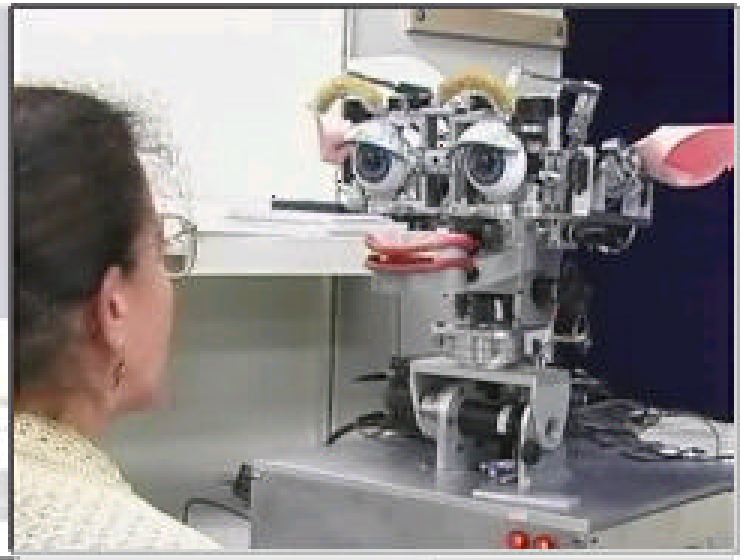
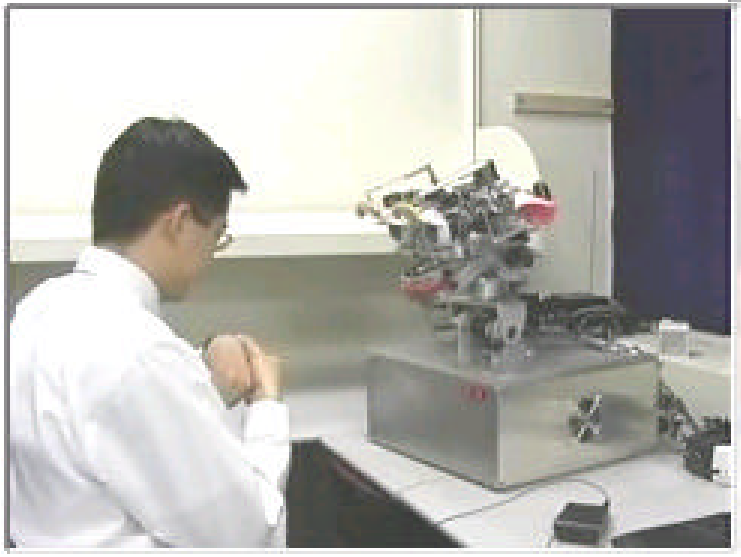
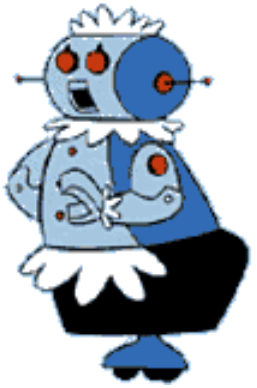
Social Constraints *on* Animate Vision





