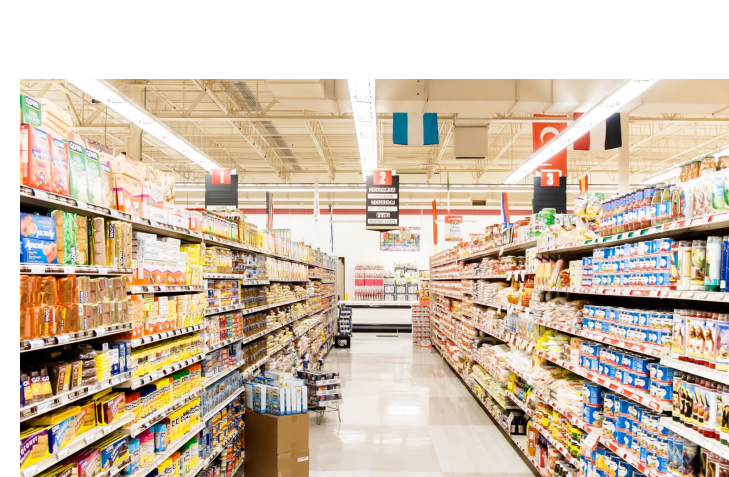


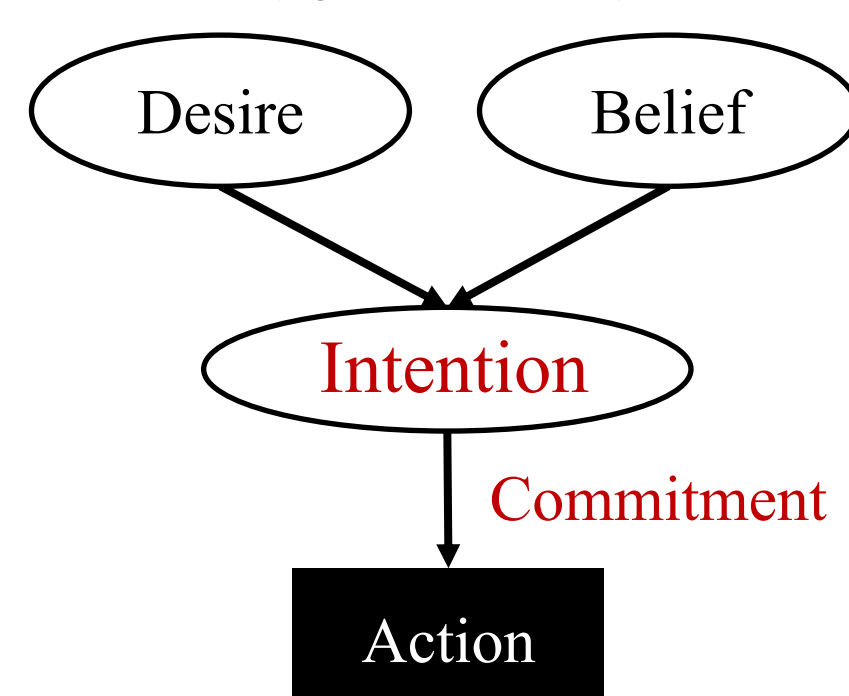
Motivation

How to act with conflicting desires



The problem of conflicting desires.

