Human-Computer Interaction

Artificial Agents

Professor Bilge Mutlu

Today's Agenda

» Topic overview: *Artificial Agents*

» Discussion

Topic overview: *Artificial Agents*

Why do we have to think about computers as agents?

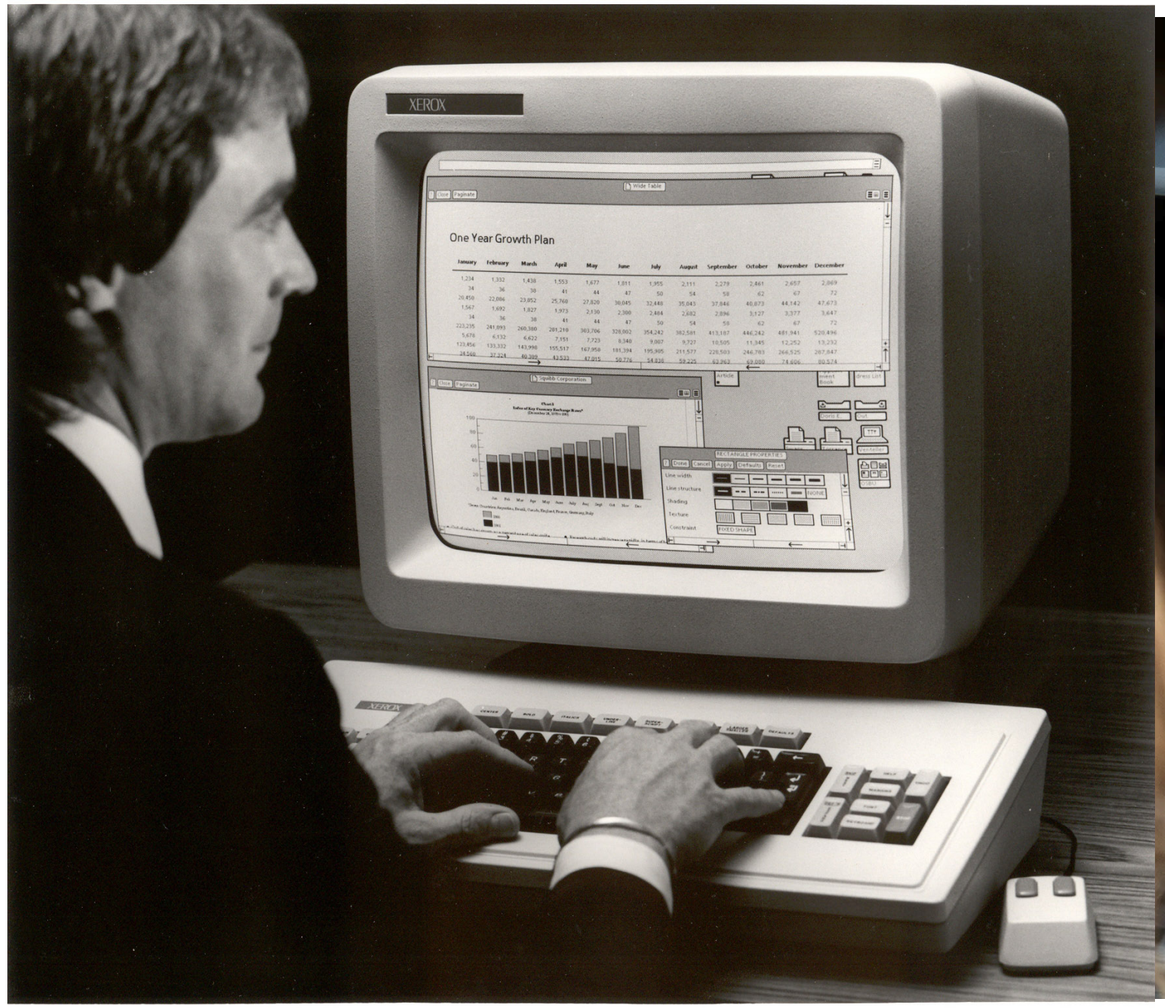
A metaphor: **Imperative vs. Declarative Programming**

