cone is taller than the heart that's green."

(Predicted) Correction:
— The cone is <u>shorter</u> than the heart that's green.
— **The cone is taller than the heart that's <u>silver</u>.**

Design Details

* Task: Correct robot's mistake by giving a sentence that would have been correct to say (starting with same NP)

* 36 subjects

* 6 conditions

  * 24 items + 48 fillers = 72 trials

* Logistic Regression (produced sentences contain target/competitor & inspection probabilities)

Results

✤ Eye-Movements: Gaze and utterance mediated eye-movements

  ✤ People follow robot gaze to objects & look at mentioned objects

✤ Produced Corrections: Robot gaze affects what people correct

**Robot Utterance:** "The cone is taller than the heart that's green."

Described Objects in Correction Statements
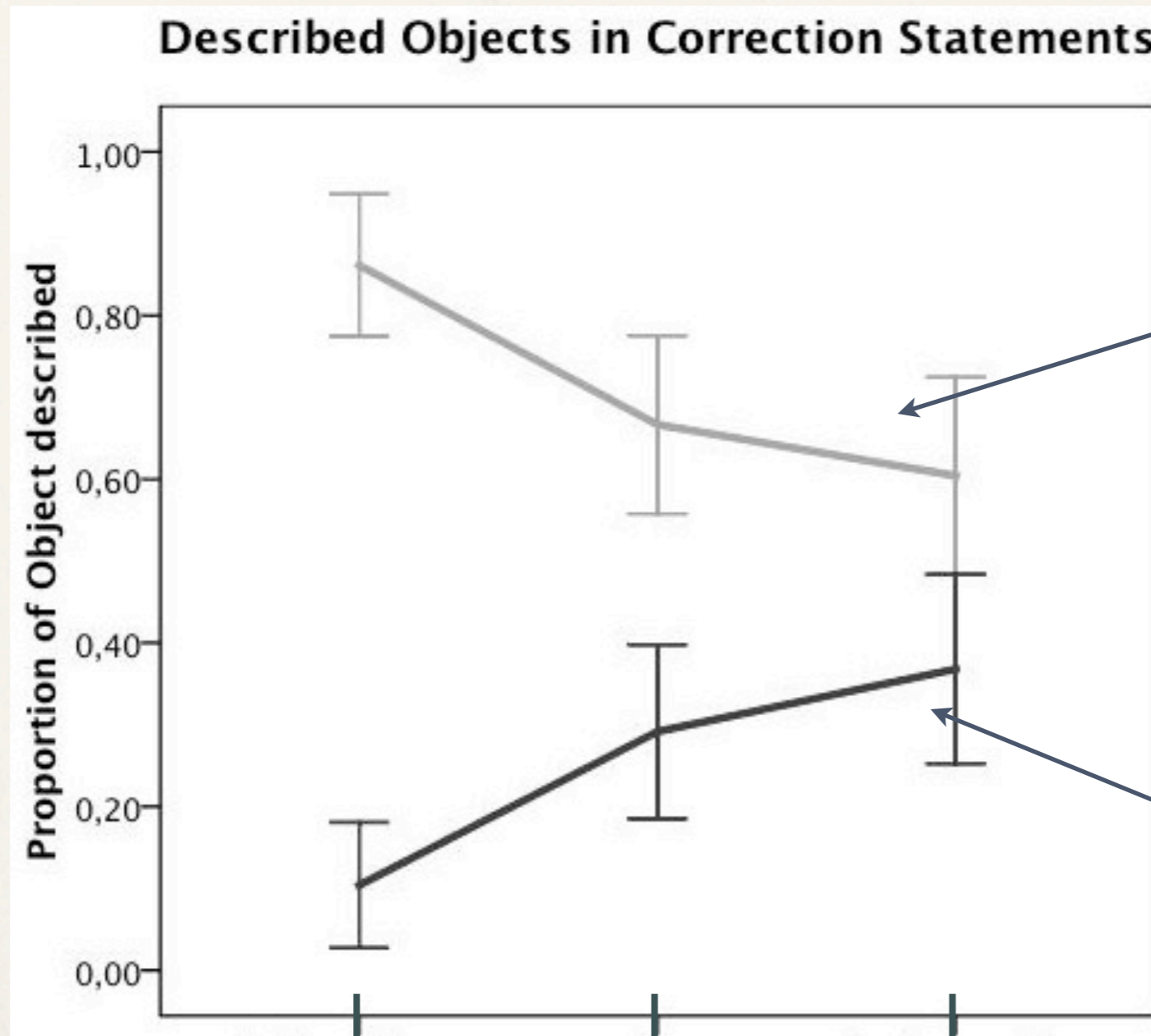
"green heart"

"silver heart"

Gaze to: <Green heart> ----- <Silver heart>

Conclusions (Exp 1 & 2)

✤ People follow robot gaze

   ✤ Consistent with results on reflexive visual orienting in response to human eyes (Driver et al '99, Langton & Bruce '99, Friesen & Kingstone '98)

✤ Robot gaze influences utterance comprehension!

✤ 'Long-term' effect of gaze-following (our RT data vs. cueing effect)

✤ Integration of multi-modal references:

   ✤ Gaze influences (intended) referent resolution!

