CogSci 2013 in Berlin

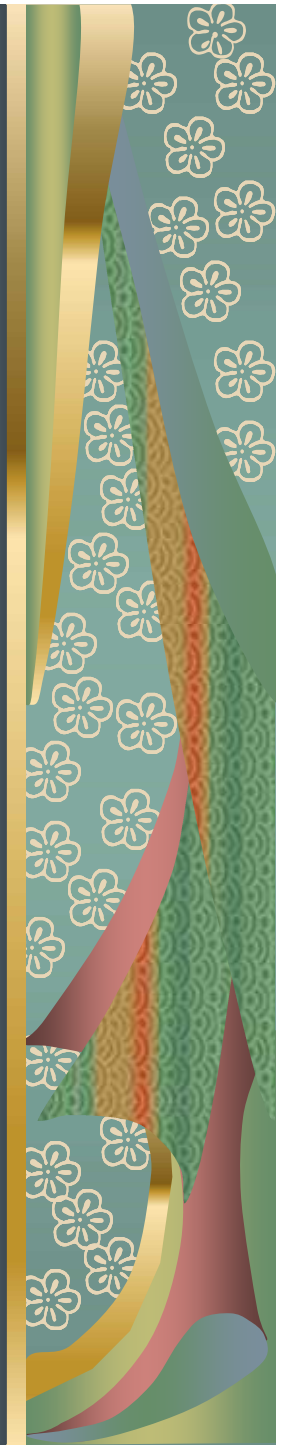
1st/August/2013

Reciprocal Ascription of Intentions Realized in Robot-human Interaction

Nagataki, S., M. Shibata, T. Konno,

and T. Hashimoto

nagataki@lets.chukyo-u.ac.jp



Background and Purpose

- There are two kinds of stream of humanoid robotics research:

(1) focusing on mimicking typical “outer” human behavior.

(2) focusing on explicating and reproducing “inner” cognitive function of humans.

➢ **We have taken the latter approach.**

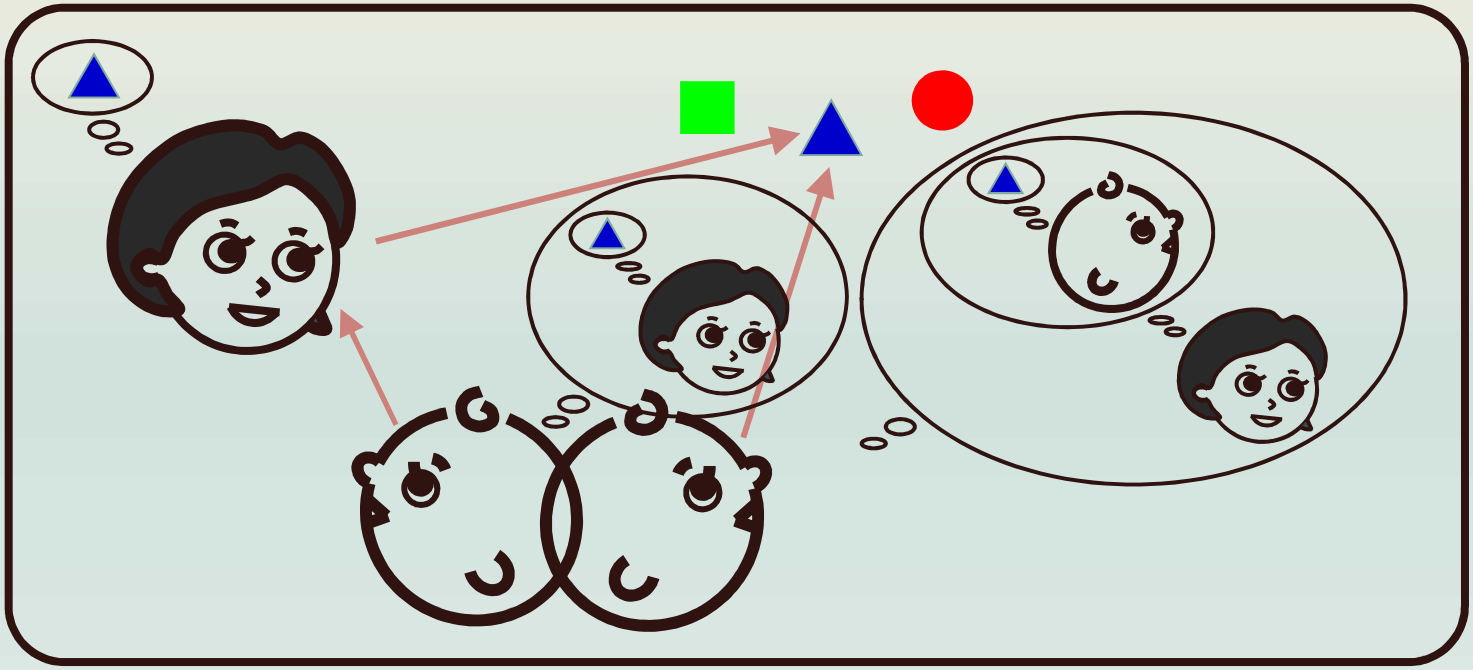
- In this general context, we have been working on the mechanism of "joint attention"(JA), the ability of which infants acquire during the earlier stage of development.
- We already succeeded in constructing a robot which can engage in JA activity of an elementary level (type A and type B).
- In a more matured stage, humans ascribe intentions to each other in JA (type C and type D).

