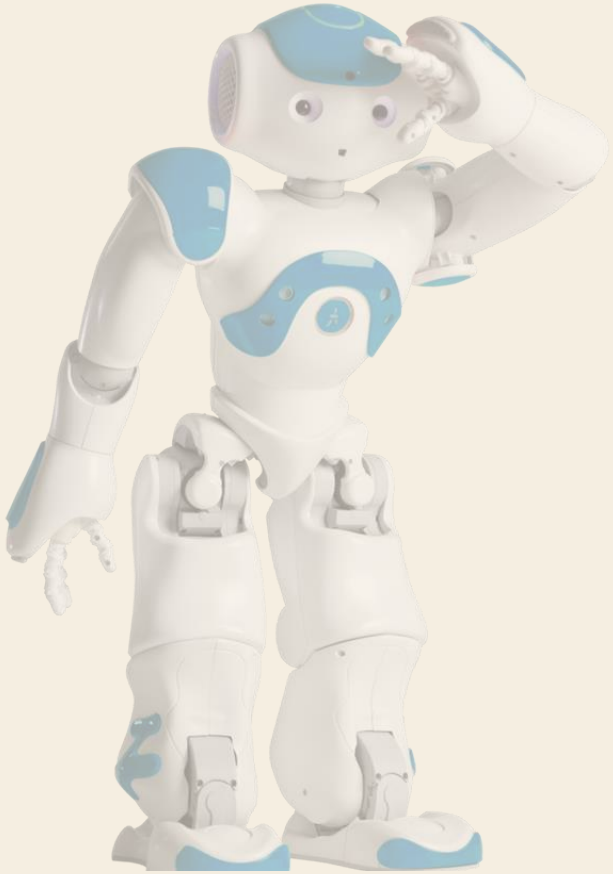


Sense of Agency in Individual and Joint Actions with Humans vs. Artificial Agents

Effects of Action Selection and Outcome Valence

Dr Zeynep Barlas

22nd November, 2022

DE MONTFORT UNIVERSITY LEICESTER

OUTLINE

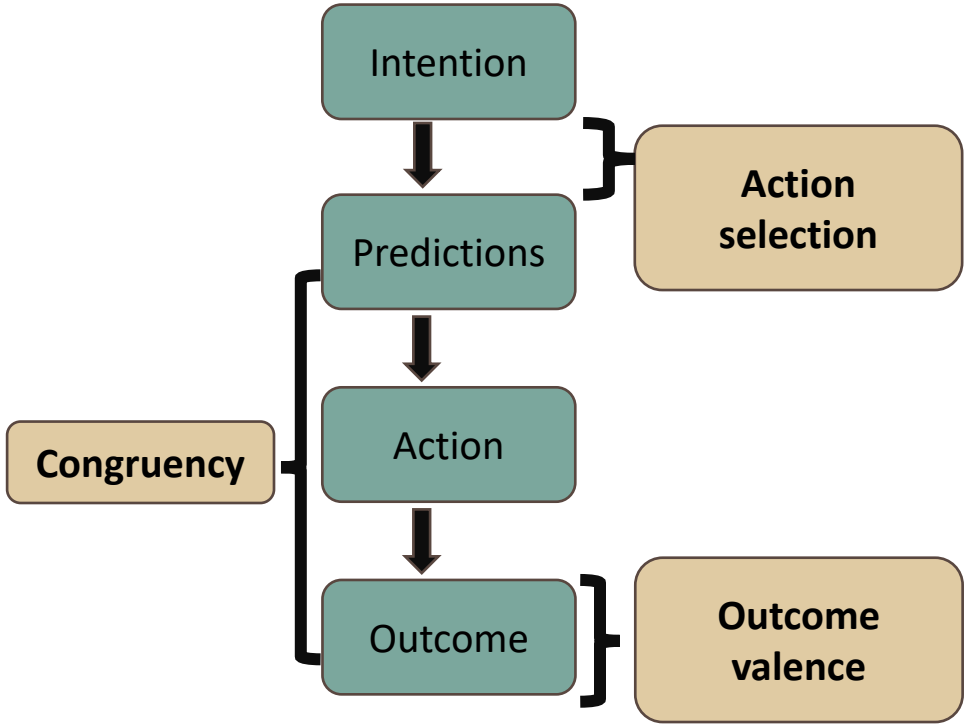
- Overview: The effects of choice, outcome valence, and action-outcome congruency on the sense of agency (SoA) in *individual* actions.

- Background: Identity of the co-actor in *joint* actions.

- Experiment 1&2 (Barlas, 2019, *Consciousness & Cognition*): **SoA in human- vs. robot-instructed actions.**

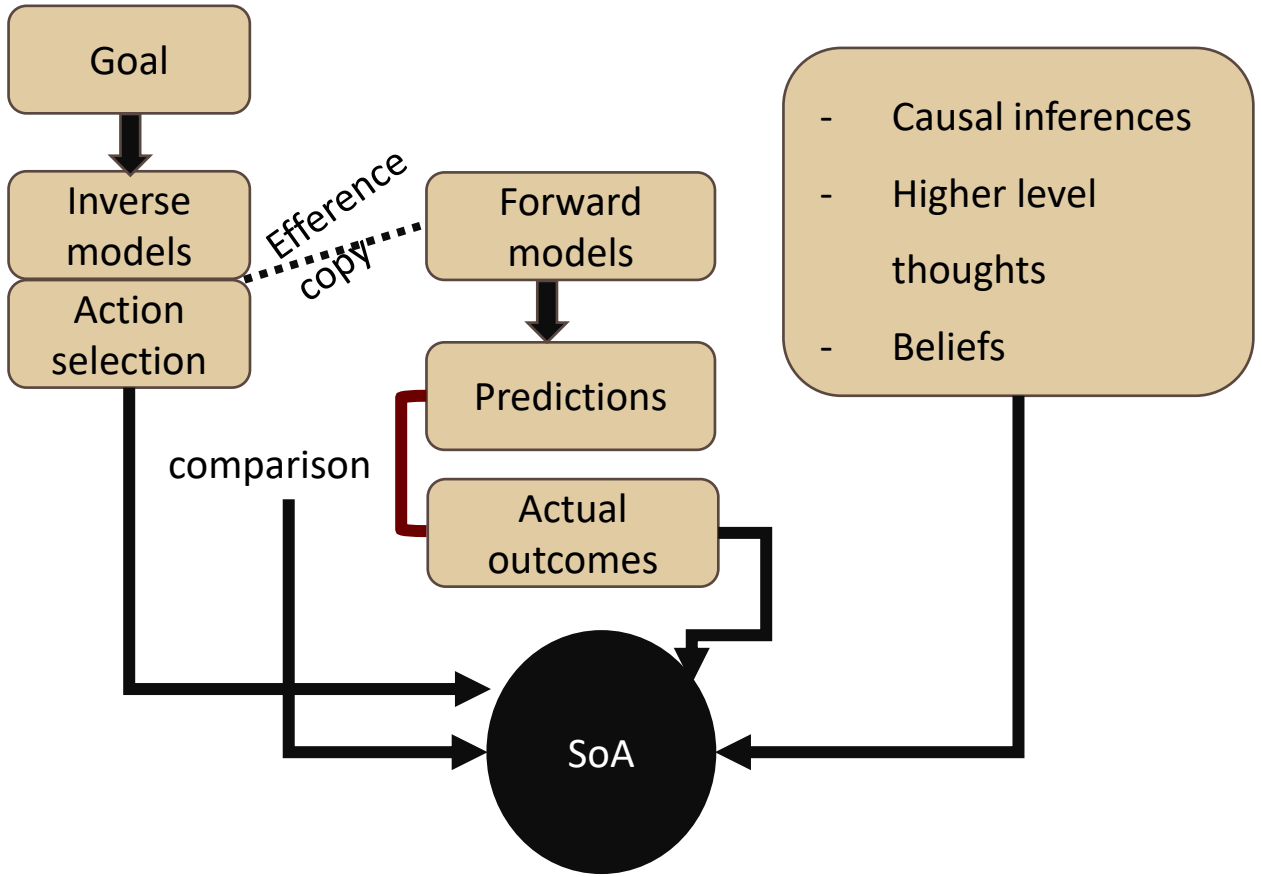
- Future directions: Examination of SoA, responsibility, and trust in HRI.

PROCESSES CONTRIBUTING TO THE SOA



Prospective (Predictive) Processes

Retrospective (Postdictive)

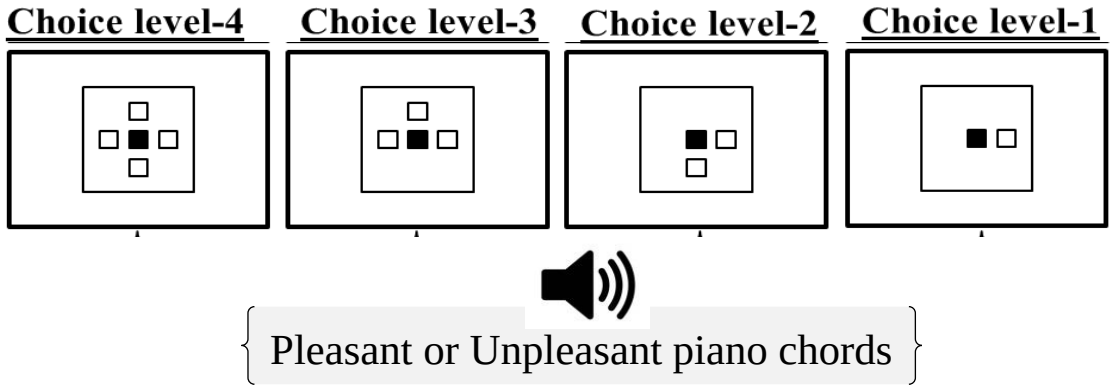


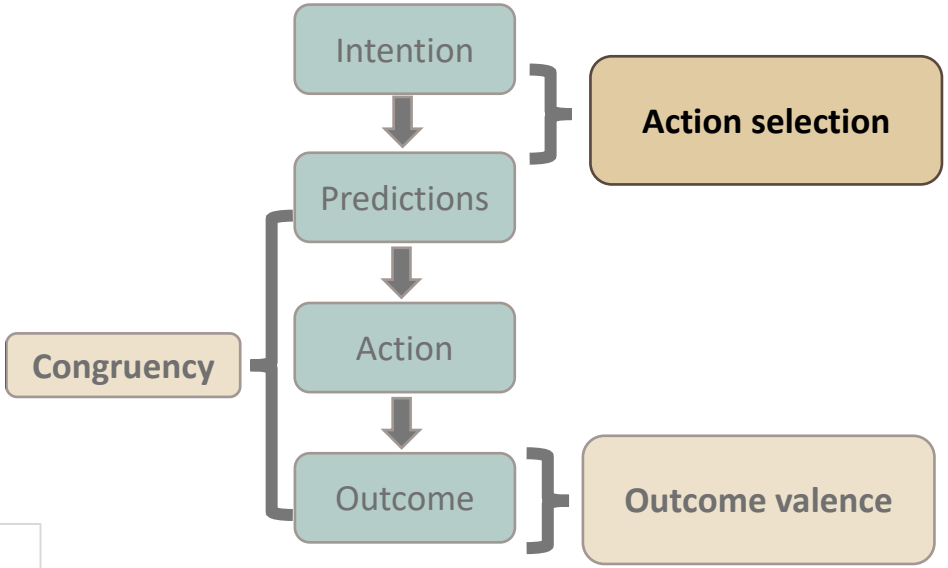
(Blakemore et al., 2002; Frith et al., 2000; Wolpert et al., 1995)

