Human-Computer Interaction

Agents, Delegation & Embodiment

Professor Bilge Mutlu

Today's Agenda

- Topic overview: *Agents, Delegation & Embodiment*
- The frame: *enduring fundamental → how AI changes it*
- Provocation: *Automate the sequence, or keep initiative returnable?*
- Discussion

This term's frame

Enduring fundamental → How AI is changing it → Depth (a claim to argue with)

Fundamentals are not a relic to move past; they are the **instrument** you use to read the AI moment.

- **Today's fundamental:** *delegation*, a **transfer of authority** to act on someone's behalf, over a *sequence* of actions
- **Depth: Deep.** Software agents and embodied robots are **one family**: reliance over action sequences

This week's readings

- **Required**
  - Mackay, *Creating Human-Computer Partnerships* (2023)
  - Newendorp et al., *Apple's Knowledge Navigator* (CHI 2024)
- **Recommended**: Gajos & Mamykina (IUI 2022) · Suchman (2007)
- **Optional**: Park et al., *Generative Agents* (UIST 2023), from Week 5

Topic overview: *Agents, Delegation & Embodiment*

Delegation as transfer of authority

The enduring question isn't "is it an agent?" It's *what did you hand over, and can you get it back?*

- Delegation = authority to act **on your behalf**, over a **sequence** of actions
- Software or embodied: **one family**, one problem, *reliance over sequences*
- The hard sub-questions: **agency · autonomy · accountability · *when* to delegate**

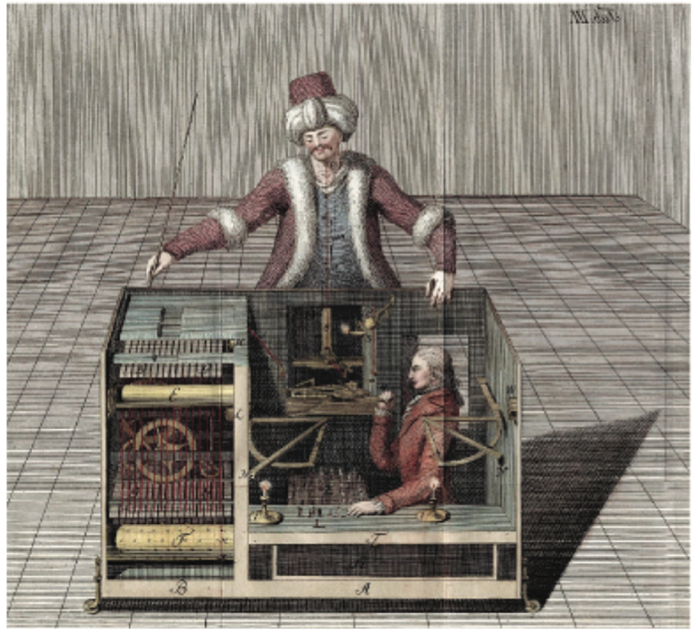
What is agency?[1]

- **Aristotle:** the agent is the *efficient cause*: the sculptor, not the stone
- **Affordances invite action;** agents choose which to take up, by goals and values
- **Agency is perceived:** a robot that comes to you and interrupts has **its own goals**
- **On someone's behalf:** a principal, and with it **accountability**

[1] Mutlu, B. (2021). The virtual and the physical: two frames of mind. *iScience*, 24(2), 101965. doi:10.1016/j.isci.2020.101965

A long lineage of "artificial others"

- **1495 / 18th c.:** Leonardo's mechanical knight; the Turk, a human hidden inside
- **1950s–90s:** programmable robot arms → autonomous software agents
- **1990s–2000s:** embodied conversational agents, intelligence *located* in a body
- **2020s:** agentic AI, LLM agents that plan, use tools, remember





Anchor: agency is *relational*[2]

Lucy Suchman, *Human-Machine Reconfigurations* (2007)

- Agency is a **relational/distributed achievement**, not a property inside the machine
- "Autonomy" is sustained by **hidden human labor**, the antidote to "the agent acts on its own"
- Every "the agent did it alone" claim → *ask what human work was made invisible*

[2] Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd ed. Cambridge Univ. Press.

Share agency, keep control[3]

Wendy Mackay, *Creating Human-Computer Partnerships*

- Humans **share agency** with systems, yet keep control
- **Reciprocal co-adaptation:** each learns from the other
- Users move between **levels of agency** (ImageSense)
- Our test: after the agent acts, can the human take over cleanly?