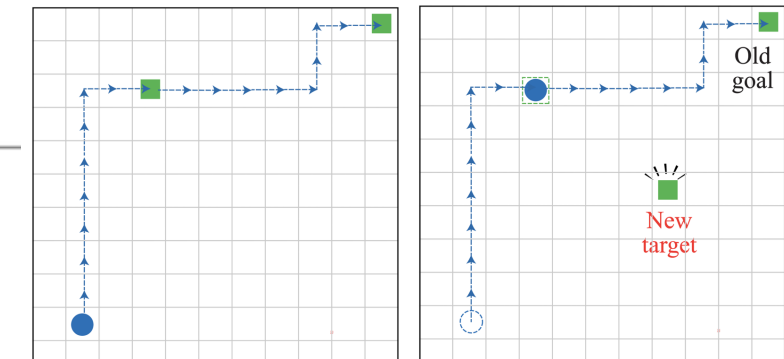
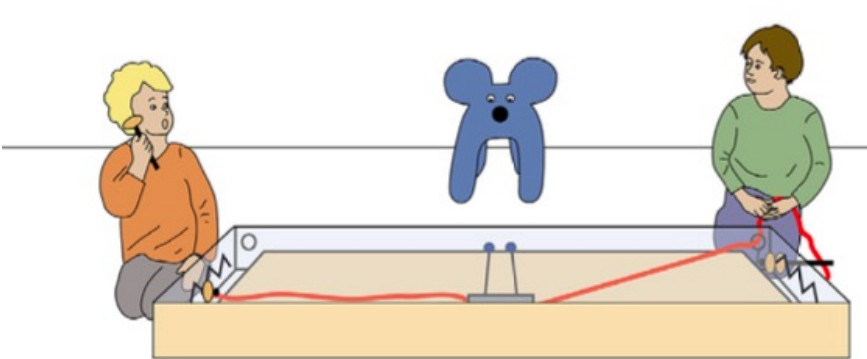
Intention theory (e.g., Bratman, 1987)



Commitment to intentions for coherent actions over time.

Visual-grounded commitment

Joint-commitment Self-commitment



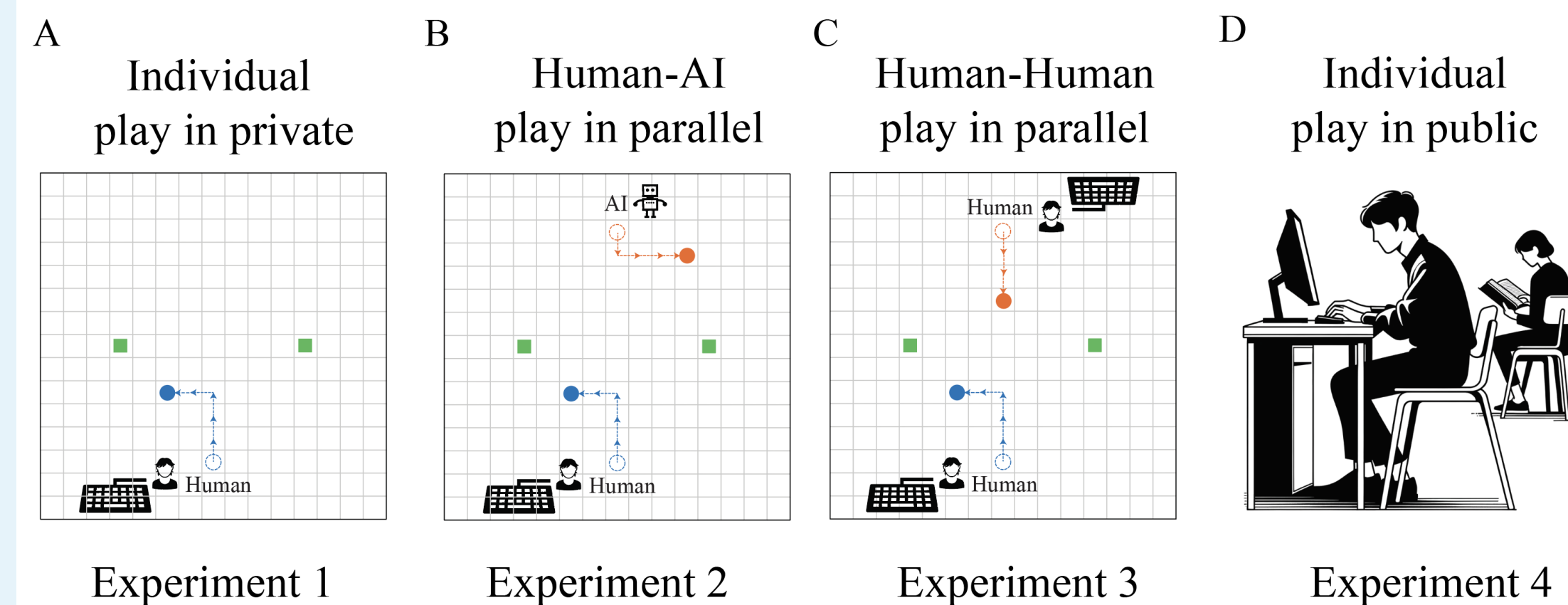
- Children commit to joint goals for cooperation.
- Earliest sign of uniquely human (Tomasello, 2019)
- Adults commit to intentions in individual decision-making.
- Qualitatively different from AI (Cheng et al., 2023)