- In order to realize this process in a robot, it is not sufficient for them merely to acquire the ability to follow others' eye direction.
- For that purpose, it is necessary to implement a relevant inferential mechanism which involves an apparatus for emotion-detection and object-categorization.
- In our presentation, we will show how this mechanism can work in our robot.

Hypothesis

- H 1 : Being accustomed to interacting with the robot which follows one's gaze merely reflexively, subjects may well feel uncomfortable with the one with a richer inner structure. For the latter is considered to have some sort of individuality.
- H 2 : The richer a robot's inner structure is, the deeper the subjects' uncomfortableness will be.
- H 3 : Some subjects may feel "intentionality" and "humanness" in the robot with a richer inner structure.

Nested Structure of Intentional Actions: in the case of joint attention

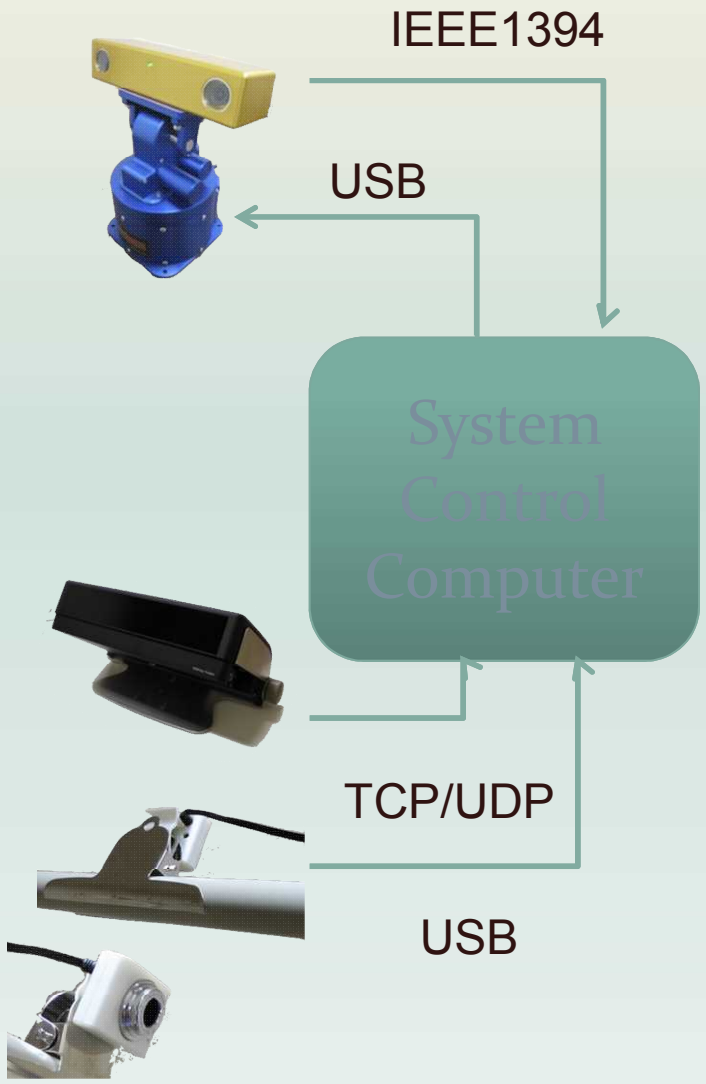


He knows that she understands that he is focusing on the object she is now looking at.

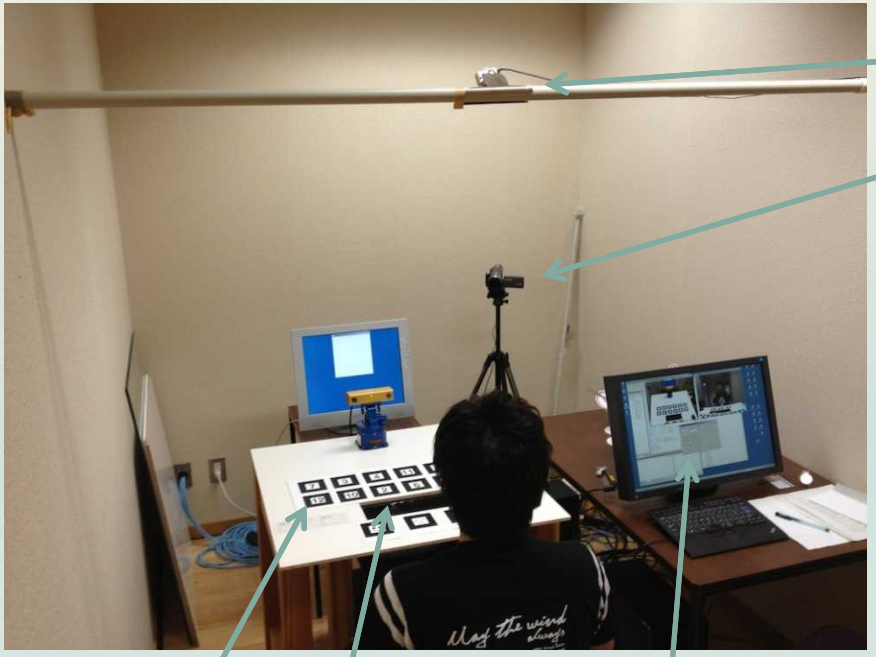
System Connection Diagram

- Robot
  - Combination of Stereo Camera （Bumblebee2) and, Pan-tilt (actuator Biclops)
- Measurement of the Gaze Focus

