and Social Cognition

Emergence of Self Awareness in Robots Based on Predictive Learning

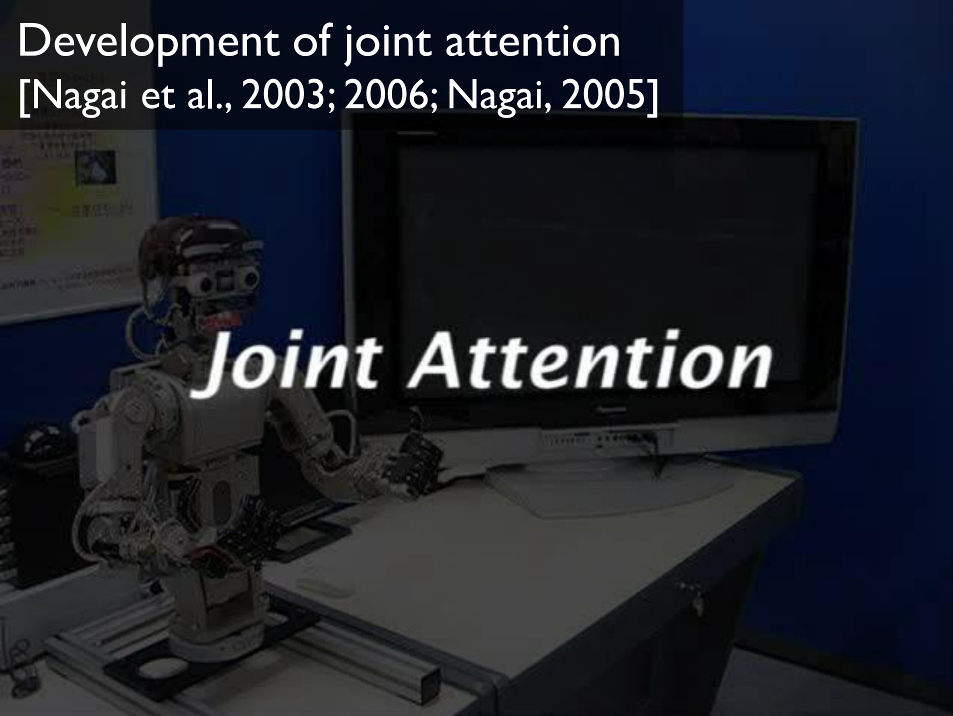
Yukie Nagai

Graduate School of Engineering, Osaka University

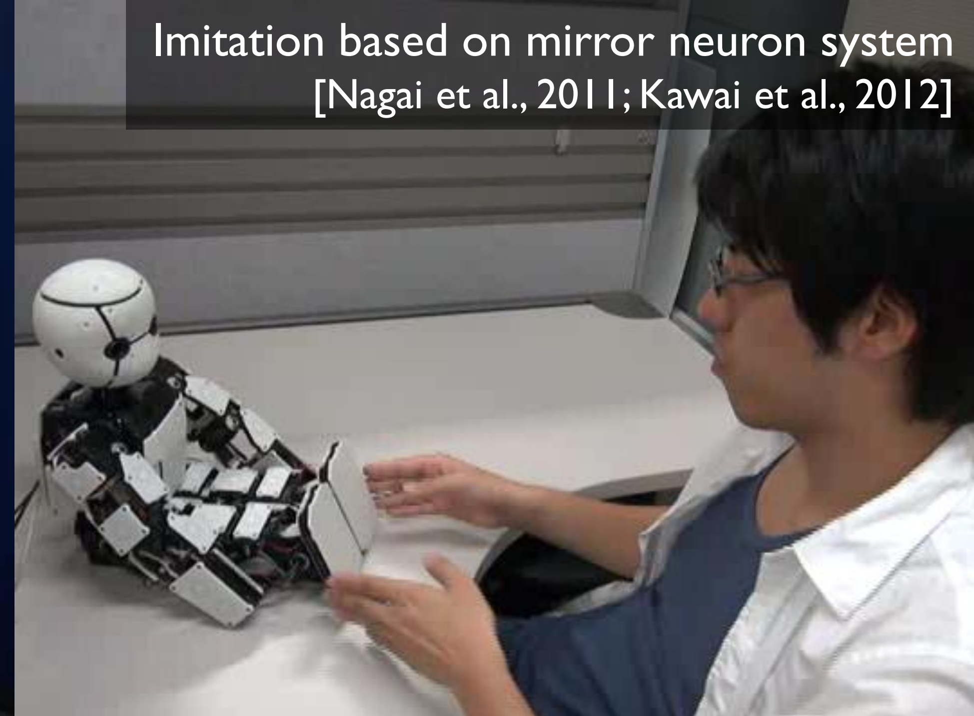
Cognitive Neuroscience Robotics

ISSA Summer School

Center for Planetary Science, Kobe, August 10, 2015



Development of joint attention
[Nagai et al., 2003; 2006; Nagai, 2005]


Imitation based on mirror neuron system
[Nagai et al., 2011; Kawai et al., 2012]


Infant-directed action
[Nagai & Rohlfing, 2009]

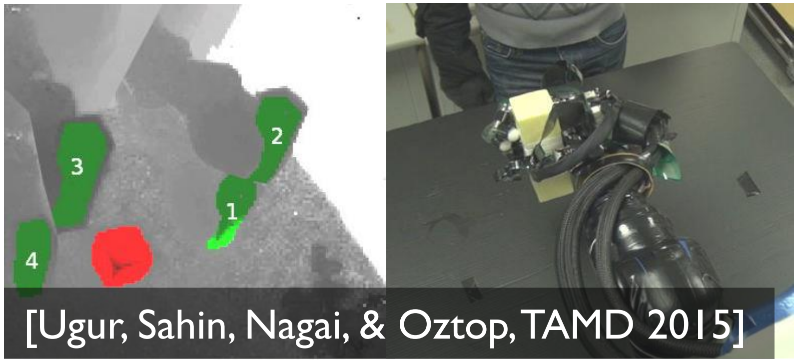

Gaze-head coordination in social interaction
[Schillingmann et al., 2015]

Cognitive Developmental Robotics

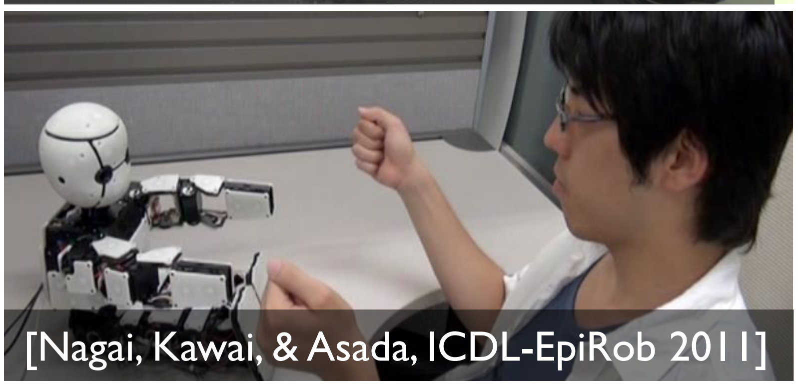
[Asada et al., 2001; 2009; Lungarella et al., 2003]

- Aim at understanding **the principle of human cognitive development** by means of **constructive approach**
  - Bridge the gap between neuroscience (micro level), and psychology and cognitive science (macro level)
  - Build human-like intelligent robots

What is the Biggest Difference Between Robot and Human Development?



[Ugur, Sahin, Nagai, & Oztop, TAMD 2015]