Questions:

- Social origin of self-commitment?
- Can social contexts impact the **strength**, **content**, and **timing** of self-commitment in individual tasks?

Design

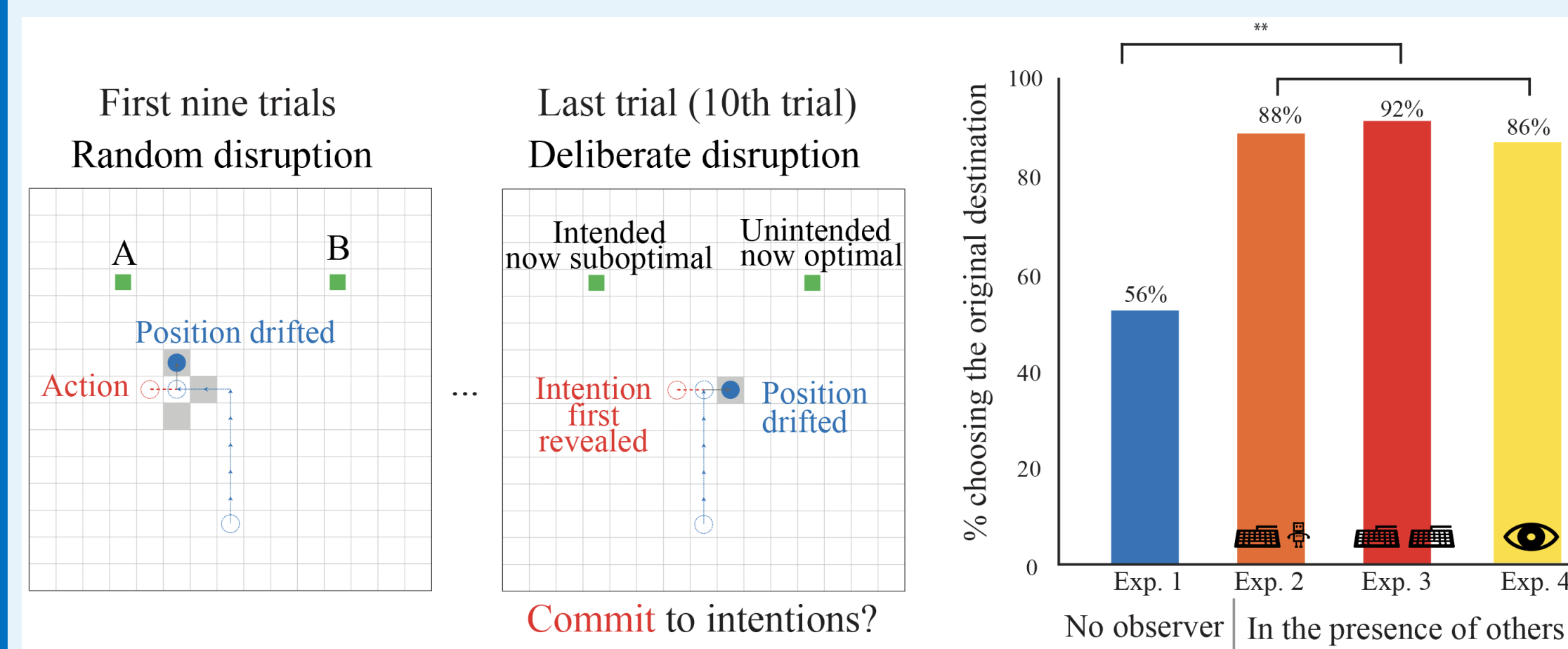
Manipulations of social contexts



Task instructions & details:

- Explicitly informed of a **purely individual task** with no need to interact with the other agent.
- The second player was either a reinforcement learning agent optimizing its own utilities (Exp. 2) or a human participant (Exp. 3); both were introduced as human players.
- $N = 200$; $n = 50$ per experiment; 10 rounds per game.

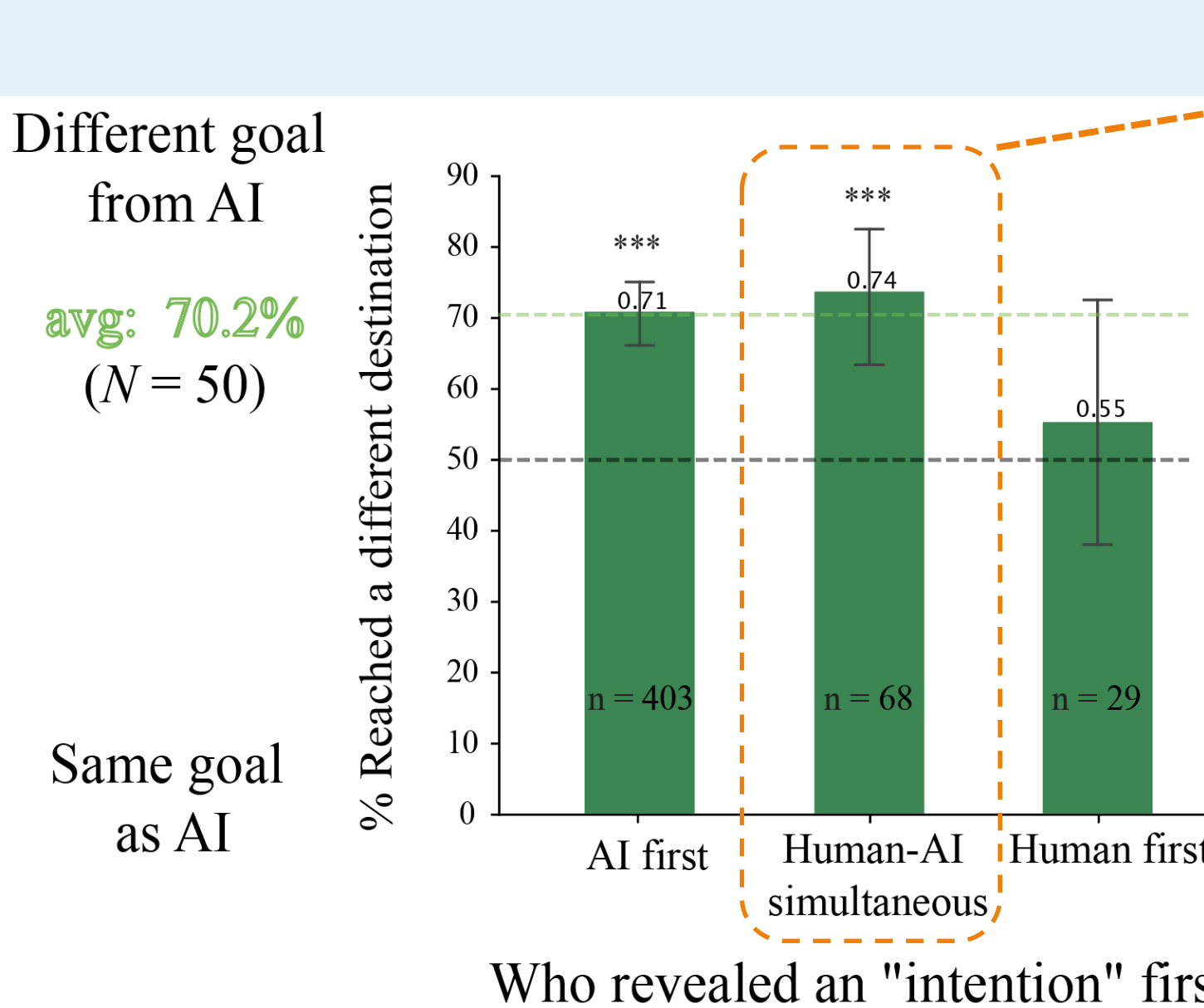
Measure commitment strength: goal perseverance