Cynthia Breazeal
Aaron Edsinger
Paul Fitzpatrick
Brian Scassellati
MIT AI Lab

Social constraints

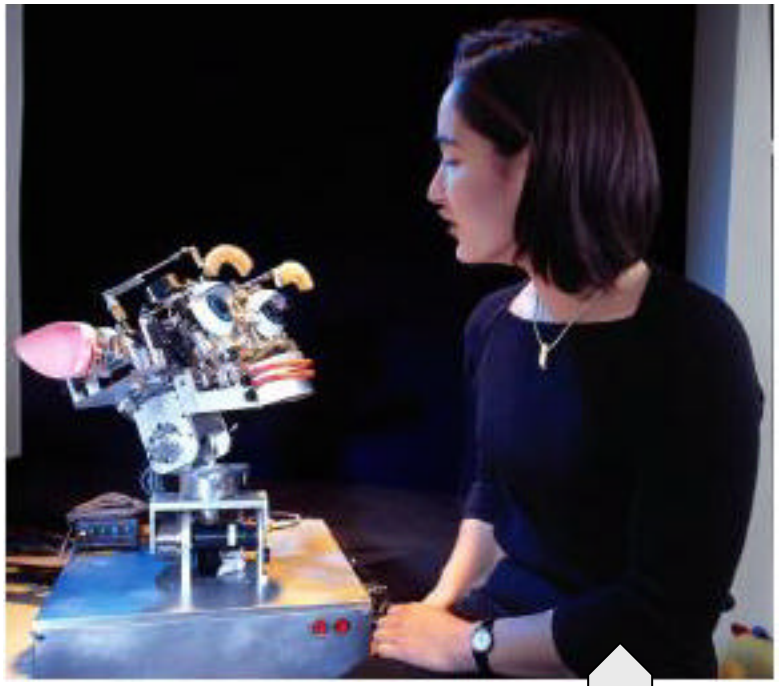


- Robots create expectations through their physical form – particularly humanoid robots
- But with careful use, these expectations can facilitate smooth, intuitive interaction
  - Provide a natural "vocabulary" to make the robot's behavior and state readable by a human
  - Provide natural frameworks for trying to negotiate a change in each other's behavior and (through readability) knowing when you have succeeded
  - These elements have their own internal logic and constraints which, if violated, lead to confusion

Visually-mediated social elements

- ✓ Readable locus of attention
- Negotiation of the locus of attention
- Readable degree of engagement
- Negotiation of interpersonal distance
- Negotiation of object showing
- Negotiation of turn-taking timing