  Measurement Device of the Gaze Focus（Tobii X120） and Top View Camera（USB Camera)



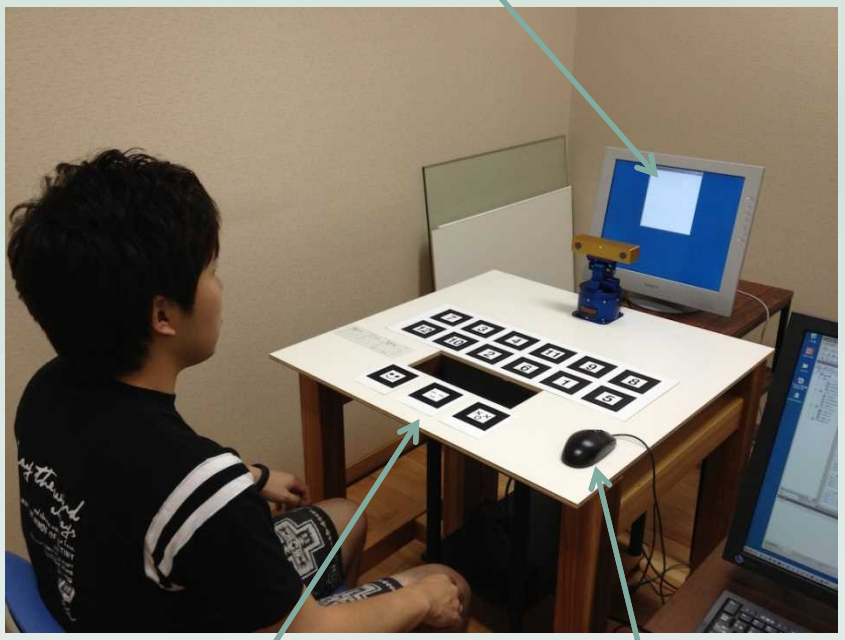
Experiment environment



Top view Camera

Vide Camera

Showing numbers by Robot



System control CP

Measurement device (Tobii)

Number cards（1-12）

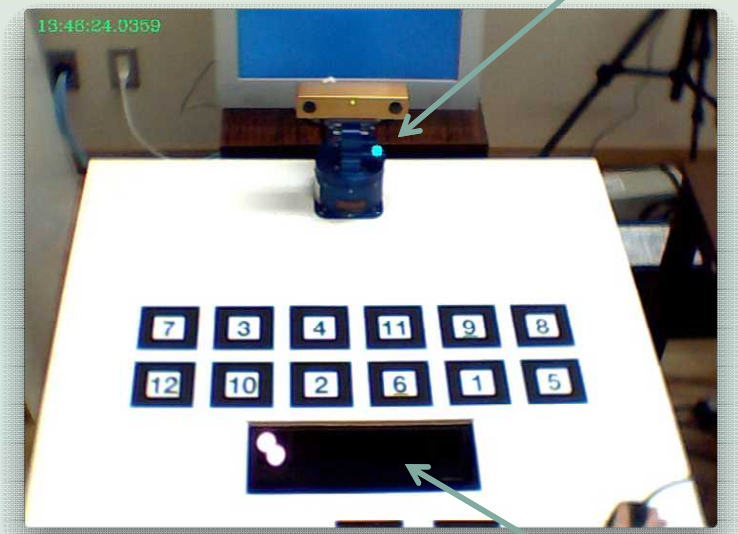
Expression cards

Mouse for Yes/No

Robot's View and Top View

- Subject shows his emotion-type by a card.

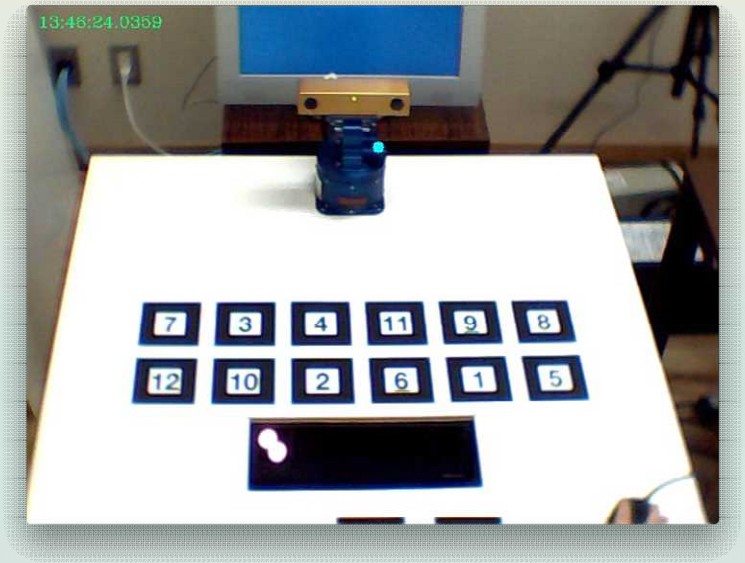
Robot's View (only right eye)



Top View Camera

Robot can recognize subject's face and card type (Type D Robot).

