THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Signaling Human Intentions to Service Robots: Understanding the Use of Social Cues during In-Person Conversations

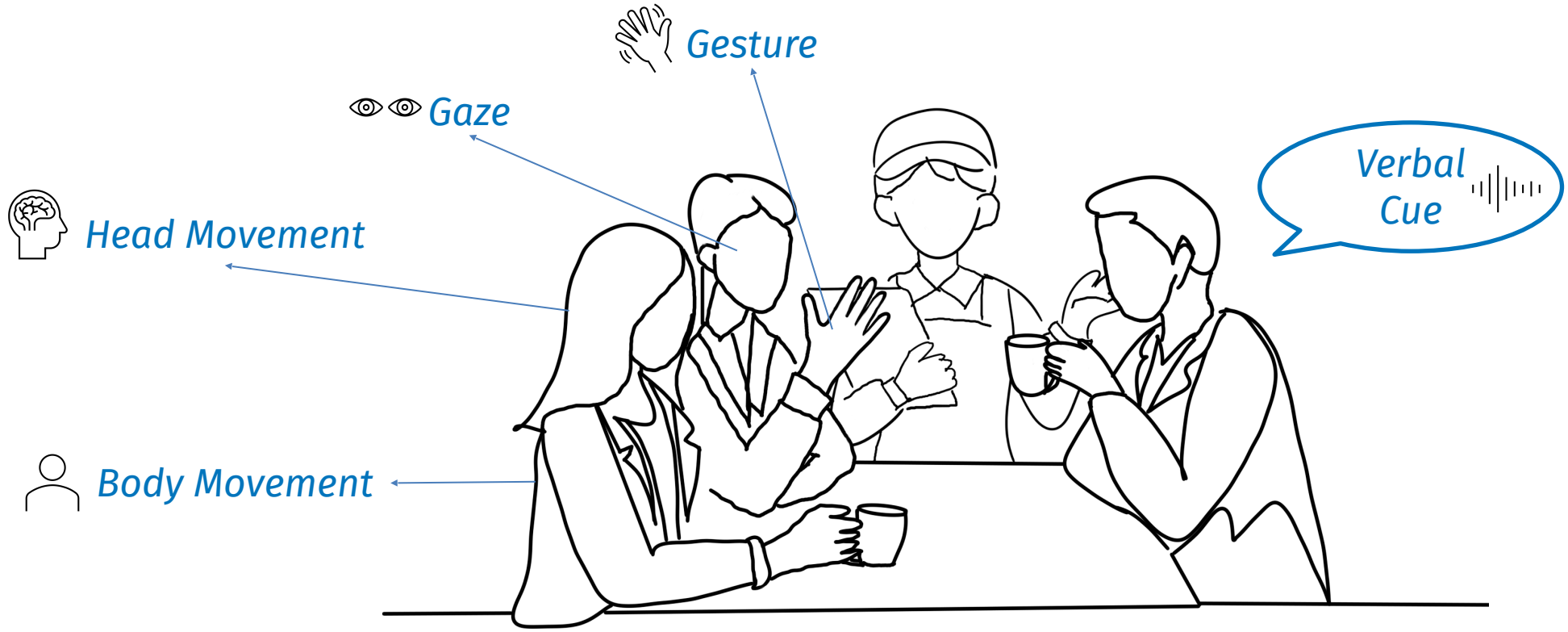
Hanfang Lyu[1], Xiaoyu Wang[1], Nandi Zhang[2], Shuai Ma[1], Qian Zhu[3], Yuhan Luo[4], Fugee Tsung[1,5], and Xiaojuan Ma[1].

[1]The Hong Kong University of Science and Technology, [2]University of Calgary, [3]Renmin University of China, [4]City University of Hong Kong, [5]The Hong Kong University of Science and Technology (Guangzhou)

CHI'25, April 26 – May 1, 2025, Yokohama, Japan

Human use different social cues every day

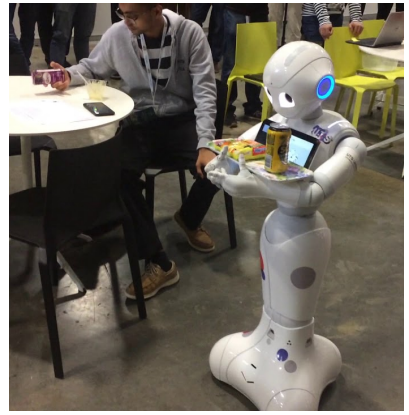
During in-person conversations, people naturally employ different modalities of social cues to communicate their intentions.



Service Robot in Social Settings

- Service robots in retail, healthcare, hospitality, etc.
- Serving during *human-human interactions* (HHI)

Aliengo
https://youtu.be/CaH-SSKAIe4?si=ZH0gik_RAIhWV11K&t=7



Pepper
https://youtu.be/KzwHOufMT_A?si=uC1XoMWu7SqLdh8b&t=56

Drone
https://www.techinasia.com/singapore-restaurant-autonomous-drone-waiters

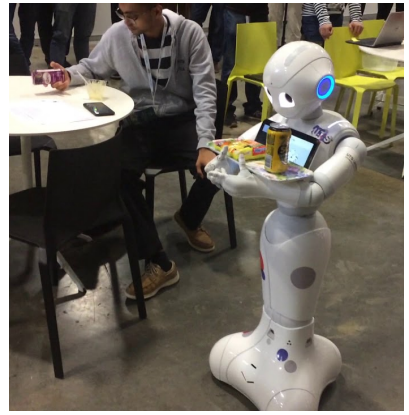
BellaBot
https://youtu.be/_kDj3PEm6aw?si=SIFz7jsa69xu114w&t=22

Service Robot in Social Settings

- Service robots in retail, healthcare, hospitality, etc.
- Serving during *human-human interactions* (HHI)

Understanding how human signal intentions to different forms of service robots, when primarily engaged in an important social encounter (HHI)

Aliengo
https://youtu.be/CaH-SSKAIe4?si=ZH0gik_RAIhWV11K&t=7

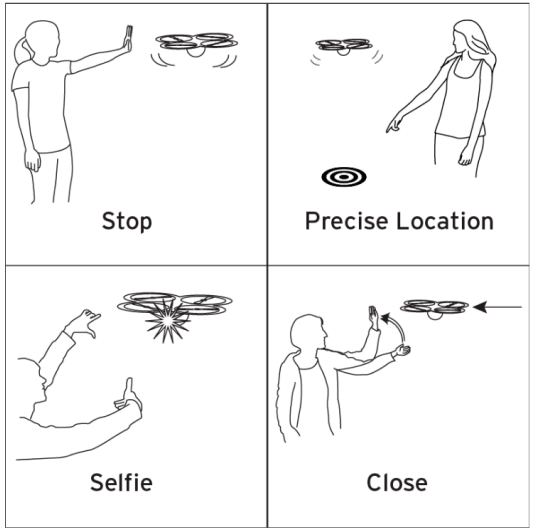


Pepper
https://youtu.be/KzwHOufMT_A?si=uC1XoMWu7SqLdh8b&t=56

Drone
https://www.techinasia.com/singapore-restaurant-autonomous-drone-waiters

BellaBot
https://youtu.be/_kDj3PEm6aw?si=SIFz7jsa69xu114w&t=22

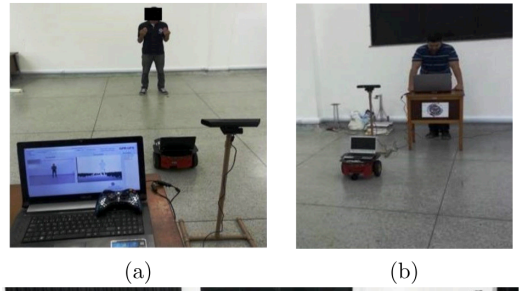
Social Cues Understanding in HRI



Gesture and sound, drone
(Cauchard et al., 2015)

Multi-modality, wearable robotic arm
(Muehlhaus et al., 2023)

Gesture, drone
(Firestone et al., 2019)



Gesture, mobile robot
(Canuto et al., 2022)

Lack a systematic understanding of human *social cue* interaction with *social service robots*

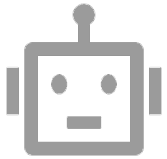
- Including a wide range of social cue **modalities**
- Comparing multiple **robot forms**
- From lab-based settings to **real-world** social scenarios

Research Questions

How humans choose and combine different modalities of social cues to signal different intentions to **a service robot** during **an engaging social encounter**?

Research Questions

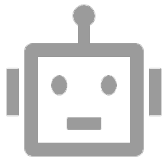
How humans choose and combine different modalities of social cues to signal different intentions to **a service robot** during **an engaging social encounter**?



How different **robot morphologies** influence the use of social cues?

Research Questions

How humans choose and combine different modalities of social cues to signal different intentions to **a service robot** during **an engaging social encounter**?



How different **robot morphologies** influence the use of social cues?

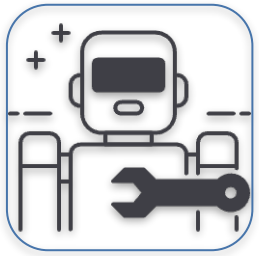
How different **human engagements in the HHI** influence the use of social cues?

Contribution

We contribute insights from An Elicitation Study on human signaling robot waiters during an important coffee chat to better understand