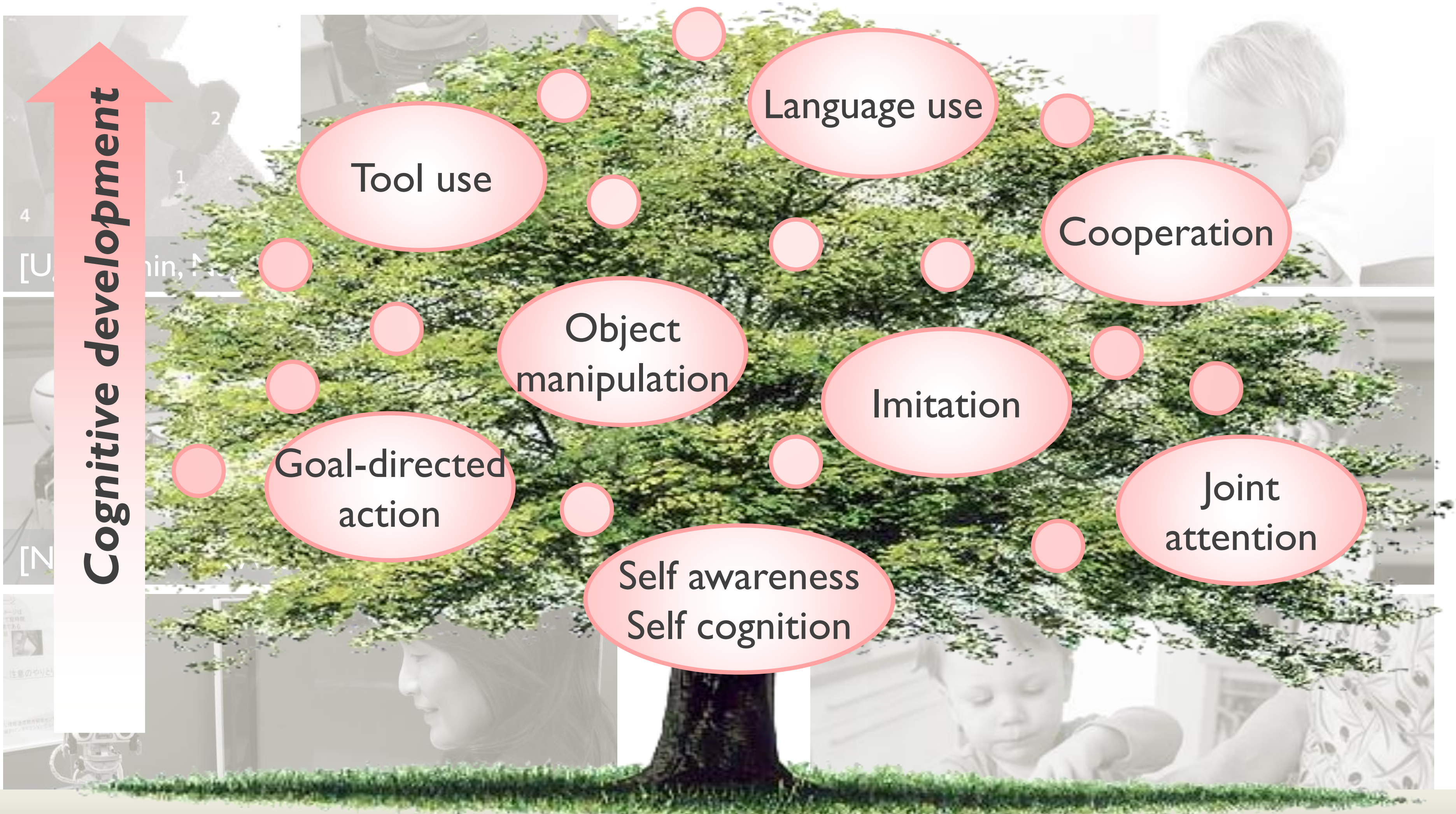


[Nagai, Kawai, & Asada, ICDL-EpiRob 2011]

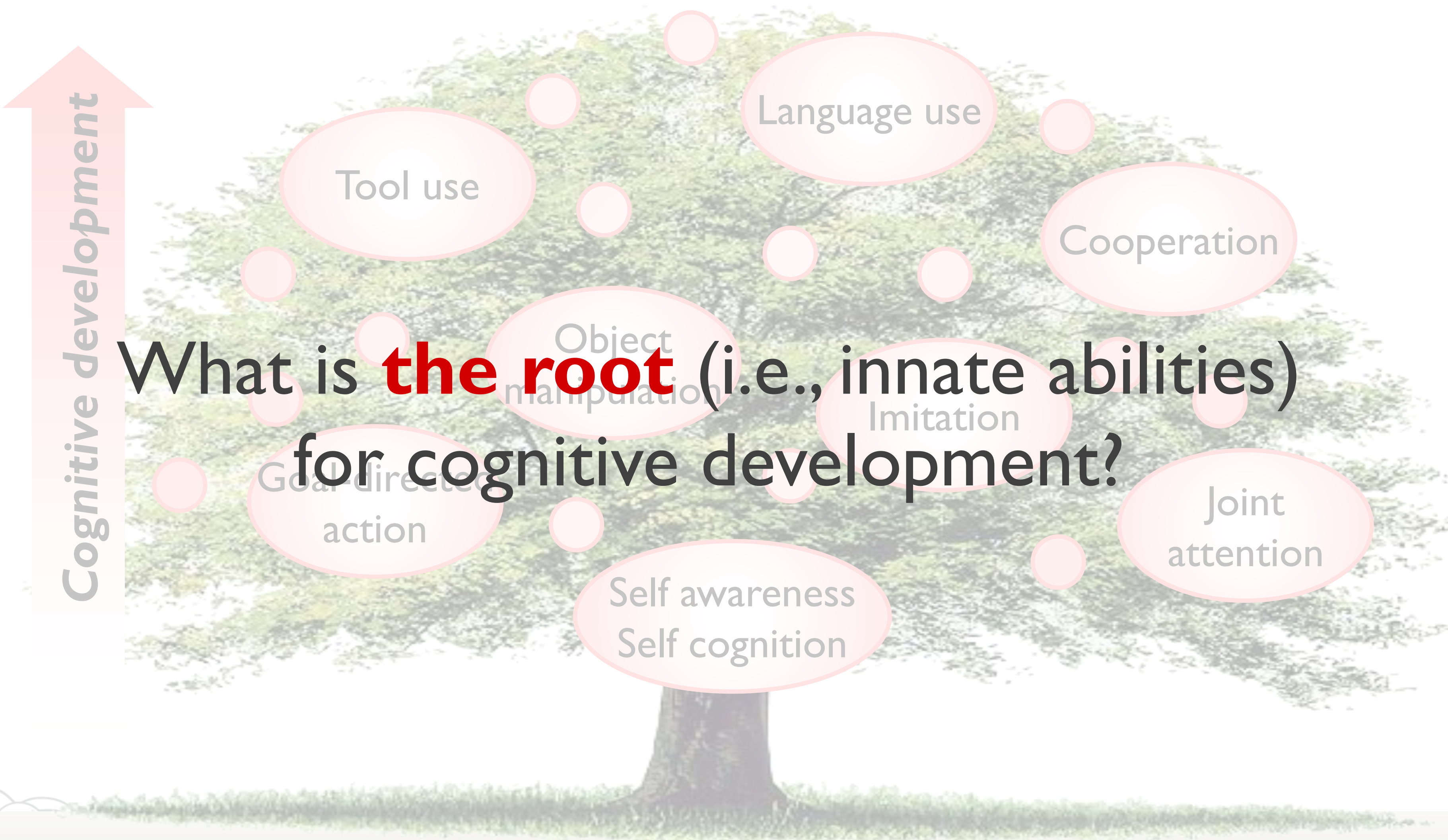
[Nagai, Hosoda, Morita, & Asada, Connection Science 2003]

≟

Human Development is a Continuous Process

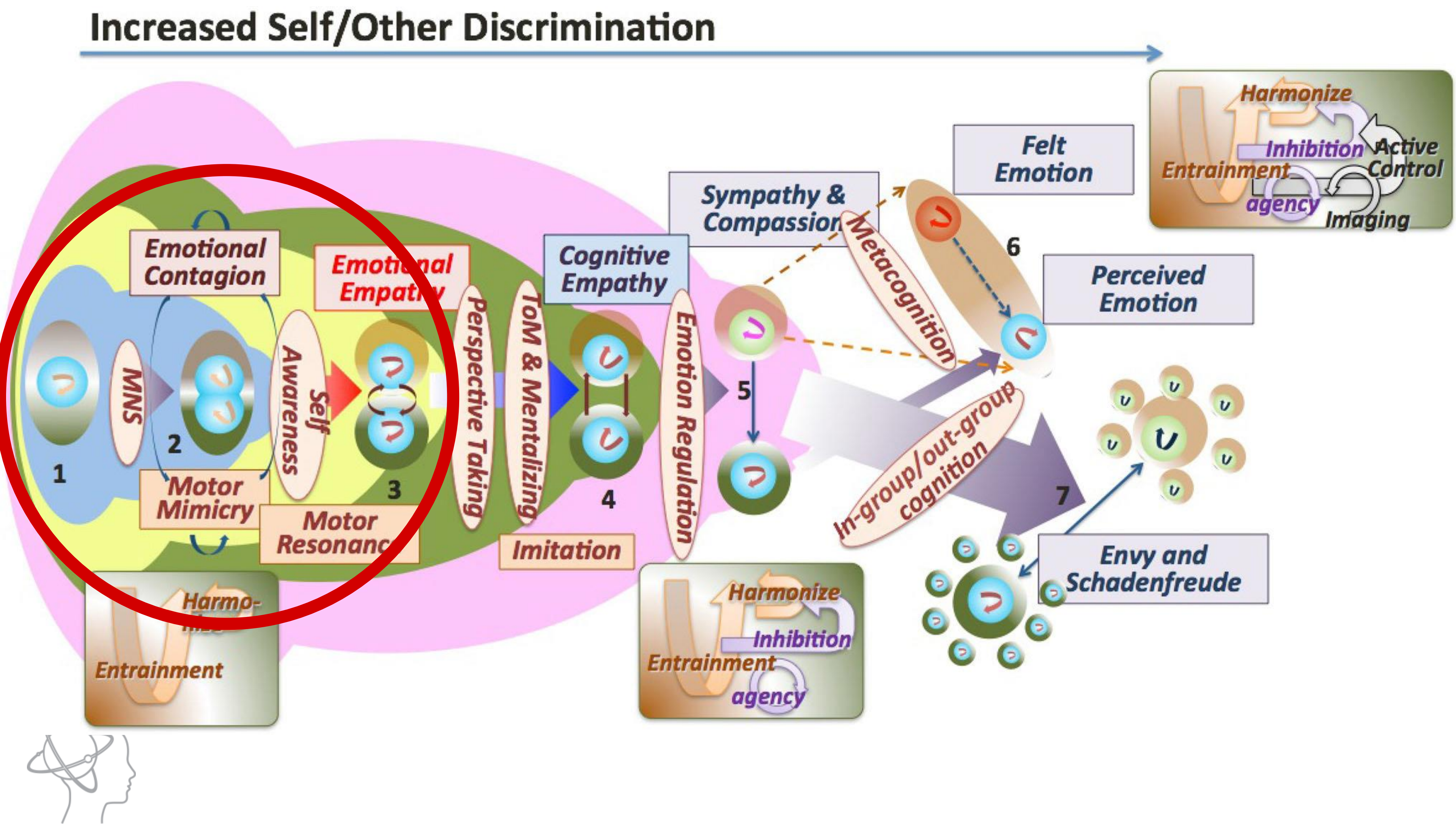


Human Development is a Continuous Process



What is **the root** (i.e., innate abilities) for cognitive development?