Human-Robot visual interaction



- A person sits in front of the robot, and looks at one of 12 numbers.
- Positions of the numbers are fixed.
- Resolution of eye-direction is 30 [deg].
- The robot gazes at the person and the number alternately.

4 Types of Robots

- **Type A:** Gaze following Model

  The robot turns his eyes to follow subject's eye direction, and then the robot gazes at the number the subject gazes at by using a measurement device of the subject's gaze focus.

- **Type B:** Intentional Agent Model I

  Robot determines a target number on the base of his memory of connections between directions of subject's gaze and locations of numbers, and then turns his eyes to the target number and gazes at it.

We have already produced the robots of these types.
(Nagataki, Shibata, and Konno 2010)

- **Type C:** Though its inner structure is the same as type B, its memory can accumulate interactive experience with the subjects.

- **Type D:** This adds two functions to type C. They are (1) recognizing three types of emotion-expression and (2) classifying 12 numbers into these three emotion-categories, thereby determining the target number.

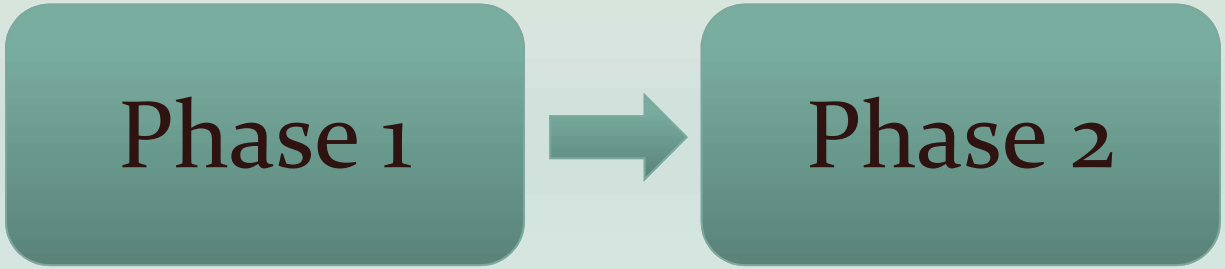
Experiment Design

- Without interaction of Yes/No
- No accumulation of experience

- With interaction of Yes/No
- With accumulation of experience

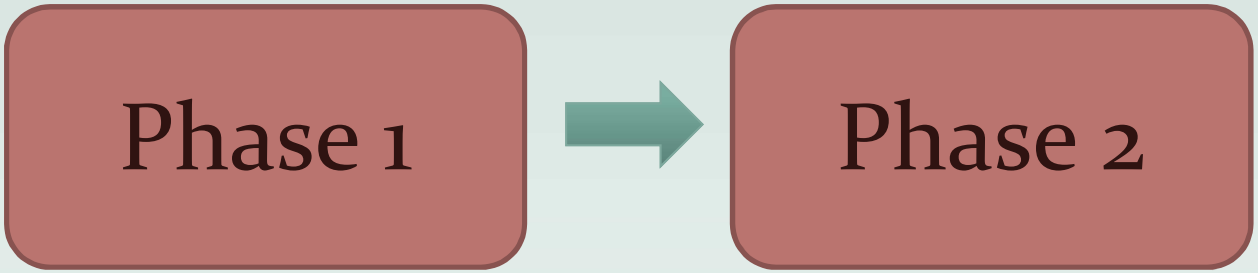
Type C

Without showing expression :
5 subjects



Type D

With showing expression :
5 subjects



Phase 1 Experiments

- **For Type C**: The robot turns his eyes to follow subject's eye direction, then gazes at the number the subject gazes at, by using his memory of only one subject within each session.
- If the subject thinks that the robot figured out what she gazed at, she clicks the mouse in order to stop the interaction.

- **For Type D:** First, each subject classifies 12 numbers into three emotion-categories: 1. like, 2. neutral and 3. dislike. In this experiment, the subject shows her preference by the emotion-expression card (smile, neutral, or disgust), and the rest condition is the same as type C.
  - Both types of robot have accumulated their experiences with all five subjects after Ph. 1 experiments.

Phase 2 Experiments

- **For "experienced" type C:** The type C determines the target number by using accumulated memories. Then the subject gives a yes/no response to it. If the answer is "no," the robot chooses the number that was the second most looked at. If an answer is "yes", it is done.