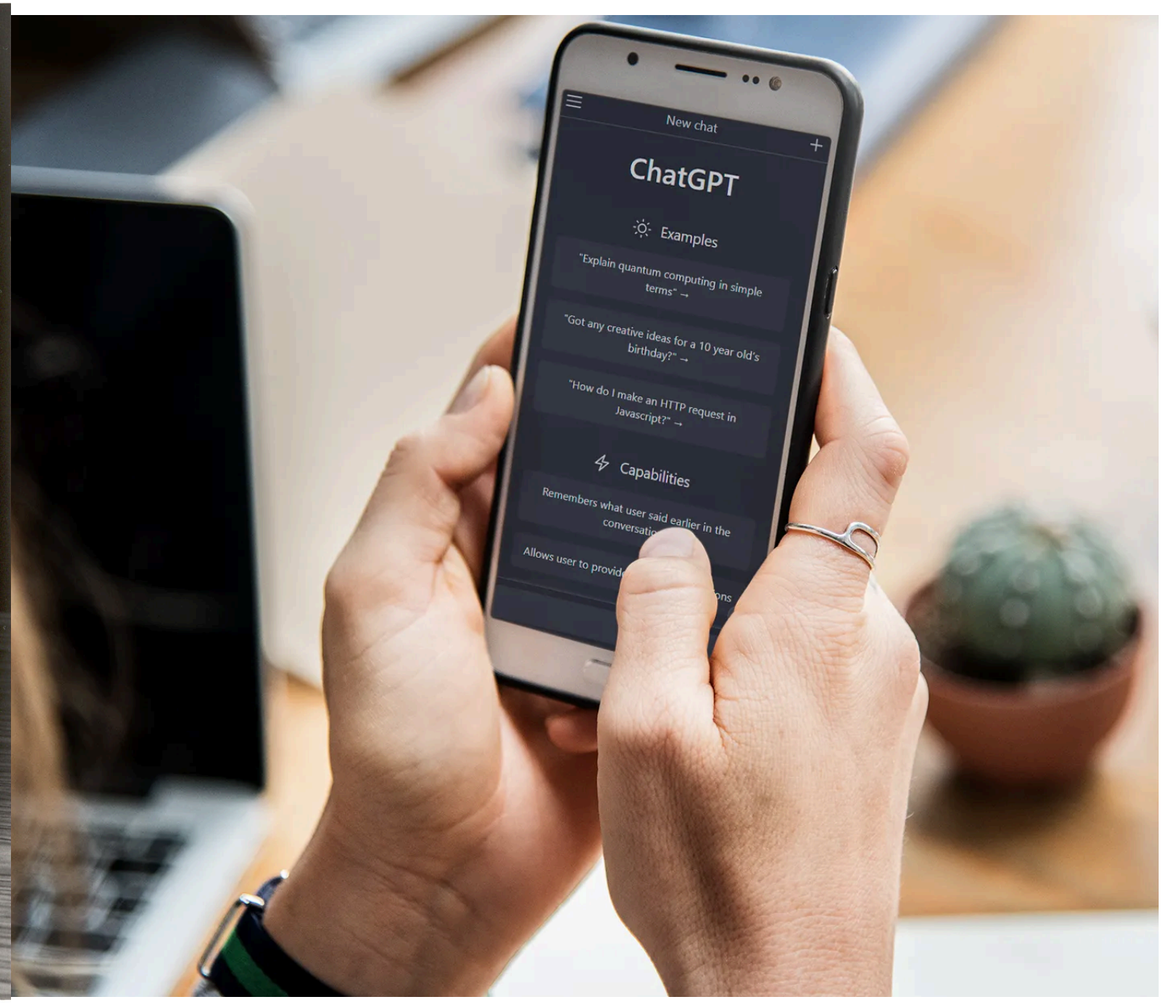
- **Imperative**
  - *Tell the computer exactly how to do things*
  - Step-by-step instructions (control flow, loops, state updates)
  - Parallel: early workstations → user drives applications to perform functions

```python
# Print numbers 1 to 5
for i in range(1, 6):
    print(i)
```

- **Declarative**
  - *Express intent, let the computer figure out how*
  - Describe desired outcomes (constraints, goals, queries)
  - Parallel: agents interpret goals, plan and act autonomously on our behalf

```sql
-- Get numbers 1 to 5
SELECT number
FROM numbers
WHERE number BETWEEN 1 AND 5;
```


XEROX
One Year Growth Plan


New chat
ChatGPT
Examples
"Explain quantum computing in simple terms" →
"Got any creative ideas for a 10 year old's birthday?" →
"How do I make an HTTP request in Javascript?" →
Capabilities
Remembers what user said earlier in the conversation
Allows user to provid

How can we define agents?

Software agents

Definition: A software agent is a computer program that acts for a user or other program in a relationship of *agency*.

Agency

Definition: An agreement to act on one's behalf.

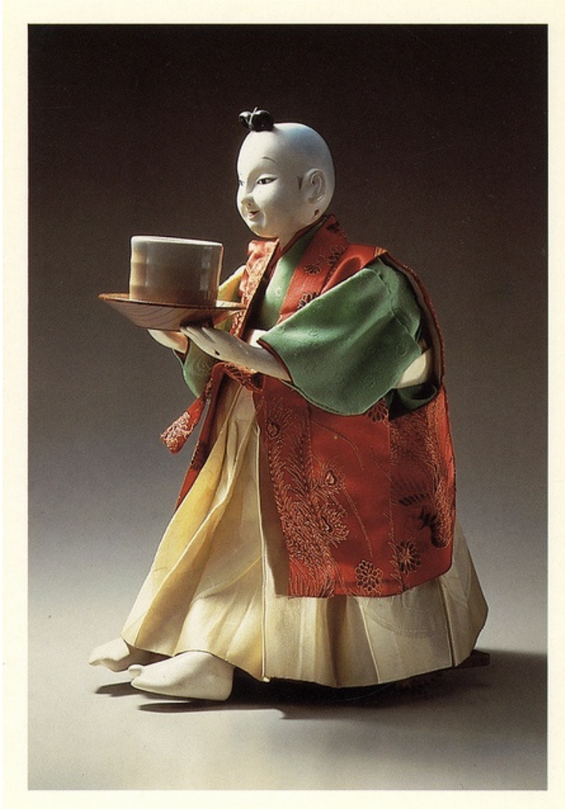
Agency implies intelligence, autonomy, decision-making

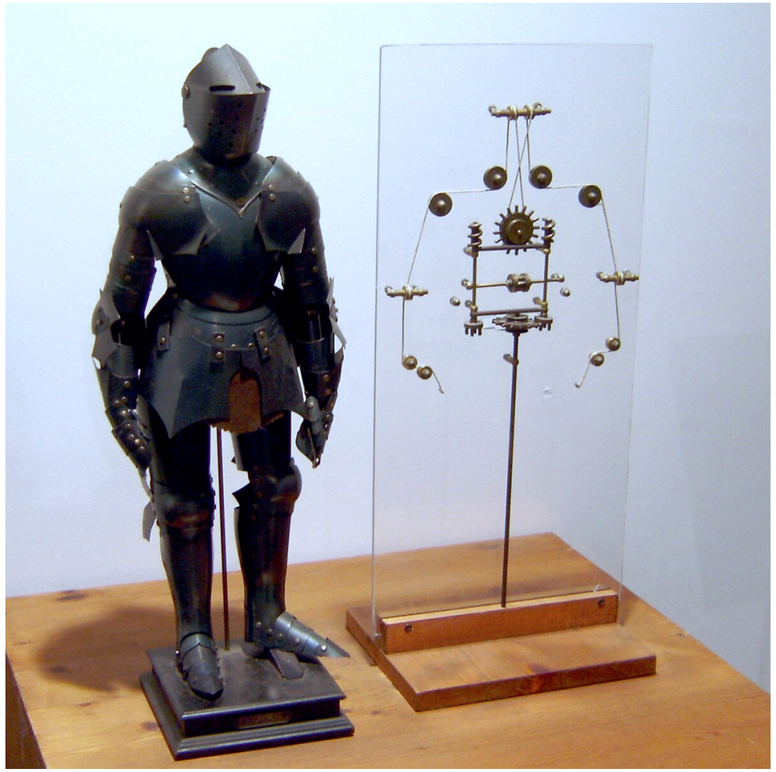
History of agents

» **Pre-modern inspirations:** Automata and puppetry (e.g., Bunraku) seed the idea of lifelike artifacts.