Gaze as Reflexive Cue

1500 ms

50 ms

50 ms/950 ms

To response

+   o
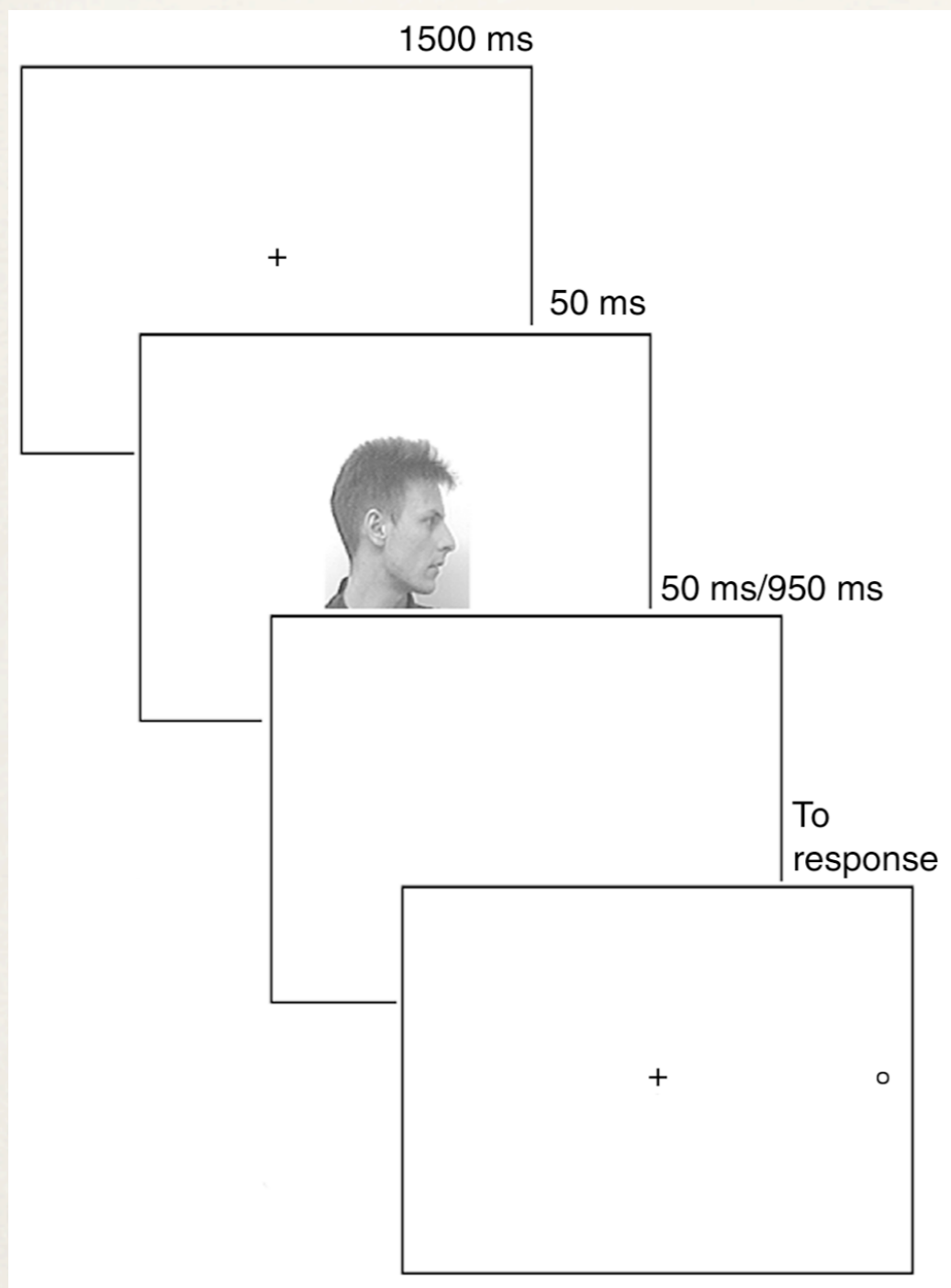
✤ Stimulus onset asychrony (100/1000ms delay before target onset)

✤ For 100ms: Response faster for cued target

✤ For 1000ms: No cueing effect

Conclusions (Exp 1 & 2)

* People robustly follow robot gaze

    * Consistent with results on reflexive visual orienting in response to gaze cues (Driver et al '99, Langton & Bruce '99, Friesen & Kingstone '98)

* Robot gaze also influences utterance comprehension!

    * 'Long-term' effect of gaze-following (our RT data vs. short-lived cueing effect)

    * Integration of multi-modal references:

        * Gaze influences resolution of (intended) referents!

Conclusions (Exp 1 & 2)

- Gaze influences resolution of (intended) referents!

  - Gaze affects people's beliefs about referential intentions, i.e. supports *Intentional Account*

 - Could a purely *attentional* explanation account for these phenomena?

Synchronicity

✤ Does gaze communicate speaker's goals and referential intentions any time or does its occurrence define its meaning?

✤ Inferred referential intentions would be expected to be more persistent than a purely attentional effect of gaze

✤ Is chronological/sequential order and temporal alignment critical?

   ✤ For indirect gaze cues (Kreysa et al '09)

   ✤ Direct versus indirect gaze cue ?
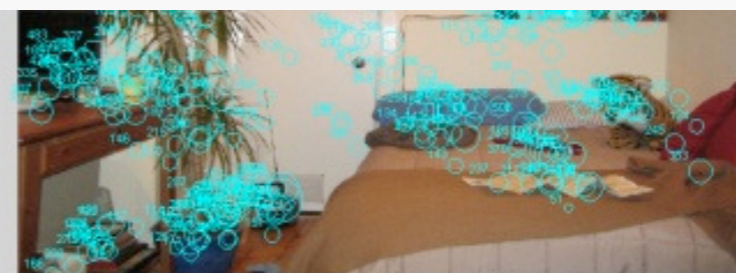
   ✤ Temporal vs. sequential alignment ?

Alignment of Indirect Cues

(Kreysa et al '09)