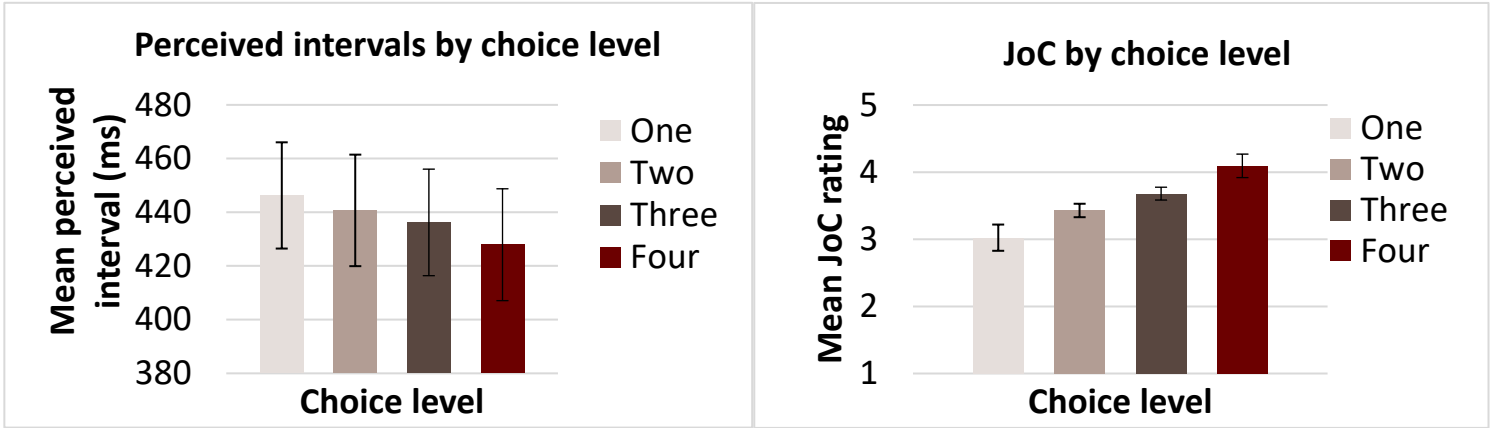
(Wegner & Wheatley, 1999; Wegner & Sparrow, 2004)

EFFECT OF ACTION CHOICE

- **The effect of choice space:** Greater number of action alternatives yield stronger SoA (binding & control ratings) compared to instructed actions (Barlas & Obhi, 2013; Barlas et al., 2017; 2018).



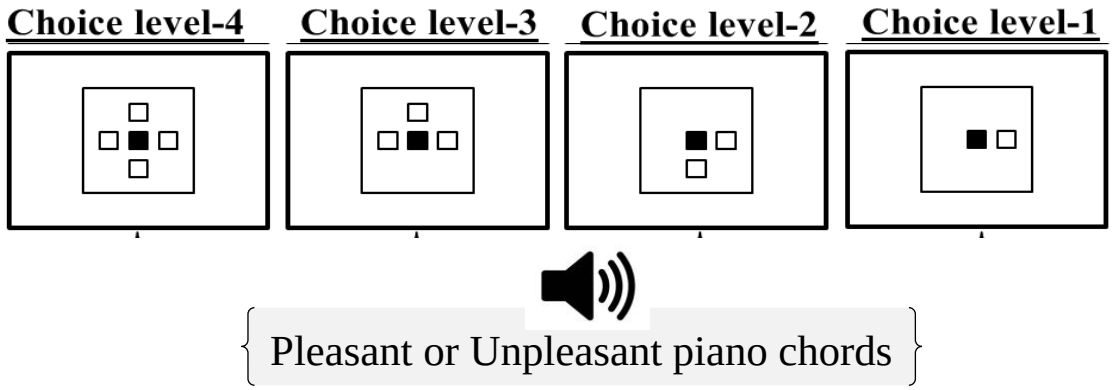


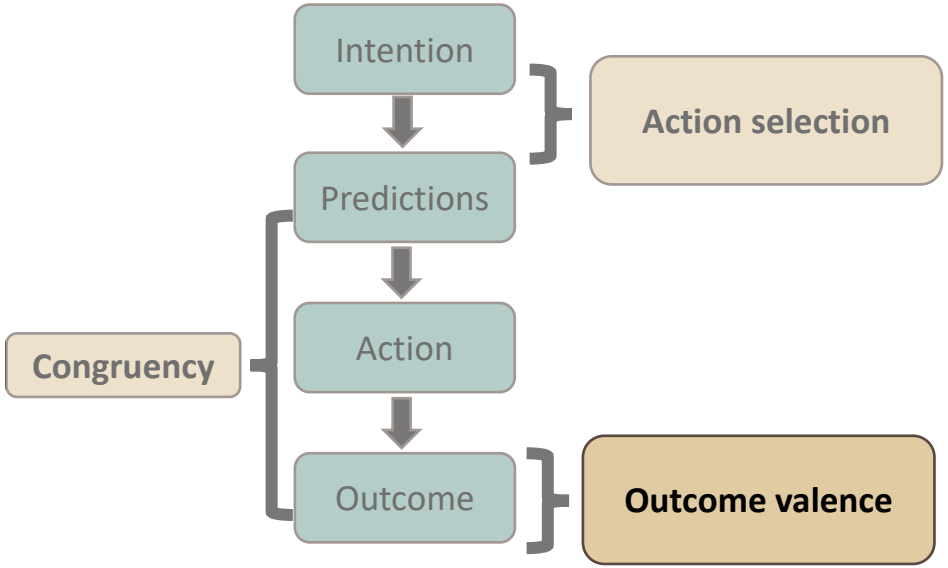


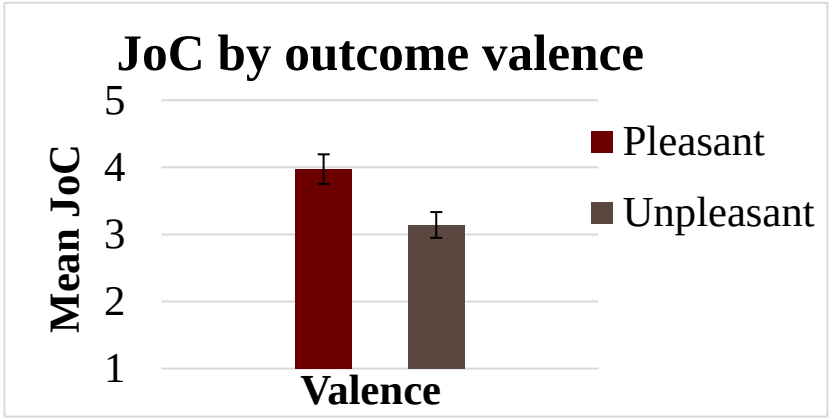
(Barlas et al., 2018)

EFFECT OF OUTCOME VALENCE

- **The effect of outcome valence:** Pleasant outcomes enhance the SoA (explicit-judgment of control) compared to unpleasant outcomes (Barlas & Obhi, 2014; Barlas et al., 2017; 2018).



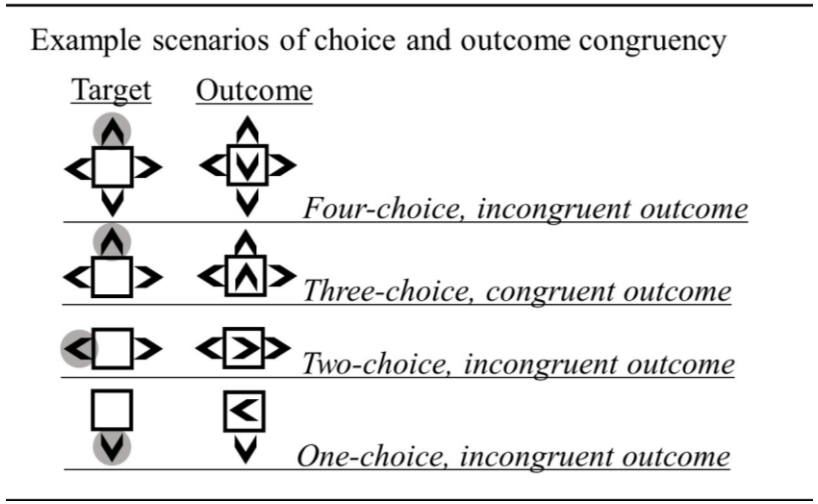


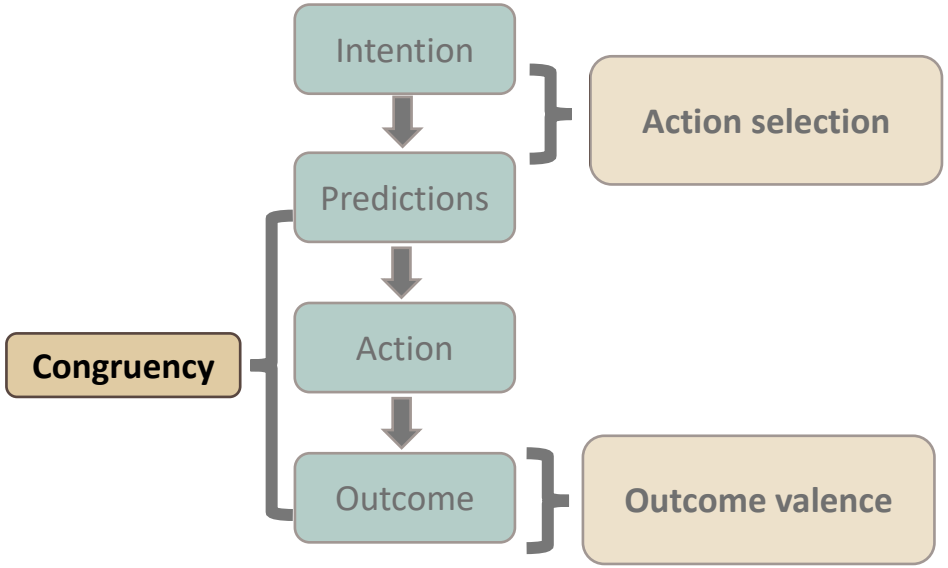


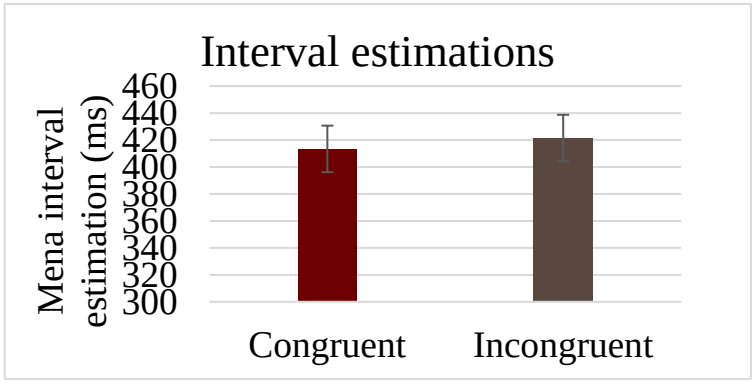
(Barlas et al., 2018)

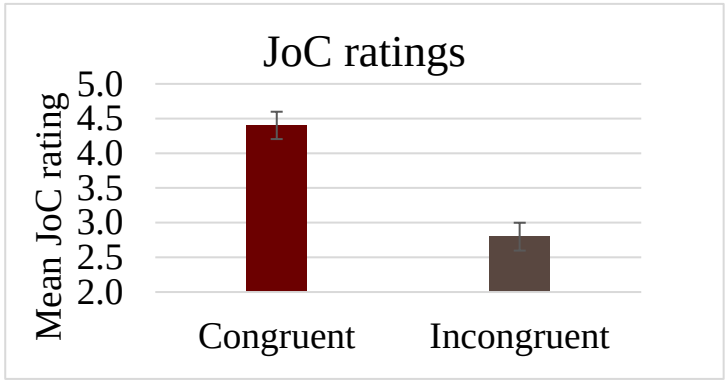
EFFECT OF ACTION-OUTCOME CONGRUENCY

- **The effect of outcome congruency:** Congruent (expected) outcomes yield stronger the SoA compared to incongruent outcomes (Barlas & Kopp, 2018).