Social contexts consistently increased the strength of commitment.

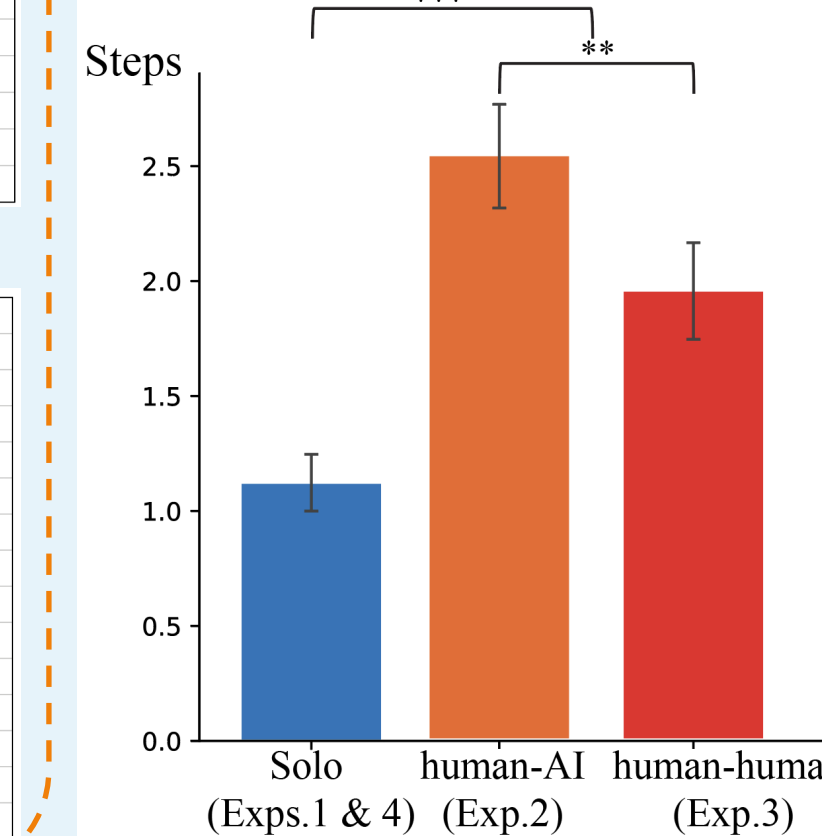
Measure commitment content: conflict avoidance

Different goal from AI
avg: 70.2%
($N = 50$)



Preferred a different goal, as long as they were not the first to reveal an intention.