* Task: Identify described objects as fast as possible

* Record initial speaker gaze

* Display to listeners

* Measure clicking latencies
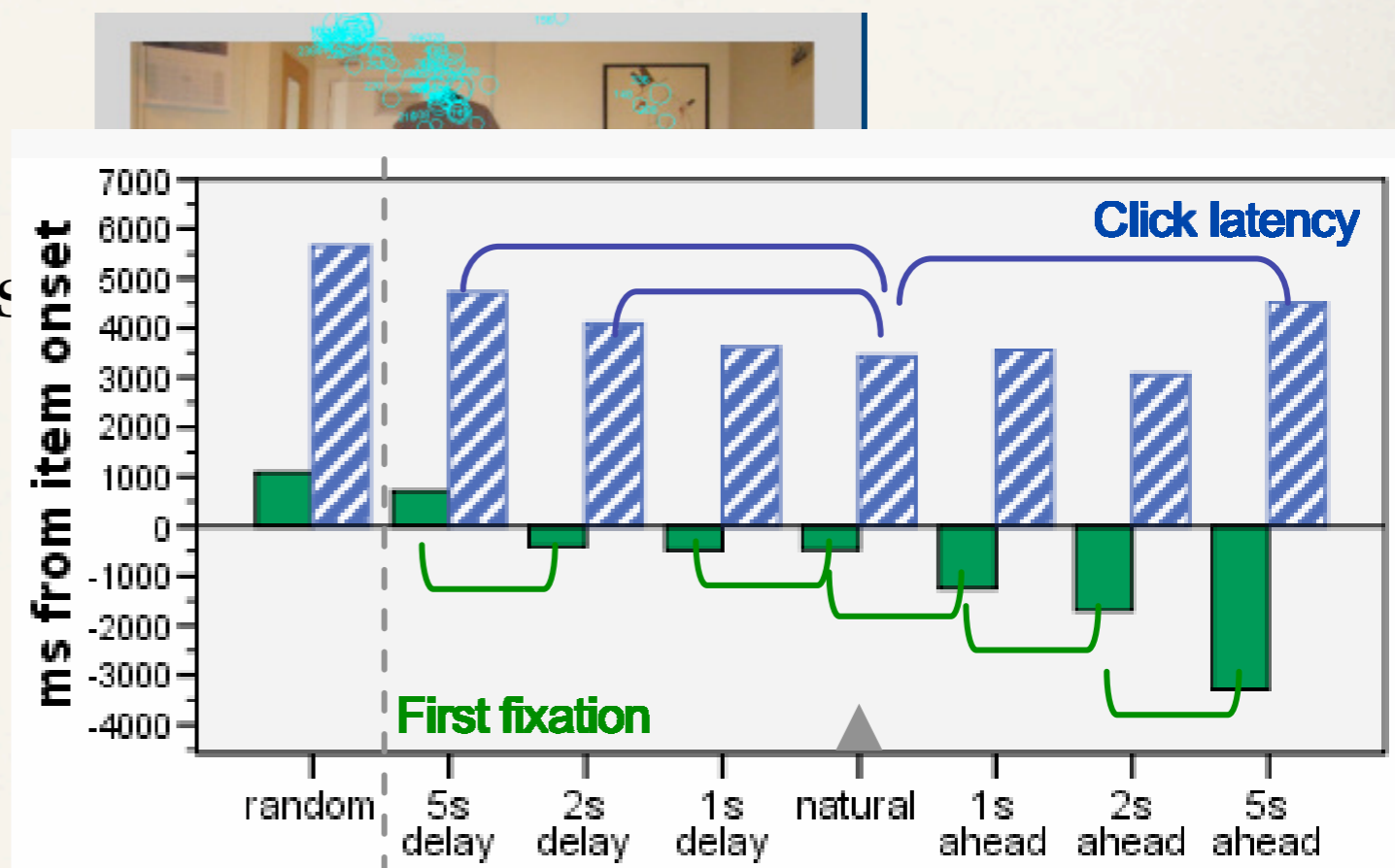
NB: brackets indicate significant planned comparisons, p < .05

Listener eye movements: no cursor

Experiment 3

Objectives

* How flexible are people in using gaze cues? What role does synchronization play?

* The temporal and sequential flexibility will reveal insights into the nature of this cue:

  * *Intentional*: Reflects referential intentions

  * Purely *Attentional*: Directs visual attention - maybe even automatically

* Temporal and sequential synchronization will be manipulated

Conditions

* Temporally synchronized vs. preceding

  * `<cylinder><pyramid>` The cylinder is taller than the pink pyramid.

  * `<cylinder>` The cylinder is taller than the `<pyramid>` pink pyramid.

* Sequentially original/congruent vs. reverse

  * `<cylinder>` The cylinder is taller than the `<pyramid>` pink pyramid.

  * `<cylinder>` The pyramid is shorter than the `<pyramid>` cylinder.