FROM INDIVIDUAL TO JOINT ACTIONS (WITH HUMANS VS ARTIFICAL AGENTS)

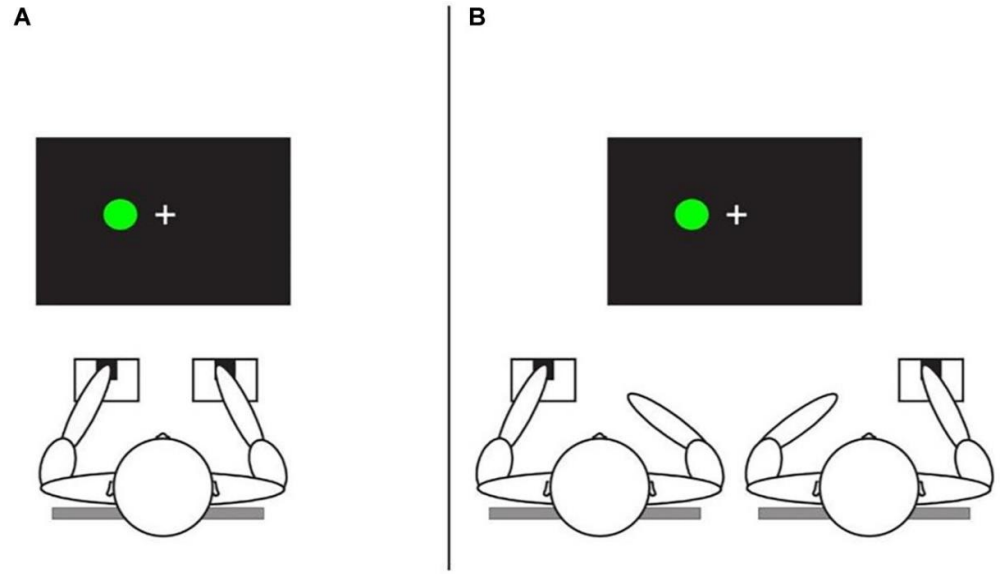
- SoA is distinctively experienced in joint actions depending on the (believed) identity of the action partner.
- We-agency in cooperative joint (human-human) actions (e.g., Obhi & Hall, 2011).
  - Comparable intentional binding for self-caused outcomes and those that were believed to be caused by a human partner's actions.
  - When participants believed that the interaction partner was a computer, intentional binding for both their own and the computer's actions was diminished.
- The failure to co-represent the intentions and motor plans of these artificial systems (Obhi & Hall, 2011; Wohlschläger et al., 2003)

FROM INDIVIDUAL TO JOINT ACTIONS (WITH HUMANS VS ARTIFICAL AGENTS)

- Joint Simon effect: In joint actions, individuals automatically co-represent each other's actions (Sebanz et al., 2003).

- Believing that the robot partner in autonomous / intentional (vs. machine-like) yields the effect (Stenzel et al., 2014).

- Both intentional binding and Social Simon effect were diminished when participants were paired with a computer in comparison to another person (Sahaï et al., 2019).



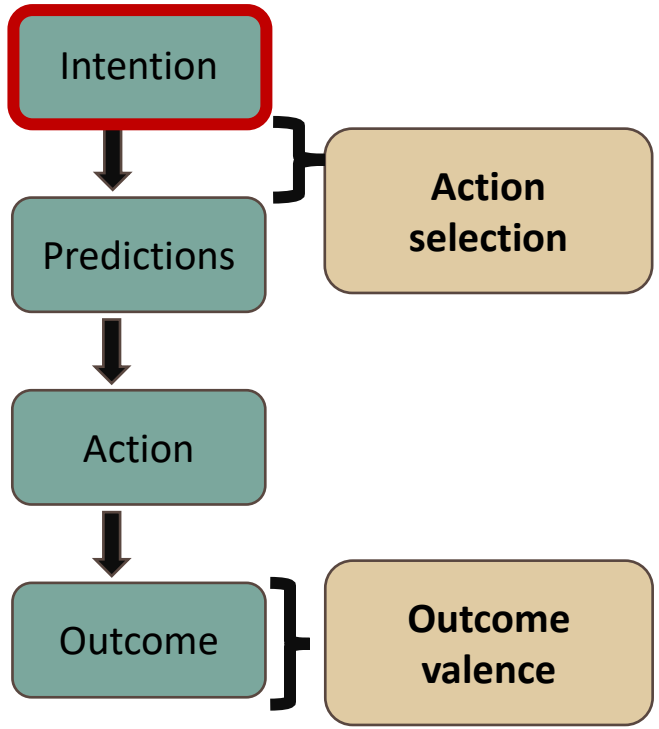
Standard Simon task (A) and joint Simon task (B) (Ruissen & de Bruijn, 2016).

FROM INDIVIDUAL TO JOINT ACTIONS (WITH HUMANS VS ARTIFICAL AGENTS)

- SoA is distinctively experienced in joint actions with humans vs. artificial agents.
- Based on observation of the co-actor's (human or machine) actions.

- Co-representing the intentions of action partners, whose role is to give action instructions.

- Does belief in the intentionality of artificial agents affect one's SoA when performing their instructions?



THE ROLE OF AA: CO-ACTING VS GUIDING

- In many applications of artificial agents, the major role of these agents is to guide human activities.

Navigation systems advising directions

iRobi and Cafero (Yujin Robot):assistive robots that remind patients to take medications and provide cognitive stimulation (Broadbent, 2017)

SOA IN HUMAN VS. ROBOT INSTRUCTED ACTIONS: AIMS

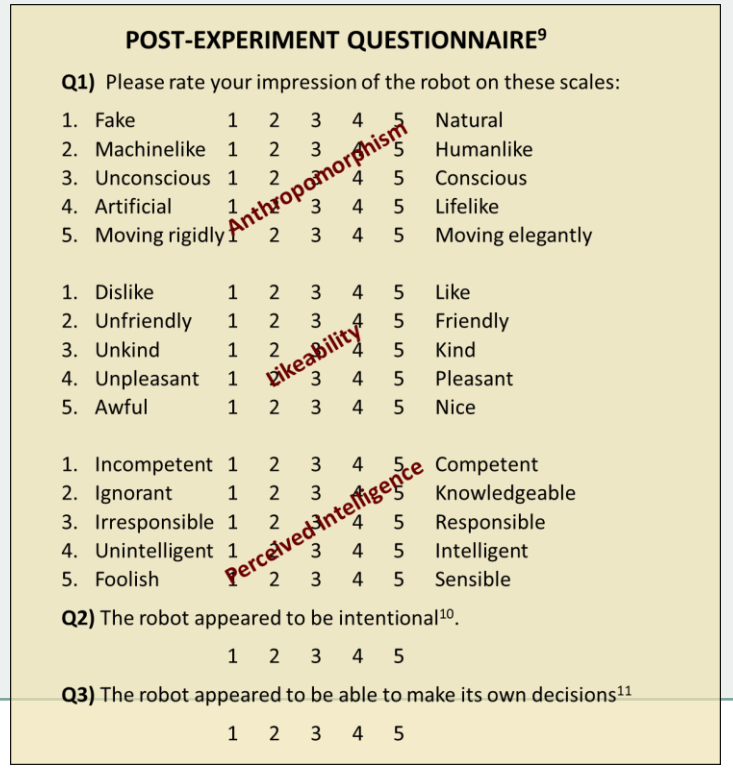
1. to examine how receiving action instructions from a humanoid robot as compared to a human would influence SoA (intentional binding & judgment of control) also compared to when actions are freely selected.
2. to investigate whether belief in robot autonomy affects one's SoA in robot-instructed actions.

Human-instructed

Robot-instructed

EXPERIMENT 1: DESIGN

IVs	DVs
1. Choice (blocked) - Free (right or left key press) - Human-instructed - Robot-instructed - Passive **2. Action-outcome delay (mixed)** 200, 400, 600, 800 (ms) **3. Robot autonomy (between-subjects)** -Autonomous (n=30) -Non-autonomous (n=30)	1. Interval estimations 2. Judgment of control ratings 3. Post experiment questionnaire



POST-EXPERIMENT QUESTIONNAIRE[9]

Q1) Please rate your impression of the robot on these scales:

Anthropomorphism

1.	Fake	1	2	3	4	5	Natural
2.	Machinelike	1	2	3	4	5	Humanlike
3.	Unconscious	1	2	3	4	5	Conscious
4.	Artificial	1	2	3	4	5	Lifelike
5.	Moving rigidly	1	2	3	4	5	Moving elegantly

Likeability

1.	Dislike	1	2	3	4	5	Like
2.	Unfriendly	1	2	3	4	5	Friendly
3.	Unkind	1	2	3	4	5	Kind
4.	Unpleasant	1	2	3	4	5	Pleasant
5.	Awful	1	2	3	4	5	Nice

Perceived Intelligence

1.	Incompetent	1	2	3	4	5	Competent
2.	Ignorant	1	2	3	4	5	Knowledgeable
3.	Irresponsible	1	2	3	4	5	Responsible
4.	Unintelligent	1	2	3	4	5	Intelligent
5.	Foolish	1	2	3	4	5	Sensible

Q2) The robot appeared to be intentional[10].