» **1495:** Leonardo's mechanical knight (self-moving armor via pulleys).

» **Late 1700s:** The Turk (chess "automaton," actually human-operated) sparks public imagination.



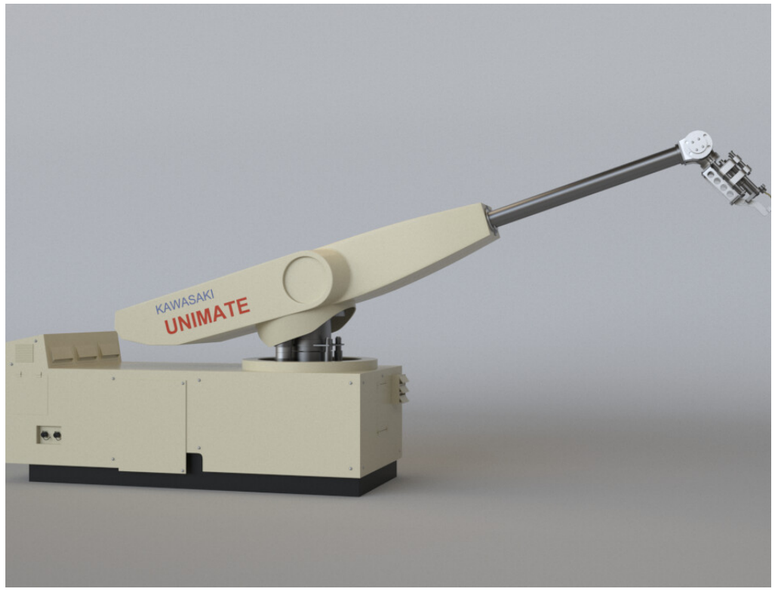


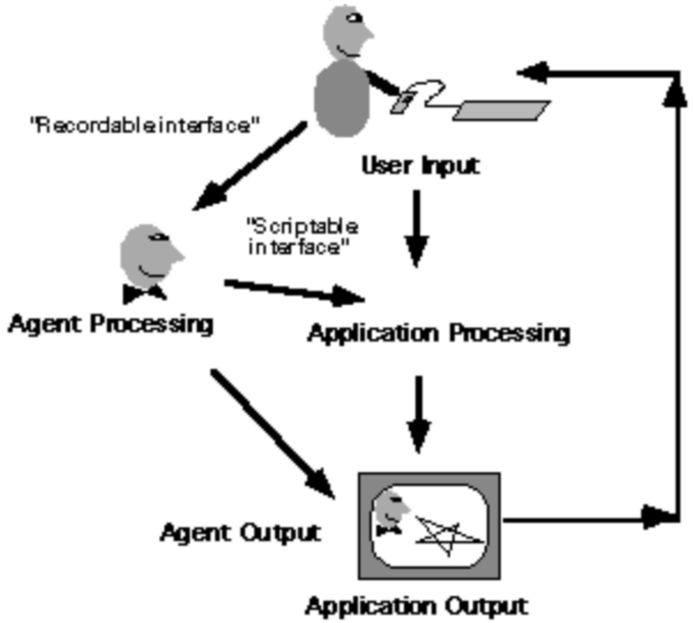
» **1950:** Turing formalizes computation; foundations for autonomous machines.

» **1950s–1960s:** First digitally operated programmable robot arm (industrial robotics).

» **1990s:** Autonomous software agents.

» **1990s–2000s:** Embodied conversational agents — "locating" intelligence in bodies to solve invisibility; cues for social interaction.



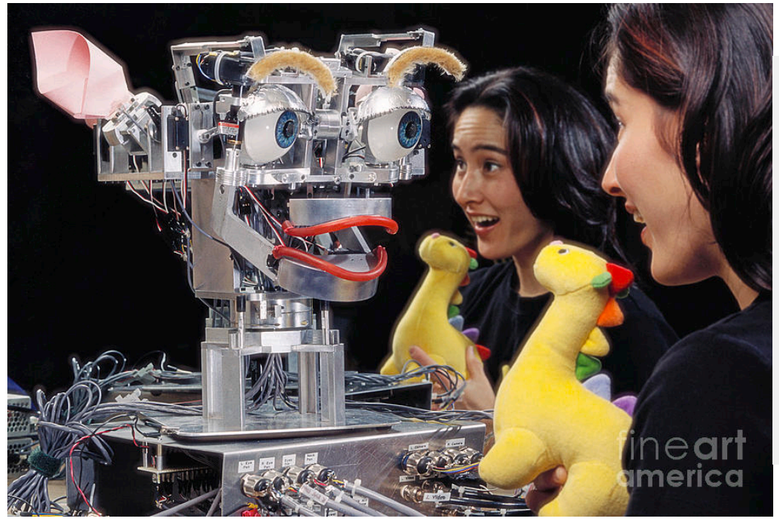


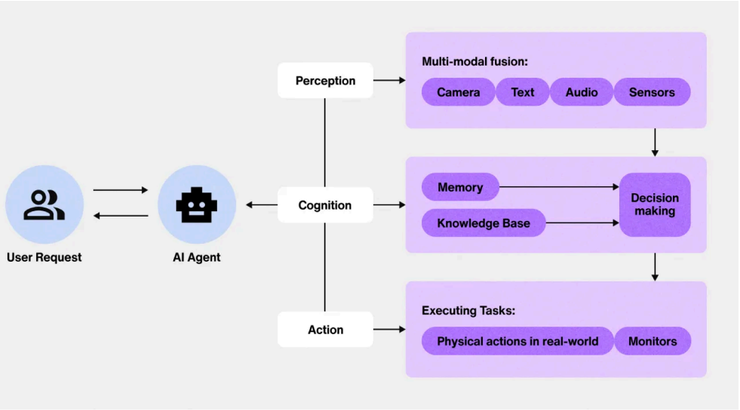
» **2003:** "Toward Sociable Robots”"catalyzes socially interactive robot design.