Original vs. Reverse
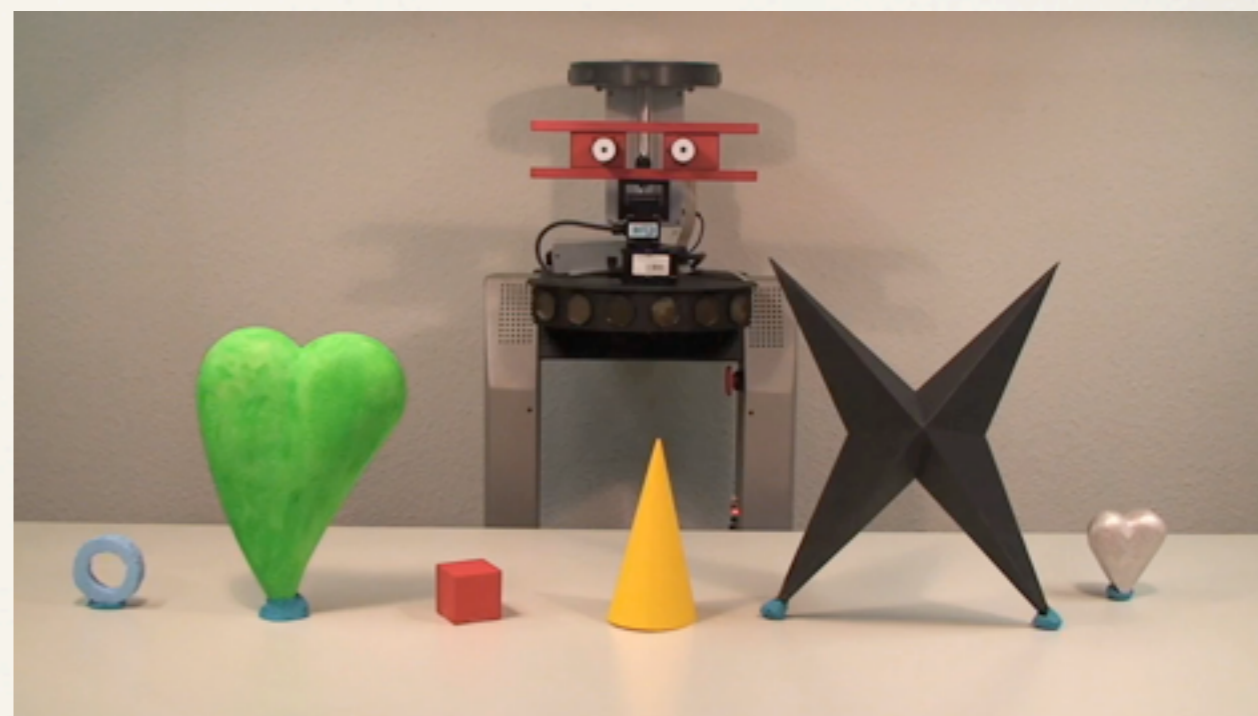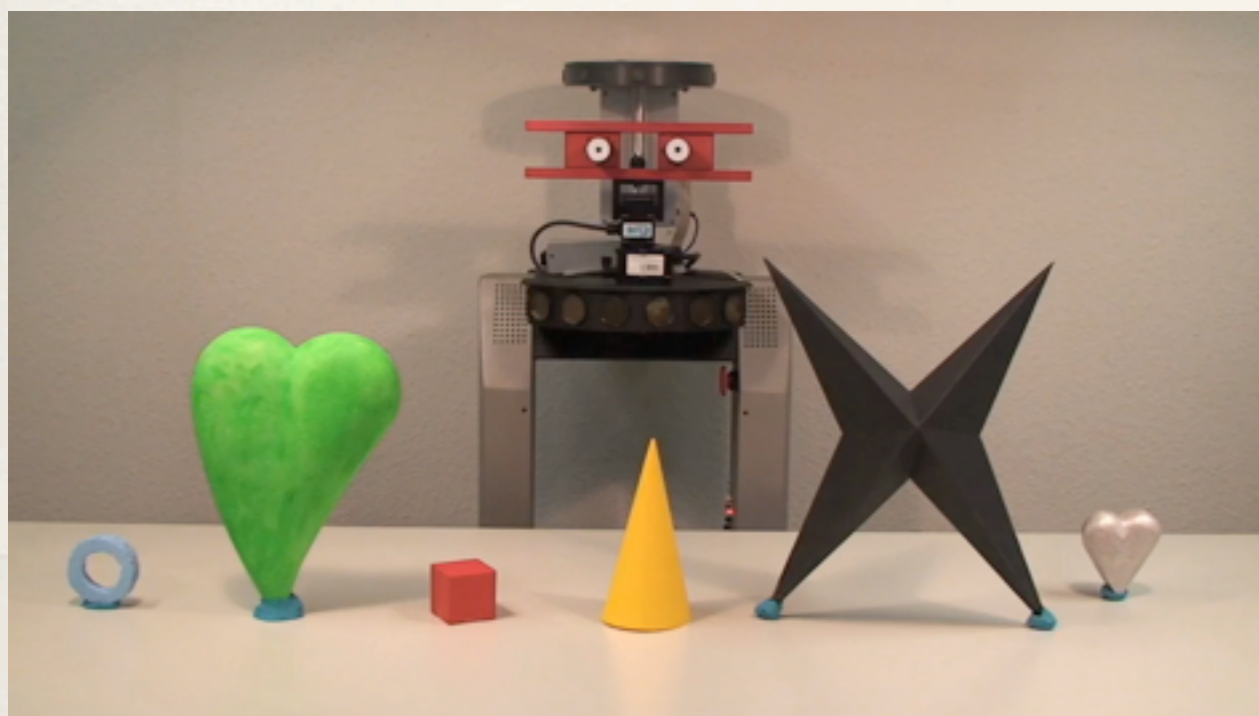
Synchronized vs. Preceding
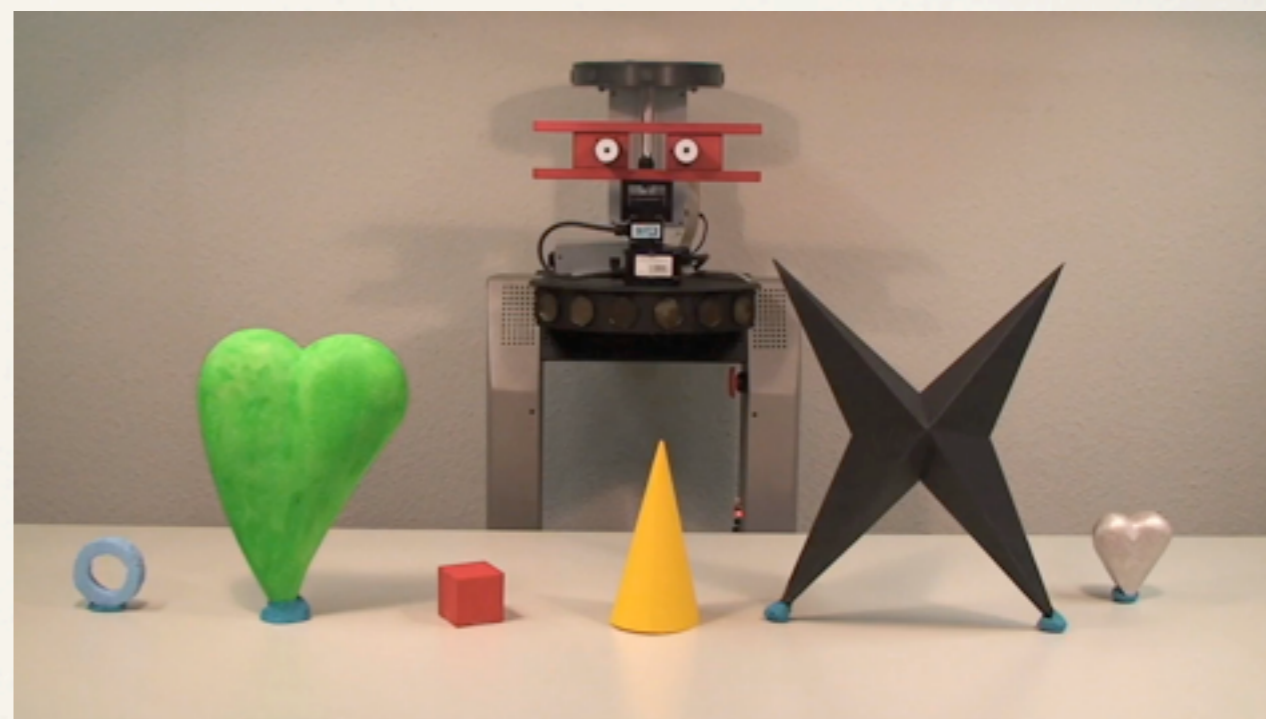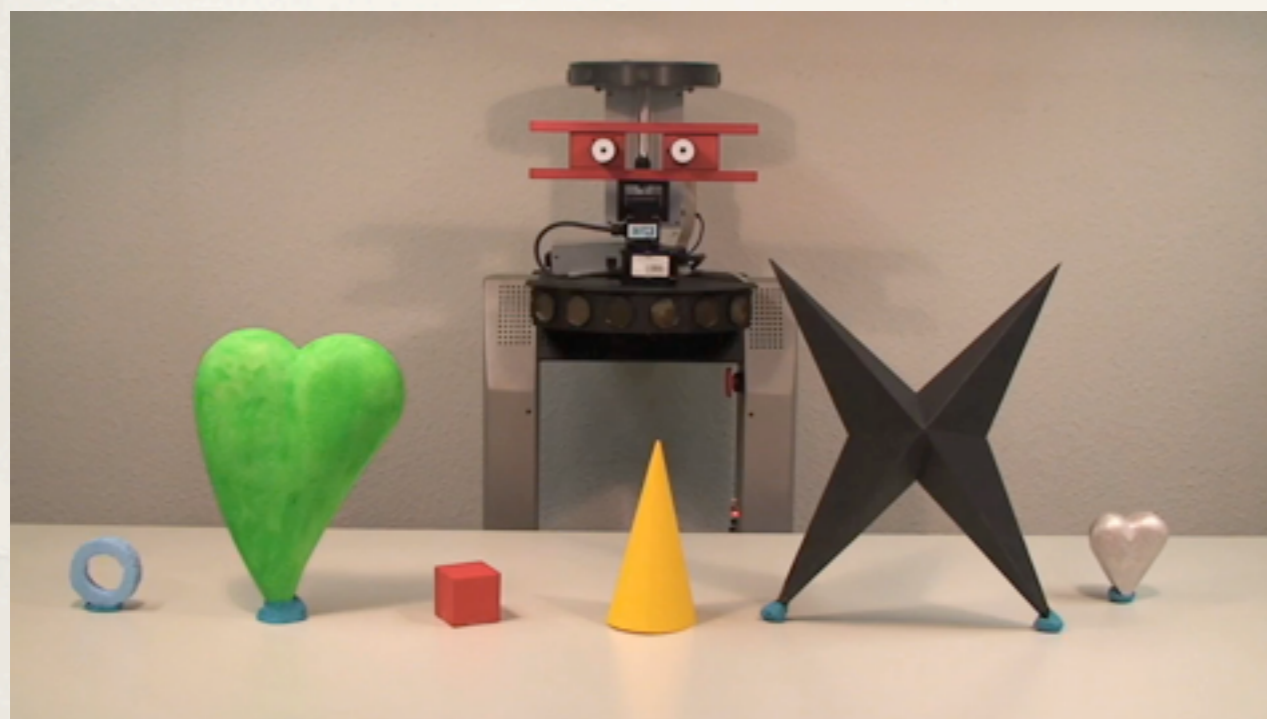
Design Details