1 2 3 4 5

Q3) The robot appeared to be able to make its own decisions[11]

1 2 3 4 5

EXPERIMENT 1: MANIPULATION OF ROBOT AUTONOMY

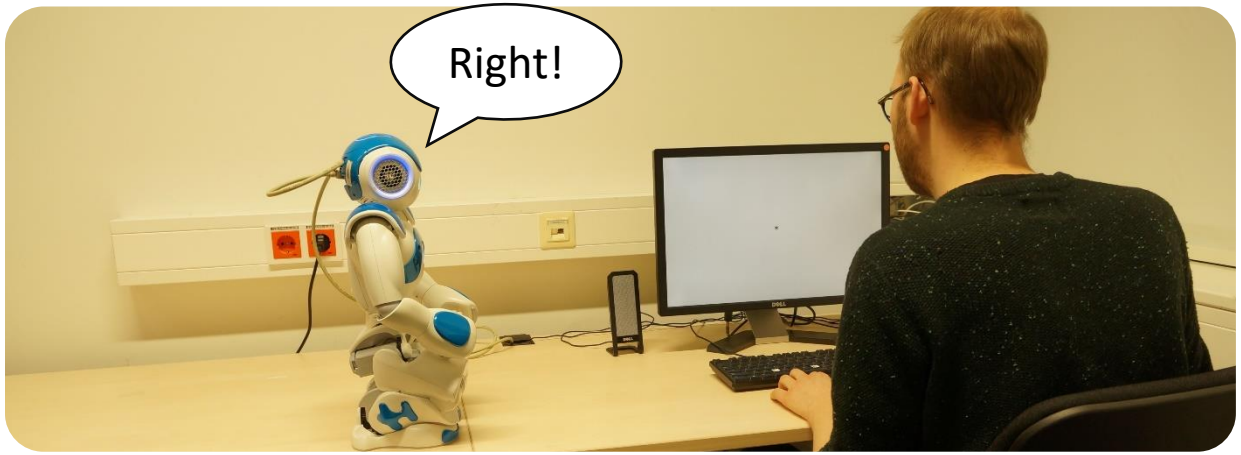
Autonomous

“Zora can make its own decisions by modeling how humans determine their actions and thus, Zora would actively decide in each trial which key you should press“

Non-Autonomous

“Zora’s key press instructions were pre-programmed and will simply tell you which key to press.”

NAO robot (Softbank-Aldebaran Robotics)

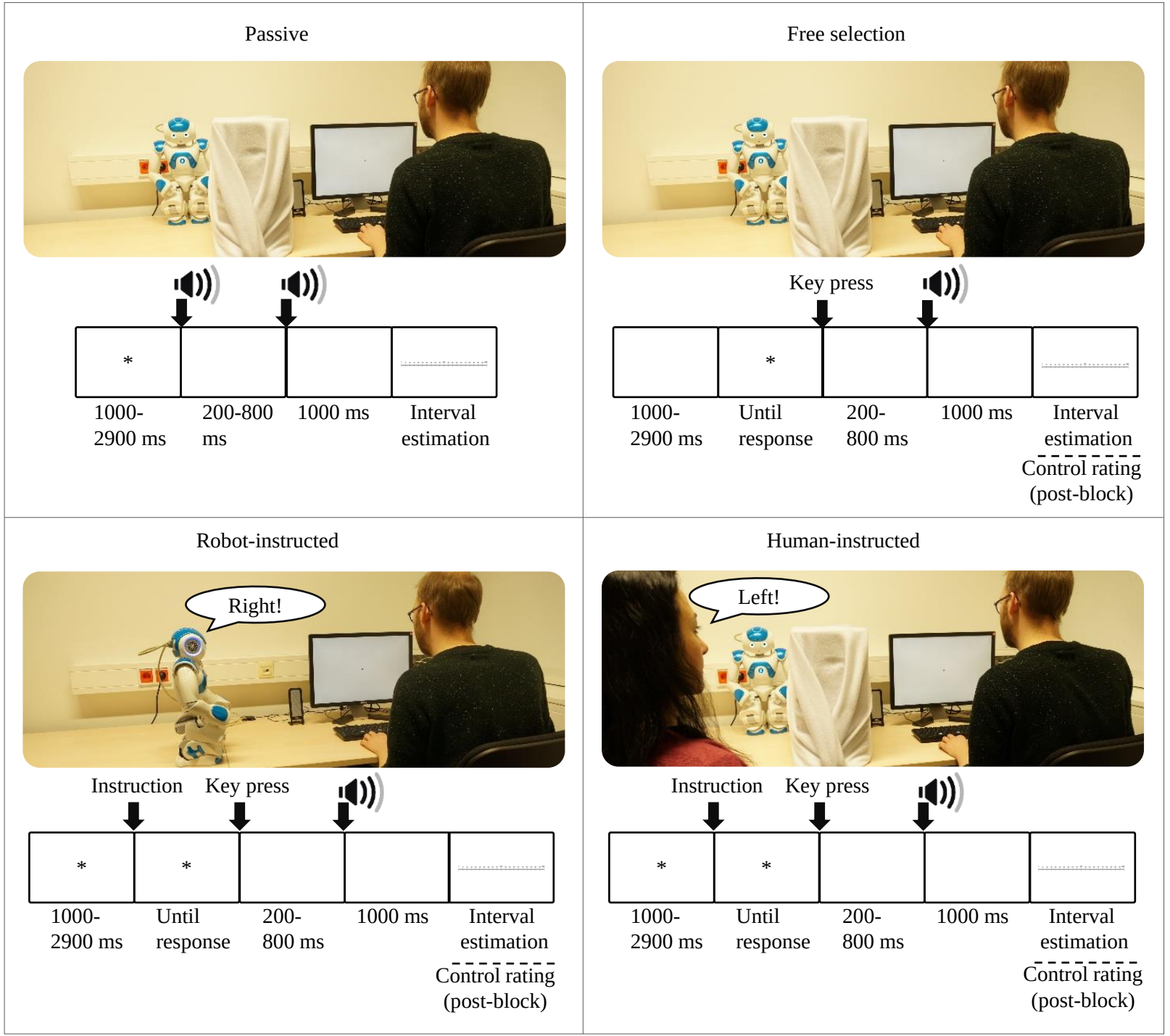


EXPERIMENT 1: PROCEDURE

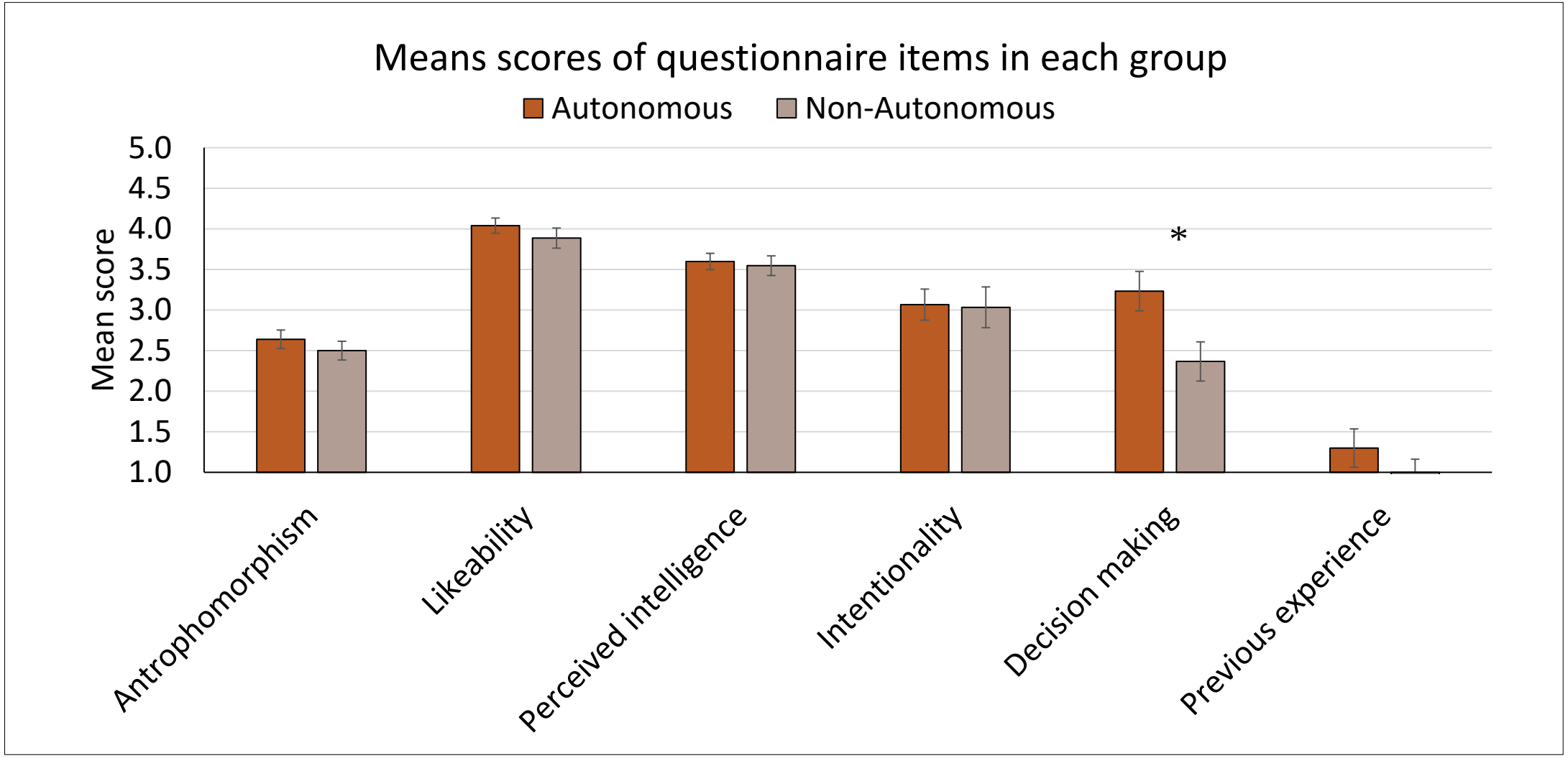
Interval estimation
320 trials in total (80 per block)

JoC rating
Once at the end of each block (except the passive block)