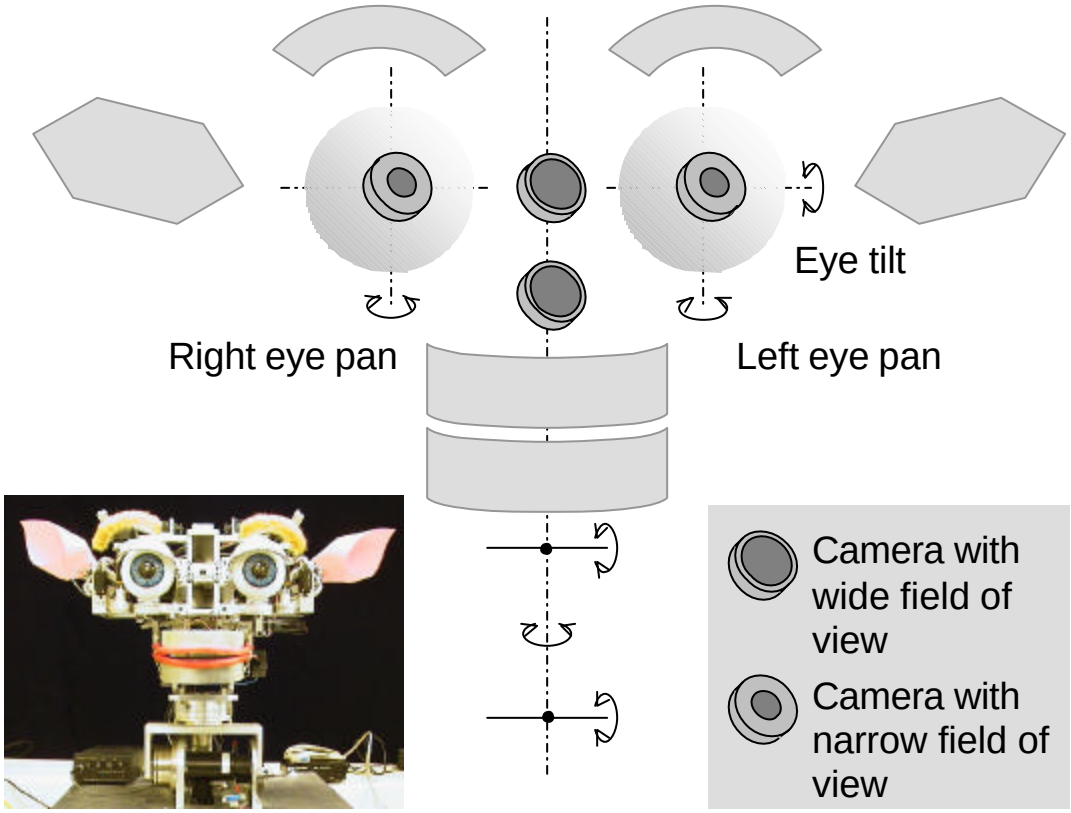
Readable locus of attention



Attention can be deduced from behavior

Or can be expressed more directly

Kismet – a readable robot



- Designed to evoke infant-level social interactions
  - Eibl-Eiblsfeldt “baby scheme”
  - physical size, stature
- But not exactly human infant
  - caricature that is readable
- Naturally elicit scaffolding acts characteristic of parent-infant scenarios
  - directing attention
  - affective feedback, reinforcement
  - simplified behavior, suggested to make perceptual task easier
  - slow down, go at infant’s pace

Visually-mediated social elements

- ☑ Readable locus of attention
- ☑ Negotiation of the locus of attention
- ☐ Readable degree of engagement
- ☐ Negotiation of interpersonal distance
- ☐ Negotiation of object showing
- ☐ Negotiation of turn-taking timing

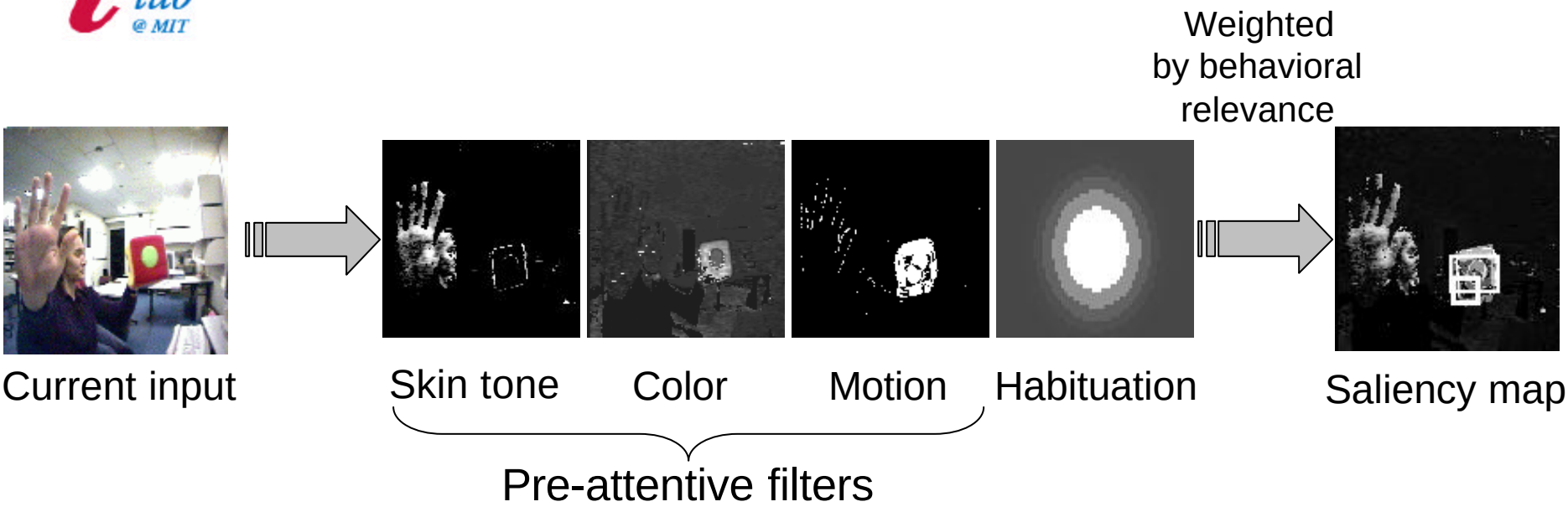
Negotiating the locus of attention

One person's strategies

Another's strategies

- For object-centered activities, attention is fundamental
- There are natural strategies people use to direct attention
- The robot's attention must be receptive to these influences, but also serve the robot's own agenda