Development from Self-Other Discrimination to Higher Cognition [Asada, 2014; 2015]



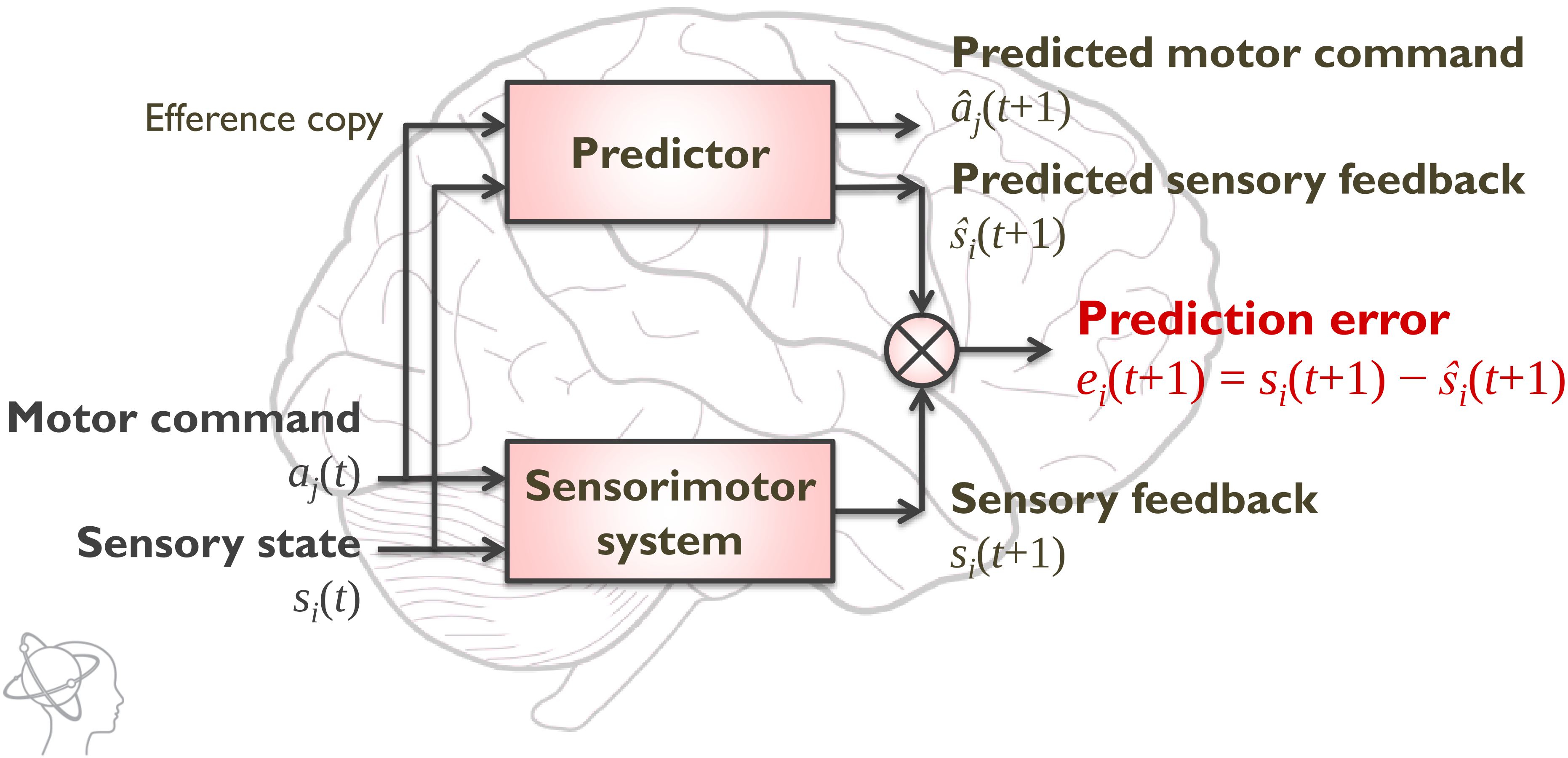
Outline

1. Our theory: *predictive learning of sensorimotor information* as a key for cognitive development
2. Cognitive development in robots based on *predictive learning*
   – Self-other cognition & imitation
   – Altruistic behavior
3. Autism spectrum disorder (ASD) caused by *atypical tolerance for prediction error*
   – Simulator of atypical perception
   – Local processing bias caused by neural imbalance

Our Theory about Cognitive Development

[Nagai, in press]

Predictive learning of sensorimotor information (i.e., minimizing prediction error $e_i(t+1)$) leads to cognitive development.



Our Theory about Cognitive Development

[Nagai, in press]

Predictive learning of sensorimotor information (i.e., minimizing prediction error $e_i(t+1)$) leads to cognitive development.

(1) *Learn the predictor* through sensorimotor experiences

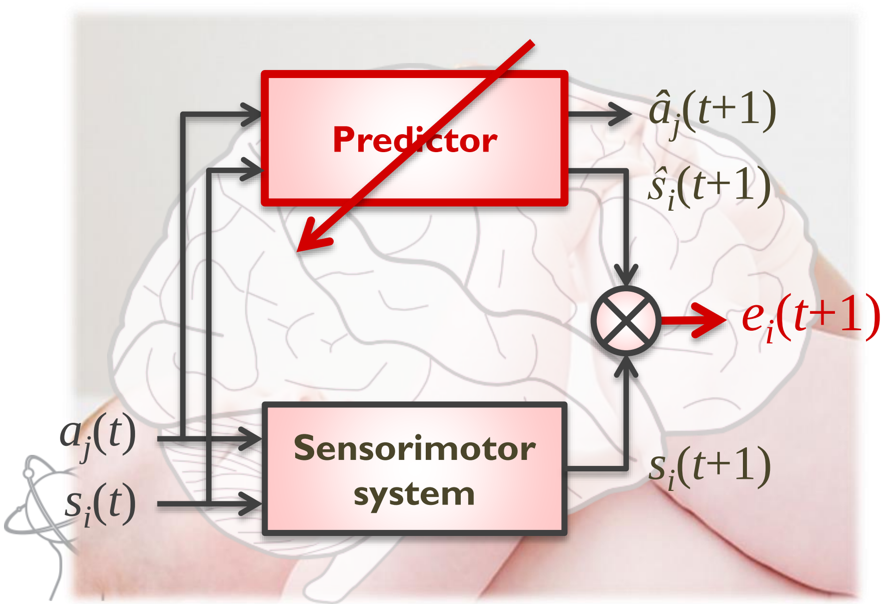
→ Self-other cognition

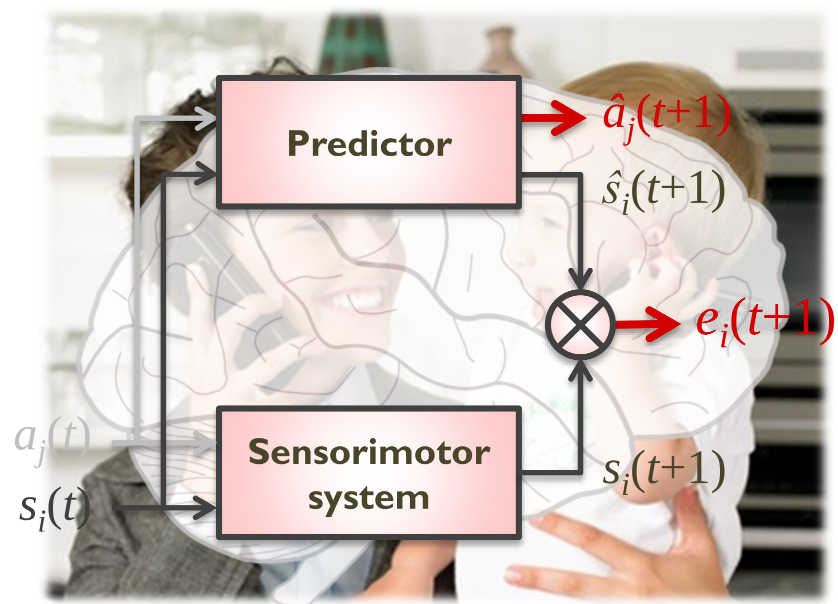
→ Goal-directed action, etc.

(2) *Produce an action* in response to other's action

→ Imitation

→ Altruistic behavior, etc.