- **For "experienced" type D:** The subject shows her preference when she looks at the target number. The type D determines by using its emotion-category connected with her preference. Its emotion-category has been formed after ph. 1 experiments. Then the subject gives a yes/no response to it. If the answer is "no," the robot chooses the number which is the second most looked-at one in the same category.

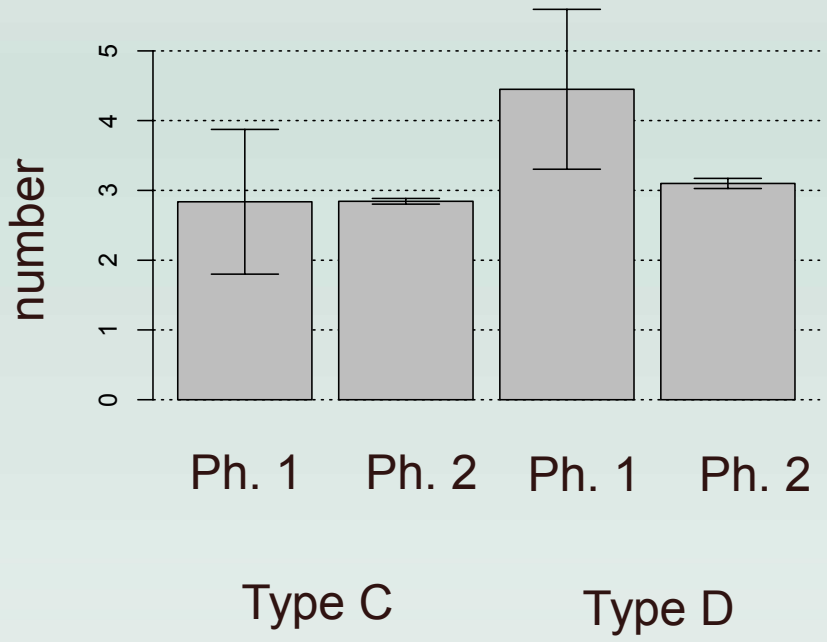
Participants and Procedure

- Graduate students and researchers of JAIST
  - 10 people （Three are women)
  - Average age : 26.7 years old
- Subjects of type C is five (two are women), subjects of type D is five (one is woman).
- Subjects participated on different two days and their interval is about 4.7 days.
- One trial consists of the following procedure: the subject looks at one number and the robot determines it.

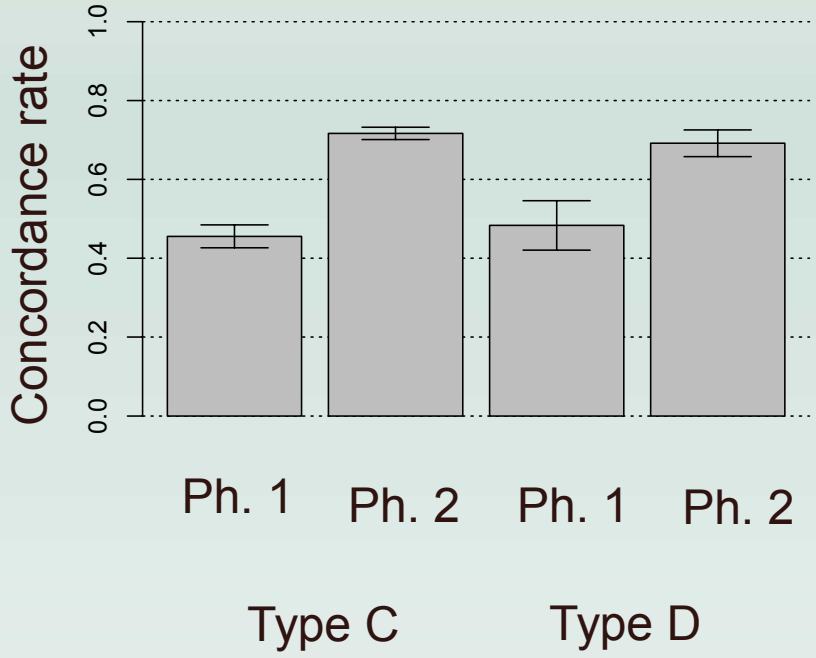
  Each session has 24 trials.

Result 1

The average number of gaze alternation



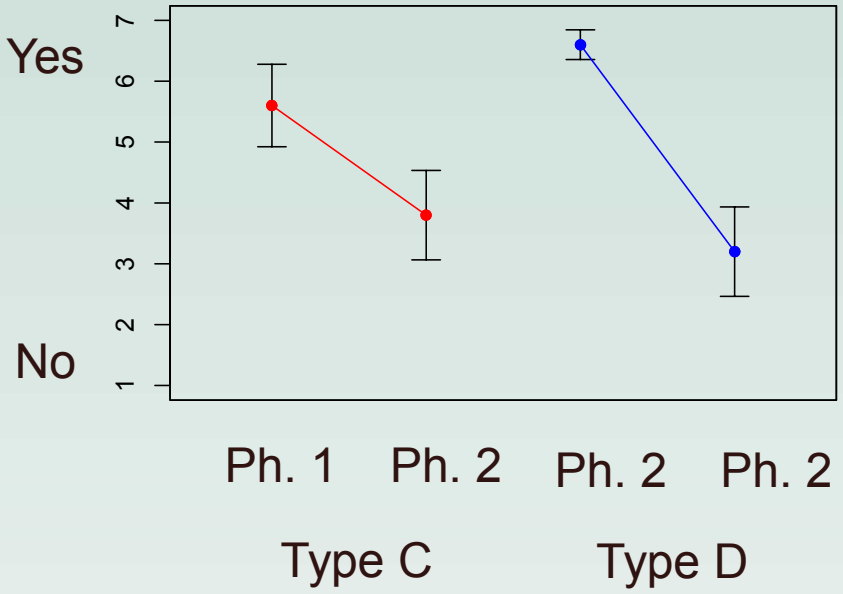
the concordance rate between subjects and robots