* Task: Validate the robot's statement by pressing a button (correct/false) as fast and as accurate as possible

* 32 subjects

* 4 conditions

    * 20 items + 36 fillers = 56 trials

* Logistic Regression (inspection probability)

* Linear Mixed-Effects Models (response time)

Conditions

	Sentence Order **original**	Sentence Order **reverse**
Synchronized	The cone is taller than the silver heart. <cone> <heart>	The silver heart is shorter than the cone. <cone> <heart>
Preceding	The cone is taller than the .... <cone> <heart>	The silver heart is shorter than ... <cone> <heart>

(Confounded) Conditions

	Sentence Order **original / +adjective**	Sentence Order **reverse / -adjective**
Synchronized	The cone is taller than the silver heart. <cone> <heart>	The silver heart is shorter than the cone. <cone> <heart>
Preceding	The cone is taller than the …. <cone> <heart>	The silver heart is shorter than…. <cone> <heart>

Response Times

(Noun Onset - Button Press)

	original	reverse
synchronized	1331.8	1548.6
preceding	1289.1	1586.6

good <* reverse

Eye-movements
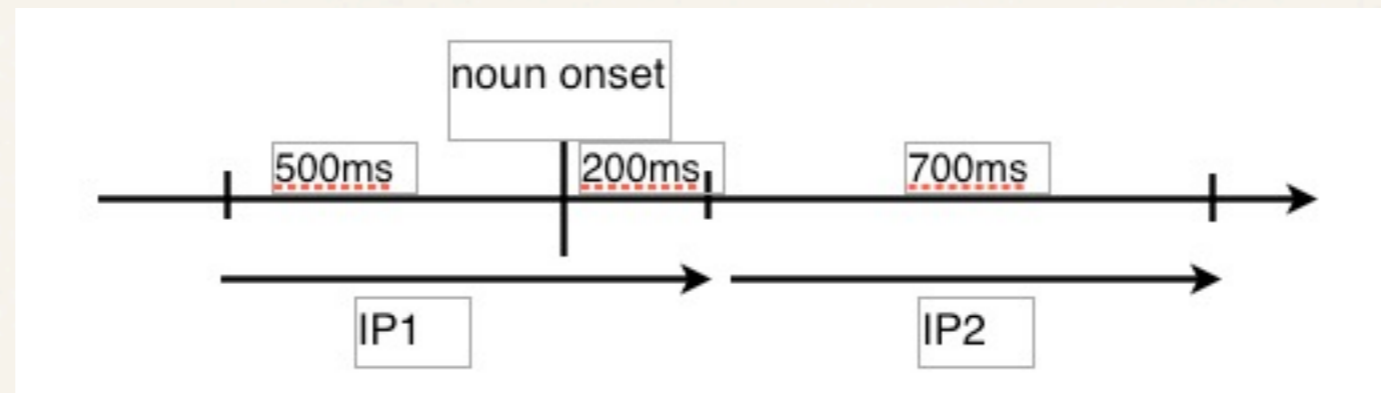
IP1	original	reverse
aligned	.57	.55
preceding	.46	.43