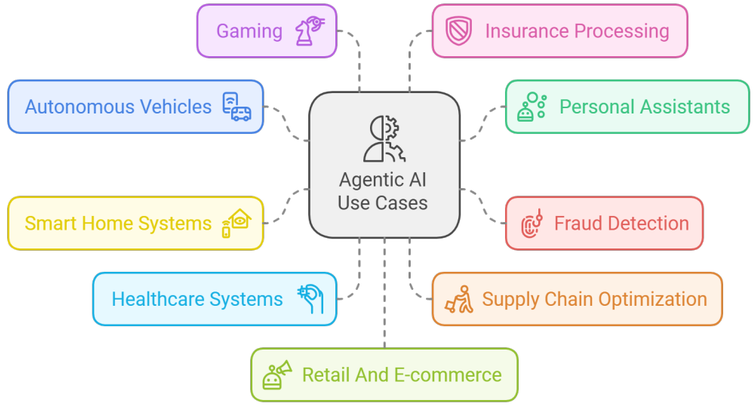
» **2010s:** Evidence that *physical* presence changes outcomes (commitment, compliance).

» **2020s: Agentic AI** emerges — LLM-based agents with planning, tool use, memory; mostly virtual, sometimes embodied.

» **Today:** Convergence: software agents + embodiment + agentic behaviors → tutors, copilots, assistive robots.







What makes an agent "feel" agentic to people?

» We rapidly ascribe minds to nonhumans; perception of mind is a *psychological* phenomenon you can design for (cues, behaviors), not just a technical property.

» **Implication:** design levers (intentional movement, contingent response, goal-directedness) can increase perceived agency even with simple capabilities.

Note: Sets up why "agentic AI" experiences can work with minimal mechanics.

Where does agency "live"?

» Users look for a *locus* for intelligence; embodiment is one answer, but not the only one (UI focus, avatar, chat bubble can also serve).

» Cassell's provocation: locating intelligence in the *body* solves "invisible computer" problems.

Why do agents need bodies?

We need to locate intelligence, and this need poses problems for the invisible computer. The best example of located intelligence, of course, is the body.

— Cassell, 2001[1]

[1] Cassell, 2001, Embodied conversational agents: representation and intelligence in user interfaces

What does a body give us?

» A *locus of attention*—a target toward which we can our attention and behavior

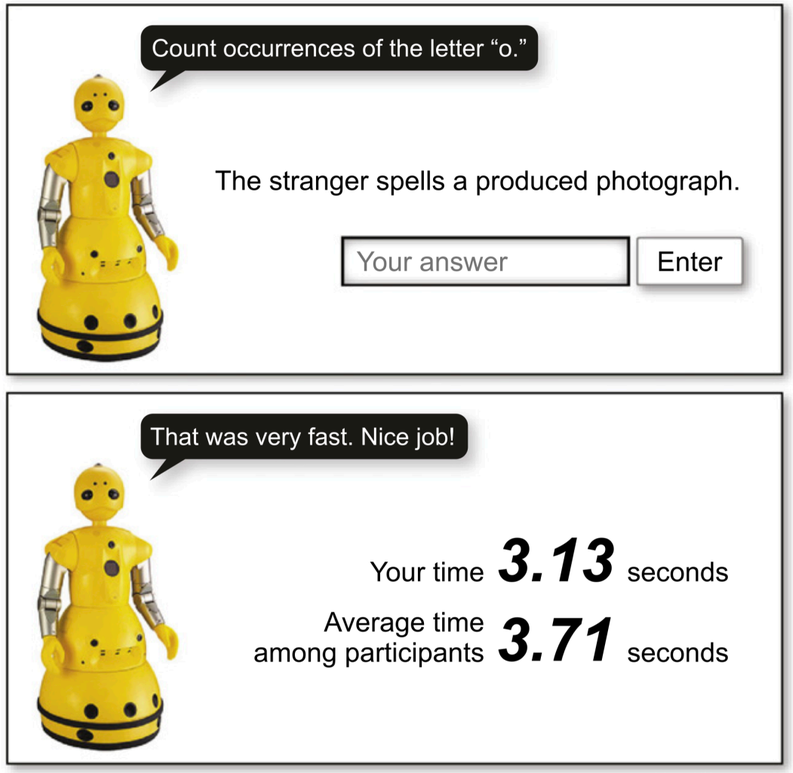
» Cues about the agent's status (e.g., functioning, not broken, speaking, waiting)

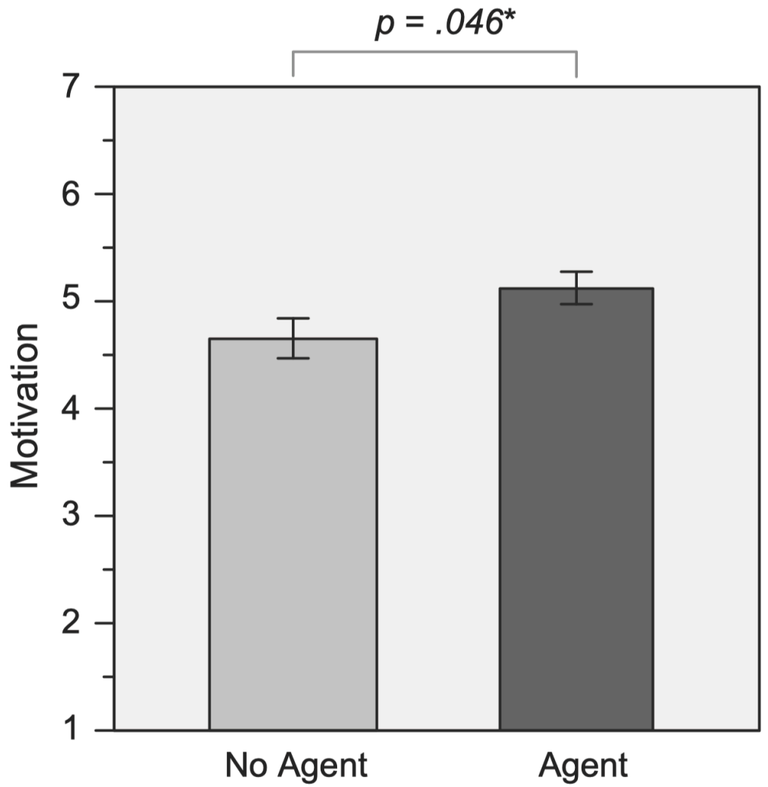
» Opportunity to create plausible, coherent characters that signal the agent's role (e.g., a butler, a personal assistant, a collaborator)