External influences on attention

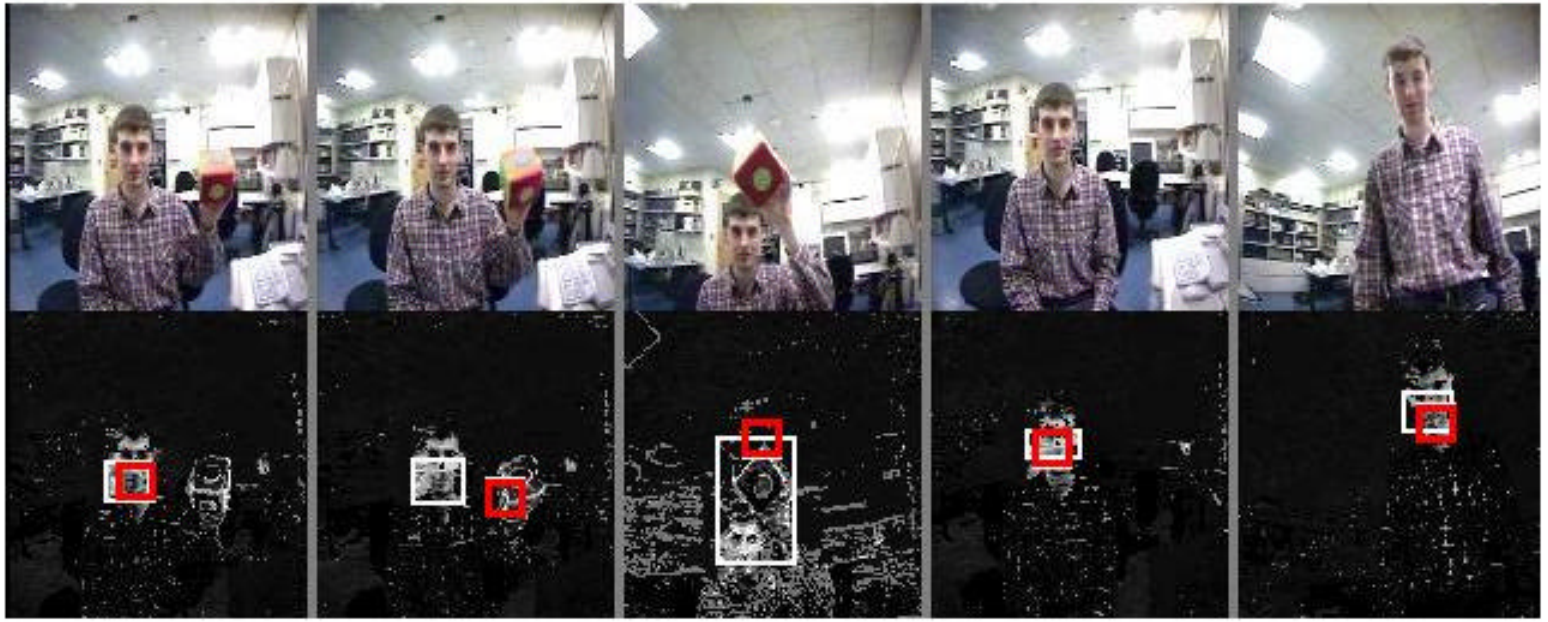


- Attention is allocated according to salience
- Salience can be manipulated by shaking an object, bringing it closer, moving it in front of the robot's current locus of attention, object choice, hiding distractors, …

Tuned to natural cues

stimulus category	stimulus	presentations	average time (s)	commonly used cues	commonly read cues
color and movement	yellow dinosaur	8	8.5	*motion across centerline,* *shaking,* *bringing object close*	*change in visual behavior,* *face reaction,* *body posture*
	multi-colored block	8	6.5		
	green cylinder	8	6.0		
movement only	black&white cow	8	5.0		
skin toned and movement	pink cup	8	6.5		
	hand	8	5.0		
	face	8	3.0		
Overall		56	5.8		