>*

IP2	original	reverse
aligned	.34	.68
preceding	.39	.73

<*

Conclusions

* Kreysa's results suggest that temporally shifted cues (>2sec) are not so useful anymore, even similar to random cues

* Our results suggest no influence of temporal synchronization, i.e., effect of gaze seems to be persistent and even when preceding it is:

    * beneficial, when in congruent sequential order with spoken reference

    * not so beneficial, when in reversed order

Experiment 4

Objectives

* Our results suggest no influence of temporal synchronization, i.e., effect of gaze seems to be persistent and even when preceding it is:

    * beneficial, when in congruent sequential order with spoken reference

    * not so beneficial, when in reversed order

* Reversed gaze and speech cues are not as beneficial as congruently ordered cues, but do they disrupt comprehension? -> compare to neutral gaze

Conditions

	Sentence Order **original**	Sentence Order **reverse**
Synchronized	The yellow cone is taller than the silver heart. \<cone\> \<heart\>	The silver heart is shorter than the yellow cone. \<cone\> \<heart\>
Neutral	The yellow cone is taller than the silver heart.	The silver heart is shorter than the yellow cone.

Conditions

	Sentence Order **original**	Sentence Order **reverse**
Synchr onized	The yellow **cone** is taller than the silver **heart**. <cone> <heart>	The silver **heart** is shorter than the yellow **cone**. <cone> <heart>
Neutral	The yellow cone is taller than the silver heart.	The silver heart is shorter than the yellow cone.

Conditions

	Sentence Order **original**	Sentence Order **reverse**
Synchr onized	**Original** The yellow **cone** is taller than the silver **heart**. <cone> <heart>	**Reverse** The silver **heart** is shorter than the yellow **cone**. <cone> <heart>
Neutral	The yellow cone is taller than the silver heart.	The silver heart is shorter than the yellow cone.

Design Details

* Task: Validate the robot's statement by pressing a button (correct/ false) as fast and as accurate as possible

* 32 subjects

* 4 conditions

    * 20 items + 32 fillers = 52 trials

* Logistic Regression (inspection probability)

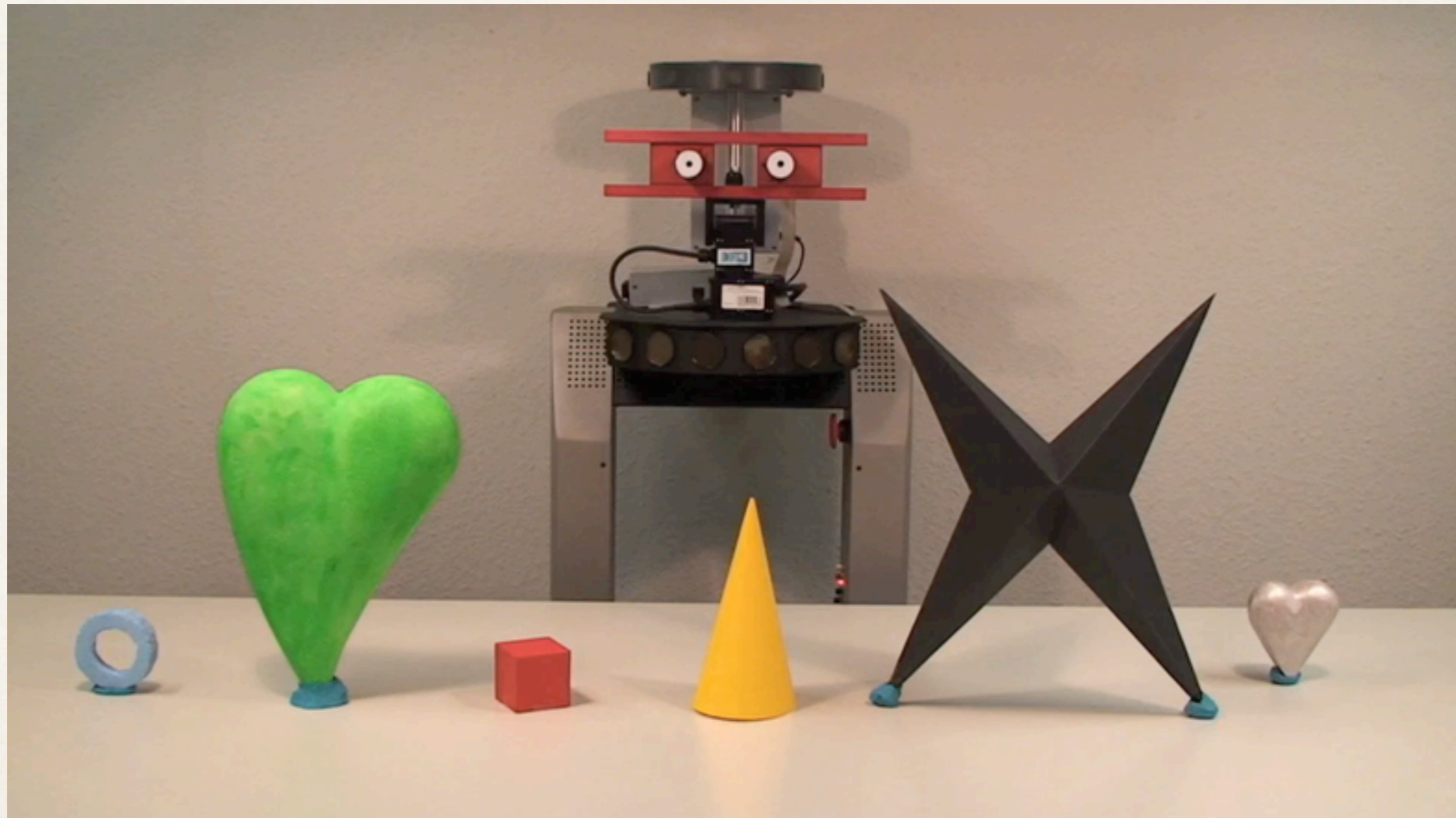* Linear Mixed-Effects Models (response time)

Response Times

(Adjective Onset - Button Press)

	original	reverse
synchronized	1459.8	1567.6
neutral	1544.1	1404.2
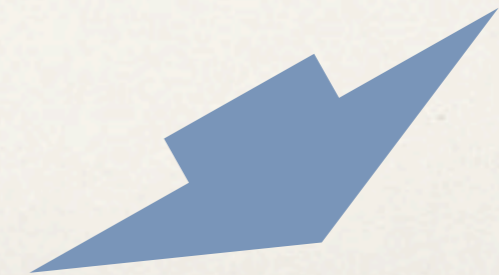
Interaction!