Post experiment questionnaire

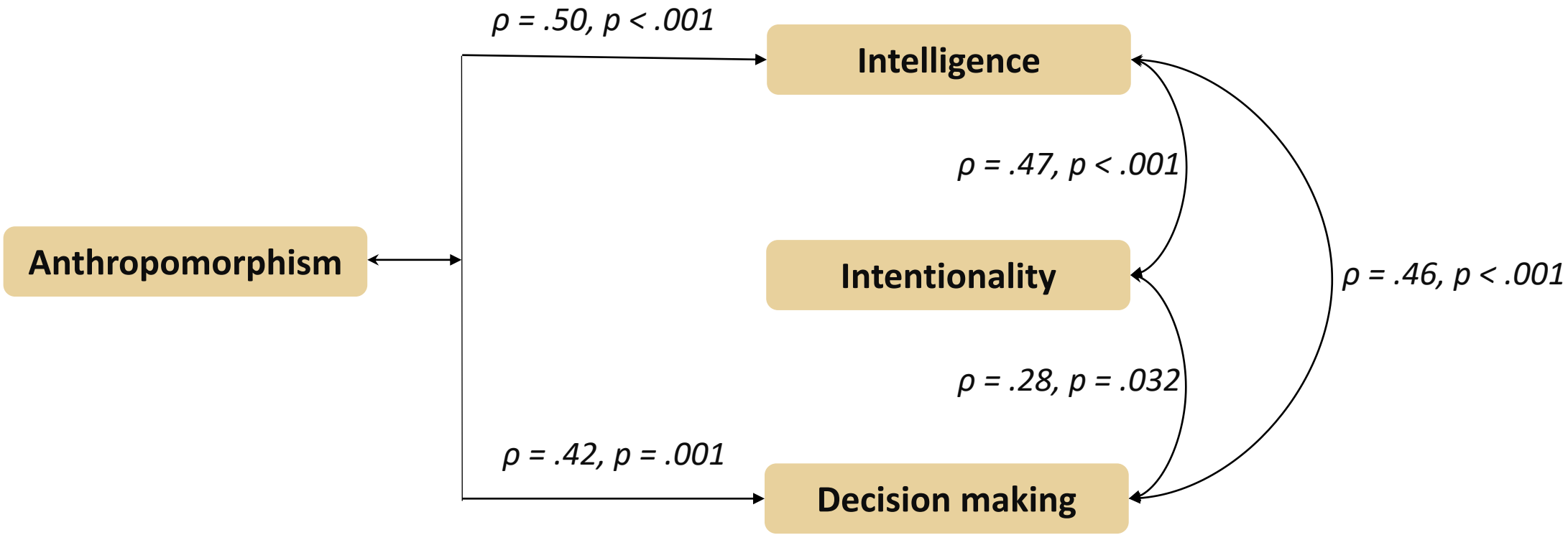


EXPERIMENT 1: RESULTS

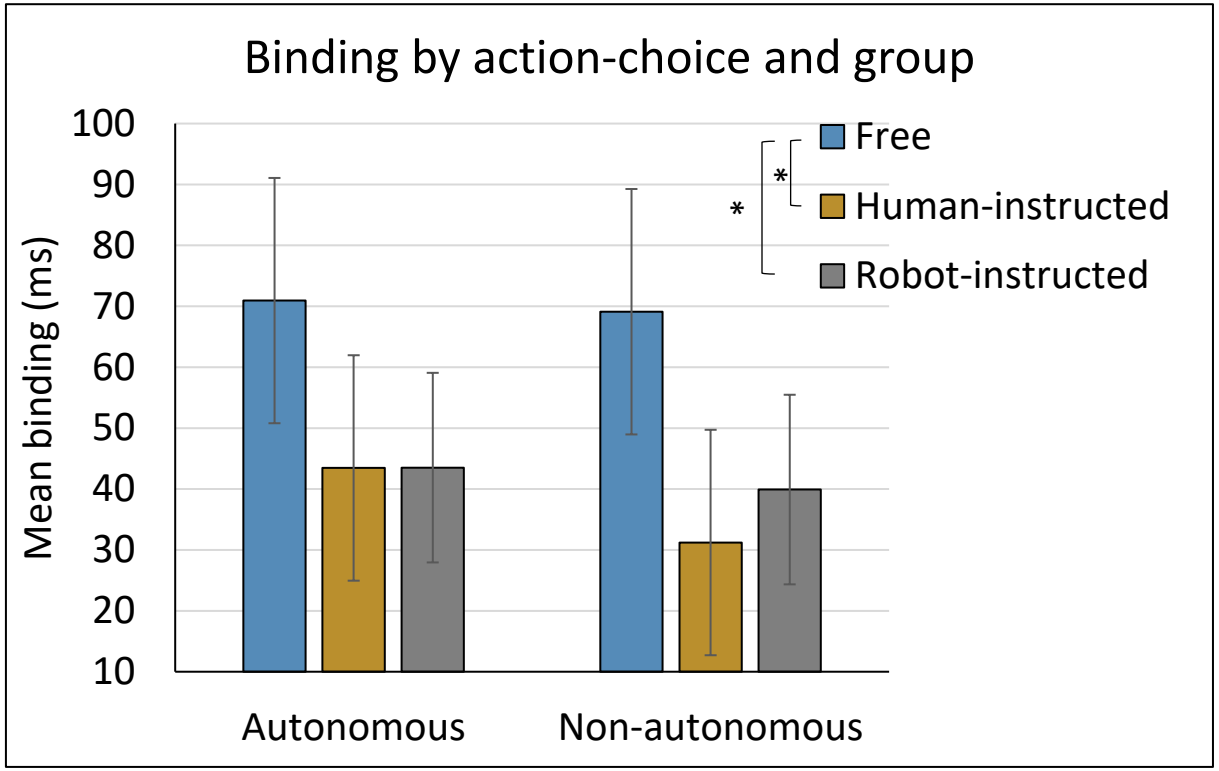


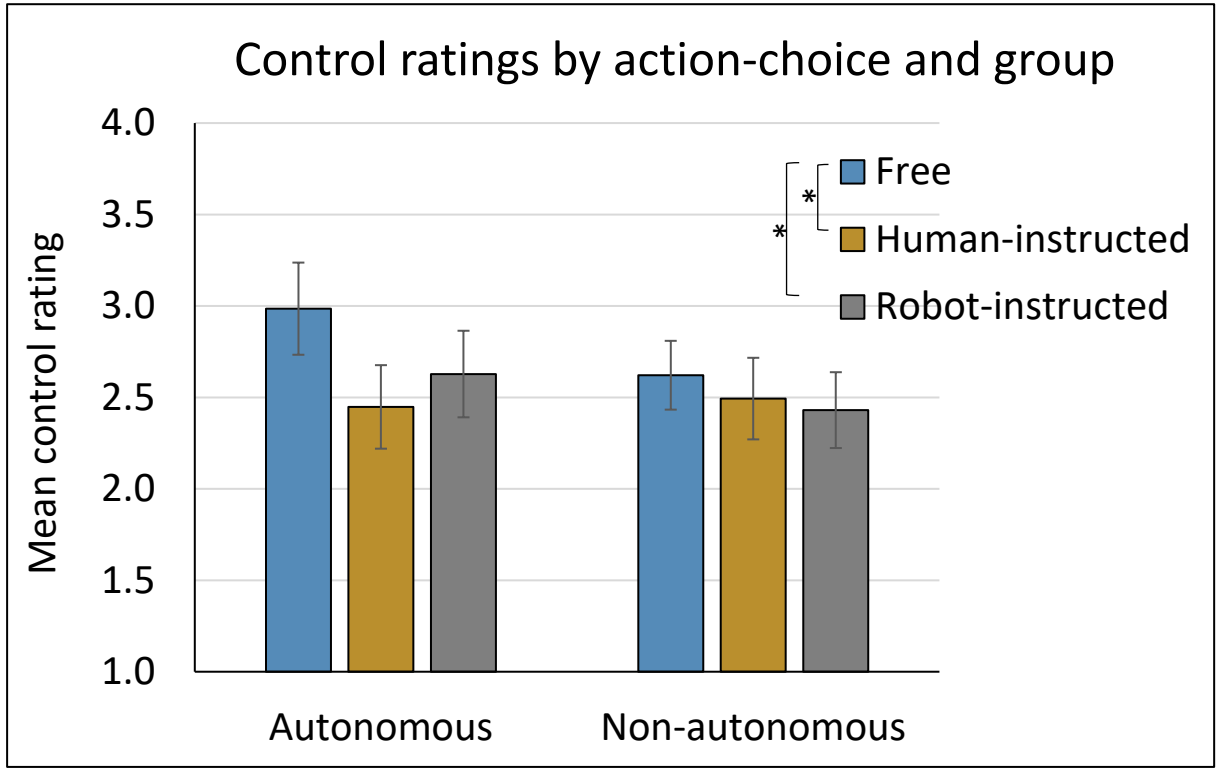
EXPERIMENT 1: RESULTS

- **Relationship among the questionnaire items**



EXPERIMENT 1: RESULTS





Choice: p = .008, η_p^2 = .09

Choice: p = .024, η_p^2 = .06

Binding= Difference in interval estimation between passive and action conditions

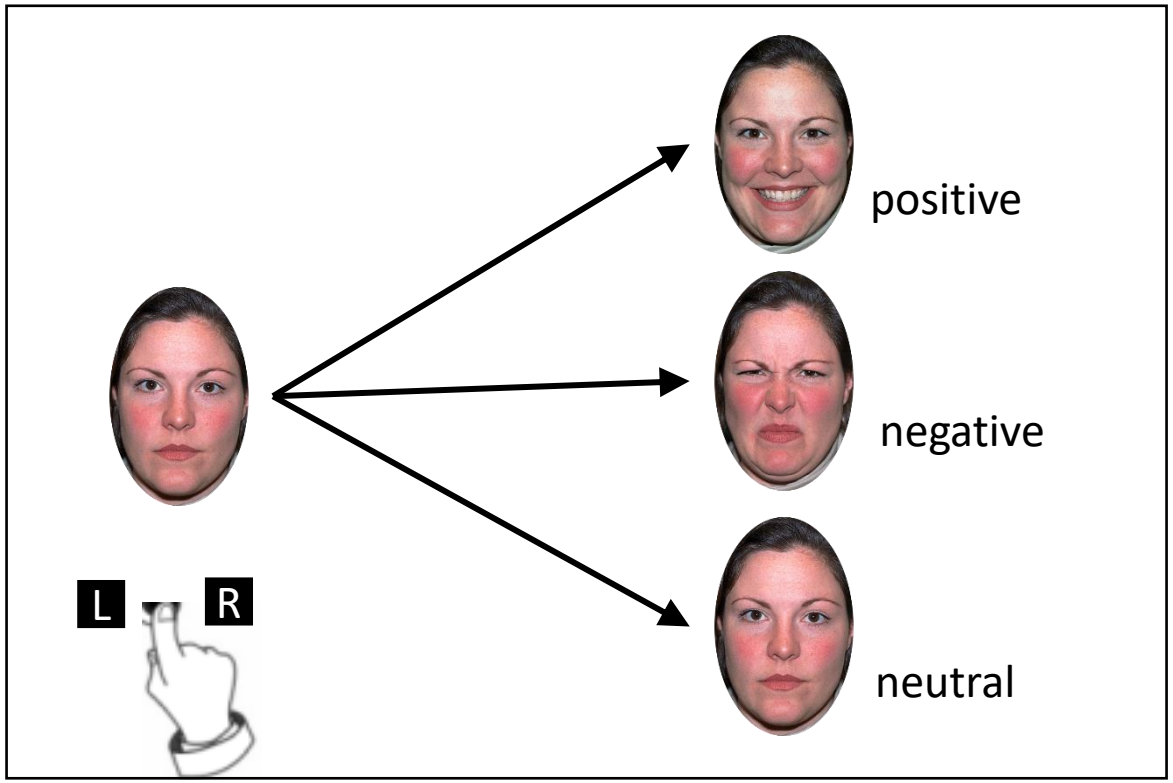
EXPERIMENT 1: RECAP

- **The source of actions:**
  - Free choice of actions yielded stronger binding and explicit judgment of control.
- **Effects of the identity of the instructor or perceived autonomy of the robot**
  - No difference in SoA between human vs. robot instructed actions.
  - No effect of perceived autonomy in robot-instructed actions.

- **Limitations**
  - One data point of control judgments per condition
  - Specificity of outcomes

EXPERIMENT 2: AIM

- Examine the SoA in a similar context in which participants performed free and instructed actions (as in Experiment 1) that produced positive, negative, or neutral outcomes.