Increasing Interest in Predictive Learning

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2009) **364**, 1211–1221
doi:10.1098/rstb.2008.0300

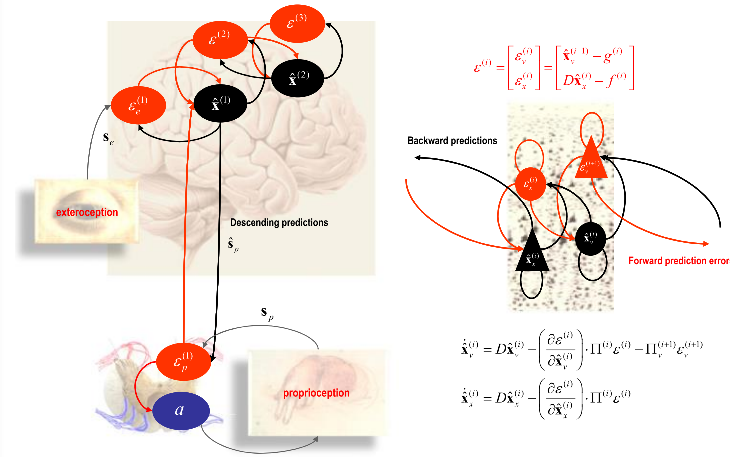
Predictive coding under the free-energy principle

Karl Friston* and Stefan Kiebel

The Wellcome Trust Centre of Neuroimaging, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK

This paper considers prediction and perceptual categorization as an inference problem that is solved by the brain. We assume that the brain models the world as a hierarchy or cascade of dynamical systems that encode causal structure in the sensorium. Perception is equated with the optimization or inversion of these internal models, to explain sensory data. Given a model of how sensory data are generated, we can invoke a generic approach to model inversion, based on a free energy bound on the model's evidence. The ensuing free-energy formulation furnishes equations that prescribe the process of recognition, i.e. the dynamics of neuronal activity that represent the causes of sensory input. Here, we focus on a very general model, whose hierarchical and dynamical structure enables simulated brains to recognize and predict trajectories or sequences of sensory states. We first review hierarchical dynamical models and their inversion. We then show that the brain has the necessary infrastructure to implement this inversion and illustrate this point using synthetic birds that can recognize and categorize birdsongs.

Keywords: generative models; predictive coding; hierarchical; birdsong



Nature Reviews Neuroscience | AOP, published online 13 January 2010; doi:10.1038/nrn2787

REVIEWS

The free-energy principle: a unified brain theory?

Karl Friston

Abstract | A free-energy principle has been proposed recently that accounts for action, perception and learning. This Review looks at some key brain theories in the biological (for example, neural Darwinism) and physical (for example, information theory and optimal control theory) sciences from the free-energy perspective. Crucially, one key theme runs through each of these theories — optimization. Furthermore, if we look closely at what is optimized, the same quantity keeps emerging, namely value (expected reward, expected utility) or its complement, surprise (prediction error, expected cost). This is the quantity that is optimized under the free-energy principle, which suggests that several global brain theories might be unified within a free-energy framework.

Cogn Process (2007) 8:159–166
DOI 10.1007/s10339-007-0170-2

REVIEW

Predictive coding: an account of the mirror neuron system

James M. Kilner · Karl J. Friston · Chris D. Frith

Received: 21 February 2007 / Revised: 19 March 2007 / Accepted: 21 March 2007 / Published online: 12 April 2007

Abstract Is it possible to understand the intentions of other people by simply observing their actions? Many believe that this ability is made possible by the brain's mirror neuron system through its direct link between action and observation. However, precisely how intentions can be inferred through action observation has provoked much debate. Here we suggest that the function of the mirror system can be understood within a predictive coding framework that appeals to the statistical approach known as empirical Bayes. Within this scheme the most likely cause of an observed action can be inferred by minimizing the prediction error at all levels of the cortical hierarchy that

used to execute that same action (Jeannerod 1994; Prinz 1997). Interest in this idea has grown recently, in part due to the neurophysiological discovery of ''mirror'' neurons. Mirror neurons discharge not only during action execution but also during action observation, which has led many to suggest that these neurons are the substrate for action understanding.

Mirror-neurons were first discovered in the premotor area, F5, of the macaque monkey (Di Pellegrino et al. 1992; Gallese et al. 1996; Rizzolatti et al. 2001; Umilta et al. 2001) and have been identified subsequently in an area of inferior parietal lobule, area PF (Gallese et al. 2002;

Outline

1. Our theory: *predictive learning of sensorimotor information as a key for cognitive development*

2. Cognitive development in robots based on *predictive learning*
   - Self-other cognition & imitation
   - Altruistic behavior

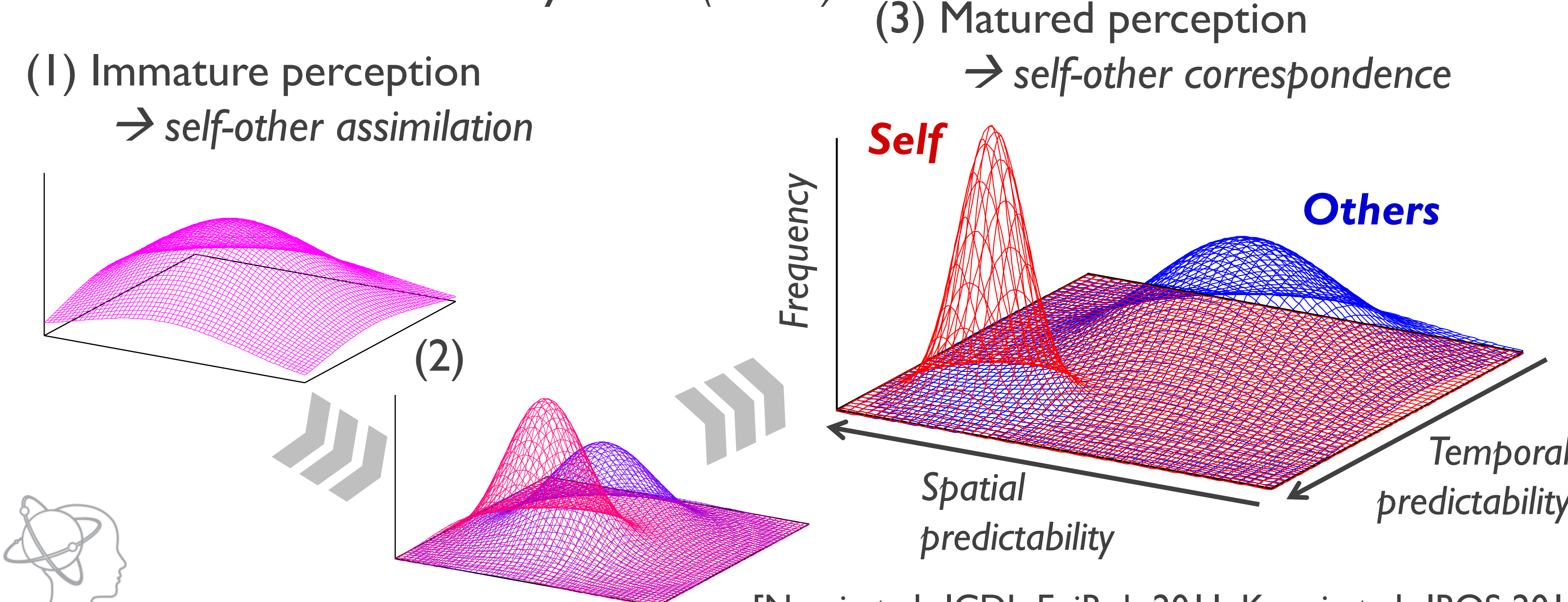
3. Autism spectrum disorder (ASD) caused by *atypical tolerance for prediction error*
   - Simulator of atypical perception
   - Local processing bias caused by neural imbalance

Young Infants Cannot Recognize Self – Why?

(Adapted from "The Baby Human 2" Discovery Channel)

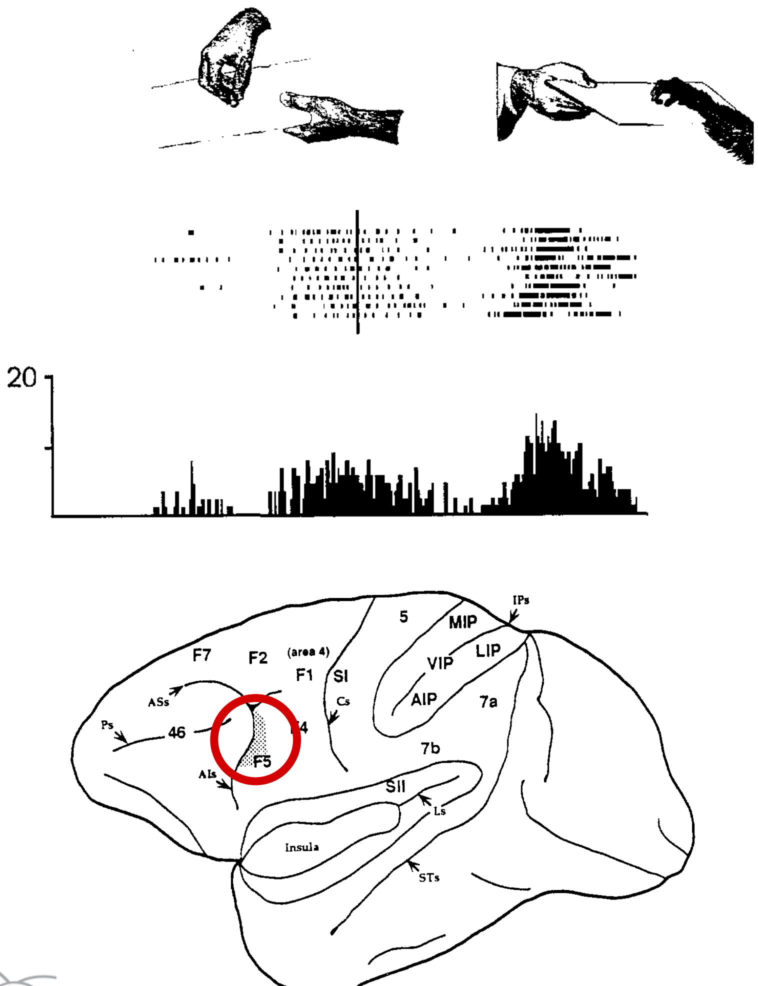
Our Hypothesis about Self-Other Cognition