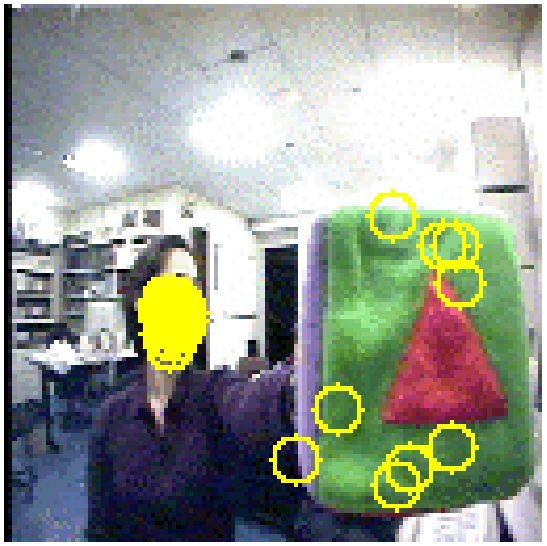
Can shape an interaction



- The robot's attention can be manipulated repeatedly
- So caregiver can shape an interaction into the *form* of an object-centered game, or a teaching session

Internal influences on attention

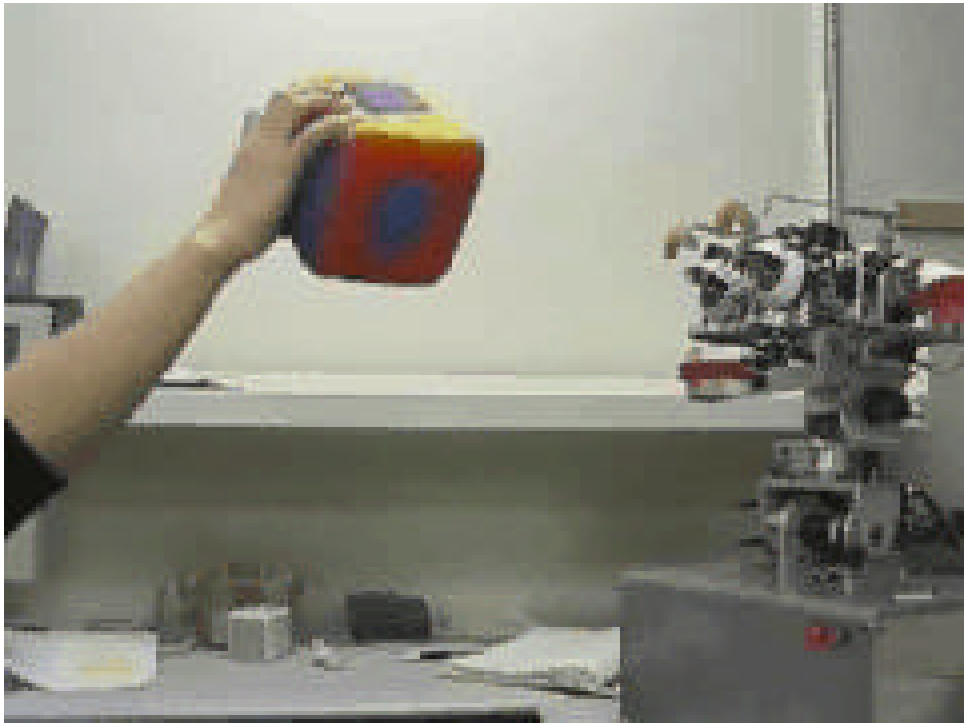
“Seek toy” –
low skin gain, high saturated-color gain
Looking time 28% face, 72% block



“Seek face” –
high skin gain, low color saliency gain
Looking time 28% face, 72% block

- Internal influences bias how salience is measured
- The robot is not a slave to its environment

Maintaining visual attention



- Want attention to be persistent enough to permit coherent behavior
- Must be able to maintain fixation on an object, when behaviorally appropriate
- Attention system interacts closely with tracker to support this robustly

Visually-mediated social elements

- ☑ Readable locus of attention
- ☑ Negotiation of the locus of attention
- ☑ Readable degree of engagement
- ☐ Negotiation of interpersonal distance
- ☐ Negotiation of object showing
- ☐ Negotiation of turn-taking timing