» Ability to utilize social mechanisms in interaction design

Why do we need a locus of attention?

Increased presence of, arousal toward, and commitment to another entity with agency.[3]

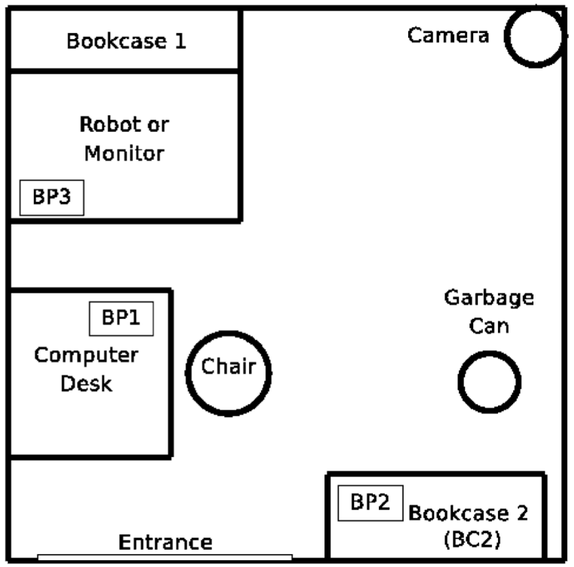


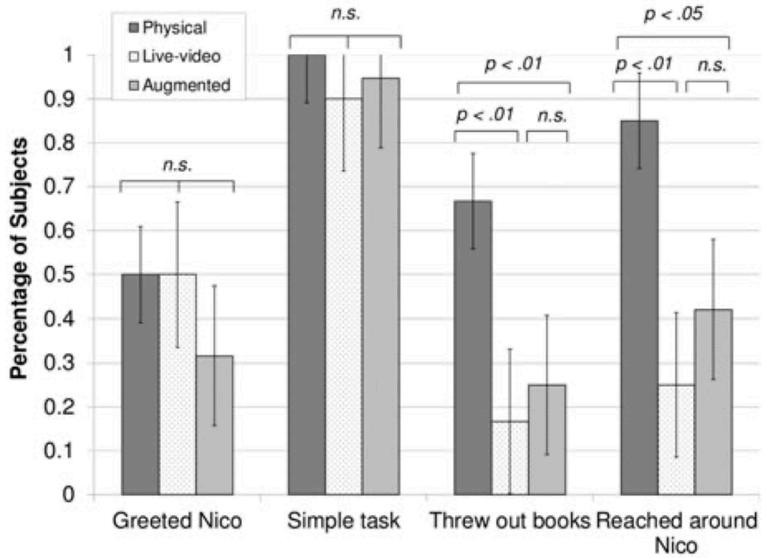


[3] Mumm & Mutlu, 2011, Designing motivational agents

Where should the body be?

Physical bodies further improve social outcomes.[4]





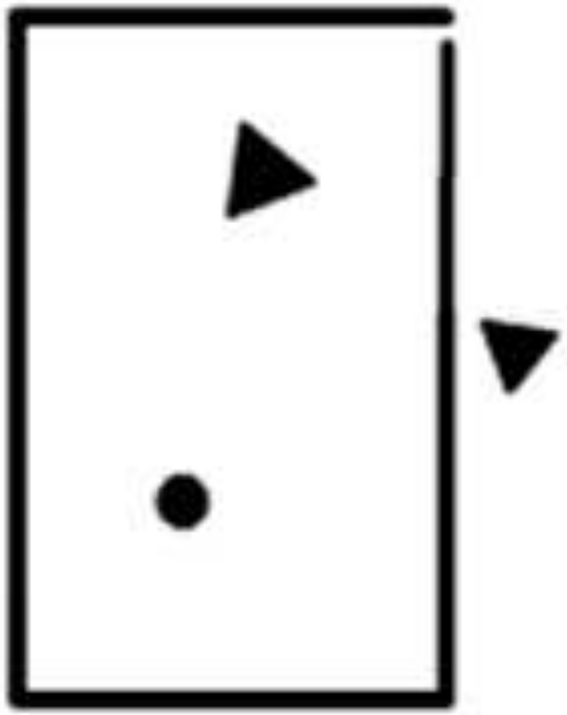
[4] Bainbridge et al, 2011, The benefits of interactions with physically present robots over video-displayed agents

Why do agents need human-like (or animal-like) bodies?

Faced with non-living things of sufficient complexity (i.e., when the observable behavior is not easily understood in terms of its underlying mechanisms), we often apply a social model to explain, understand, and predict their behavior.