[3] Mackay, W.E. (2023). Creating Human-Computer Partnerships. In *CHIRA 2023*, CCIS vol. 1996, pp. 3–17. Springer. doi:10.1007/978-3-031-49425-3_1

Think *with* AI, or defer to it?[4]

Gajos & Mamykina, *Do People Engage Cognitively with AI?* (IUI 2022)

- AI recommendation + explanation: **better decisions, no learning** beyond no assistance
- Explanation only, the person decides: **better decisions and learning**
- Delegate the decision, and you may delegate the learning too

[4] Gajos, K.Z. & Mamykina, L. (2022). Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. *IUI '22*, 794–806. doi:10.1145/3490099.3511138

Agents as a research instrument[5]

Park et al., *Generative Agents* (UIST 2023)

- Software agents as a *research instrument* for delegation & multi-agent behavior
- Read in Week 5; re-read through the delegation lens
- Caution: *believable* is not *valid*; simulated people are **not** evidence about real people

[5] Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P. & Bernstein, M.S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *UIST '23*.

Neither assistant nor butler[6]

The 1987 video from Week 1: *why doesn't this exist yet?*

- **3 of 4 barrier families are HCI problems:** privacy, social, trust
- **Asymmetric authority:** Phil interrupts Mike and speaks for him; Mike mutes Phil by *tapping his face*
- **No word for it:** not assistant, not butler; Mike needn't care about Phil's welfare

[6] Newendorp, A.K., et al. (2024). Apple's Knowledge Navigator: Why Doesn't that Conversational Agent Exist Yet? *CHI '24*, Art. 776. doi:10.1145/3613904.3642739

Why does embodiment matter?[7]

"We need to locate intelligence… The best example of located intelligence, of course, is the body." (Cassell, 2001)

What a body can give an agent:

- A **locus of attention**
- Visible **status**: working, speaking, waiting
- A **character** that signals its role, and social cues to design with

[7] Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine*, 22(4).

Embodiments → frames of mind[1]

Physical frame

Co-situated; proxemic norms; an **independent agent** that can come to you and interrupt.

Virtual frame

User decides when to engage; enters a crafted plot; conventions are learned.

Also: blended, mediated, immersive frames. Embodiment is the **high-stakes axis**: physical, irreversible action.

[1] Mutlu, B. (2021). The virtual and the physical: two frames of mind. *iScience*, 24(2), 101965. doi:10.1016/j.isci.2020.101965

Physical vs. virtual: the tradeoffs[1]

Mechanism	Physical	Virtual
Situativity	Co-situated in *your* environment	You are brought into *its* environment
Agency	Independent agent, own goals	Engagement at the user's discretion
Proxemics	Dynamic, co-managed to human norms	Constrained, learned conventions
Believability	Real-world, self-relevant	Safe space to experience emotion

[1] Mutlu, B. (2021). The virtual and the physical: two frames of mind. *iScience*, 24(2), 101965. doi:10.1016/j.isci.2020.101965

Method callback: Wizard-of-Oz[8]

(From Methods 3, Sep 25)

- **The** method for prototyping agent behavior *before the model exists*
- A human "wizard" drives the agent; users experience the interaction, not the plumbing
- Lets you study delegation & initiative-return **before** committing to autonomy

[8] Kelley, J.F. (1984). An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications. *ACM Transactions on Office Information Systems*, 2(1), 26–41. doi:10.1145/357417.357420

But what about *AI*?

LLM agents change the *cost* of delegation

- **Cheap:** delegate whole sequences with a prompt
- **Recursive:** agents spawn agents; authority forwards without oversight
- **Opaque:** the sequence of actions is hard to see, harder to *return* from

→ Software agents and robots: **one problem**, reliance over action sequences

Guardrails coda

- **Suchman:** "autonomy" conceals the human-in-the-loop labor that sustains it
- **Accountability** must hold over a *sequence*, not a single decision
- **Embodiment** raises the stakes: physical, *irreversible* action demands a returnable path

Your provocation (*post one slide*)

Pick one:

1. **Provocation:** argue the tradeoff, *efficiency through delegation* vs. *human agency & accountability*
2. **Critical artifact:** a real AI agent. Mark where authority transfers, whether initiative can be taken back, and where "autonomy" hides human labor
3. **Design / policy:** one principle that keeps oversight & accountability over a *sequence* of actions

Discussion

Today's roles

Reporter	Skeptic	Connector
Seat 1	**Seat 2**	**Seat 3**

Same in every pod. **Seats 4, 5 and 6** have no role today; you are still in the argument.

Discussion Format

In your pod, on your pod monitor: ~30 minutes.

- **Everyone shows their slide**, about 2 minutes each, cast from your own laptop
- **Skeptic** puts a counterargument on the table · **Connector** ties it to an earlier week
- **Reporter** writes the pod's three lines (agreed, split, unresolved) and delivers them to the room

What's Next?

- **Friday, Methods 8: Statistics I**
  - **Read:** the textbook **Chapter 4** (descriptive statistics) · Cumming (2014), *The New Statistics*
  - Team **project status** on Canvas by 9:30 Friday (done / blocker / next)
- *Methods handoff:* test "users delegate appropriately" by varying **agent correctness**