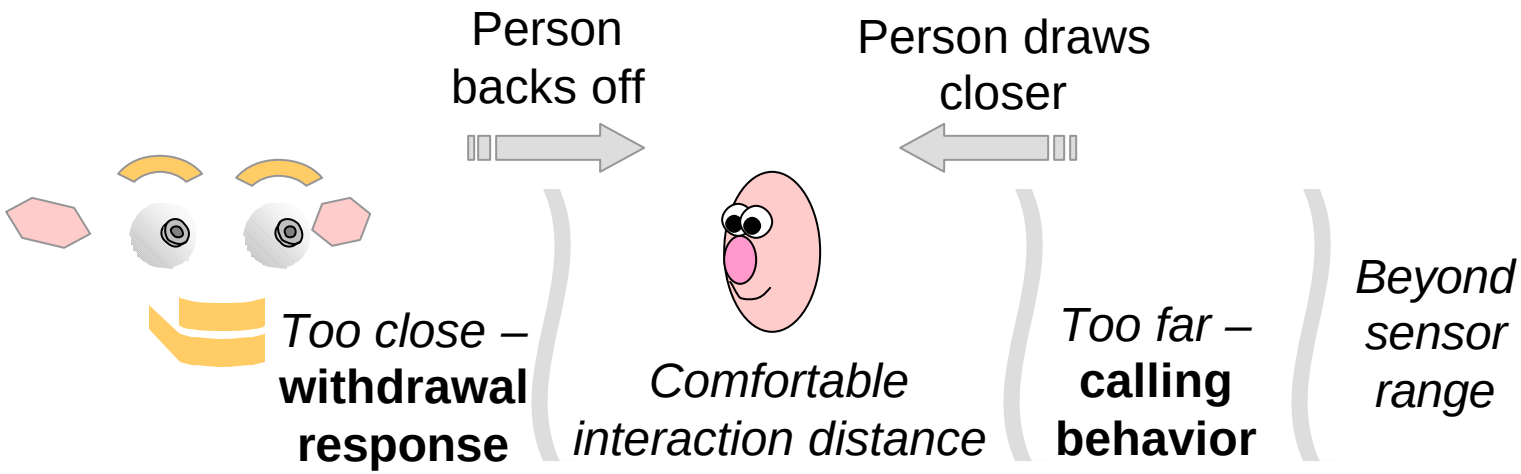
Readable degree of engagement

- Visual behavior conveys degree of commitment
  - fleeting glances
  - smooth pursuit
  - full body orientation
- Gaze direction, facial expression, and body posture convey robot's interest

Visually-mediated social elements

- ☑ Readable locus of attention
- ☑ Negotiation of the locus of attention
- ☑ Readable degree of engagement
- ☑ Negotiation of interpersonal distance
- ☐ Negotiation of object showing
- ☐ Negotiation of turn-taking timing

Negotiating interpersonal distance



- Robot establishes a “personal space” through expressive cues
- Tunes interaction to suit its vision capabilities

Negotiating interpersonal distance

“Back off buster!”

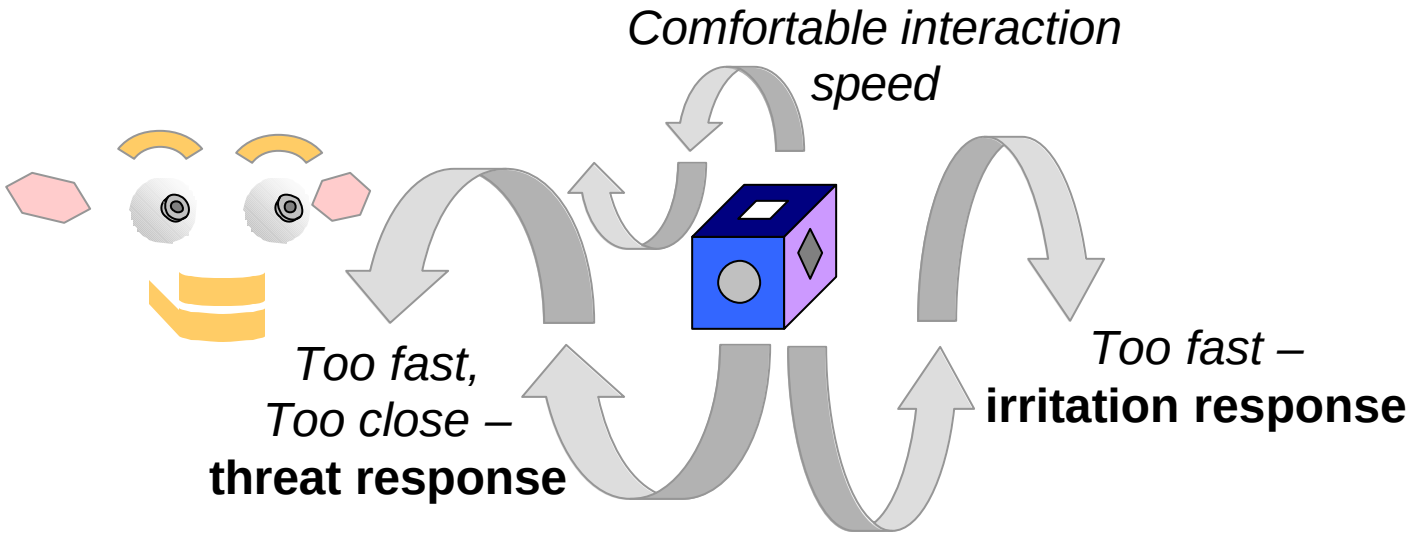
“Come hither, friend”