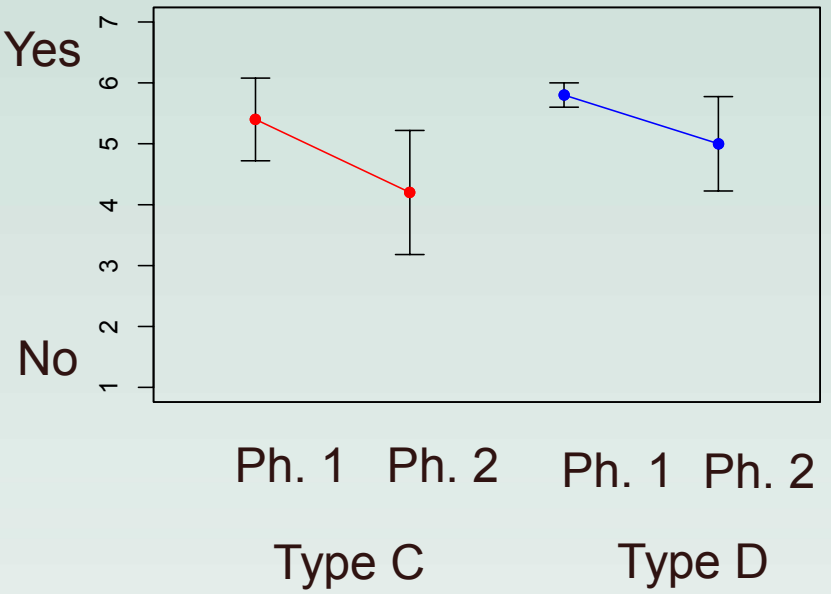
Result 2

Robot's recognition of target numbers

Q1. Do you think the robot recognized your target?

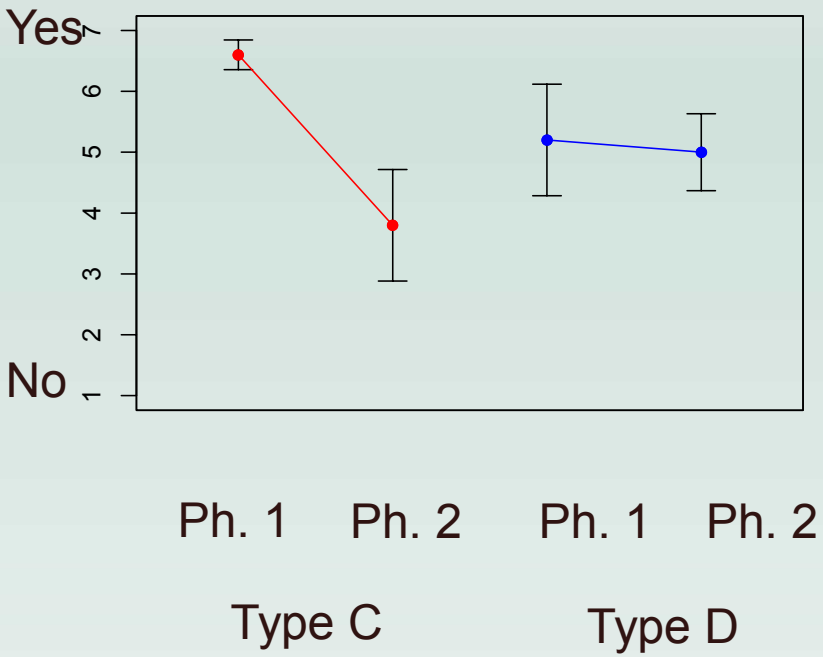


Q2. Did you recognize what robot gazed at?