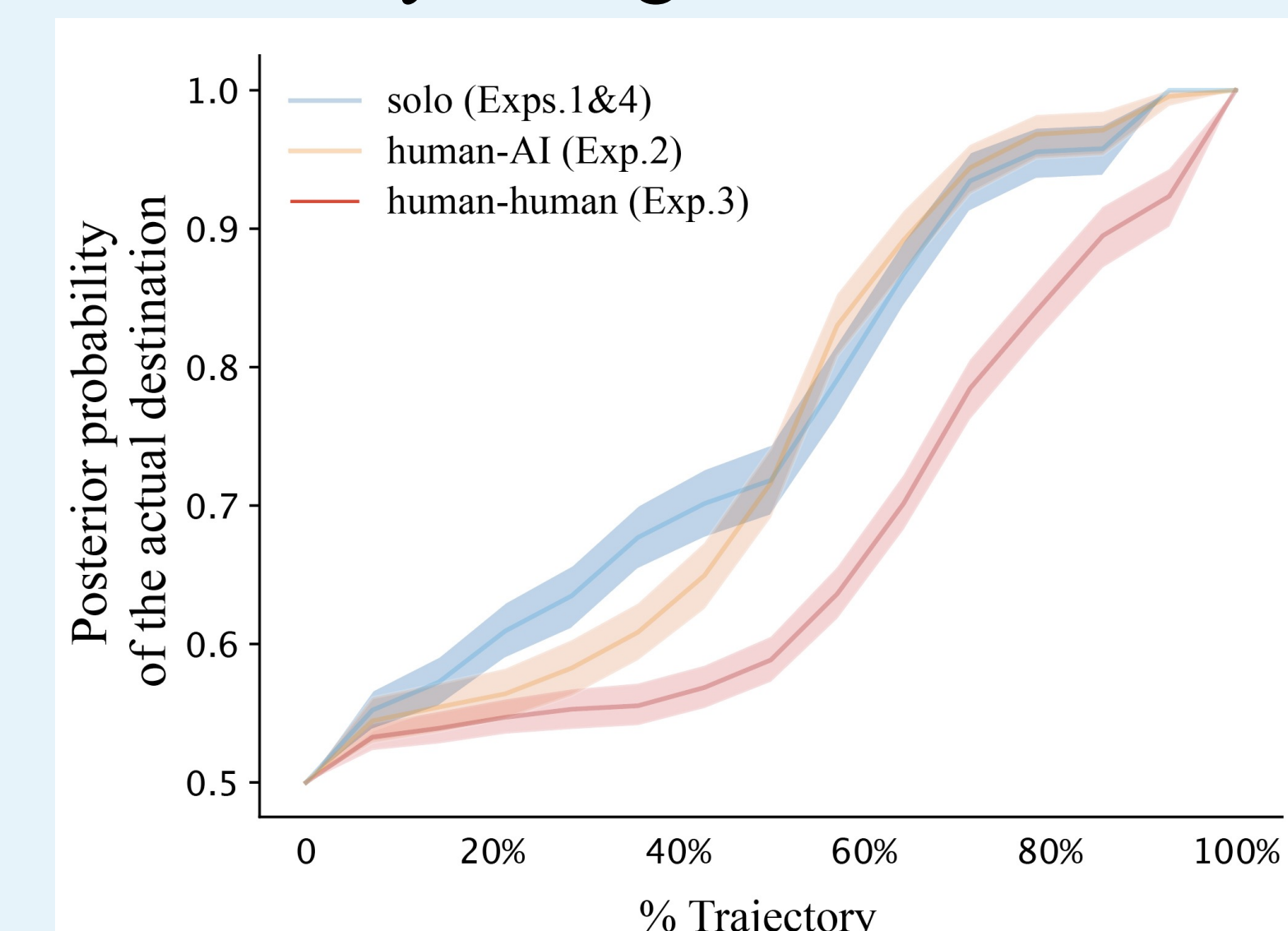
Navigation Step Performance



Took more steps than when alone.