- **Spatiotemporal predictability in sensorimotor information** discriminates the self from others.
  - Self = *perfect* predictability, others = *lower* predictability
  - Perceptual development leads to the emergence of *Mirror Neuron Systems (MNS).*



[Nagai et al., ICDL-EpiRob 2011; Kawai et al., IROS 2012]

Mirror Neurons &
Mirror Neuron System (MNS)



[Rizzolatti et al., 1996]

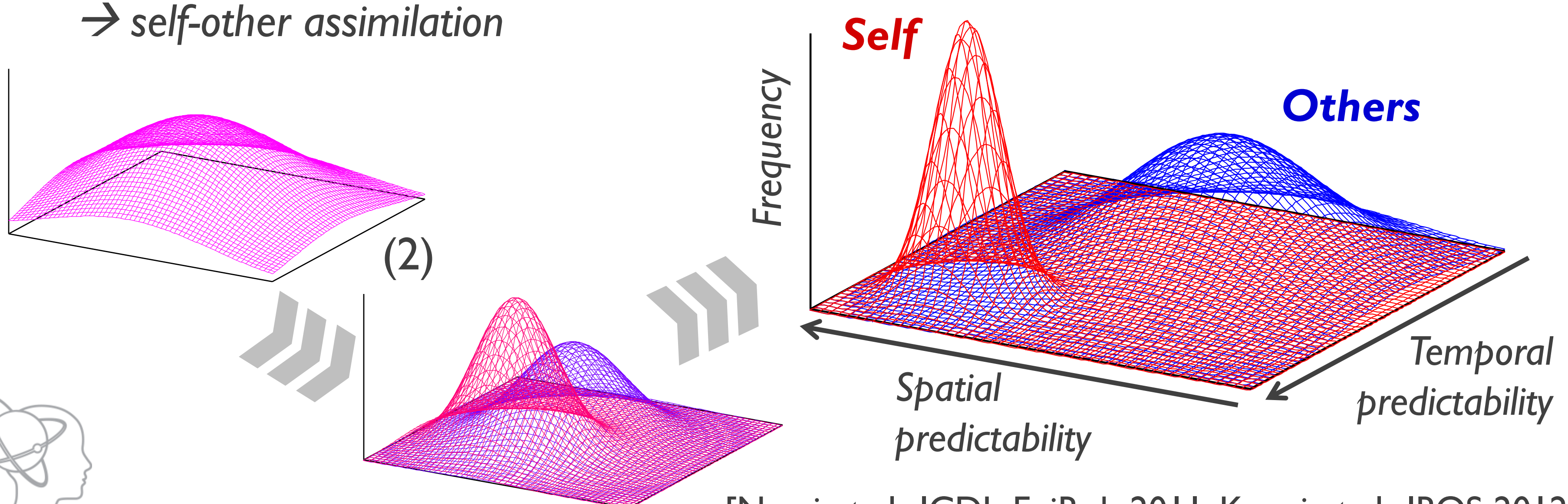
- Found in monkey's premotor cortex [Rizzolatti et al., 1996]
- Discharge both:
  - when **executing** an action
  - when **observing** the same action performed by **other individuals**
- Roles of MNS
  - Understanding the goal and intention of others' action
  - Imitation
  - etc.

Our Hypothesis about Self-Other Cognition

- **Spatiotemporal predictability in sensorimotor information** discriminates the self from others.
  - Self = *perfect* predictability, others = *lower* predictability
  - Perceptual development leads to the emergence of *Mirror Neuron Systems (MNS).*

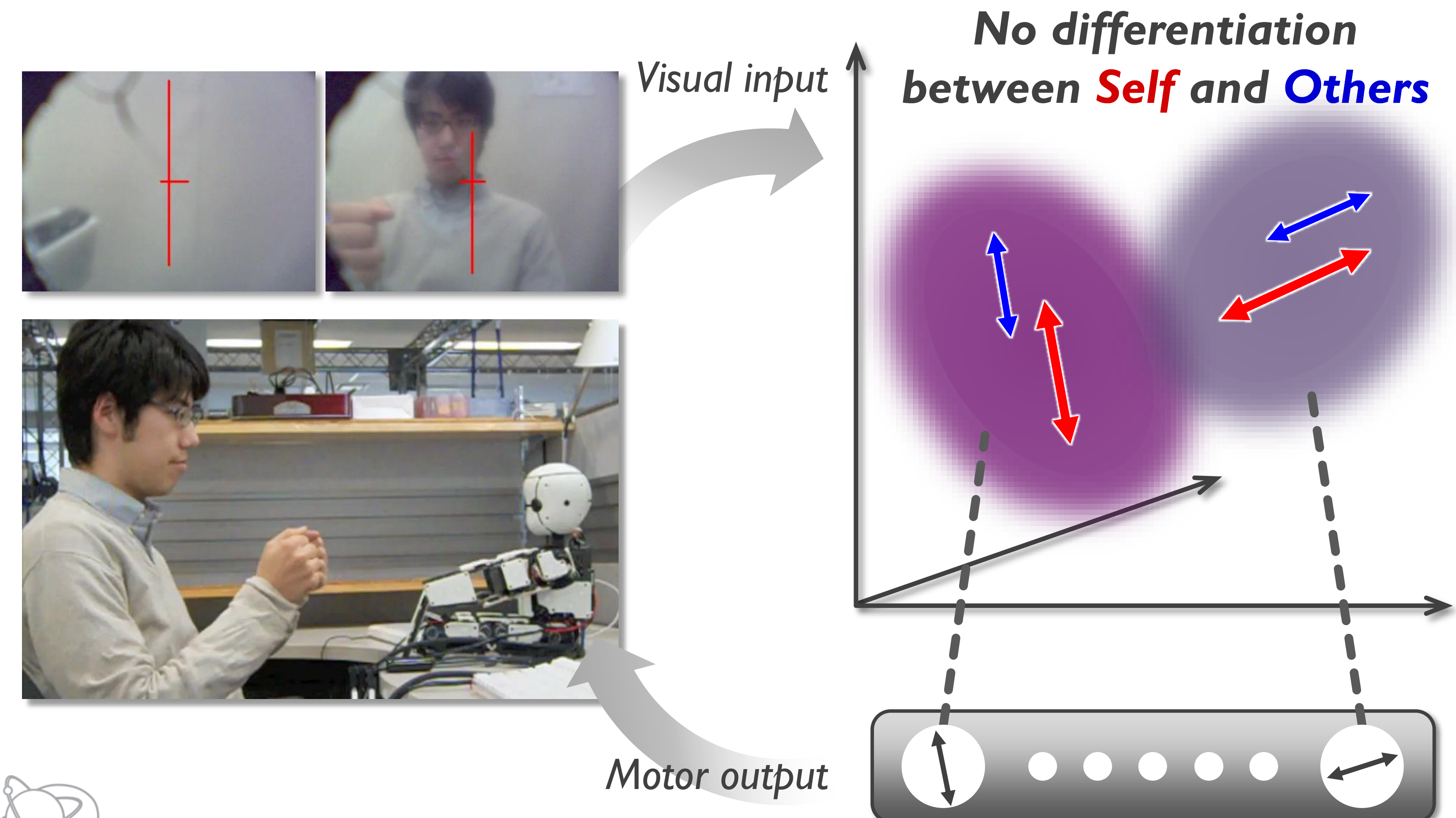
(1) Immature perception
→ *self-other assimilation*

(3) Matured perception
→ *self-other correspondence*



[Nagai et al., ICDL-EpiRob 2011; Kawai et al., IROS 2012]

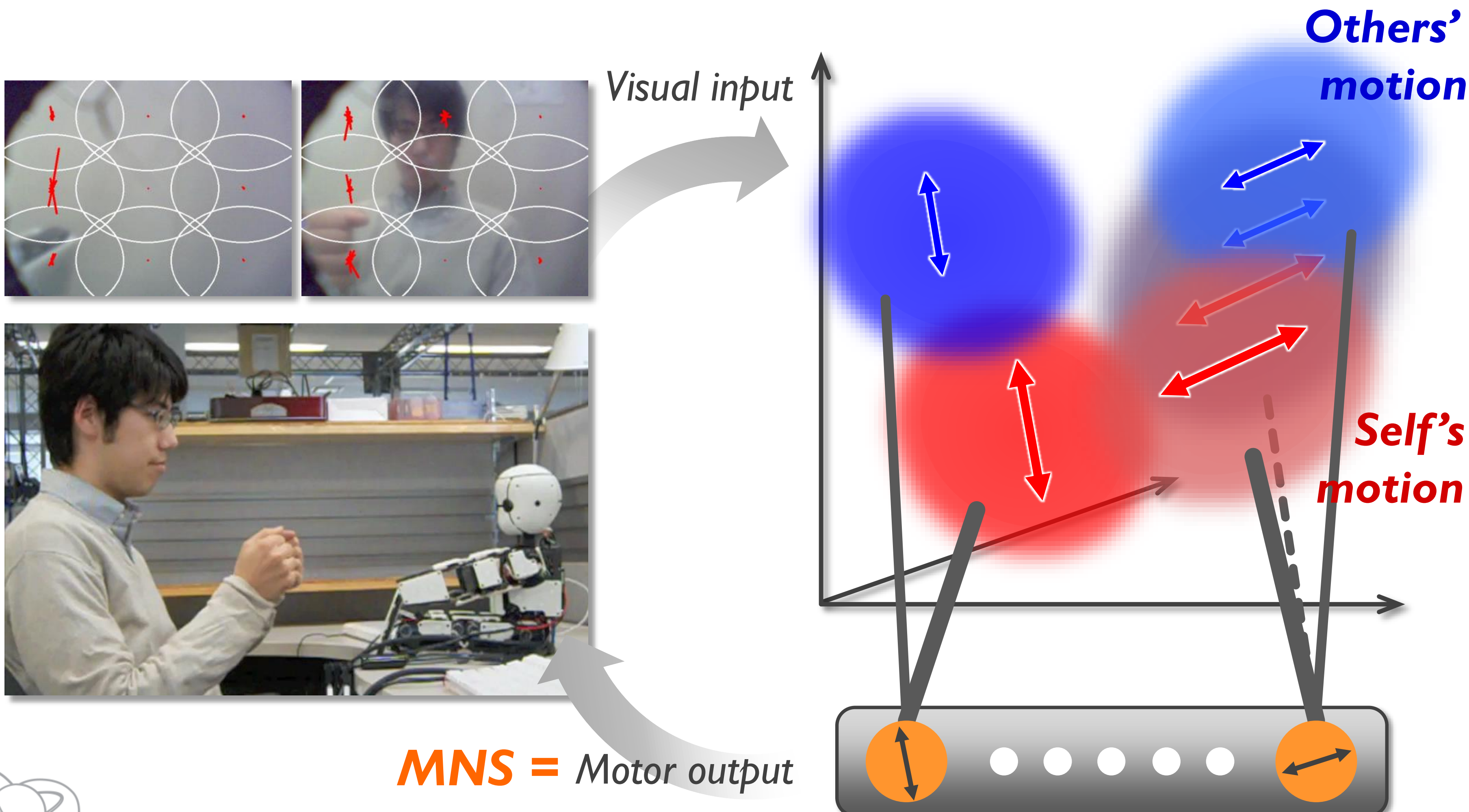
Computational Model for Emergence of MNS
− *Early* Stage of Development −