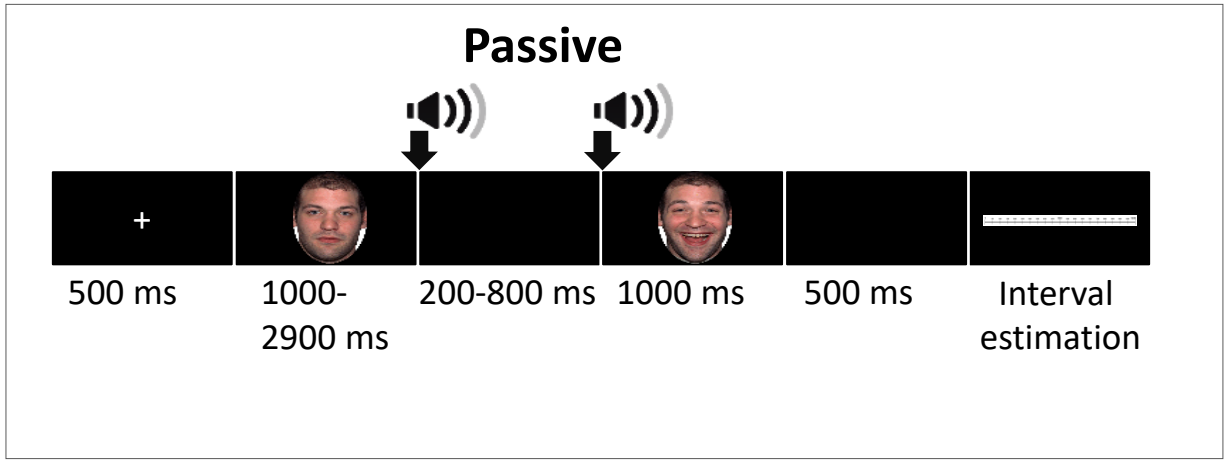
EXPERIMENT 2: DESIGN

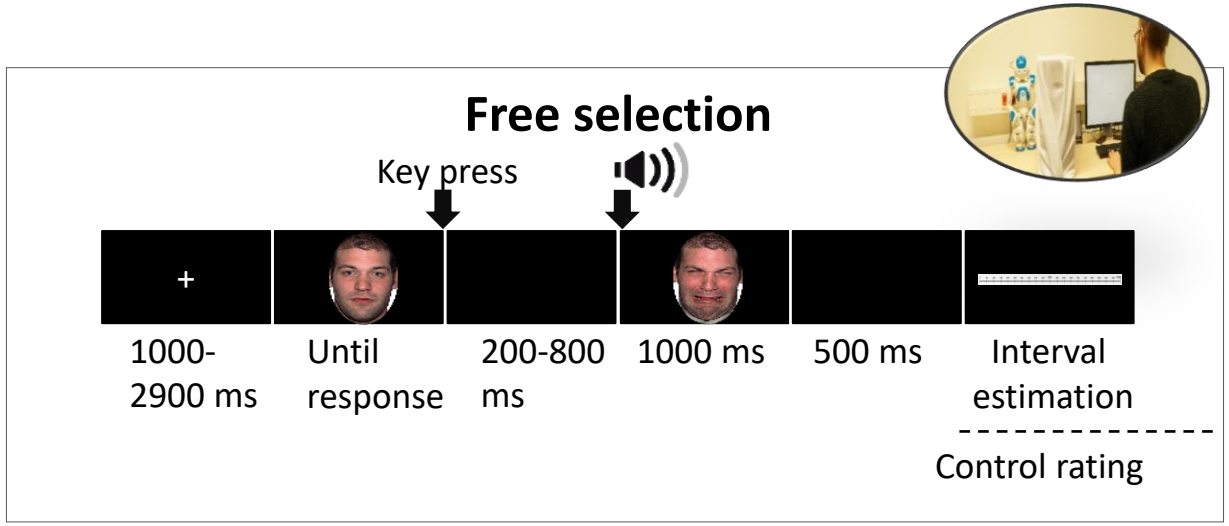
IVs	DVs
1. Choice (blocked) - Free (right or left key press) - Human-instructed - Robot-instructed - Passive **2. Action-outcome delay (mixed)** 200, 500, 800 (ms) **3. Outcome valence (mixed)** Neural, positive, negative **4. Robot autonomy (between-subjects)** -Autonomous (n=24) -Non-autonomous (n=24)	1. Interval estimations 2. Judgment of control ratings 3. Post experiment questionnaire

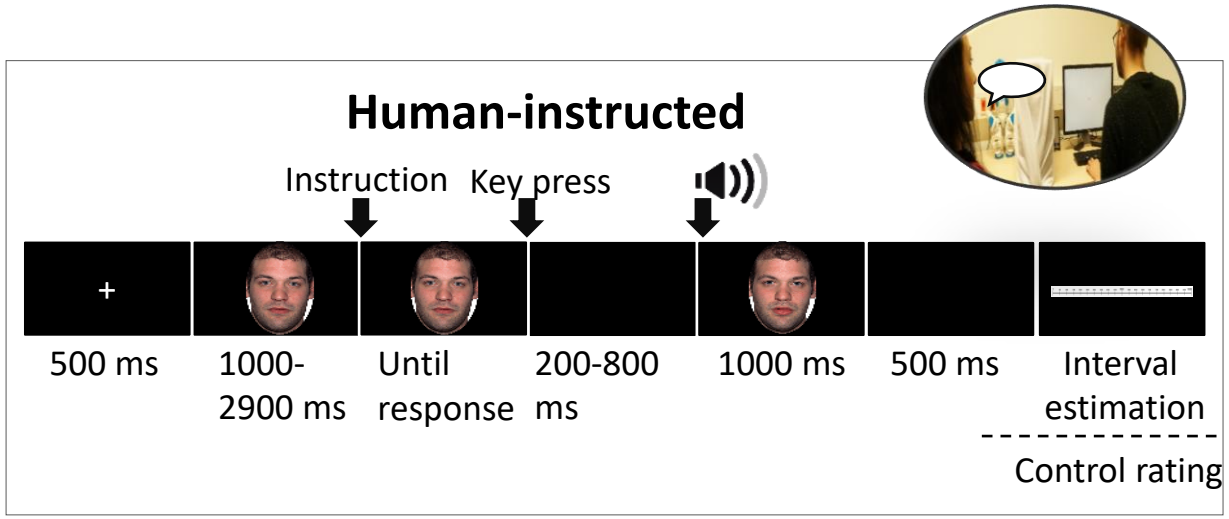
EXPERIMENT 2: Changes in robot description

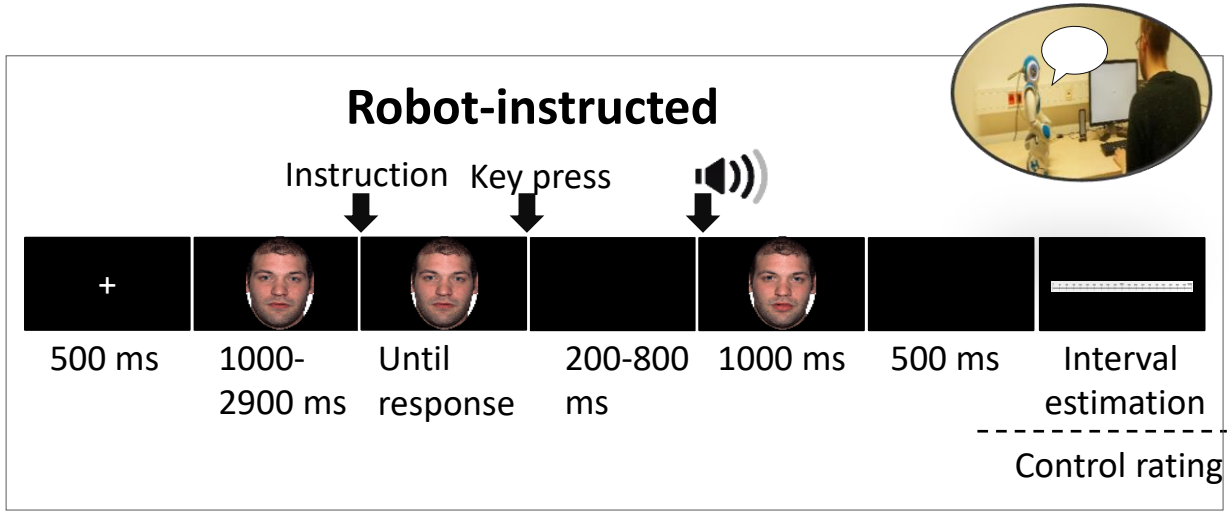
- **Autonomous :** “Zora is an autonomous robot capable of making its own decisions, and when Zora tells you which key to press, it knows how pressing that key would change the expression on the face.“
- **Non-Autonomous:** “Zora’s key press instructions were pre-programmed and will simply tell you which key to press.”