Usage of social cues

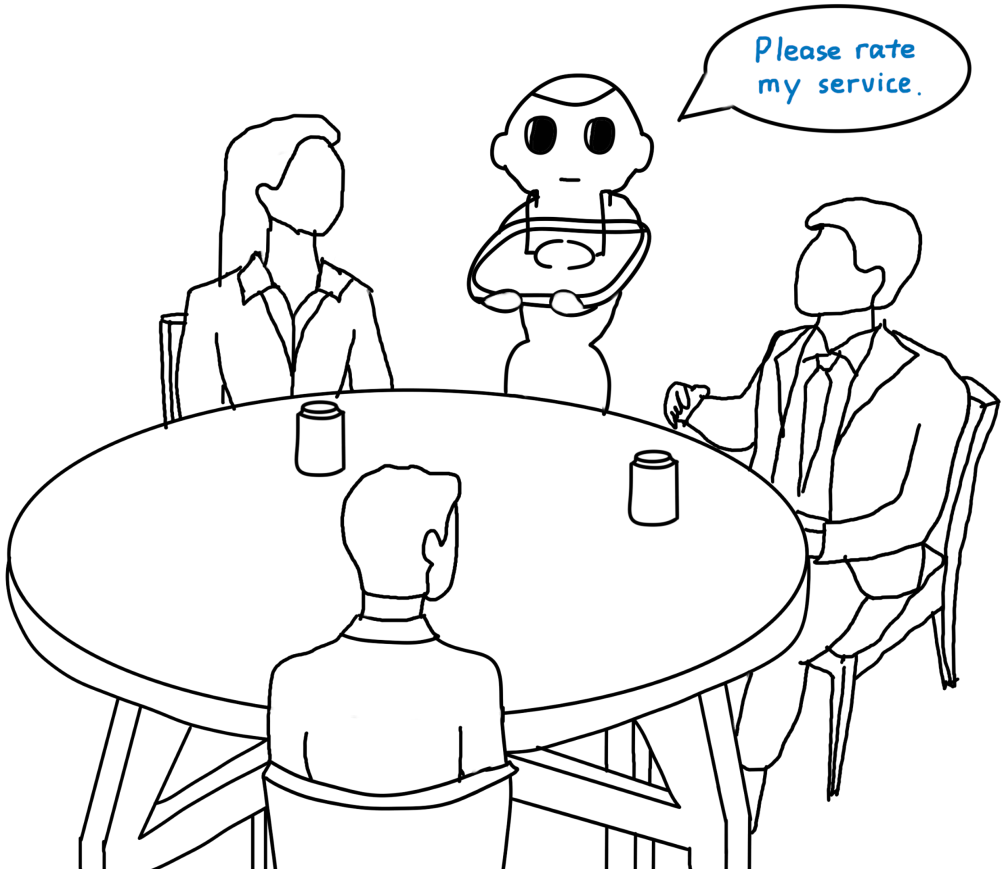
Influencing factors and rationales



Design implications

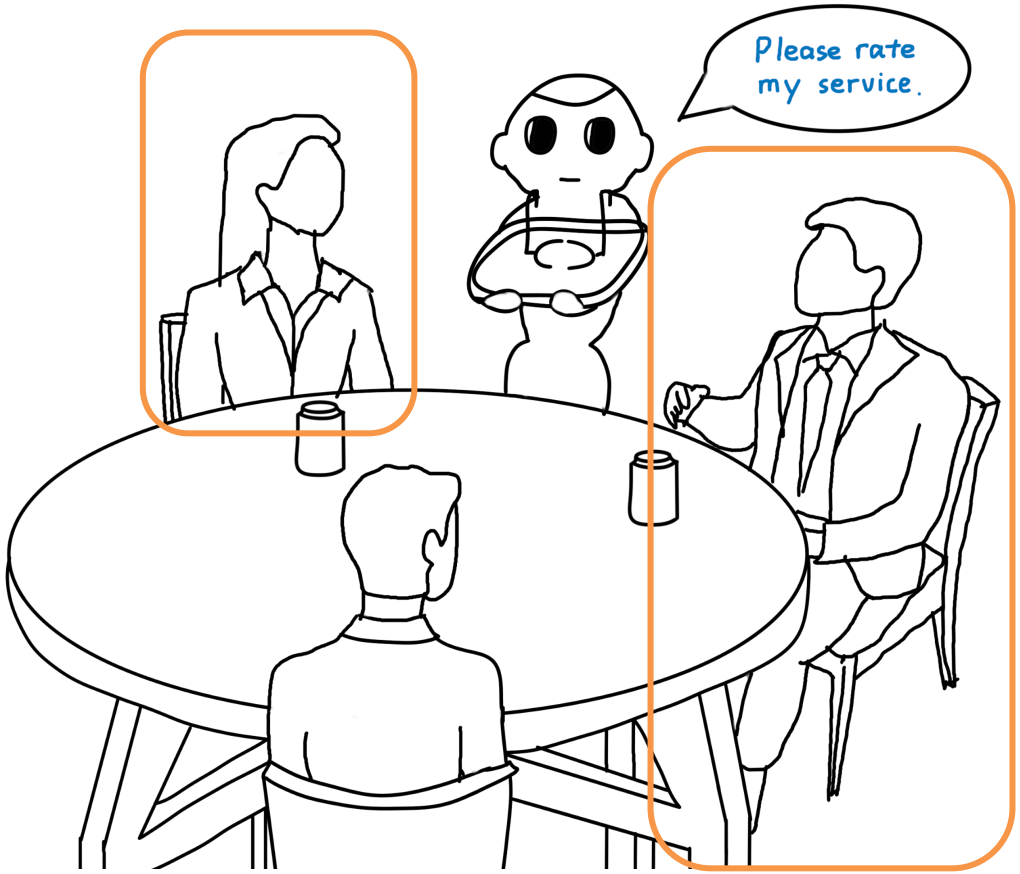
Experiment Simulation: Social Encounter

- Coffee chat with *potential employers*



Experiment Simulation: Social Encounter

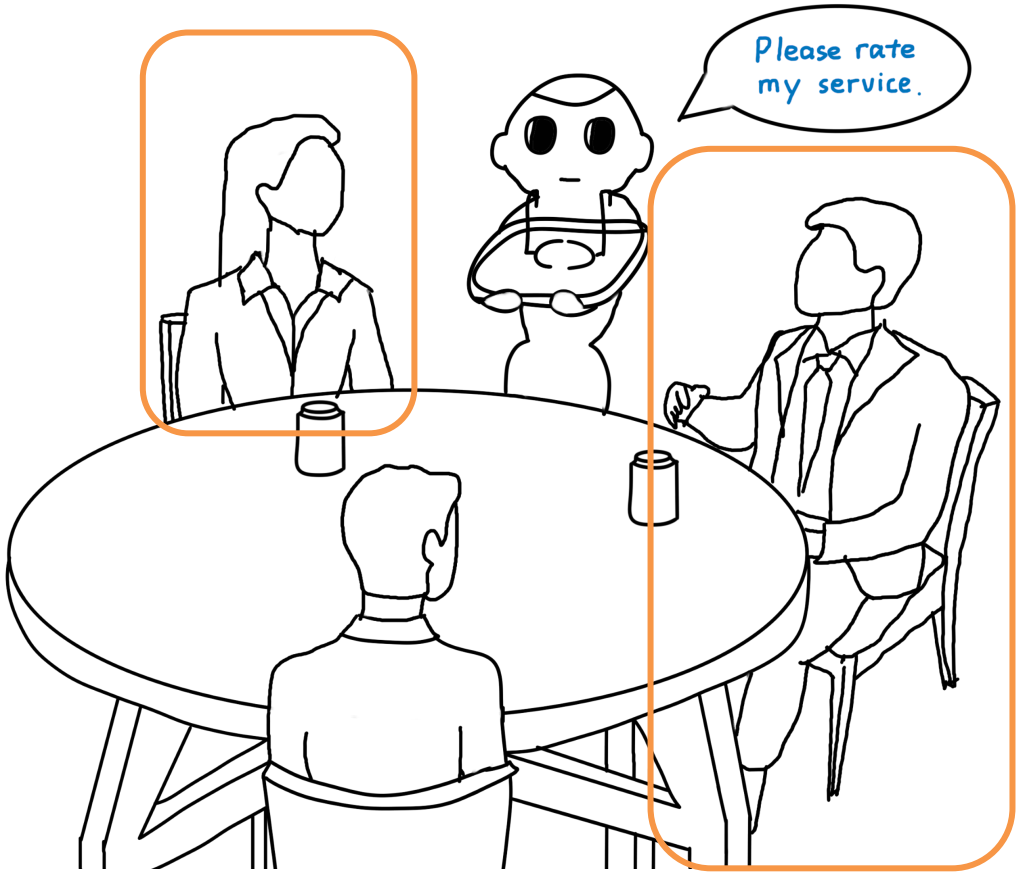
- Coffee chat with *potential employers*



Two actors trained to play the roles of two potential employers

Experiment Simulation: Social Encounter

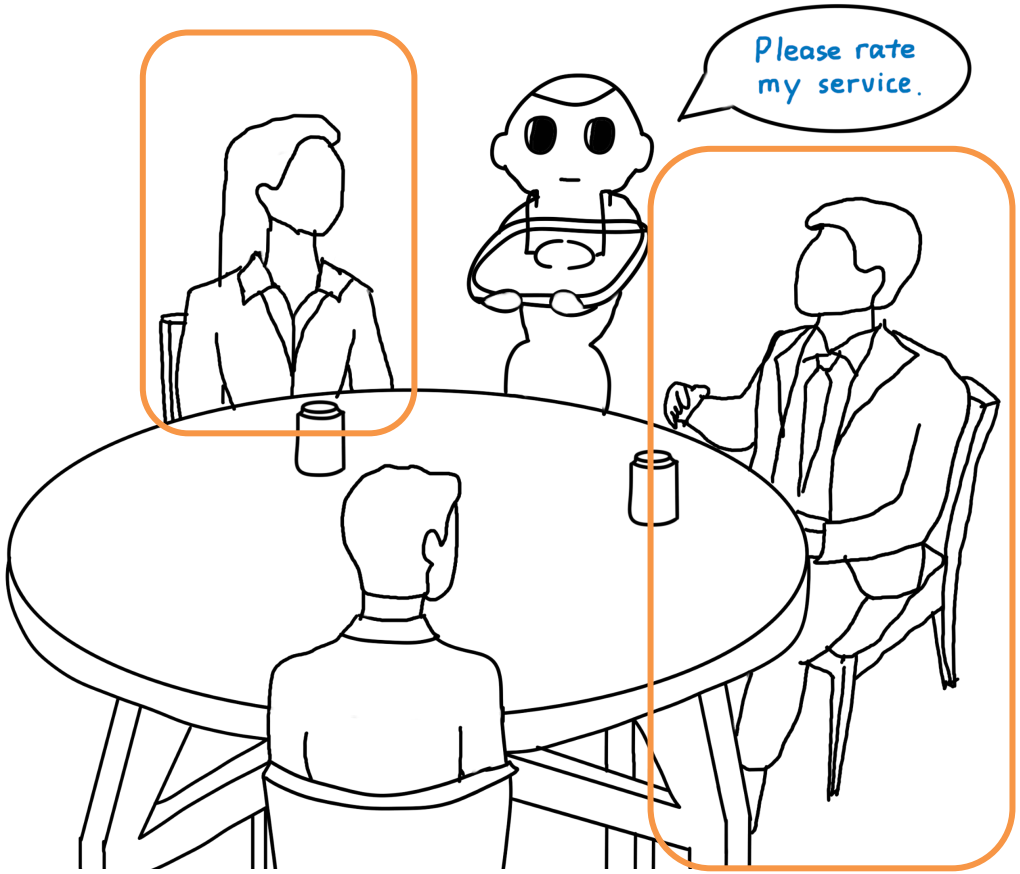
- Coffee chat with *potential employers*
- ***ConversationRole***



Two actors trained to play the roles of two potential employers

Experiment Simulation: Social Encounter

- Coffee chat with *potential employers*
- ***ConversationRole***
  - *speaker* (participants lead the conversational flow)