Measure commitment timing

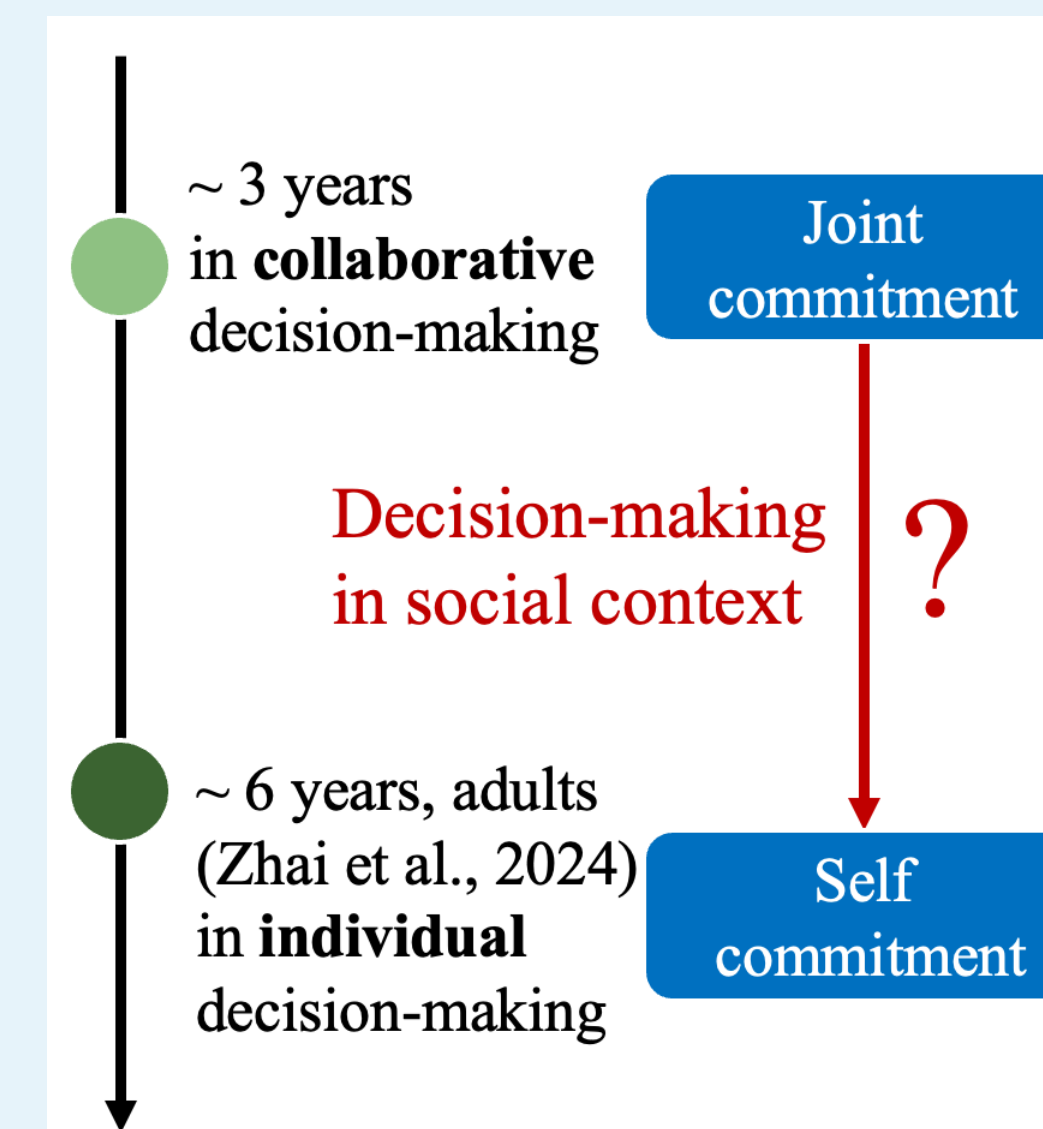
Bayesian goal inference



Delayed commitment in dual play—strongest in human-human games.