- Robot backs away if person comes too close
- Cues person to back away too – social amplification
- Robot makes itself salient to call a person closer if too far away

Visually-mediated social elements

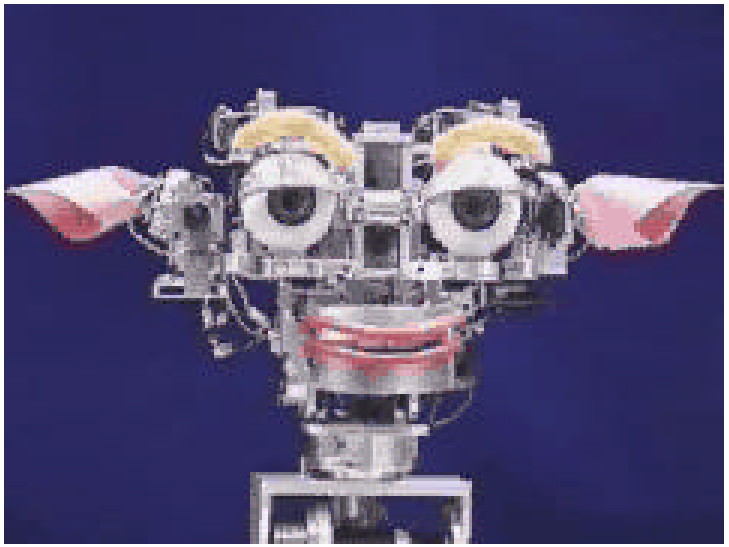
- ☑ Readable locus of attention
- ☑ Negotiation of the locus of attention
- ☑ Readable degree of engagement
- ☑ Negotiation of interpersonal distance
- ☑ Negotiation of object showing
- ☐ Negotiation of turn-taking timing

Negotiating object showing

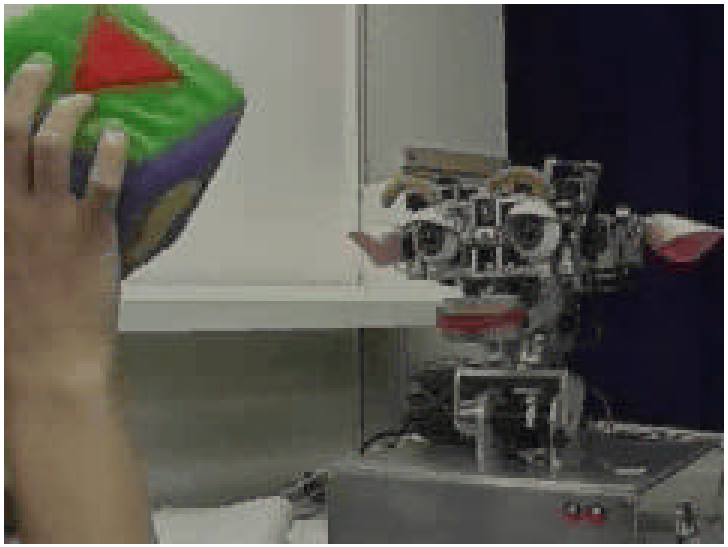


- Robot conveys preferences about how objects are presented to it through irritation, threat responses
- Again, tunes interaction to suit its limited vision
- Also serves protective role