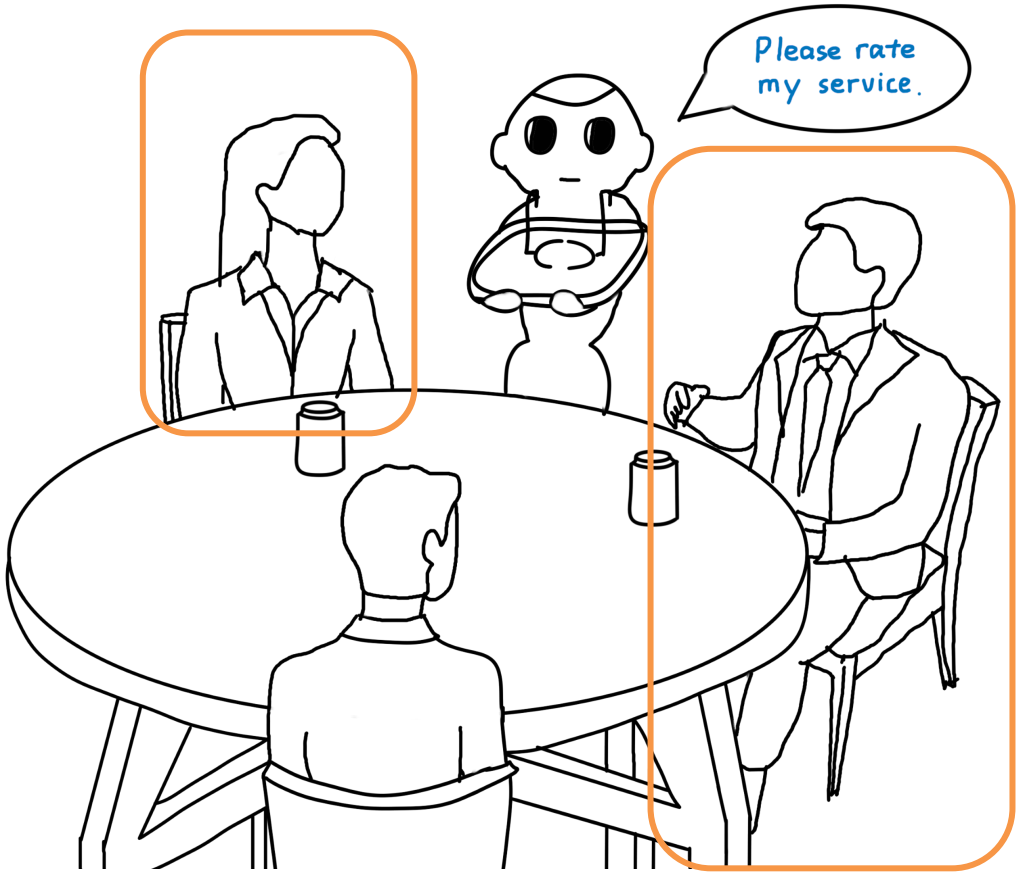


Two actors trained to play the roles of two potential employers

Experiment Simulation: Social Encounter

- Coffee chat with *potential employers*
- ***ConversationRole***
  - *speaker* (participants lead the conversational flow)
  - *listener* (participants listen to the conversation lead by others)

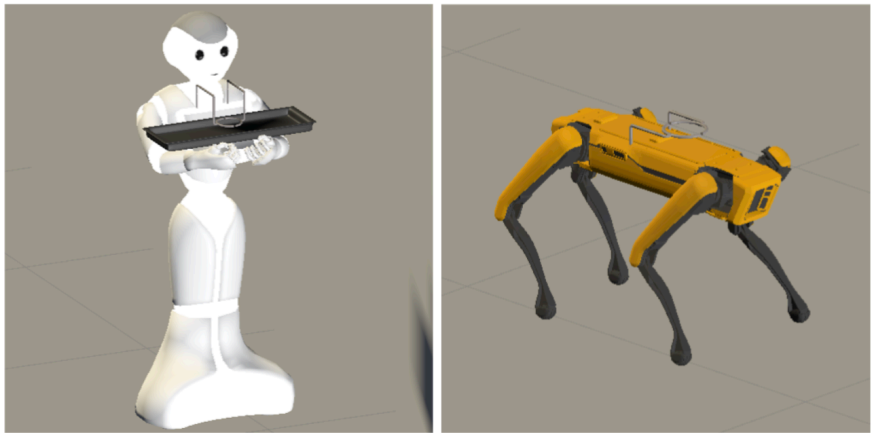
Two actors trained to play the roles of two potential employers



Experiment Simulation: AR Simulated Robot

- ***RobotMorphology*** (Onnasch and Roesler, 2021)

Experiment Simulation: AR Simulated Robot



(a) Anthropomorphic Robot

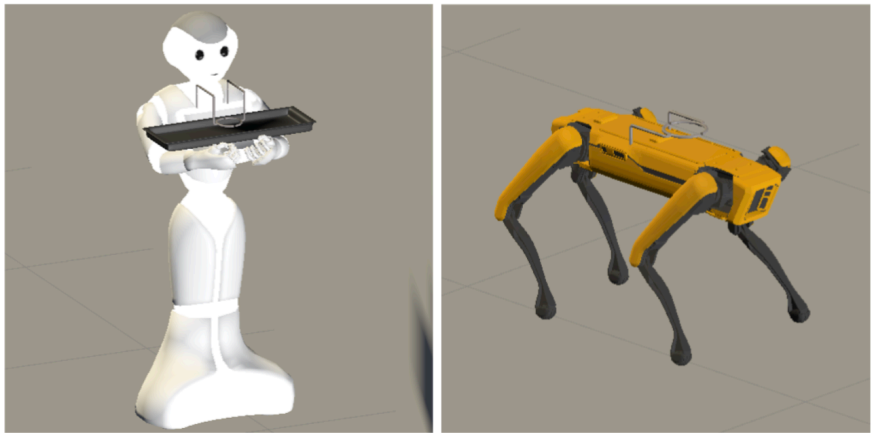
(b) Zoomorphic Robot

(c) Grounded Technical Robot

(d) Aerial Technical Robot

- ***RobotMorphology*** (Onnasch and Roesler, 2021)

Experiment Simulation: AR Simulated Robot



(a) Anthropomorphic Robot

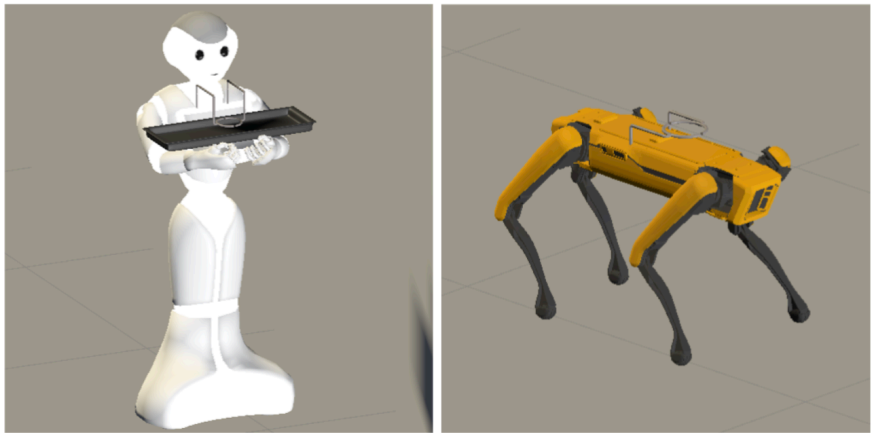
(b) Zoomorphic Robot

(c) Grounded Technical Robot

(d) Aerial Technical Robot

- ***RobotMorphology*** (Onnasch and Roesler, 2021)
  - *Anthropomorphic* → Pepper

Experiment Simulation: AR Simulated Robot



(a) Anthropomorphic Robot

(b) Zoomorphic Robot

(c) Grounded Technical Robot

(d) Aerial Technical Robot

- ***RobotMorphology*** (Onnasch and Roesler, 2021)
  - *Anthropomorphic* → Pepper
  - *Zoomorphic* → Spot

Experiment Simulation: AR Simulated Robot

(a) Anthropomorphic Robot

(b) Zoomorphic Robot

(c) Grounded Technical Robot

(d) Aerial Technical Robot

- ***RobotMorphology*** (Onnasch and Roesler, 2021)
  - *Anthropomorphic* → Pepper
  - *Zoomorphic* → Spot
  - *Grounded Technical* → A PuduBot2-like restaurant delivery robot

Experiment Simulation: AR Simulated Robot

(a) Anthropomorphic Robot

(b) Zoomorphic Robot

(c) Grounded Technical Robot

(d) Aerial Technical Robot

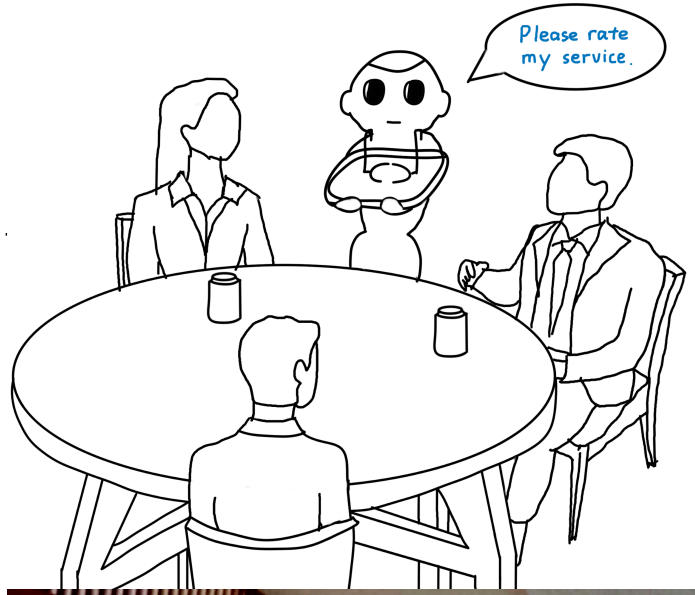
- ***RobotMorphology*** (Onnasch and Roesler, 2021)
  - *Anthropomorphic* → Pepper
  - *Zoomorphic* → Spot
  - *Grounded Technical* → A PuduBot2-like restaurant delivery robot
  - *Aerial Technical* → A typical delivery drone

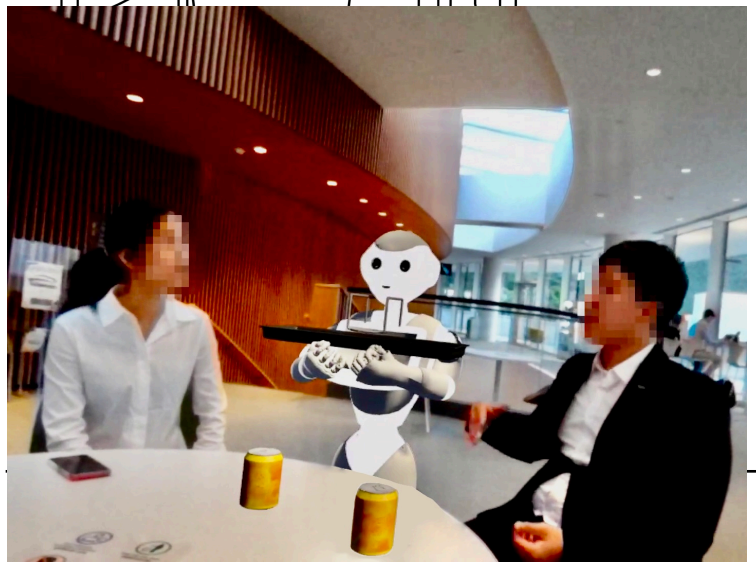
Elicitation Referents

5 Necessary HRI Types in Social Settings -> 13 interaction referents in the scenario

Robot *Active* Seeking for Human Input	i. When the robot is not sure	
	ii. When the robot asks for evaluation	
	iii. When the human signals awareness	
Robot *Passive* Receiving Human Input	When the robot has an error	iv. Performance error
		v. Social error