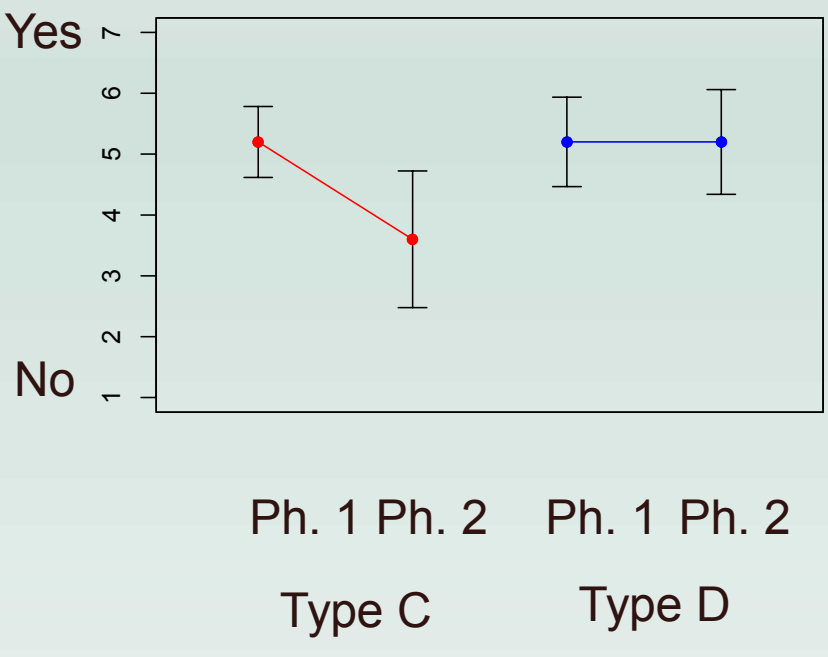


Subject's detection of robot's intention

Q3. Did you feel the robot tried to understand what you were looking at ?

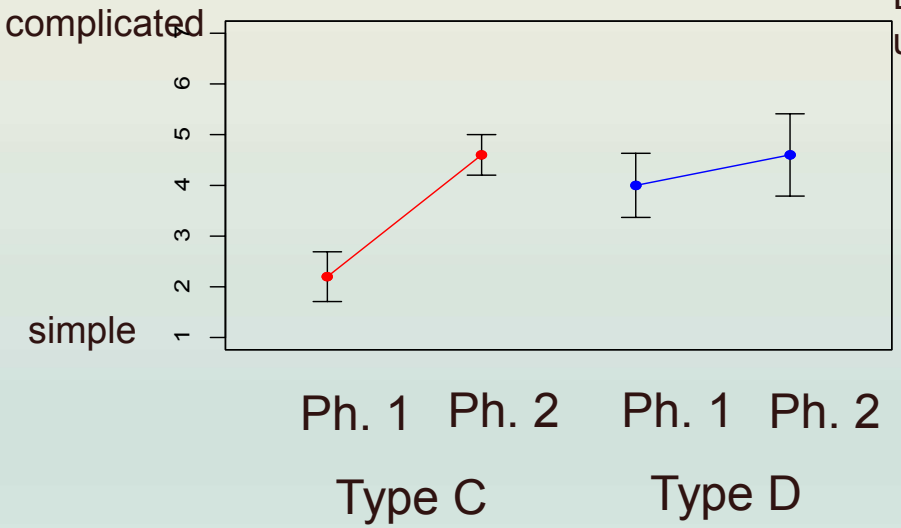


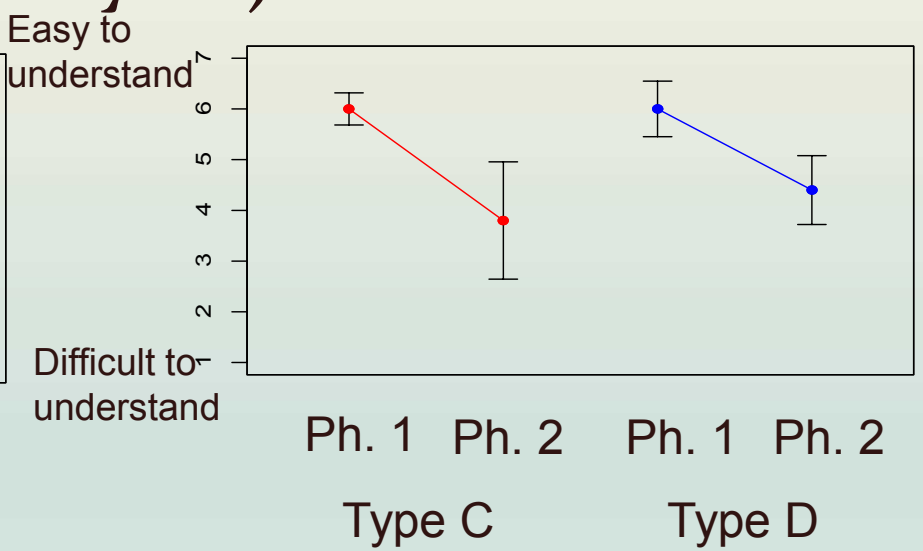
Q4. Did you feel the robot has an intention to do something ?