Response Times

Summary

- Human behaviour in HRI is consistent with that in HHI

  - Utterance-mediated & gaze-mediated gaze

  - Congruent robot gaze influences comprehension time

  - Incongruent robot gaze disrupts comprehension

    - Reference mismatch & sequential mismatch

- Interpretation and integration of both visual (gaze) and linguistic reference to resolve *intended referent*

- People align visual attention with robot's "visual attention": *Joint Attention*?

Gaze Spectrum (?)
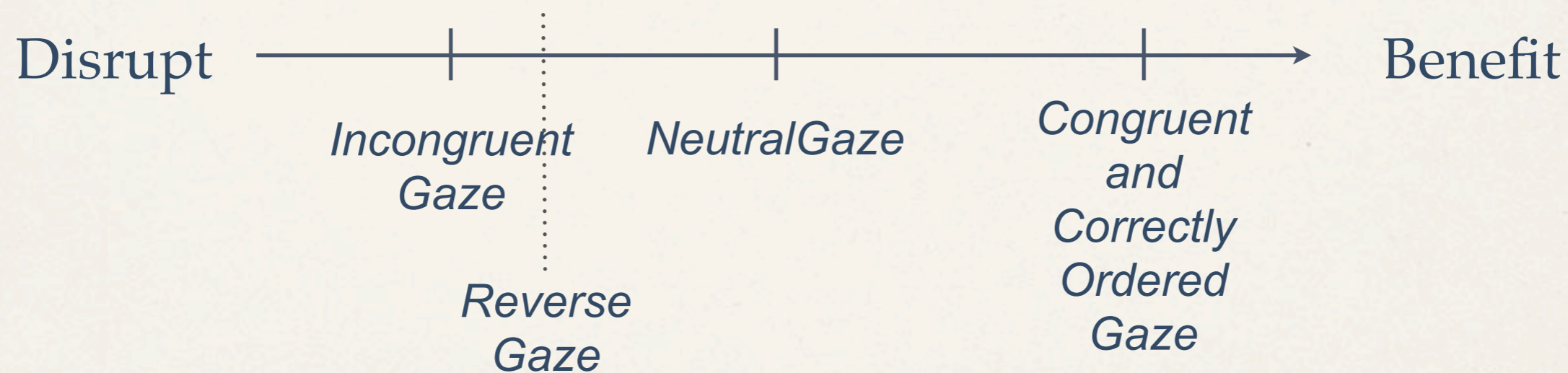
Outlook

✤ Moving to Virtual Agents

✤ More complex scenes
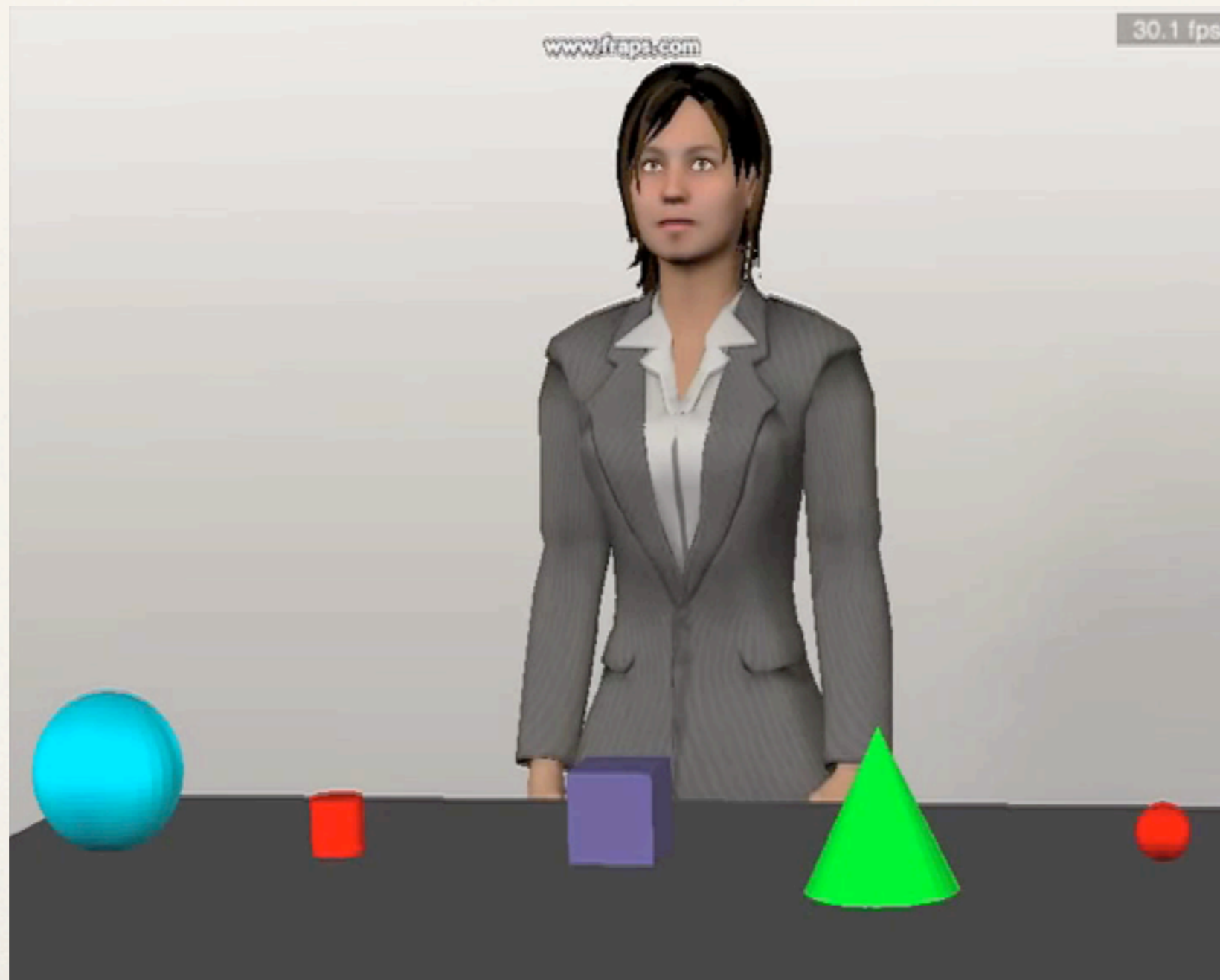
✤ Adding other multimodal cues, e.g. pointing