Elicitation Referents





ocial Settings -> 13 interaction referents in the scenario

i. When the robot is not sure

ii. When the robot asks for evaluation

iii. When the human signals awareness

When the robot has an error

iv. Performance error

v. Social error

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic 12

zoomorphic 12

grounded technical 12

aerial technical 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic 12

zoomorphic 12

grounded technical 12

aerial technical 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic 12

zoomorphic 12

grounded technical 12

aerial technical 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic 12

zoomorphic 12

grounded technical 12

aerial technical 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic 12

zoomorphic 12

grounded technical 12

aerial technical 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic — 12

zoomorphic — 12

grounded technical — 12

aerial technical — 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic 12

zoomorphic 12

grounded technical 12

aerial technical 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic — 12

zoomorphic — 12

grounded technical — 12

aerial technical — 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Elicitation Study

ConversationRole:

listener + speaker

24

Each participant

2 sessions:

- ***RobotMorphology***
  (mixed-designed variable, counter-balanced)
- ***ConversationRole***
  (within-subject variable)

× 13 ***Referent***

RobotMorphology:

anthropomorphic 12

zoomorphic 12

grounded technical 12

aerial technical 12

- retrospective think-aloud
  → *explicit* cues
- semi-structured interview

× 24 participants
× 2 sessions
× 13 referents
= 624 social cues

Data Analysis: social cue coding

1. Iterative social cue coding: by **Modality** (gesture, verbal, eye, head, upper body, leg and foot)

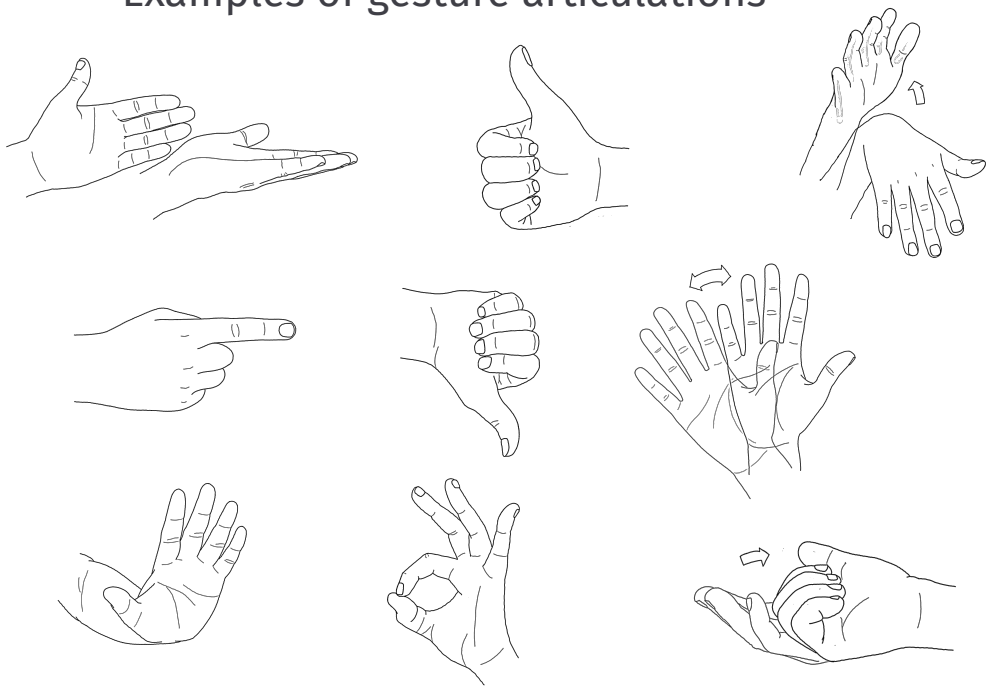
Data Analysis: social cue coding

1. Iterative social cue coding: by **Modality** (gesture, verbal, eye, head, upper body, leg and foot)

   - Articulation codes: social cue elements

3387

   - Verbal: *exact content, speech act*
   - Examples of gesture articulations



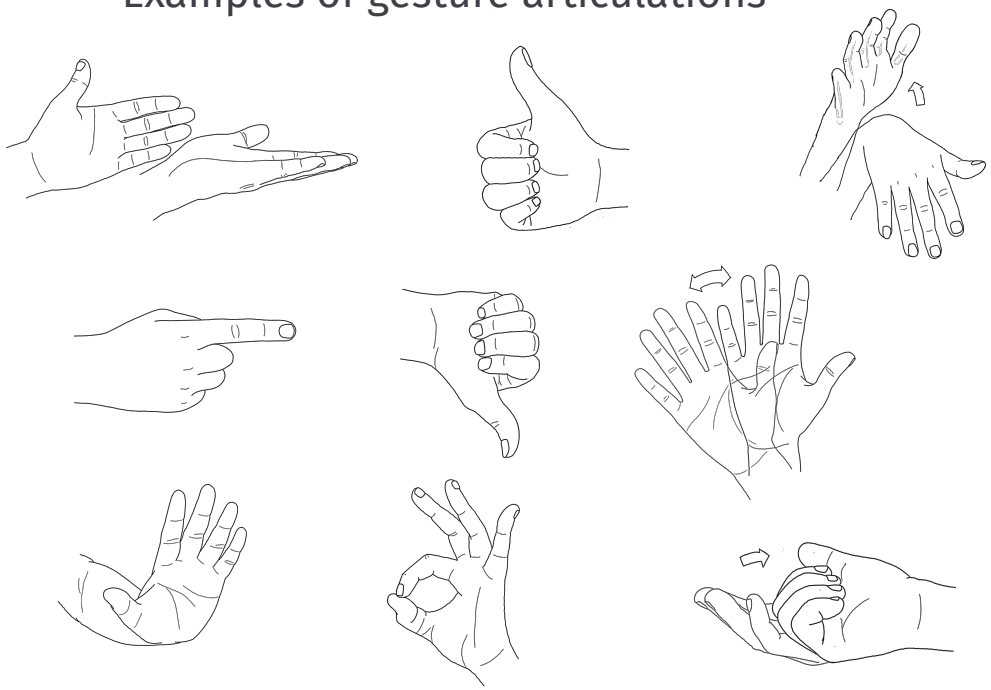
Data Analysis: social cue coding

1. Iterative social cue coding: by **Modality** (gesture, verbal, eye, head, upper body, leg and foot)

- Articulation codes: social cue elements

3387

- Verbal: *exact content, speech act*
- Examples of gesture articulations



- Feature codes: social cue characteristics
  - Gesture 3318
    - *number of hands*
    - *hand height*
    - *repetition*
  - Verbal 823
    - *volume*
    - *unclear reference*
    - *politeness*

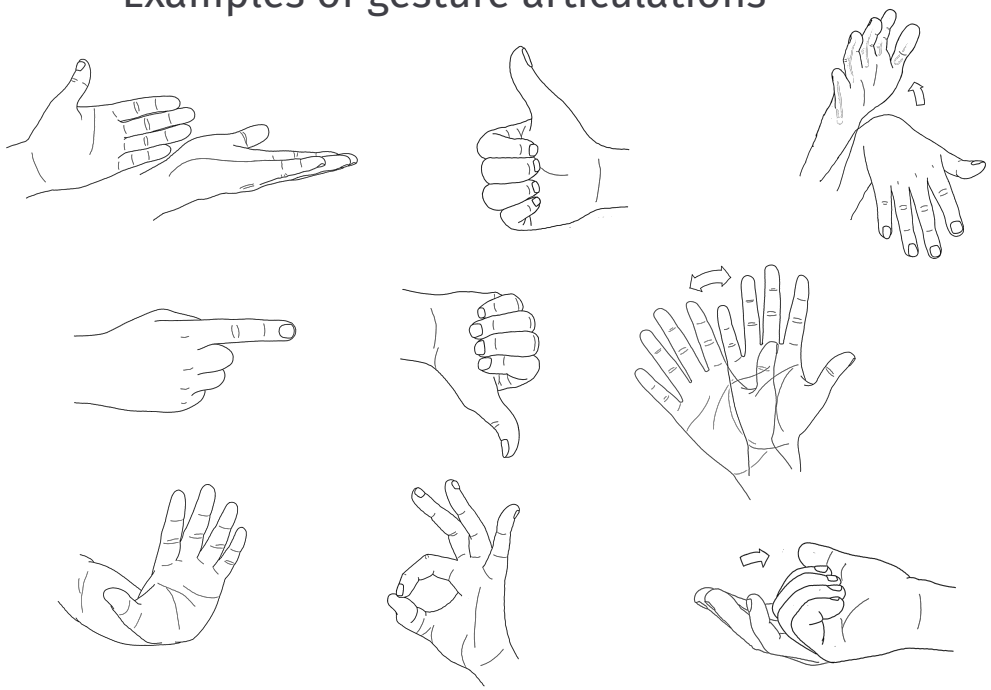
Data Analysis: social cue coding

1. Iterative social cue coding: by **Modality** (gesture, verbal, eye, head, upper body, leg and foot)

- Articulation codes: social cue elements

3387

- Verbal: *exact content, speech act*
- Examples of gesture articulations



- Feature codes: social cue characteristics
  - Gesture 3318
    - *number of hands*
    - *hand height*
    - *repetition*
  - Verbal 823
    - *volume*
    - *unclear reference*
    - *politeness*

Explicit Cues: reported during interview
Implicit Cues: observed but not reported
Observed Cues: Explicit + Implicit Cues

Data Analysis: agreement, statistical and interview

2. Agreement Rate

$$AR(r) = \frac{|P|}{|P| - 1} \sum_{P_i \in P} \left(\frac{|P_i|}{P} \right)^2 - \frac{1}{|P| - 1}$$

- $AR \leq 0.1$ for low agreement
- $0.1 < AR \leq 0.3$ for medium agreement
- $0.3 < AR \leq 0.5$ for high agreement
- $AR > 0.5$ for very high agreement

3. Cumulative Link Mixed Model:

forward selection of the following terms

$$Y(Modality\ or\ Feature\ Codes) \sim$$

$$ConversationRole + RobotMorphology$$

$$+ ConversationRole \times RobotMorphology$$

(with *participants nested within Referent* treated as random effect)