[Nagai et al., ICDL-EpiRob 2011; Kawai et al., IROS 2012]

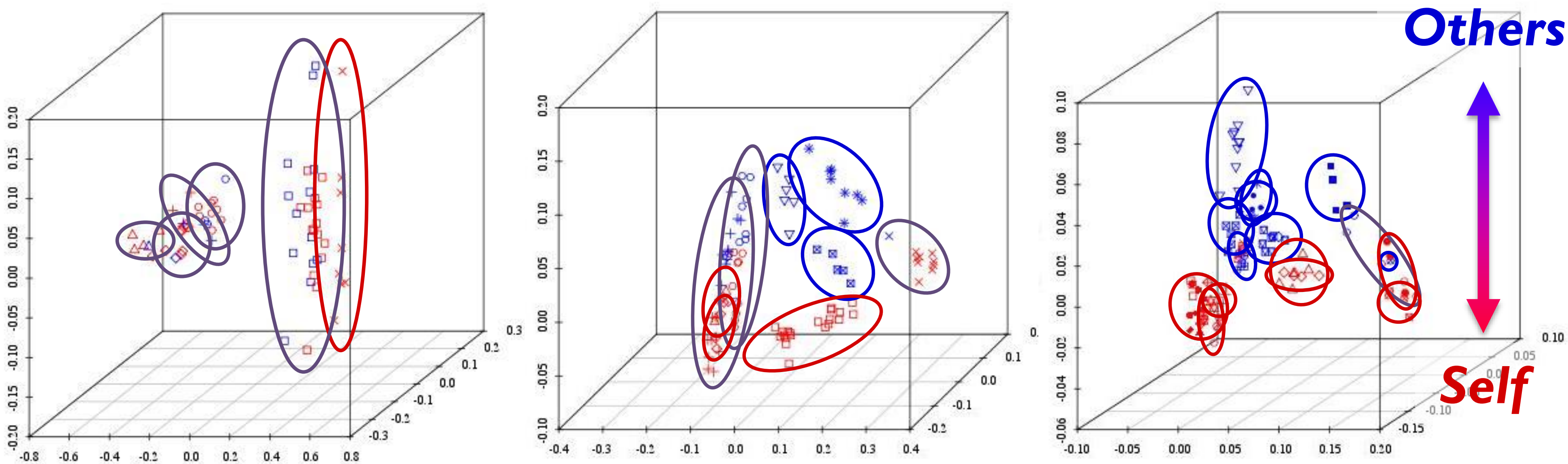
Computational Model for Emergence of MNS

− *Later* Stage of Development −



[Nagai et al., ICDL-EpiRob 2011; Kawai et al., IROS 2012]