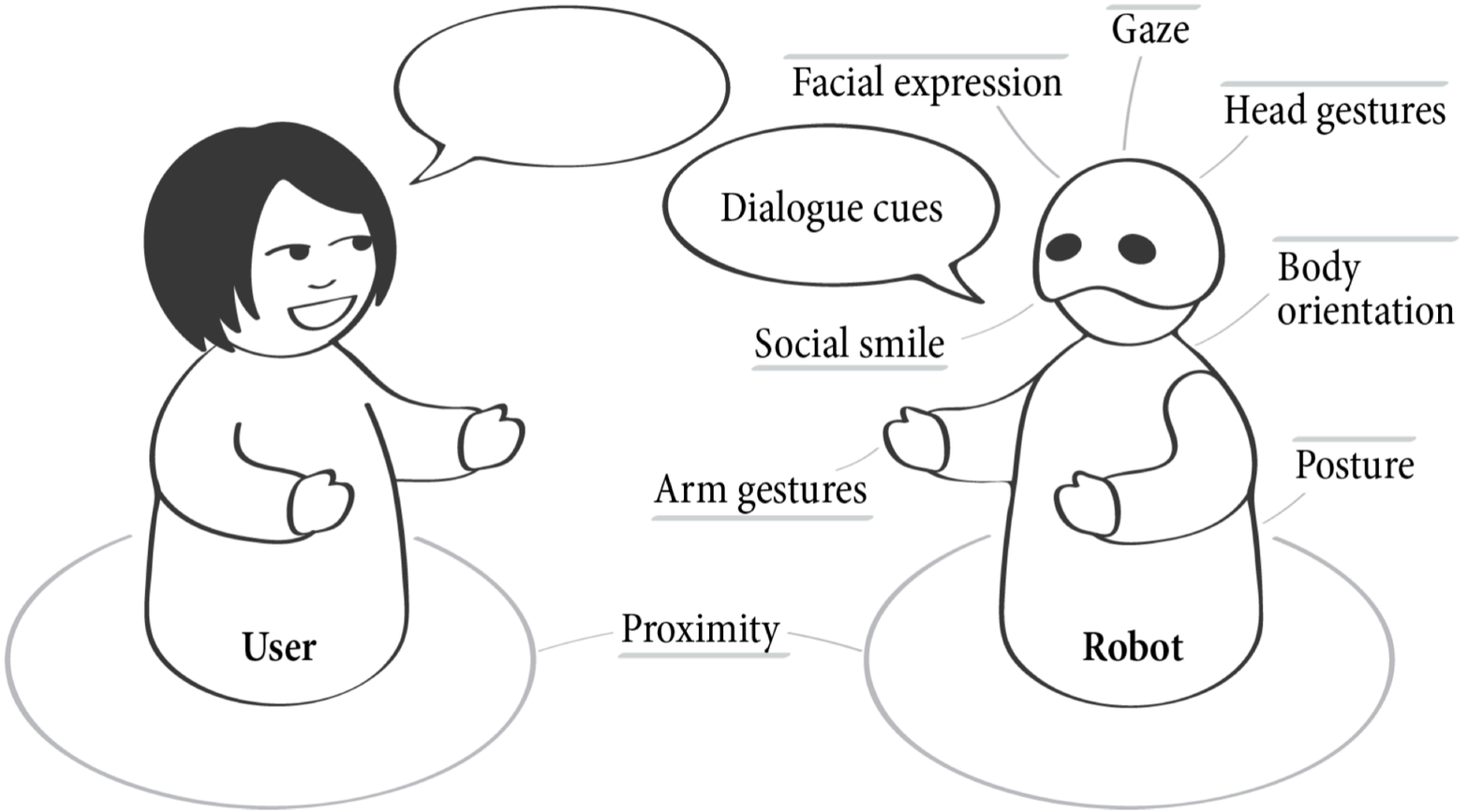
— Breazeal, 2003[5]

[5] Breazeal, C. (2003). Toward sociable robots. Robotics and autonomous systems, 42(3-4), 167-175.



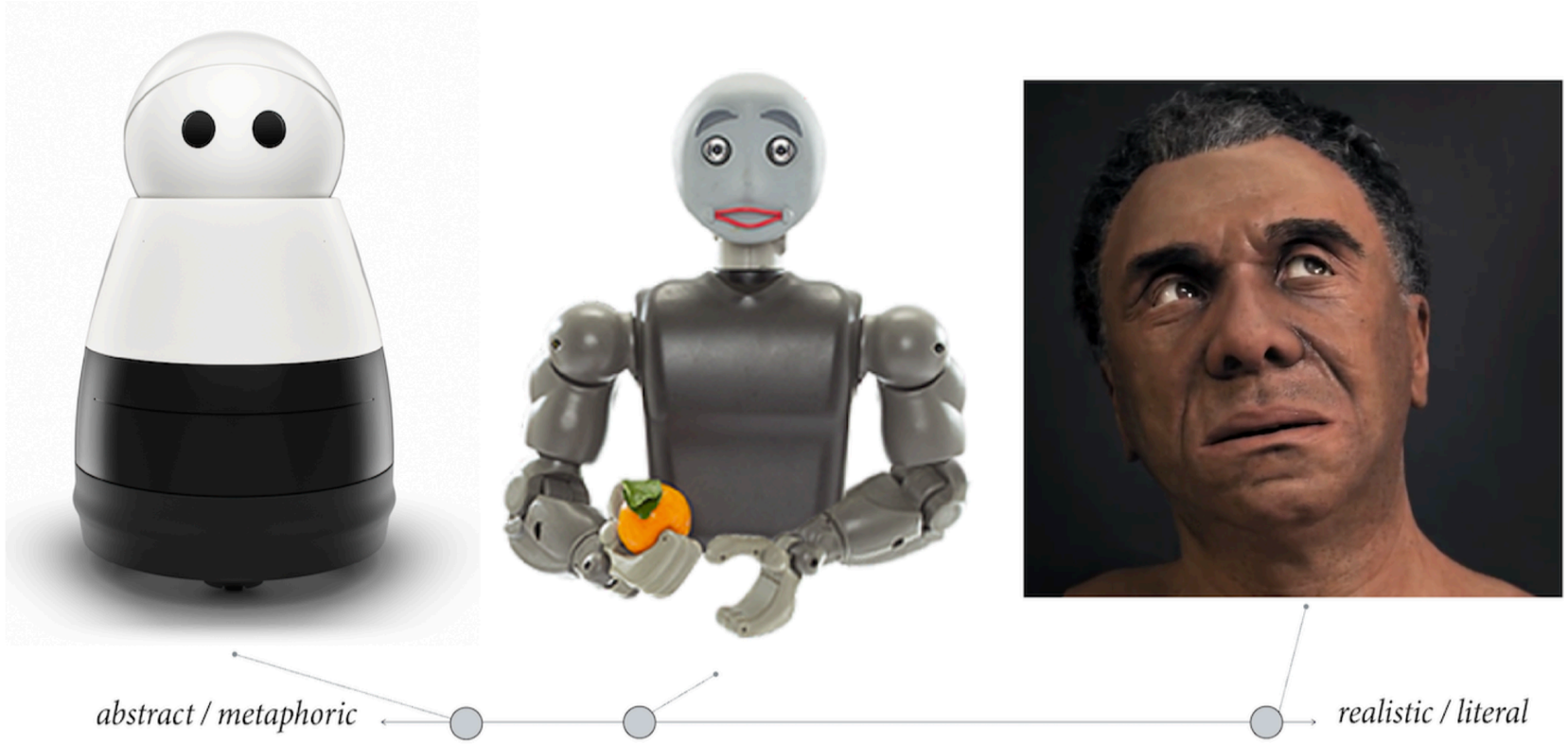
el, 1944, An experimental study of apparent behavior