Outlook

* Replication with virtual agents

  * Will two single eye/head movements that the robot made in the particular condition appear strange for a human-like character?

* Gaze reveals something about intentions. Can it also trigger affordances or even override mentioned actions?

  * What relevance are looked-at-objects assigned when involved but not mentioned in a situation.

  * Interaction of object affordances with mentioned action verbs?

Outlook
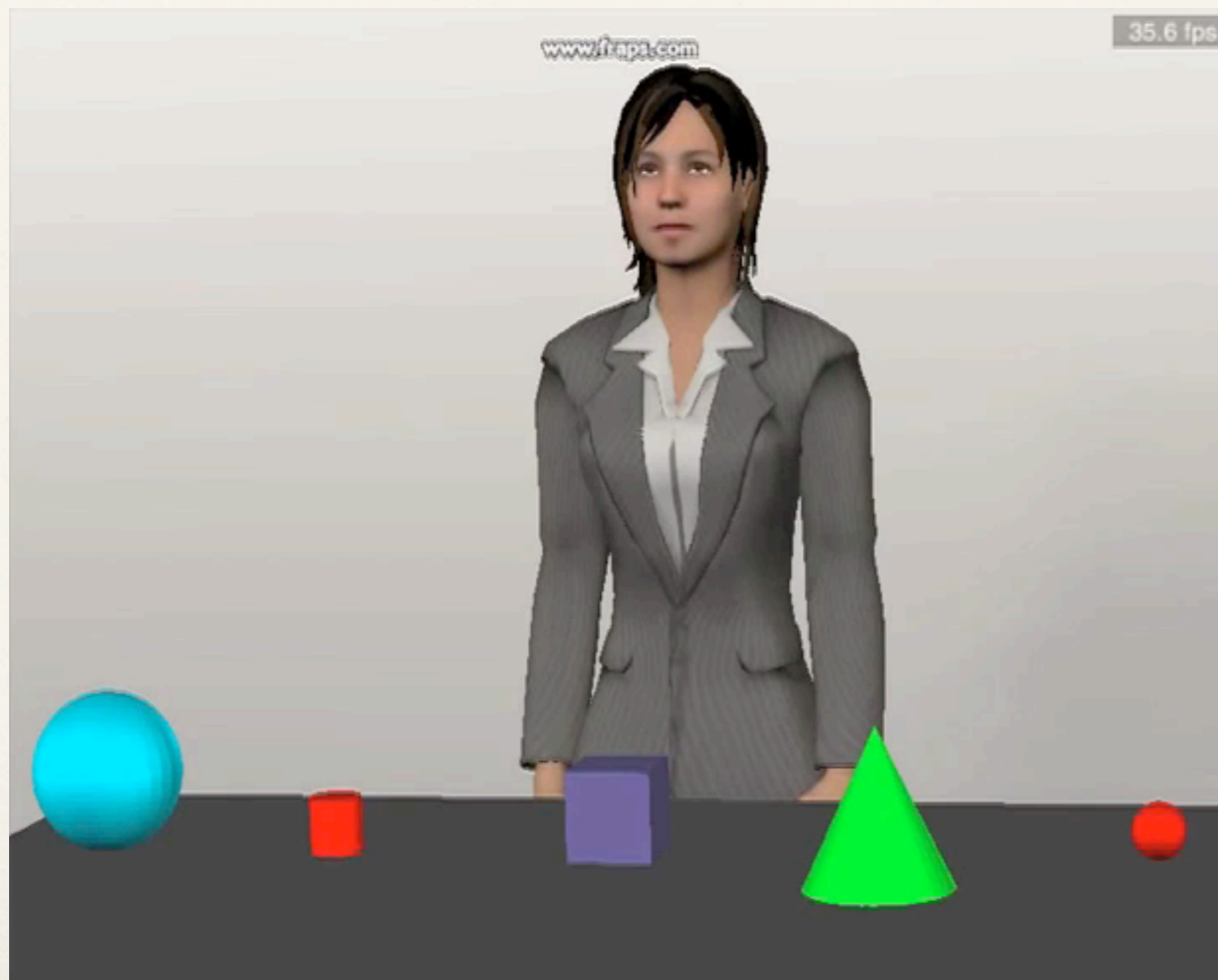
Outlook

Outlook

✤ Replication with virtual agents

 ✤ Will two single eye/head movements that the robot made in the particular condition appear strange for a human-like character?

✤ Gaze reveals something about intentions. Can it also trigger affordances or even override mentioned actions?

 ✤ What relevance are looked-at-objects assigned when involved but not mentioned in a situation.

 ✤ Interaction of object affordances with mentioned action verbs?

Object Affordances

Organizational Things

* Student Presentation

    * 1 hour talk

    * 20-30 min. for discussion (maybe take up issues/questions explicitly)

* For questions, feedback for slides etc. come and see us!

    * Make individual appointments in the week before the talk

    * Florian -> Maria

    * Katerina & Tristan -> Matt