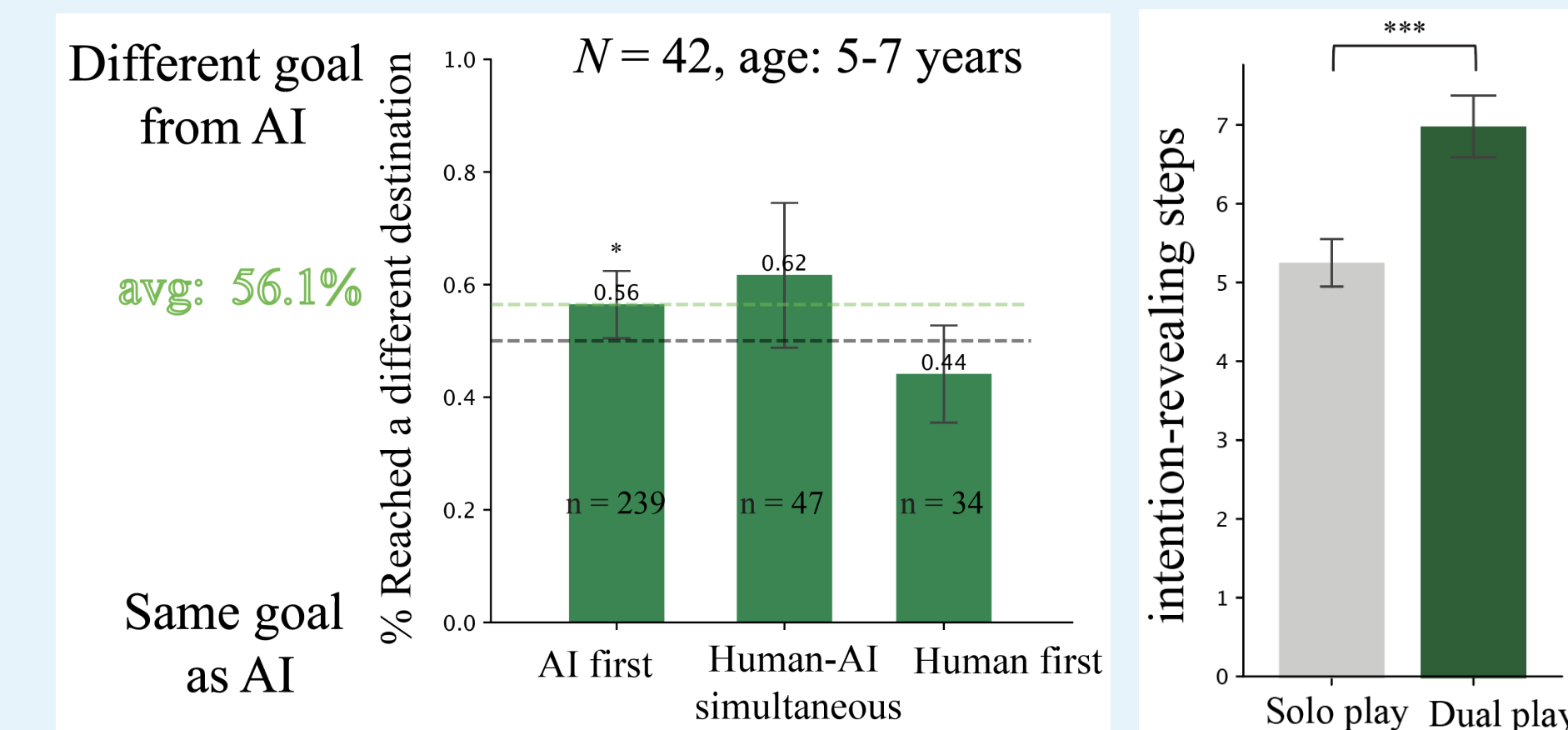
Ongoing studies: Children-AI

Developmental Hypothesis



Emerge before 6 years?

Within-subj design: solo + dual



5-7 yrs showed spontaneous social coordination.

- preferred a different goal when others signaled commitment first.
- took longer to decide on a goal in social contexts.

Discussions & future work

- Even in individual tasks, self-commitment is deeply intertwined with social context, influencing how people manage their goals and interactions with others.
- What if humans were told they are playing with AI? What if both collaborative and individual goals exist?
- What critical cognitive processes are socialized or internalized to support the emergence of self-commitment?



Email: chengshaozhe@ucla.edu
Web: chengshaozhe.github.io