EXPERIMENT 2: PROCEDURE

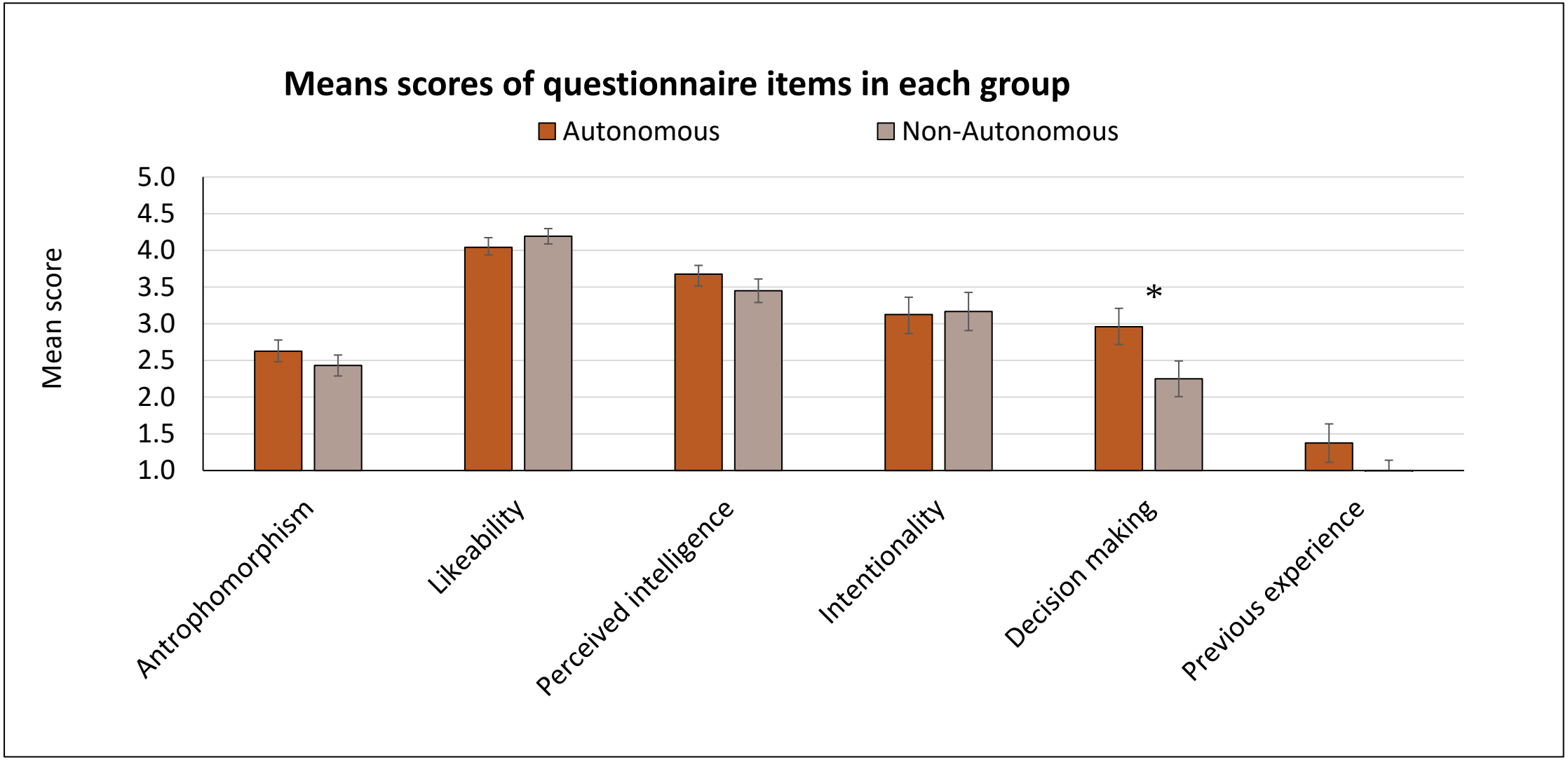






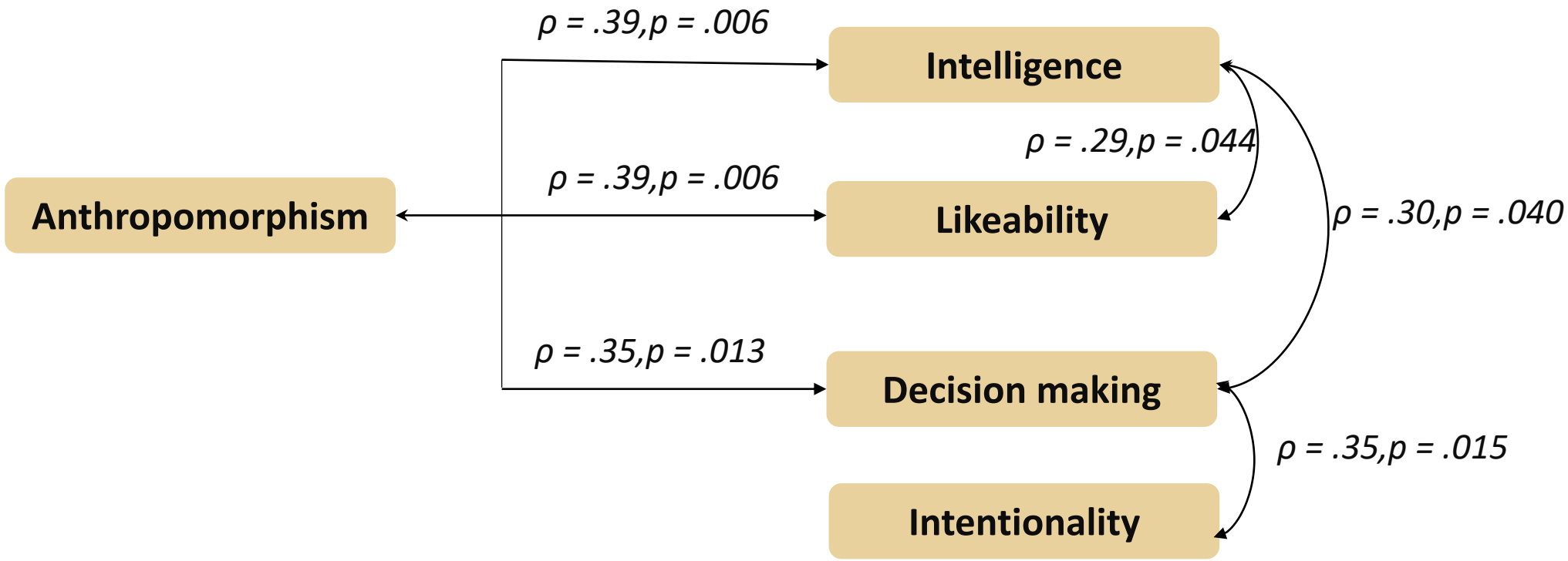


EXPERIMENT 2: RESULTS



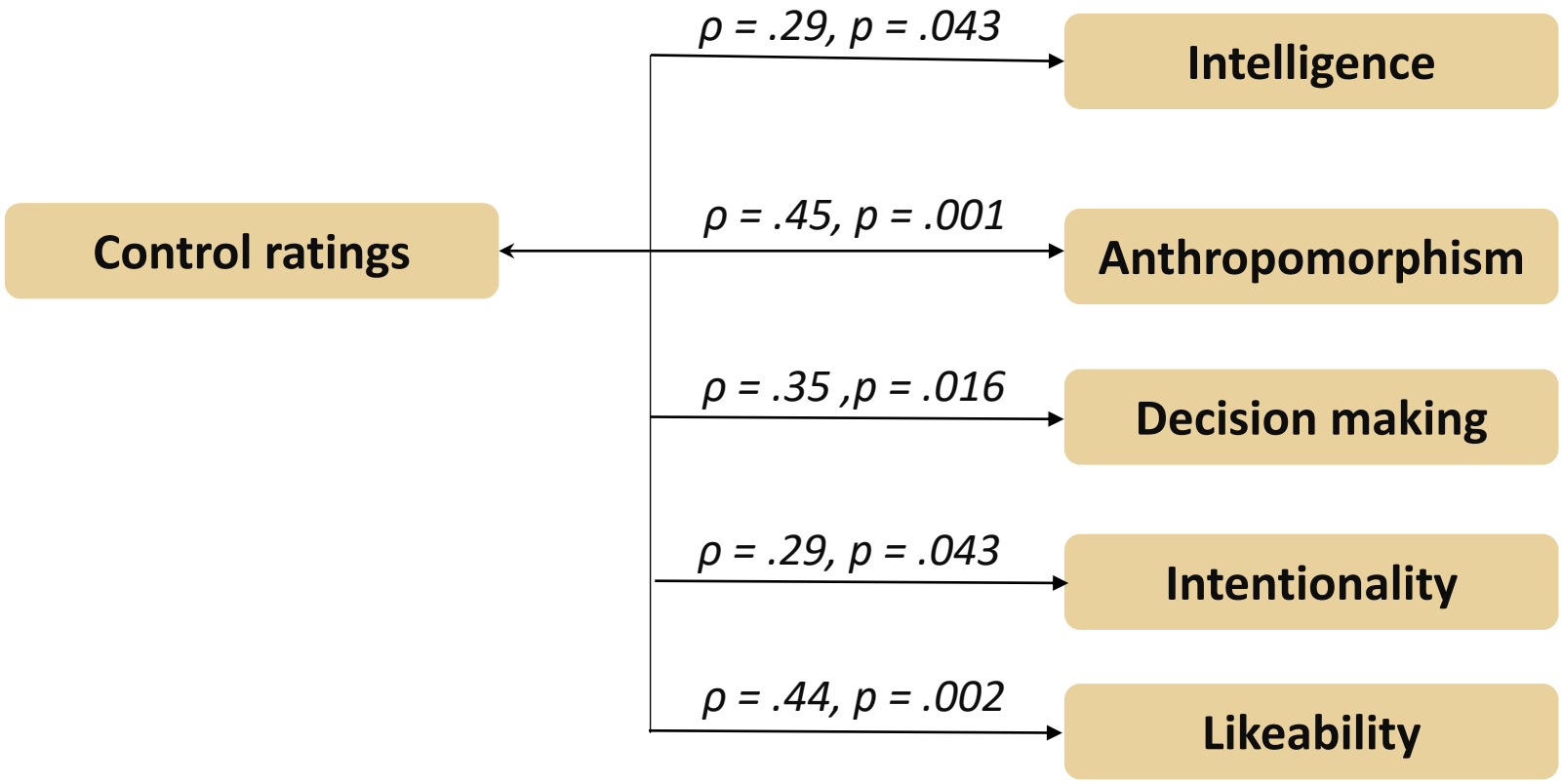
EXPERIMENT 2: RESULTS

- **Relationship among the questionnaire items**

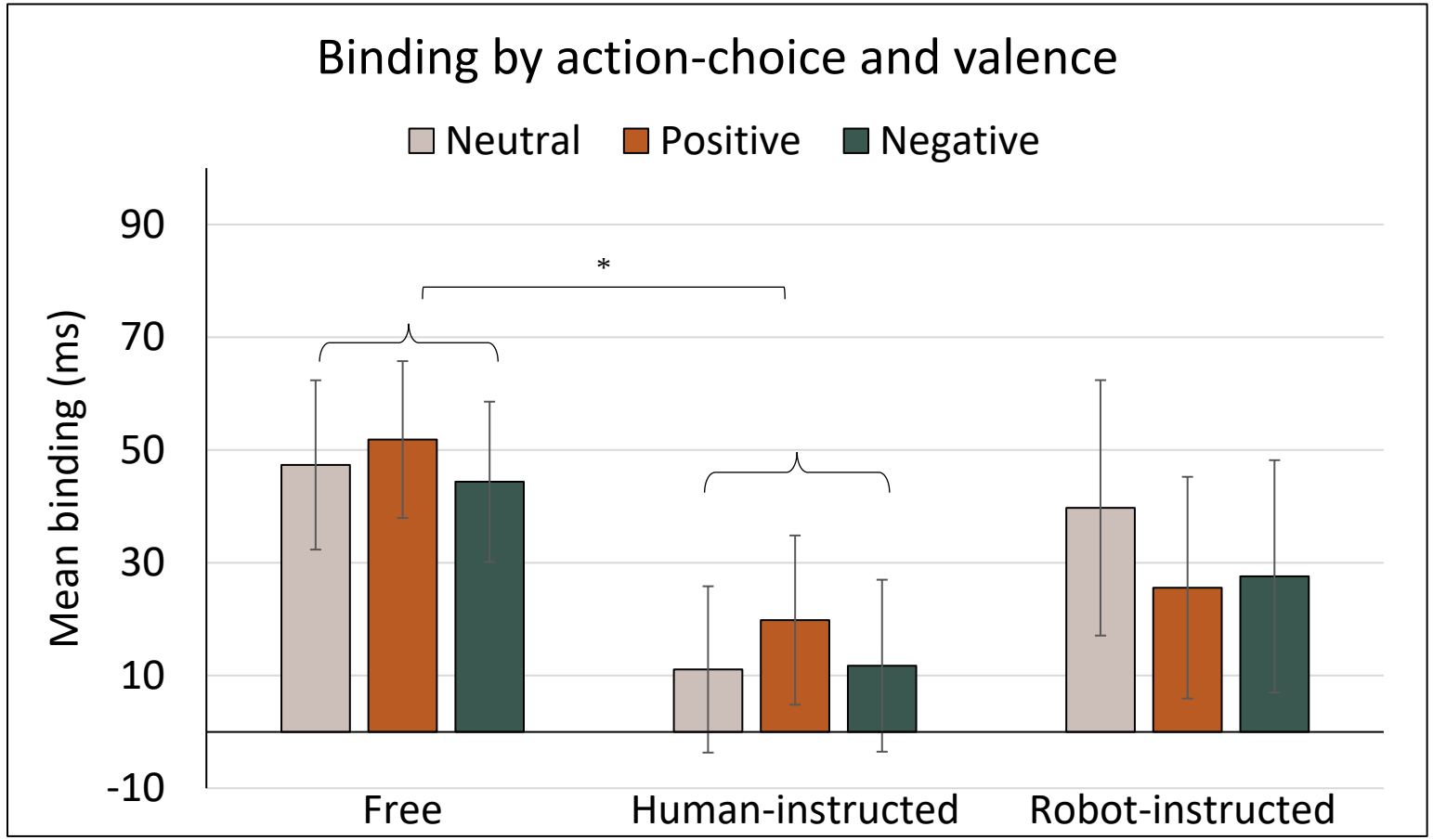


EXPERIMENT 2: RESULTS

- **Relationship between control ratings and the questionnaire items (robot-instructed condition)**