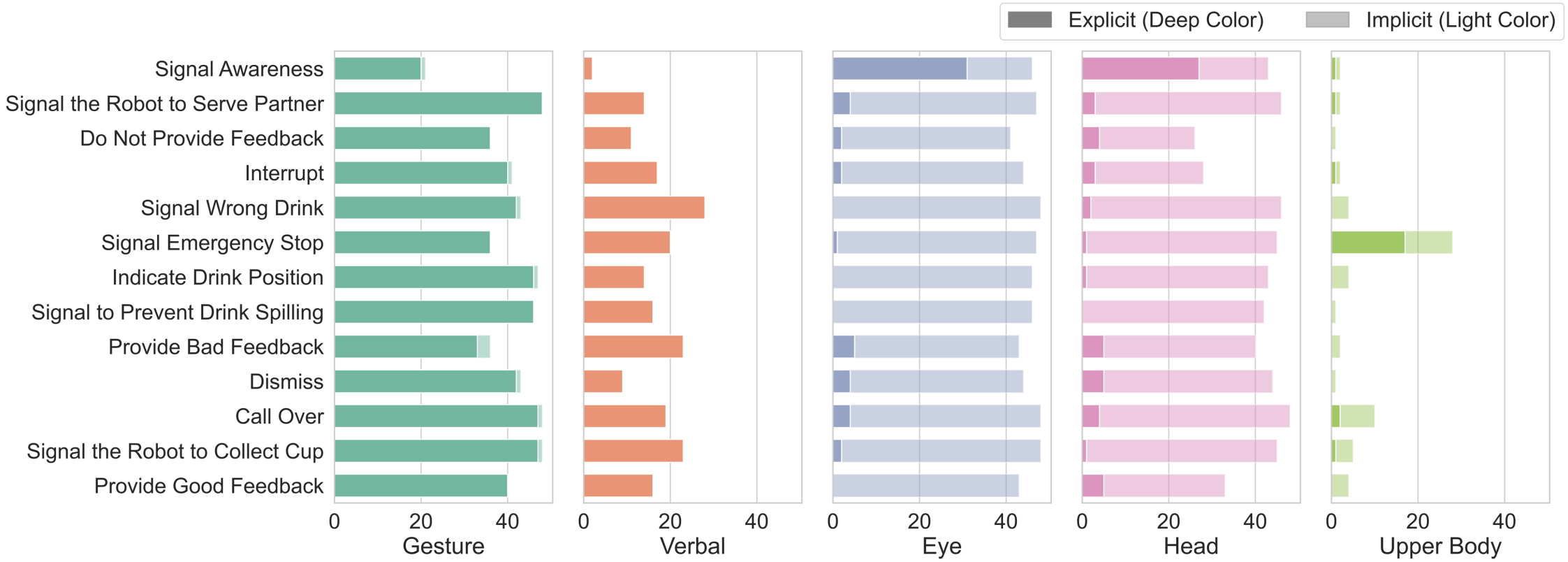
4. Interview analysis: Thematic and iterative analysis

Findings and Discussions

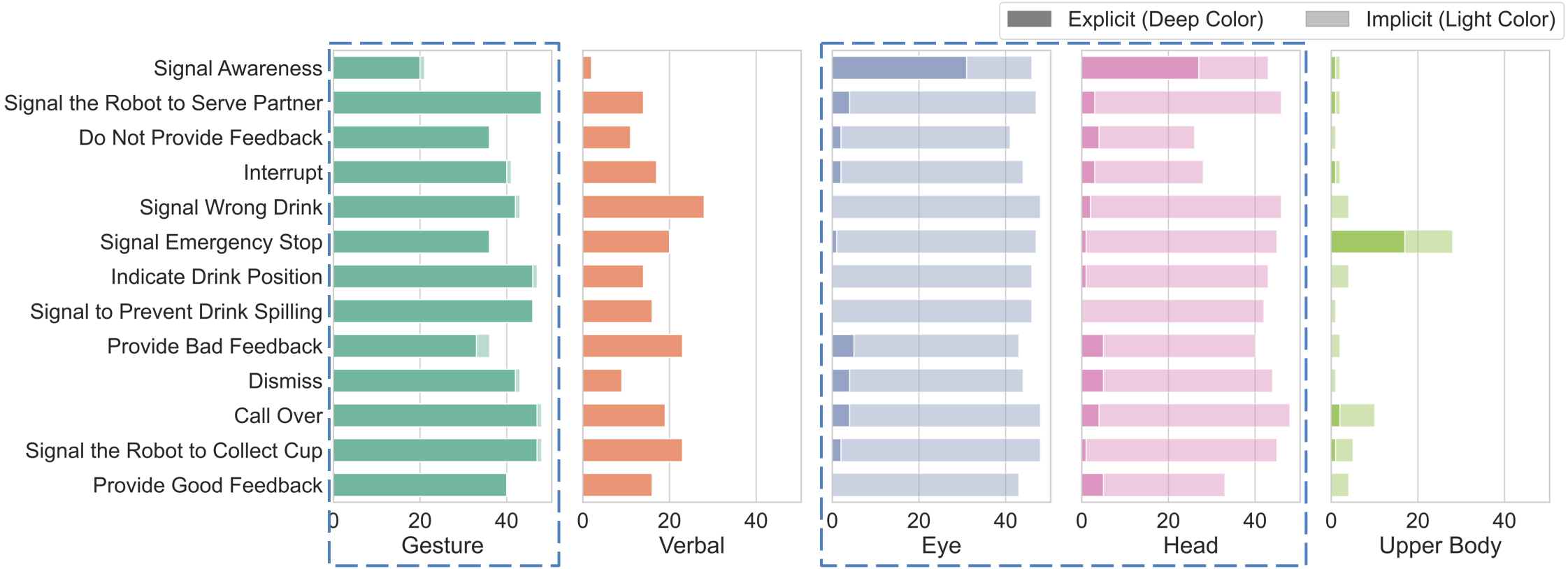
- Social Cue Patterns and Semantics
- Perception on Robots Shapes Interaction Behaviors
  - Perceived Capabilities
  - Perceived Safety
  - Perceived Intelligence
- Primary Task and Social Context Matters