Result 1: Self-Other Discrimination through Visual Development

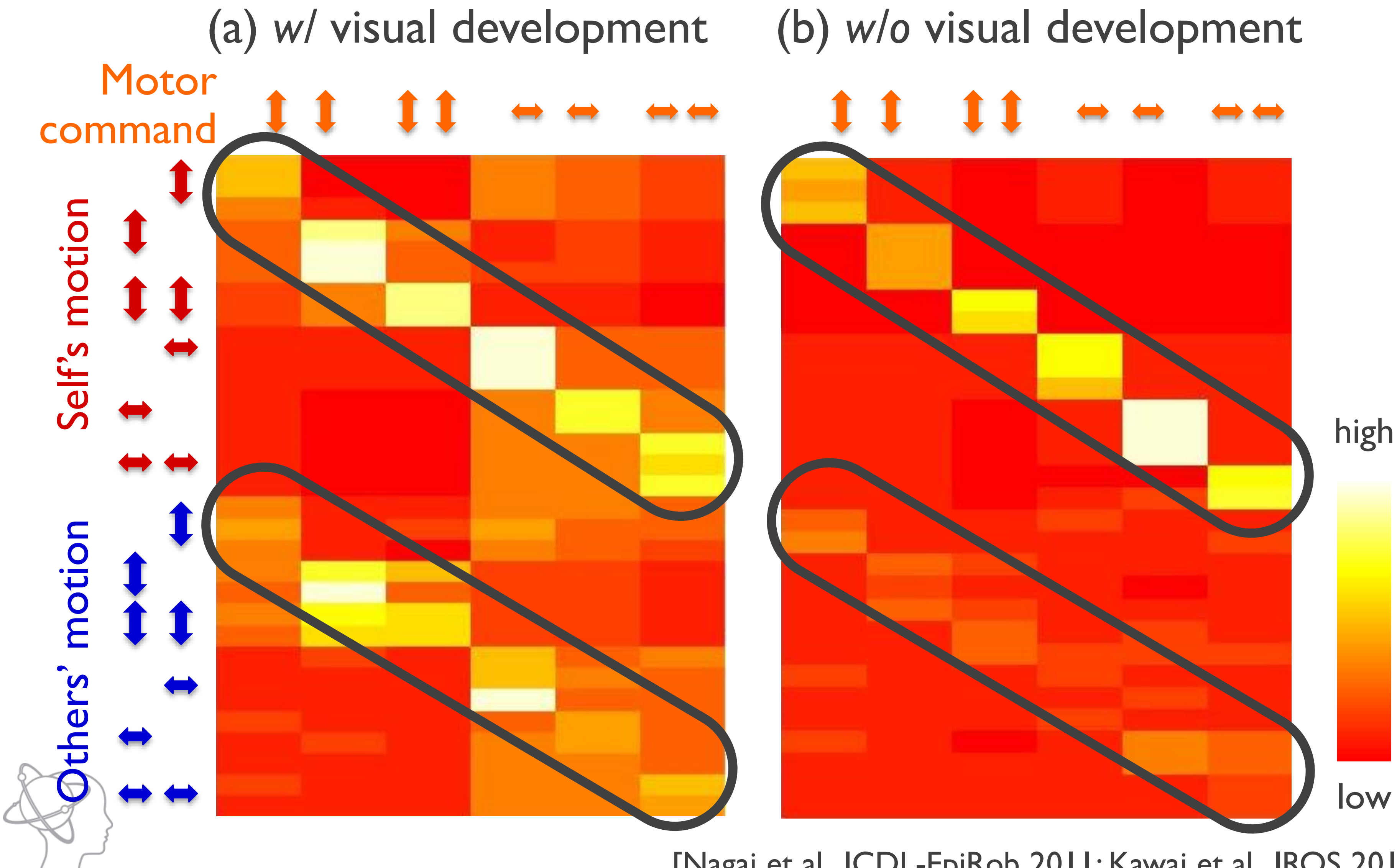


No differentiation · Self's motion · Others' motion

Visual development

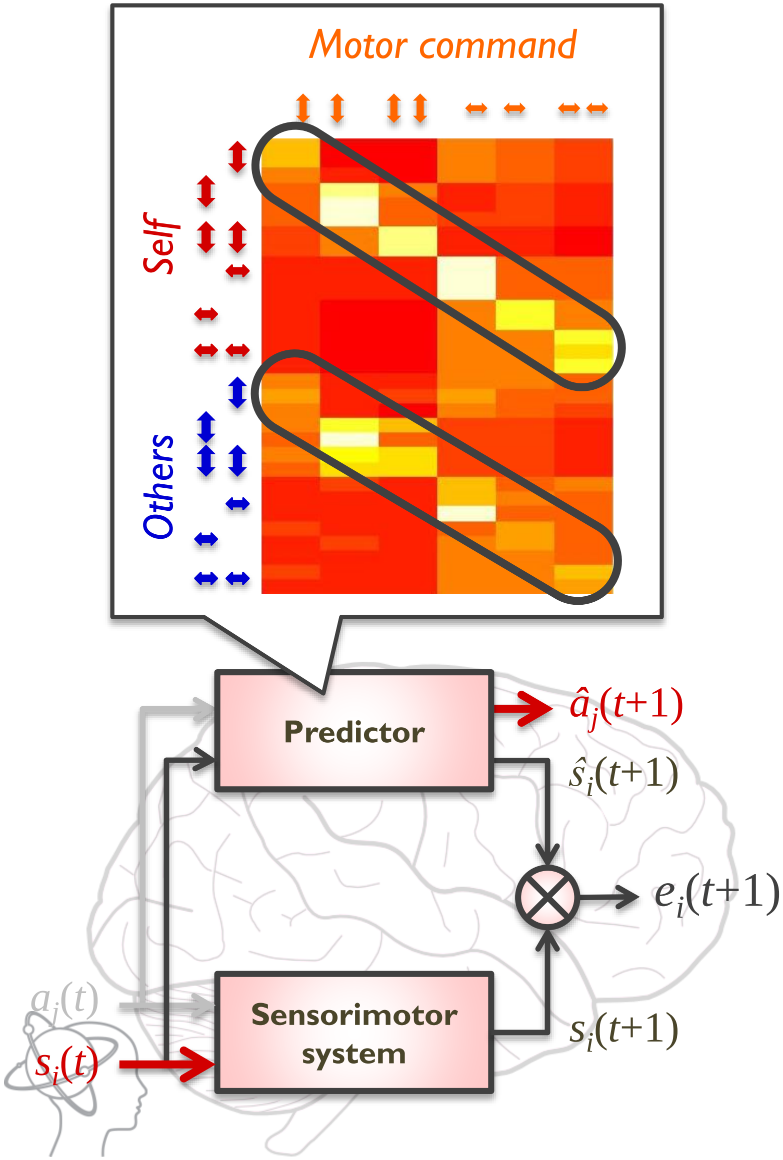
[Nagai et al., ICDL-EpiRob 2011; Kawai et al., IROS 2012]

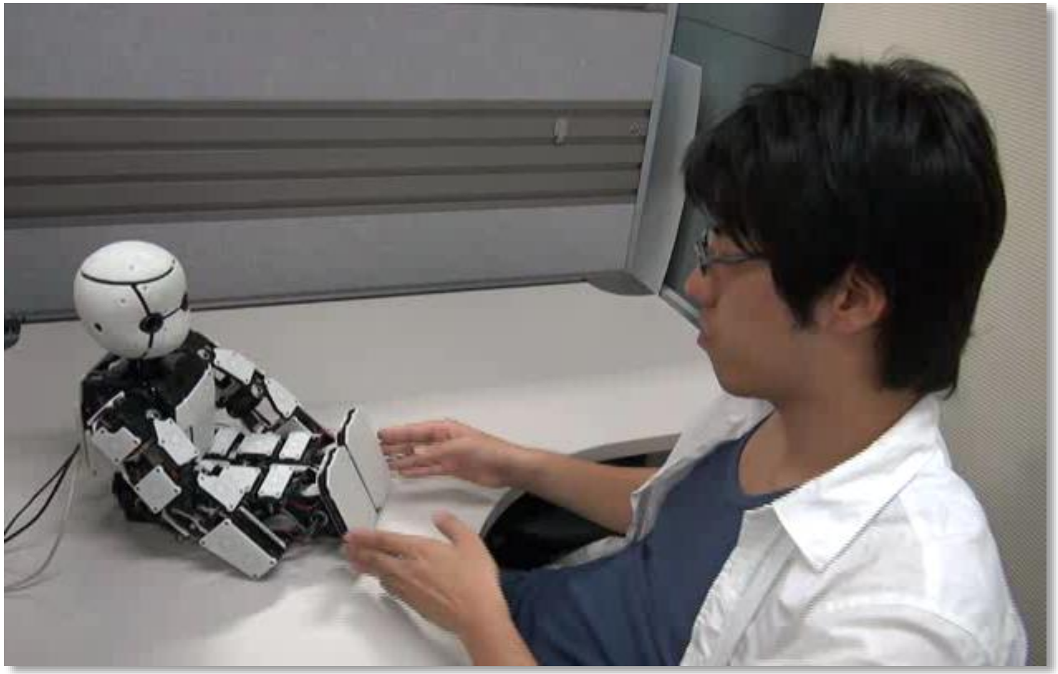
Result 2: MNS Acquired in Sensorimotor Mapping



[Nagai et al., ICDL-EpiRob 2011; Kawai et al., IROS 2012]

Result 3: Imitation Using Acquired MNS





[Nagai et al., ICDL-EpiRob 2011; Kawai et al., IROS 2012]

Infants Help Others Without Reward – Why?

[Warneken & Tomasello, 2006]

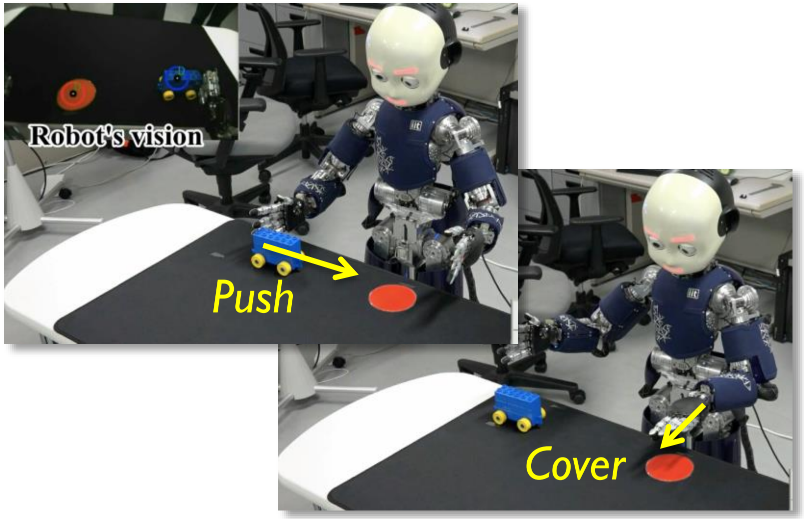
Two Theories for Altruistic Behaviors

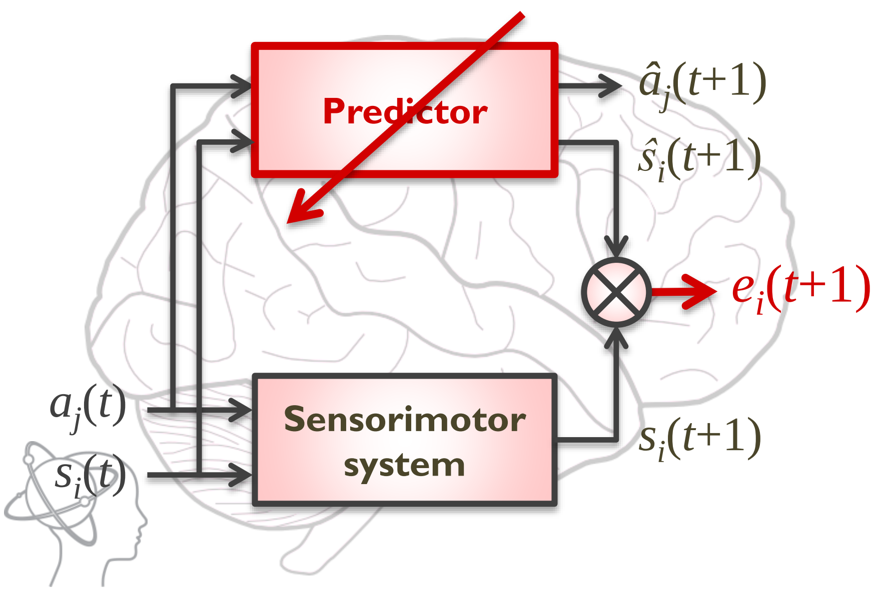
[Paulus, 2014]

[Warneken & Tomasello, 2006]

- **Emotion-sharing theory**
  - Understand other person as an *intentional agent* [Batson, 1991]
  - Be motivated to help other based on *empathic concern for other's needs* [Davidov et al., 2013]
  - Self-other differentiation
- **Goal-alignment theory**
  - Understand *other's goal*, but not his/her intention [Barresi & Moore, 1996]
  - *Take over other's goal* as if it were infant's own
  - *No* self-other discrimination