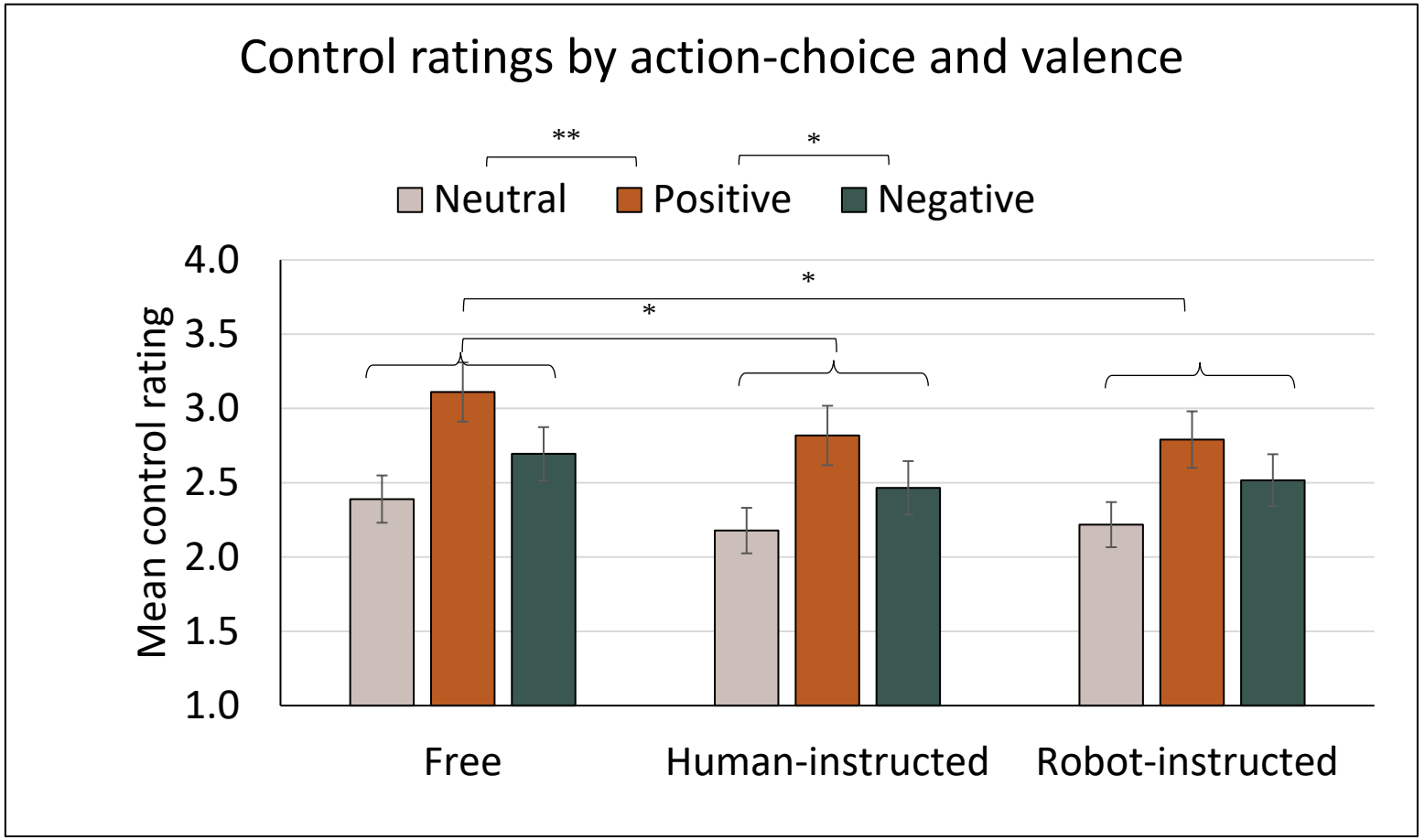


EXPERIMENT 2: RESULTS



Choice: $p = .006$, $\eta_p^2 = .11$

EXPERIMENT 2: RESULTS



Choice: p = .020, η_p^2 = .09 **Valence:** p = .001, η_p^2 = .16

EXPERIMENT 2: RECAP

- **Free vs. instructed actions:**
  - Freedom to choose one's actions yield stronger SoA (implicit & explicit) compared to instructed actions (regardless of the outcomes).
- **Identity of the instructor**
  - No difference in SoA between human vs. robot instructed actions.
- **The effect of outcome valence** Positive outcomes enhance SoA compared to negative and neutral outcomes (explicit SoA, retrospective effect).
- **Correlation between anthropomorphism and explicit SoA**
  - The more human-like the robot is perceived, the greater explicit SoA in robot instructed actions.

FUTURE DIRECTIONS IN HRI/HCI

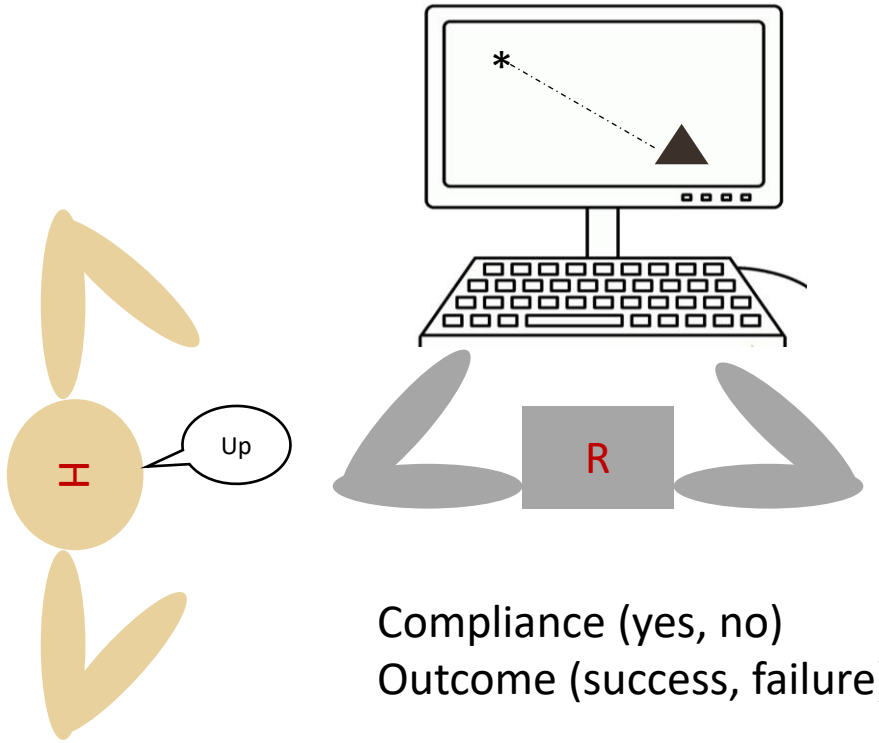
- Aim to establish the experimental scenarios depicting a wider range of use of robots in industrial production processes and aviation technology.
- The deployment of robots—and AI technology in general—has become more frequent in healthcare, education, and industry.
- Important to understand:
  - how human experience (**SoA, responsibility, and trust)** is affected in interactions with robots,
  - how **moral and ethical** considerations are to be implemented in these interactions,
  - and the issue of **accountability** in the case of adverse outcomes that may emerge out of human-robot joint tasks.

FUTURE DIRECTIONS IN HRI/HCI

IVs		DVs
Artificial agent (AA)	Contextual/Task relevant factors	
• Perceived autonomy • Human-like features • Obedience to instructions • Cooperativeness • Efficiency	• Task load: Equally shared vs. weighted • Task difficulty • Introducing disruption to human/AA actions • Outcome: Success vs failure	• SoA • Responsibility • Trust

FUTURE DIRECTIONS IN HRI/HCI

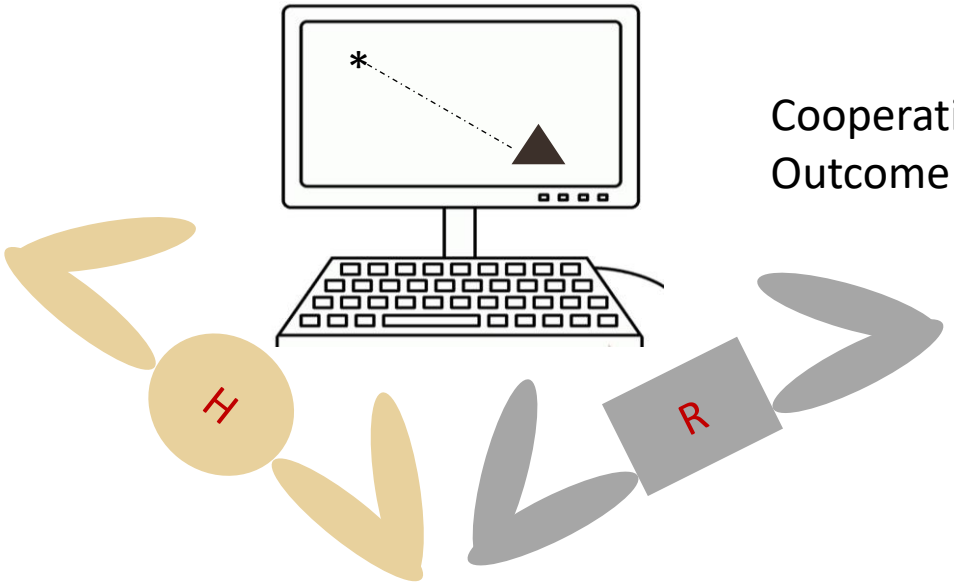
- Experiment-1 aims to examine the effect of robot-compliance and autonomy on the human partner's SoA, responsibility, and trust in their robot partner



Compliance (yes, no)
Outcome (success, failure)

FUTURE DIRECTIONS IN HRI/HCI

- Experiment-2 examine the role of robot cooperativeness and efficiency. The team performs a similar task in which the co-agents take turns to move the object towards the target.



Cooperativeness (correcting, committing error)
Outcome (success, failure)

Trajectory of human movements distorted

Compensate vs. commit another error

QUESTIONS?

THANK YOU!