How do we capitalize on social models?[8]



[8] Mutlu, 2011, Designing embodied cues for dialog with robots

How do we design for social interaction?[9]



[9] Deng et al., 2019, Embodiment in socially interactive robots

What is the design space of bodies?

Embodiments → Frames of Mind[10]

- Physical frame*
- Virtual frame*
- Blended frames
- Mediated frames
- Immersive frames

[10] Mutlu, B. (2021). The virtual and the physical: two frames of mind. *iscience*, *24*(2).

Physical frame

Co-situated, norms of proxemics, seen as independent agent; engagement is voluntary.

Physical Frame

↓

"What am I encountering?"

Virtual frame:

User enters a crafted plot; conventions are learned; interactions can be richly scripted.

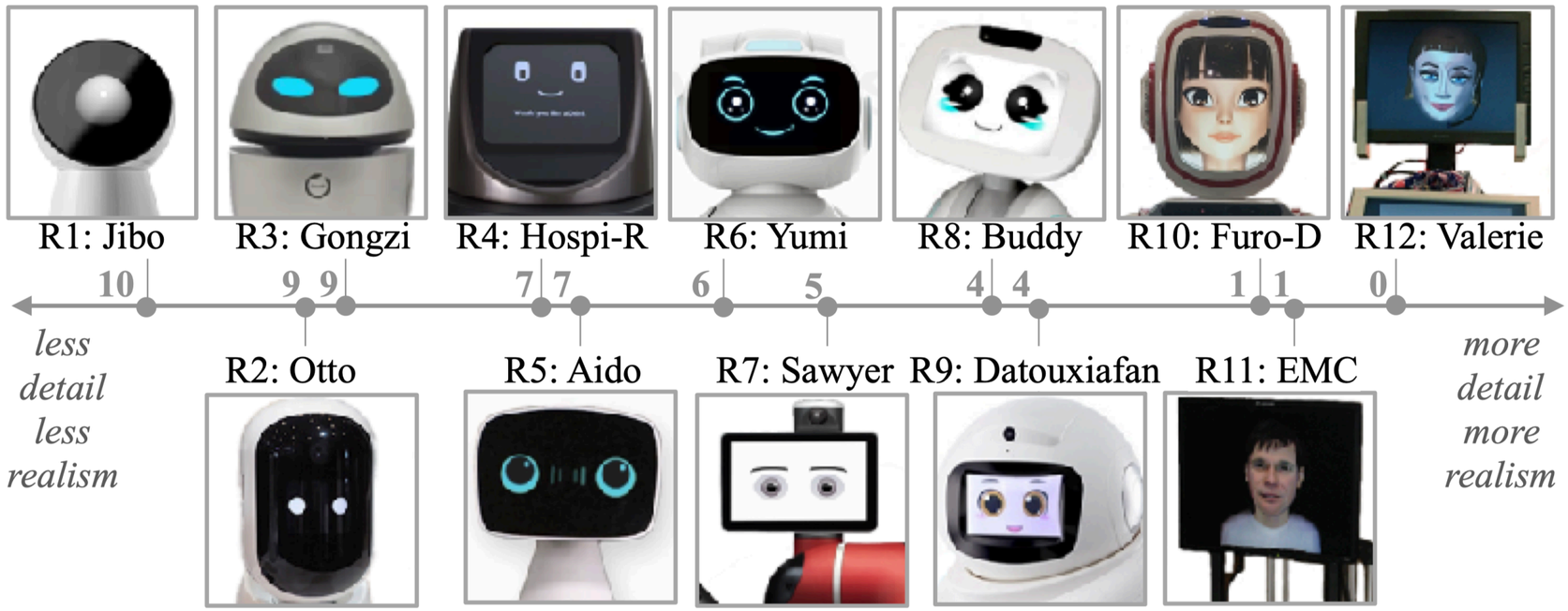
Virtual Frame

↓

"I am in a theatre play."

Blended Frames[11]

Blended: rendered robot faces, screen-on-robot, etc. (design space exploration).



[11] Kalegina et al. (2018). Characterizing the design space of rendered robot faces. *HRI 2018*.

Mediated/Immersive: telepresence bodies; metaverse-style settings.

Mediated Frame[12]

Immersive Frame

[12] Images: left, right

What are design trade offs with physical vs. virtual frames?