Check our paper for analysis details

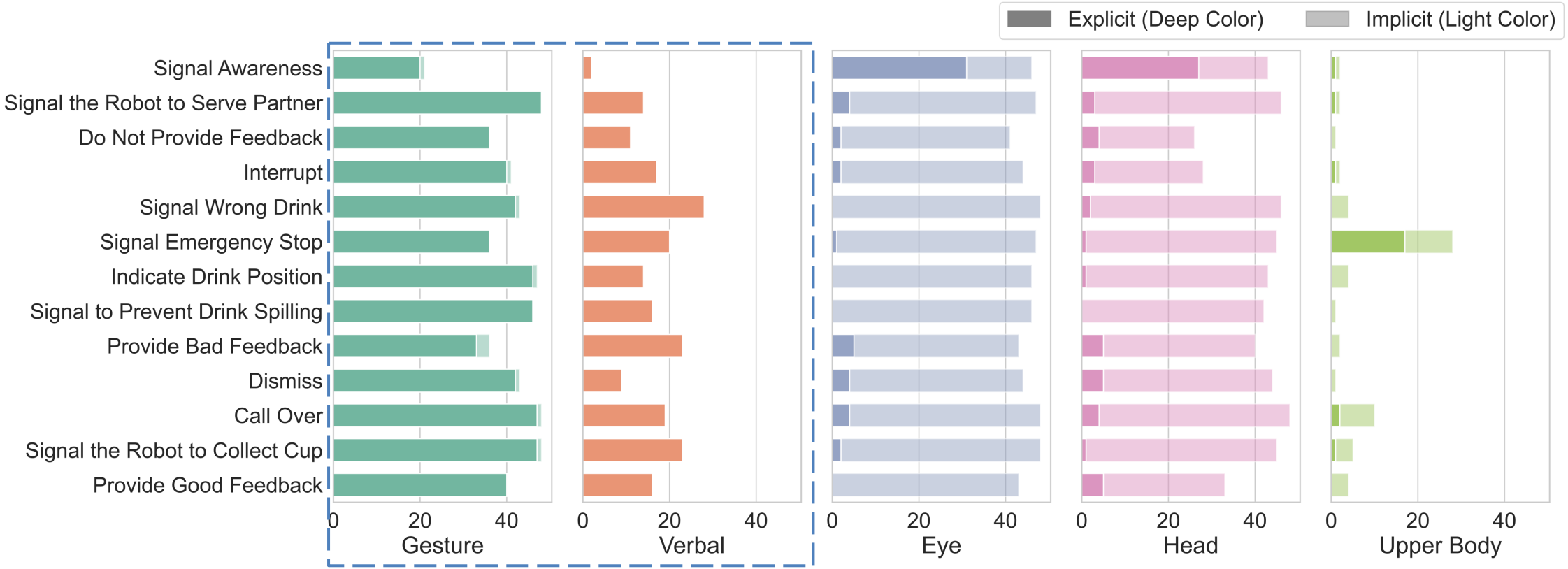
Modality Distribution by Referent



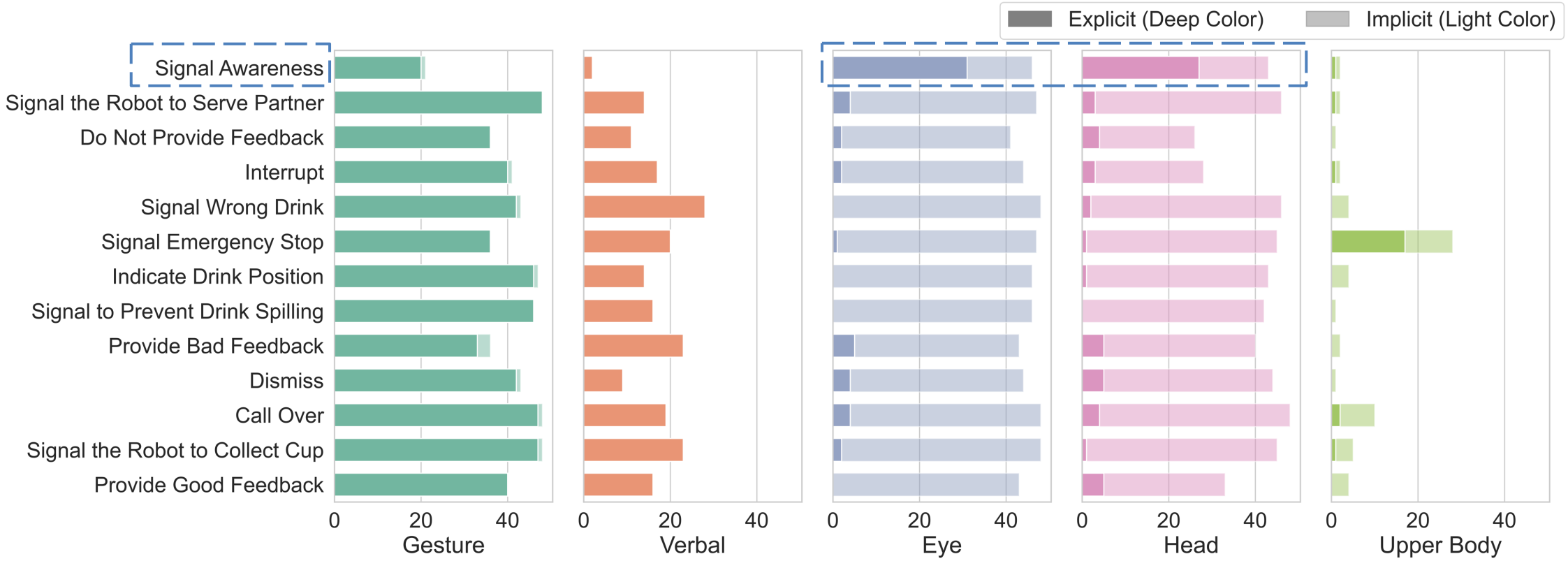
Modality Distribution by Referent



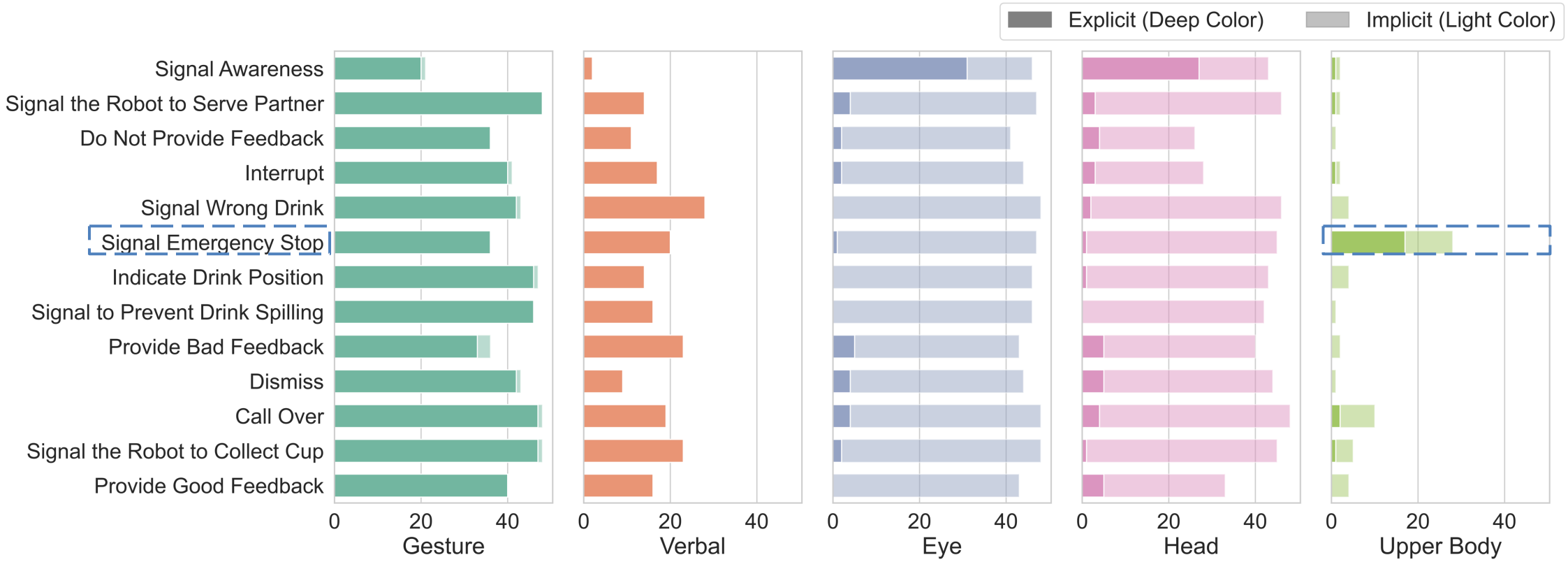
Modality Distribution by Referent



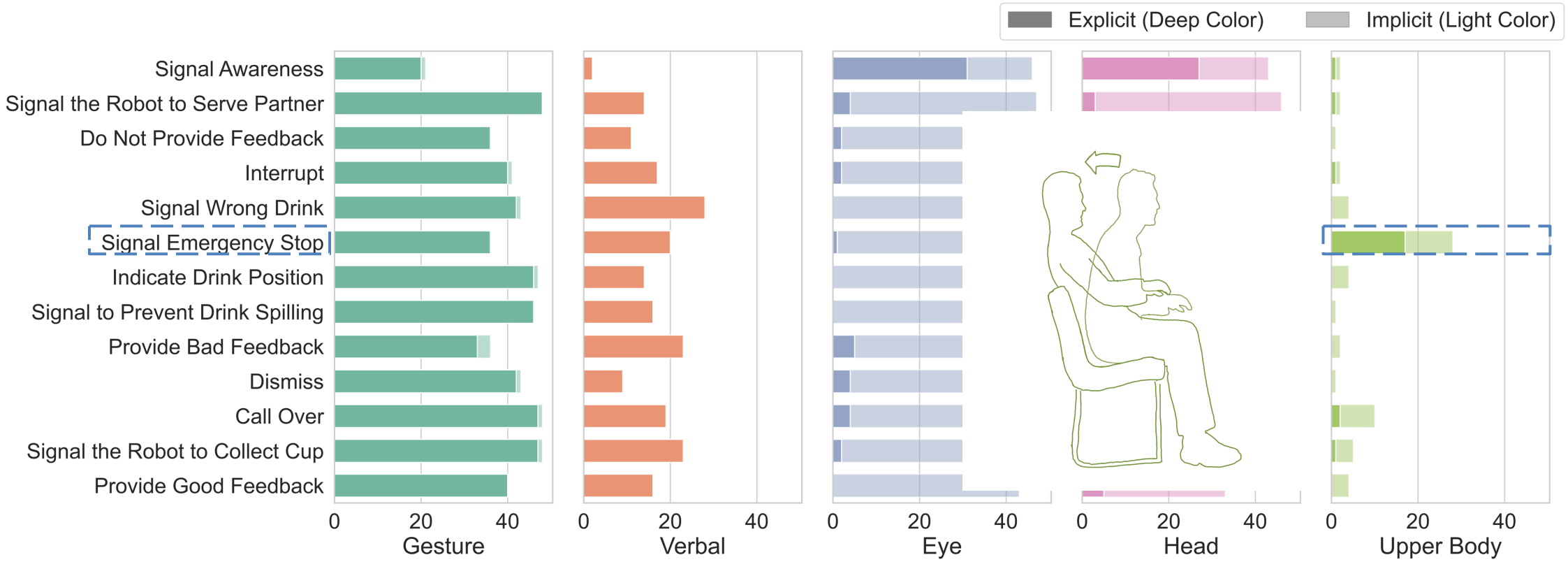
Modality Distribution by Referent



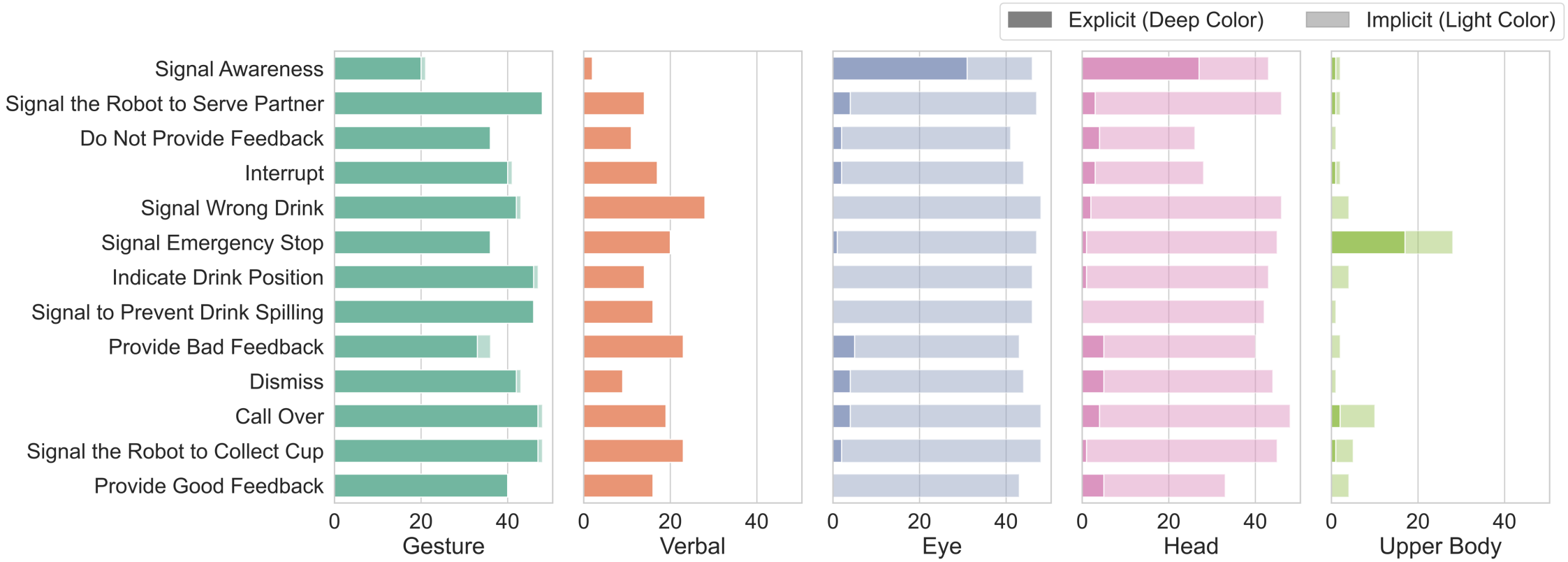
Modality Distribution by Referent



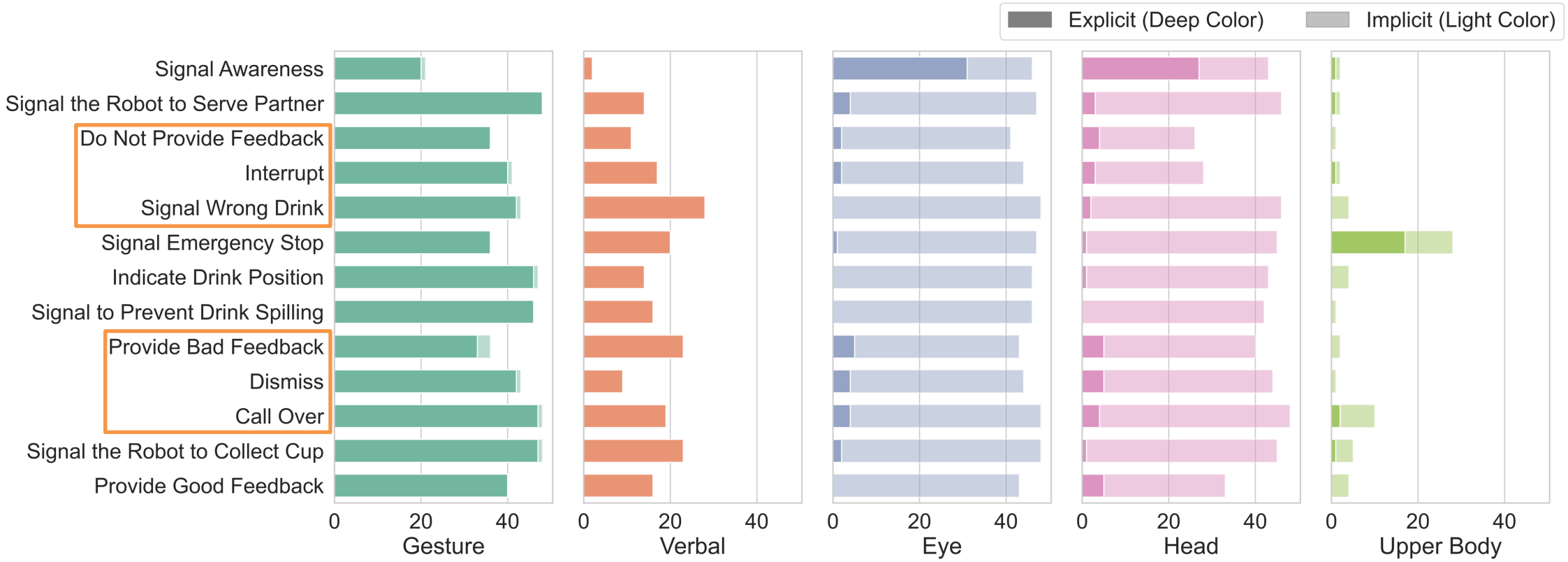
Modality Distribution by Referent



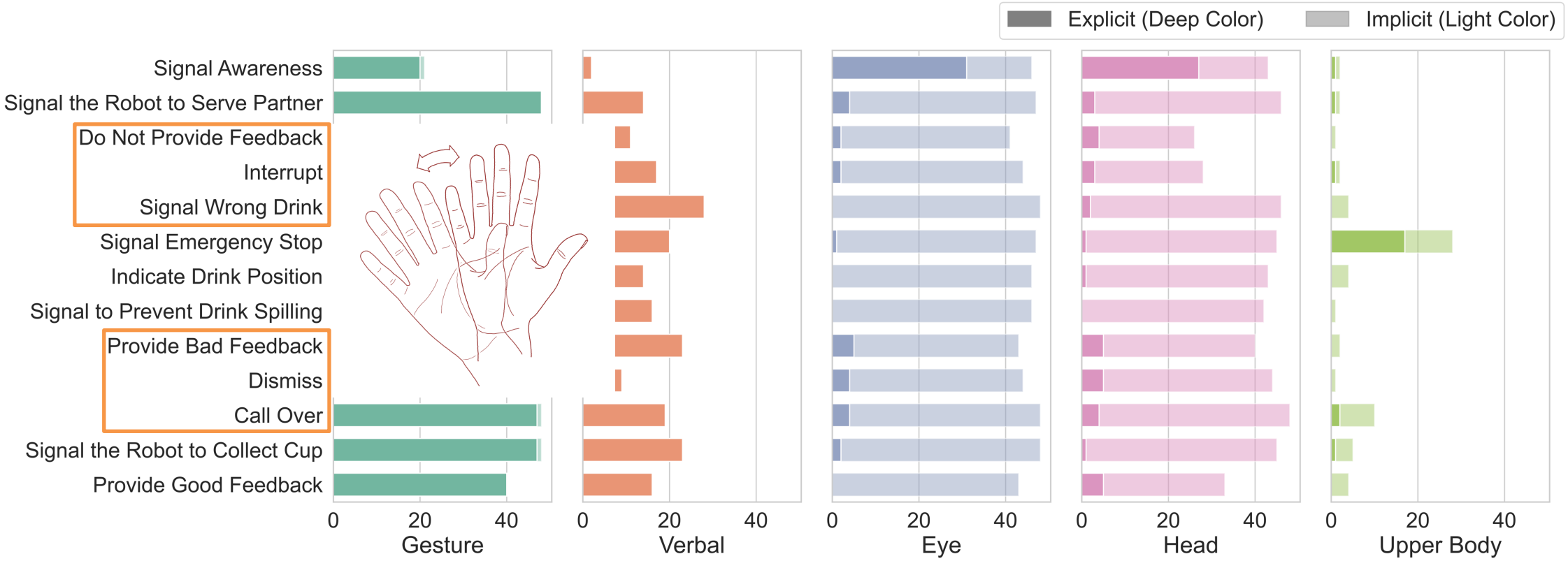