Negotiating object showing



Threat response



Withdrawal, startle

- Robot “shuts out” close, fast moving object – threat response
- Robot backs away if object too close
- Robot cranes forward as expression of interest

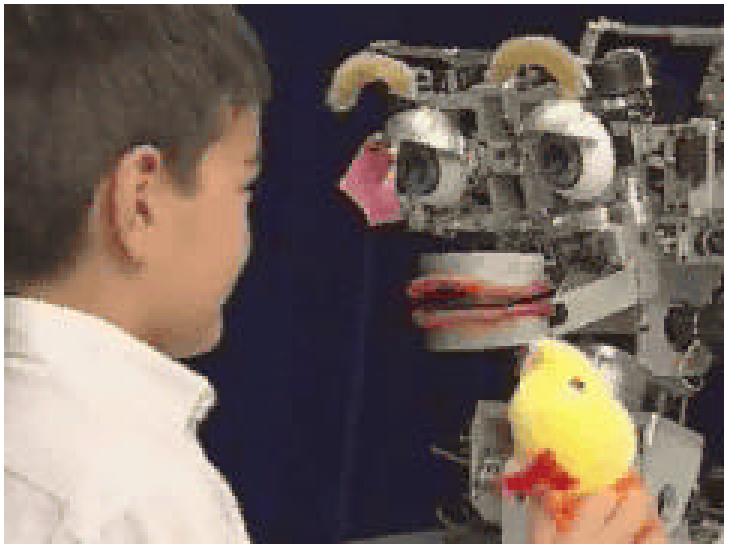
Visually-mediated social elements

- Readable locus of attention
- Negotiation of the locus of attention
- Readable degree of engagement
- Negotiation of interpersonal distance
- Negotiation of object showing
- Negotiation of turn-taking timing

Turn-Taking

- Cornerstone of human-style communication, learning, and instruction
- Four phases of turn cycle
  - relinquish floor
  - listen to speaker
  - reacquire floor
  - speak
- Integrates
  - visual behavior & attention
  - facial expression & animation
  - body posture
  - vocalization & lip synchronization

Examples of turn-taking



Kismet and Adrian

Kismet and Rick

- Turn-taking is fine grained regulation of human's behavior
- Uses envelope displays, facial expressions, shifts of gaze and body posture
- Tightly coupled dynamic of contingent responses to other

Evaluation of Performance

- Naive subjects
  - ranging in age from 25 to 28
  - All young professionals.
  - No prior experience with Kismet
  - video recorded
- Turn-taking performance
  - 82% “clean” turn transitions
  - 11% interruptions
  - 7% delays followed by prompting
- Significant flow disturbances
  - tend to occur in clusters
  - 6.5% of the time, but rate diminishes
- Evidence for entrainment
  - shorter phrases
  - wait longer for response
  - read turn-taking cues
  - 0.5—1.5 seconds between turns

		time stamp (min:sec)	seconds between disturbances
subject 1	start @ 15:20	15:20 – 15:33	13
		15:37 – 15:54	21
		15:56 – 16:15	19
		16:20 – 17:25	70
	end @ 18:07	17:30 – 18:07	37+
subject 2	start @ 6:43	6:43 – 6:50	7
		6:54 – 7:15	21
		7:18 – 8:02	44
	end @ 8:43	8:06 – 8:43	37+
subject 3	start @ 4:52 min	4:52 – 4:58	10
		5:08 – 5:23	15
		5:30 – 5:54	24
		6:00 – 6:53	53
		6:58 – 7:16	18
		7:18 – 8:16	58
		8:25 – 9:10	45
	end @ 10:40 min	9:20 – 10:40	80+

Conclusion

- Active vision involves choosing a robot's pose to facilitate visual perception.
- Focus has been on immediate physical consequences of pose.
- For anthropomorphic head, active vision strategies can be "read" by a human, assigned an intent which may then be completed beyond the robot's immediate physical capabilities.
- Robot's actions have communicative value, to which human responds.