Mechanism	Physical	Virtual
Situativity	Co-situated in the user's environment	User is brought into the agent's environment
Interactivity	Emerges from join action/intention	Invites users to participate in a crafted, patterned plot
Agency	Seen as independent agent pursuing own goals	Engagement is at the user's discretion
Proxemics	Dynamic, co-managed to follow human norms	Constrained, involving learned conventions
Believability	Real-world, self-relevant agent	Safe environment to experience emotion

Characteristics	Physical	Virtual
Applications	Physical, situated collaboration, assistance	Counseling, instruction, education, coaching
Activities	Activities interspersed across time and space	Focused, time-bound activities
Interactions	Interactions situated in day-to-day life	Metaphorical, rich, crafter interactions

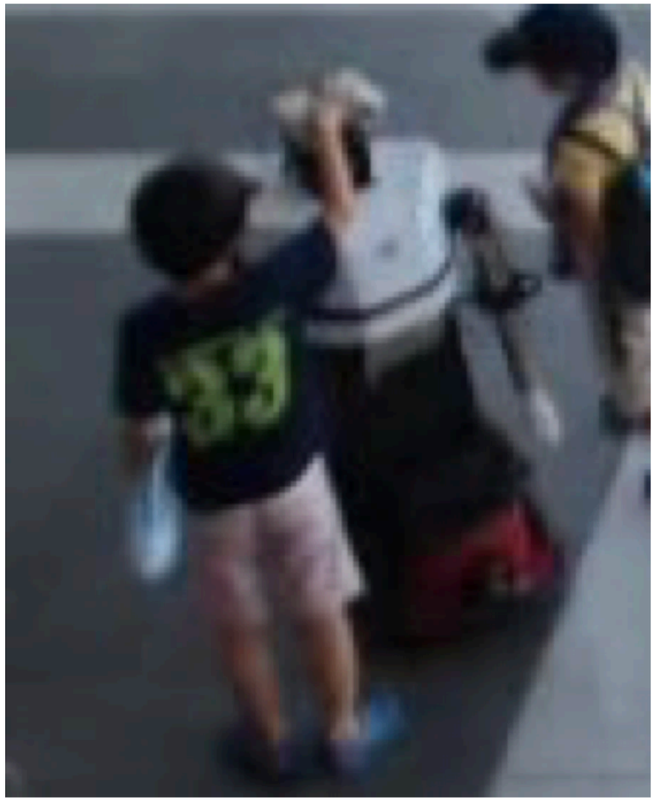
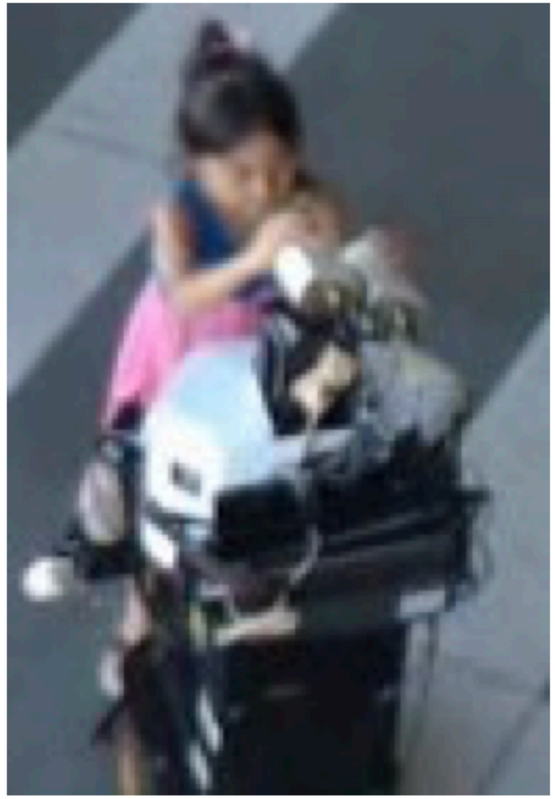
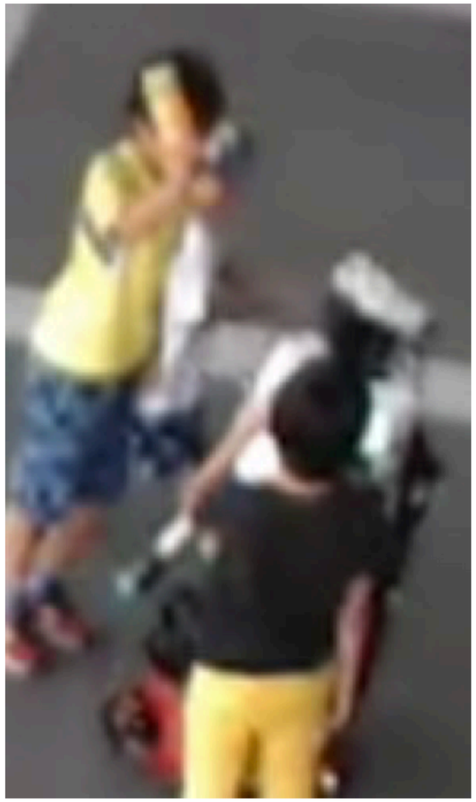
What humans reveal in interactions with robots?

» Long lineage of “artificial others”: Leonardo’s mechanical knight (1495) → The Turk automaton (18th c.) → modern computing & autonomous robots (Turing, 1950).

» Minimal cues can trigger **mind perception** (contingency, goal-directed motion, turn-taking).

» Risk: over-ascription → unrealistic expectations & disappointment; set users’ mental models early.

Ethics of Interacting with Agents[16]

[16] Nomura et al. (2015). Why do children abuse robots? HRI 2015.

17

[17] Wikipedia: Her (2013 film)

Discussion Format

- We'll let AI randomly pick 3-5 names
- In the selected order, students:
  - Present their provocation/critical artifact/policy or design recommendation (30 secs)
  - Lead class discussion (5-8 min)

What's Next?

- **Wednesday:**
  - **Methods** — Read *Gaver*[13] & *Hutchinson et al.*[14] + Textbook Chapter 6[15]
  - **Project** — Project next steps:
    - CITI training — Due Oct 3
    - Method — Due Oct 22

[13] Gaver et al. (1999). Design: cultural probes. interactions, 6(1), 21–29.

[14] Hutchinson et al. (2003). Technology probes: inspiring design for and with families. CHI 2003.

[15] Lazar et al. (2017). Chapter 6 — Diaries. Research methods in human-computer interaction. Morgan Kaufmann.