Modality Distribution by Referent



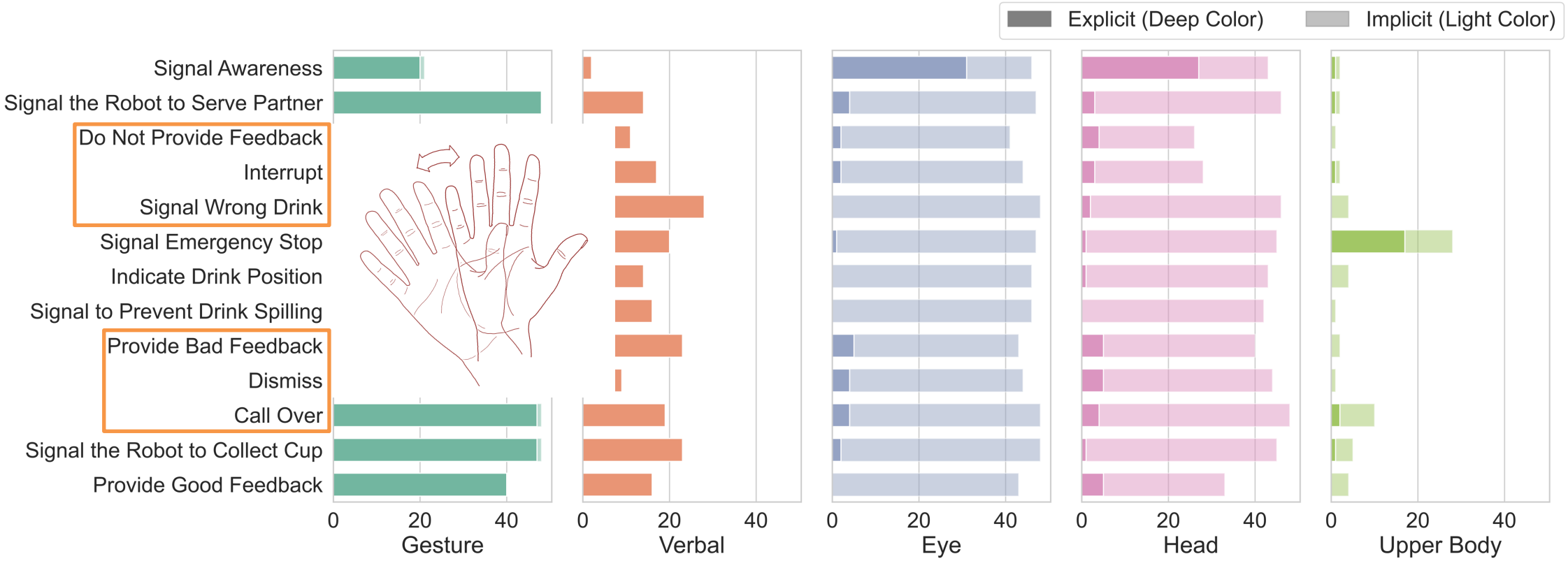
Modality Distribution by Referent



Modality Distribution by Referent

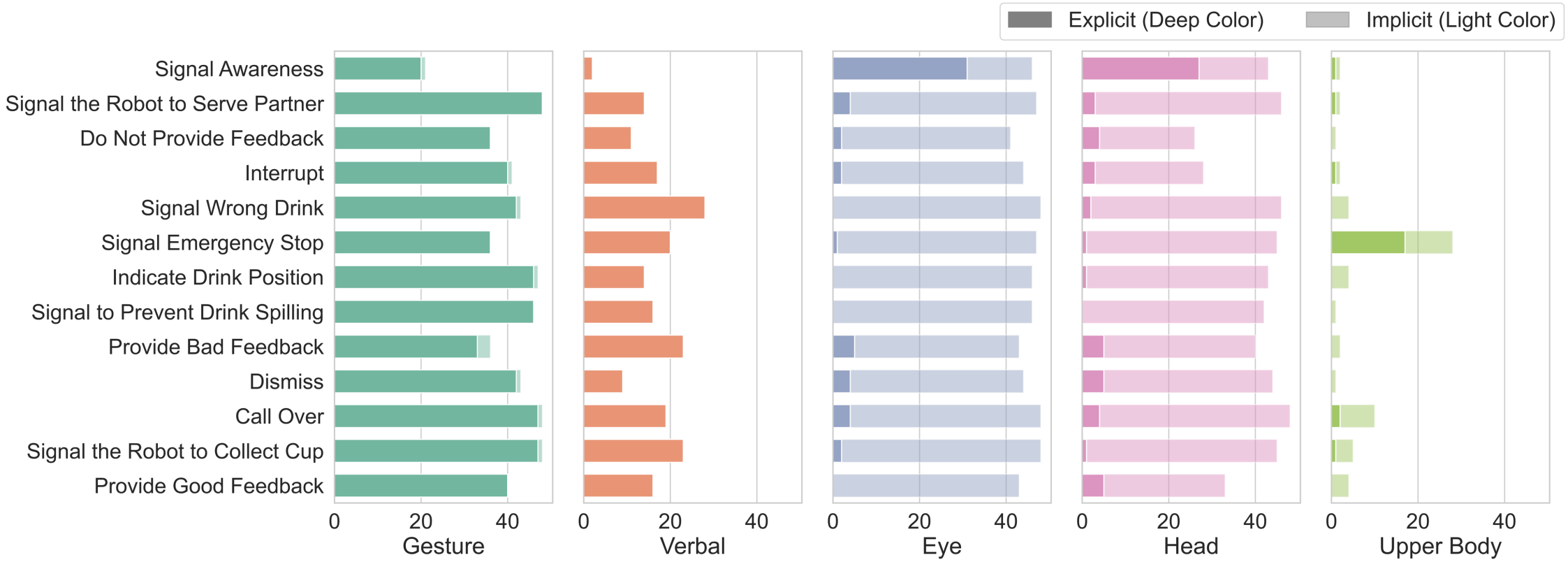


Modality Distribution by Referent



Similar *explicit cues* for different intentions → implicit cues + context to disambiguate

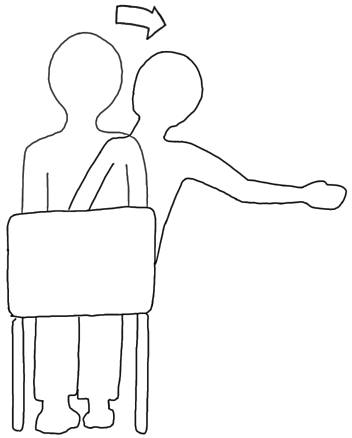
Modality Distribution by Referent

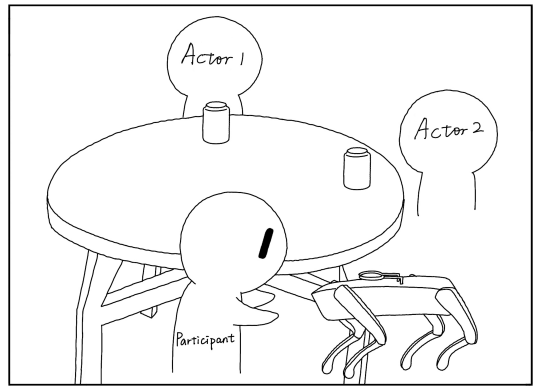


Similar *explicit cues* for different intentions → implicit cues + context to disambiguate