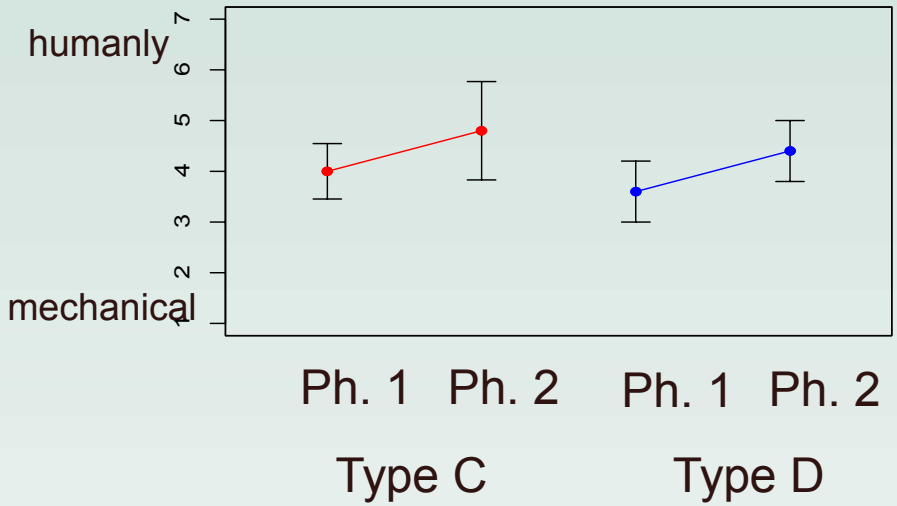


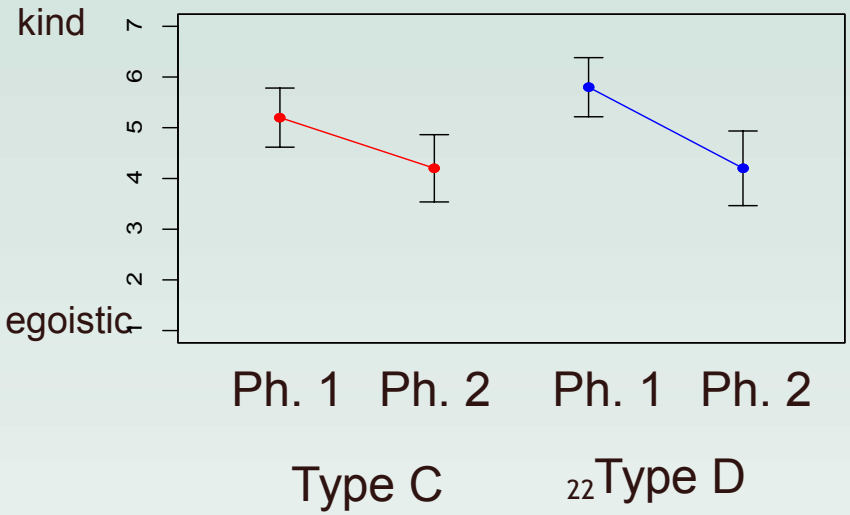
Result 3

Impression evaluation by adjectives









Discussion 1

- Result 1 shows that, in both types of robot, the concordance rate of Ph. 2 experiment, where yes/no interaction is added, is better than that of Ph. 1. Result 2 (Q1 and 2) shows, however, that the subjects' evaluation of the mutual understanding of intention in Ph. 2 becomes worse. This tendency is rather remarkable in Type D, although it has a richer inner structure.
- On the other hand, as Result 3 shows, the perceived humanness of both types in Ph. 2 is higher than in Ph. 1, while the evaluation of understandability in Ph. 2 is lower than in Ph. 1. It is notable that both types of robot in Ph. 2 are regarded more "complicated" and "egoistic" than in Ph. 1.
- This suggests that the richer the robot's inner structure is, the stronger becomes the subjects' impression that the robot has some sort of individuality or autonomous intentional agency.

Discussion 2

- Result 1 of Ph. 1 also shows that Type D does gaze alternation more frequently than Type C, despite the fact that both attain much the same concordance rate. This means that it is harder for subjects to succeed joint attention (JA) with type D than Type C. Put another way, the performance of JA in Type D with a richer inner structure is less efficient than Type C.
- By contrast, there is no such difference between C and D in Ph. 2. One reason may be that the yes/no presentation works like a linguistic aid to clarify the subjects' intention.
- We think there is a similarity between these phenomena and the development of JA in humans, because infants show, in general, a developmental U-curve in their earlier stage of the development of JA.
- We can say that we made a more humanlike mind in robot by making an inner structure richer than before.

Bibliography

- S. Nagatak, M. Shibata, T. Konno, T. Hashimoto, and H. Hattori. Joint Attetion Realized in a Robot with Intentional Agency. Proceedings of the European Conference on Complex Systems, ECCS'10, CD-ROM Track D:Complexity and Computer Science, No.59, 2010.
- G. Butterworth and N. Jarrett. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. British Journal of Developmental Psychology, 9:55–72, 1991.
- N. Emery. The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews, 24(6):581–604, 2000.
- T. Konno and T. Hashimoto. Developmental construction of intentional agency in communicative eye gaze. Proceedings of the International Conference on Development and Learning, ICDL ’06, CD– ROM, May 31 - Jun.3 2006.
- M. Tomasello. Joint attention as social cognition. In C. Moore and P. Dunham, editors, Joint Attention: Its Origins and Role in Development, pages 103–130. Lawrence Erlbaum, 1995.
- M. Tomasello. The cultural origins of human cognition. Harvard University Press, Cambridge, 2000.