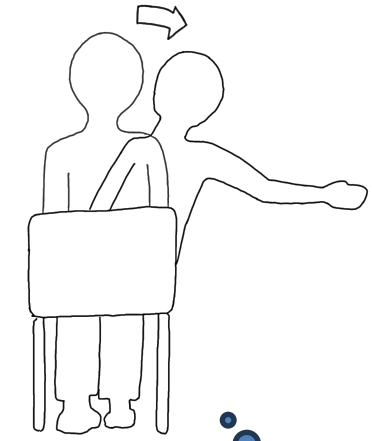
Perception on Robots

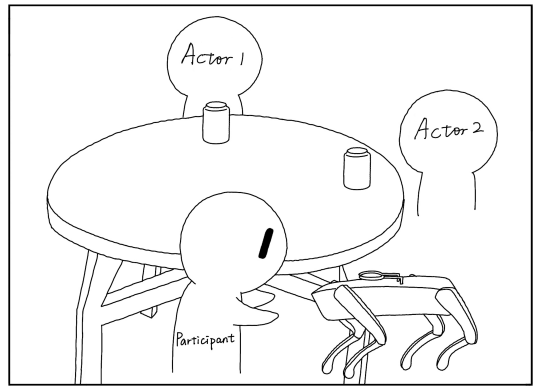
Perception on Robots





Perception on Robots

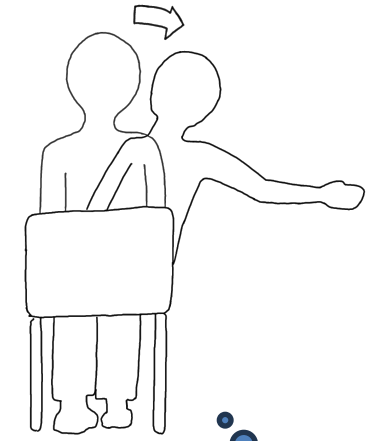


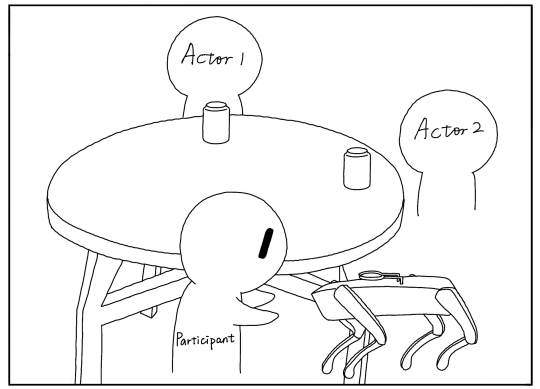


The camera and microphone are low

Perception on Robots

Alignment between Robot Capabilities and Users' Assumptions

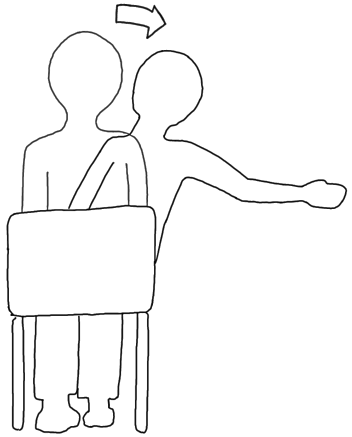


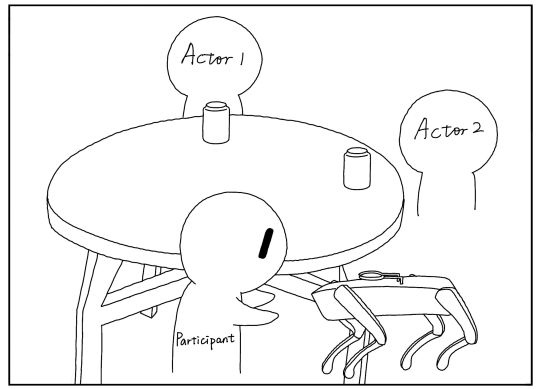


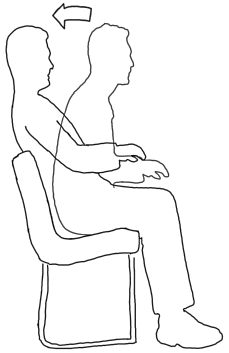
The camera and microphone are low

Perception on Robots

Alignment between Robot Capabilities and Users' Assumptions

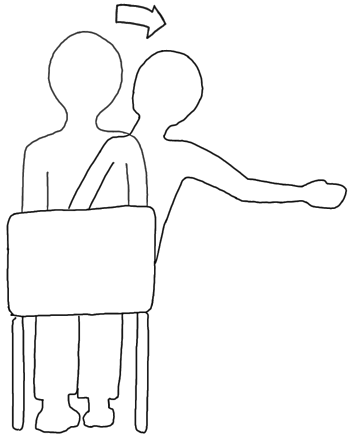


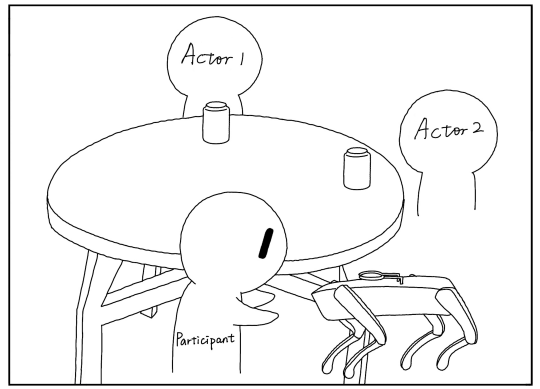


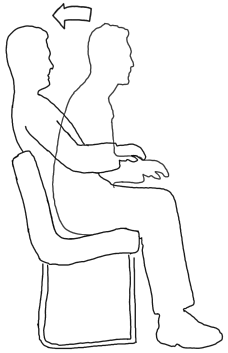


Perception on Robots

Alignment between Robot Capabilities and Users' Assumptions



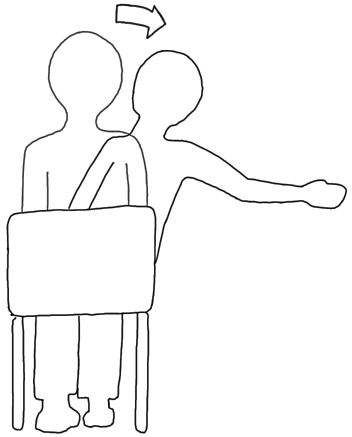


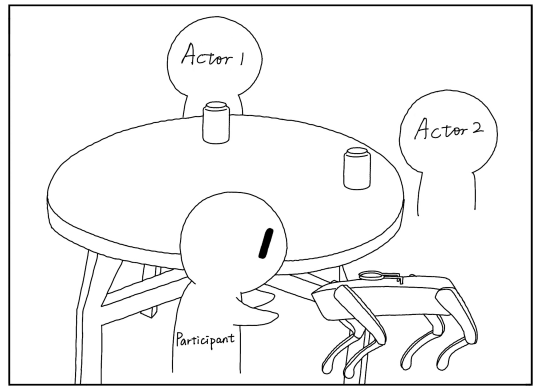


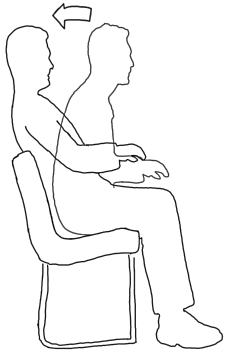
Proxemics Design

Perception on Robots

Alignment between Robot Capabilities and Users' Assumptions







Proxemics Design

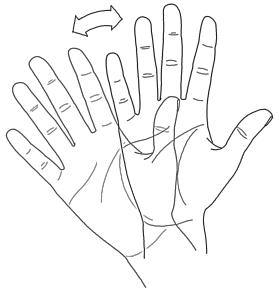
Alignment between **Robot Behaviors** and *Perceived Intelligence* of the **Robot Appearance**

Primary Task and Social Context

- Effects of ***ConversationalRole*** and Interview Results
  - More *verbal* and straightforward cues when *speaking*
  - Always choose socially appropriate cues
    - Avoid misunderstanding
    - Signal inconspicuously

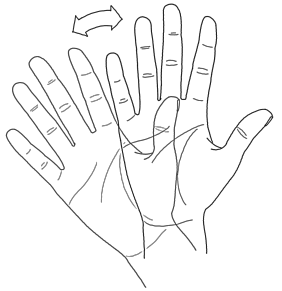
Primary Task and Social Context

- Effects of ***ConversationalRole*** and Interview Results
  - More *verbal* and straightforward cues when *speaking*
  - Always choose socially appropriate cues
    - Avoid misunderstanding
    - Signal inconspicuously



Primary Task and Social Context

- Effects of ***ConversationalRole*** and Interview Results
  - More *verbal* and straightforward cues when *speaking*
  - Always choose socially appropriate cues
    - Avoid misunderstanding
    - Signal inconspicuously



Primary Task and Social Context

- Effects of ***ConversationalRole*** and Interview Results
  - More *verbal* and straightforward cues when *speaking*
  - Always choose socially appropriate cues
    - Avoid misunderstanding
    - Signal inconspicuously

Check our paper for more details

Primary Task and Social Context

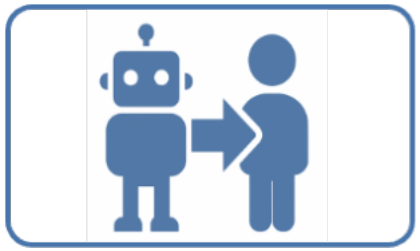
- Effects of ***ConversationalRole*** and Interview Results
  - More *verbal* and straightforward cues when *speaking*
  - Always choose socially appropriate cues
    - Avoid misunderstanding
    - Signal inconspicuously

Understanding Human Intentions Behaving Socially Appropriately

- According to the social context
- Given minimum set of social cues

Check our paper for more details

Design Implications



Robot Approaching Strategy

- Reduce unnecessary approaching
- Approach the one with less cognitive load
- Minimize interruption in necessary approaching

Human Social Cue Processing

Deduce interaction willingness
Learn common social cue semantics
Learn from implicit cues to disambiguate or understand the emphasis

Designed by Freepik

Robot Response

- Respond in a timely and socially appropriate manner
- Should be intuitive and easy to minimize interruptions
- Take proper strategies to repair failure

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Acknowledgment

Participants, actors and reviewers

Check our paper

Hanfang Lyu
hanfang.lyu@connect.ust.hk
April 2025

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Acknowledgment

Participants, actors and reviewers

Check our paper

Hanfang Lyu
hanfang.lyu@connect.ust.hk
April 2025