Our Hypothesis about Emergence of Altruistic Behavior

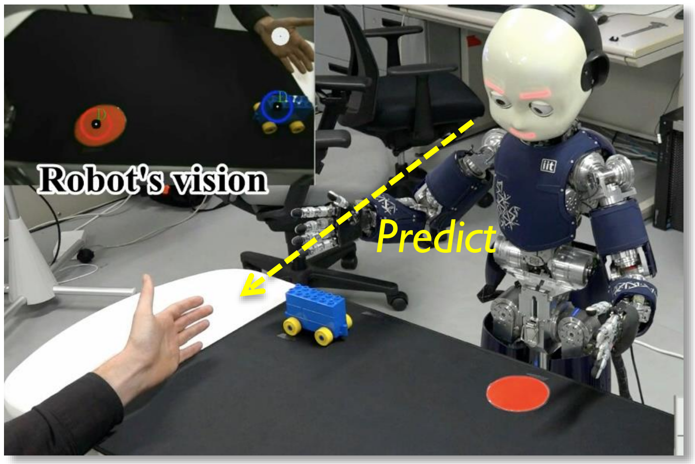


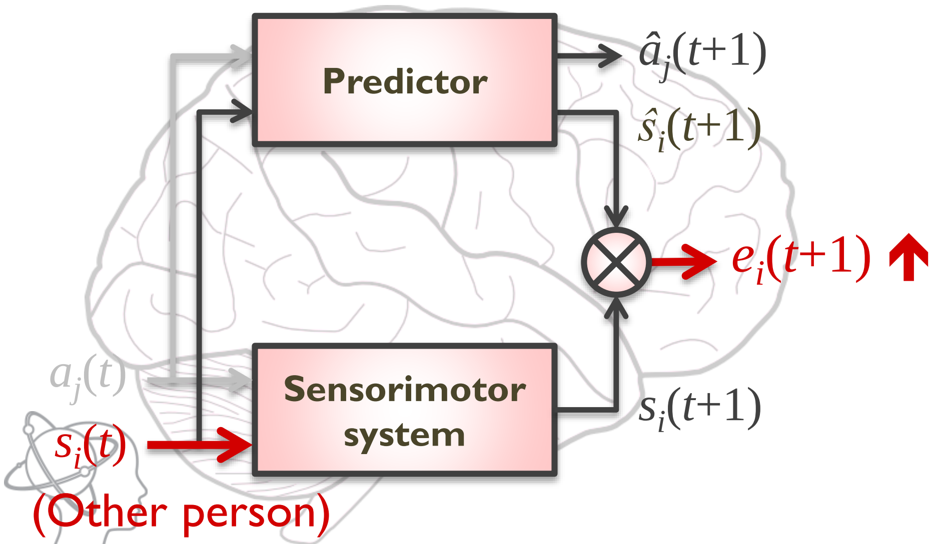


1. Learn the predictor by minimizing the prediction error $e_i(t+1)$ through the robot's own experiences

[Baraglia, Nagai, & Asada, ICDL-EpiRob 2014]

Our Hypothesis about Emergence of Altruistic Behavior



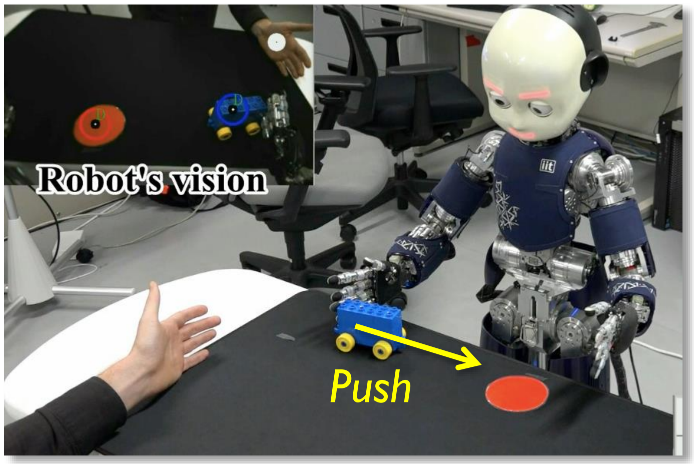


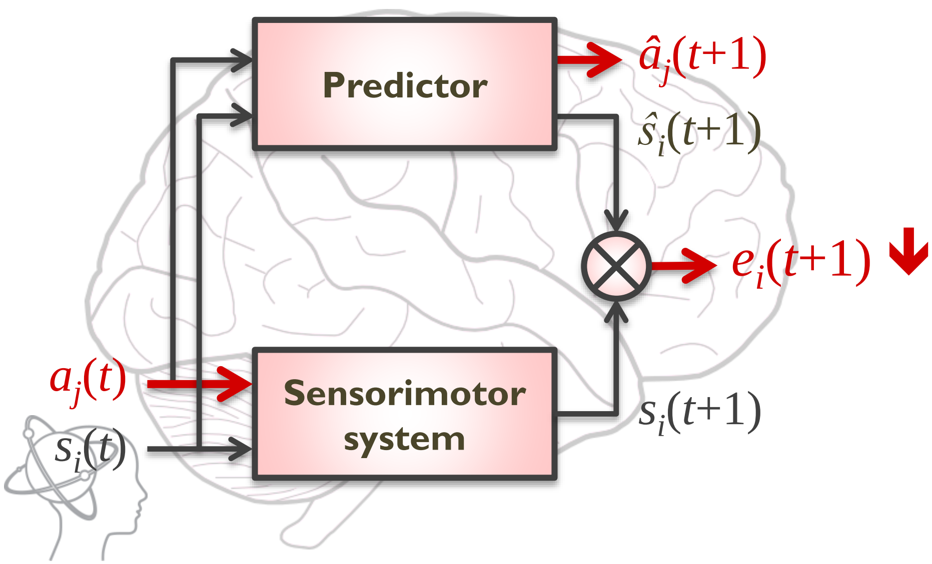
1. Learn the predictor to minimize the prediction error $e_i(t+1)$ through the robot's own experiences

2. Estimate $e_i(t+1)$ while observing other's action $s_i(t+1)$

[Baraglia, Nagai, & Asada, ICDL-EpiRob 2014]

Our Hypothesis about Emergence of Altruistic Behavior



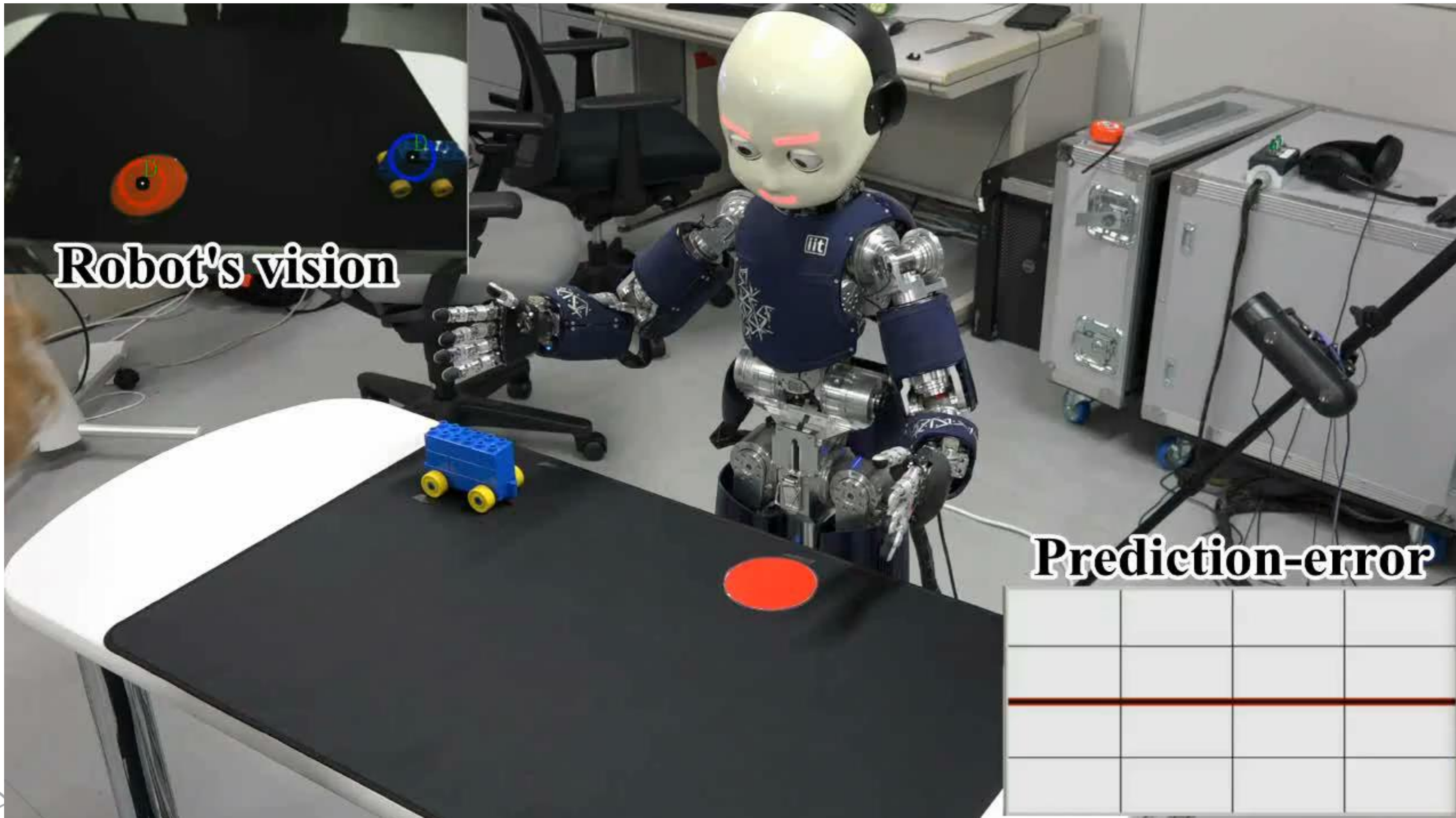


1. Learn the predictor to minimize the prediction error $e_i(t+1)$ through the robot's own experiences
2. Estimate $e_i(t+1)$ while observing other's action $s_i(t+1)$
3. Execute the action $\hat{a}_j(t+1)$ to minimize $e_i(t+1)$ if $e_i(t+1) > \text{threshold}$

→ **Altruistic behavior**

[Baraglia, Nagai, & Asada, ICDL-EpiRob 2014]

Result: Emergence of Altruistic Behavior



[Baraglia, Nagai, & Asada, ICDL-EpiRob 2014]

Outline

1. Our theory: *predictive learning of sensorimotor information* as a key for cognitive development

2. Cognitive development in robots based on *predictive learning*
   – Self-other cognition & imitation
   – Altruistic behavior

3. Autism spectrum disorder (ASD) caused by *atypical tolerance for prediction error*
   – Simulator of atypical perception
   – Local processing bias caused by neural imbalance

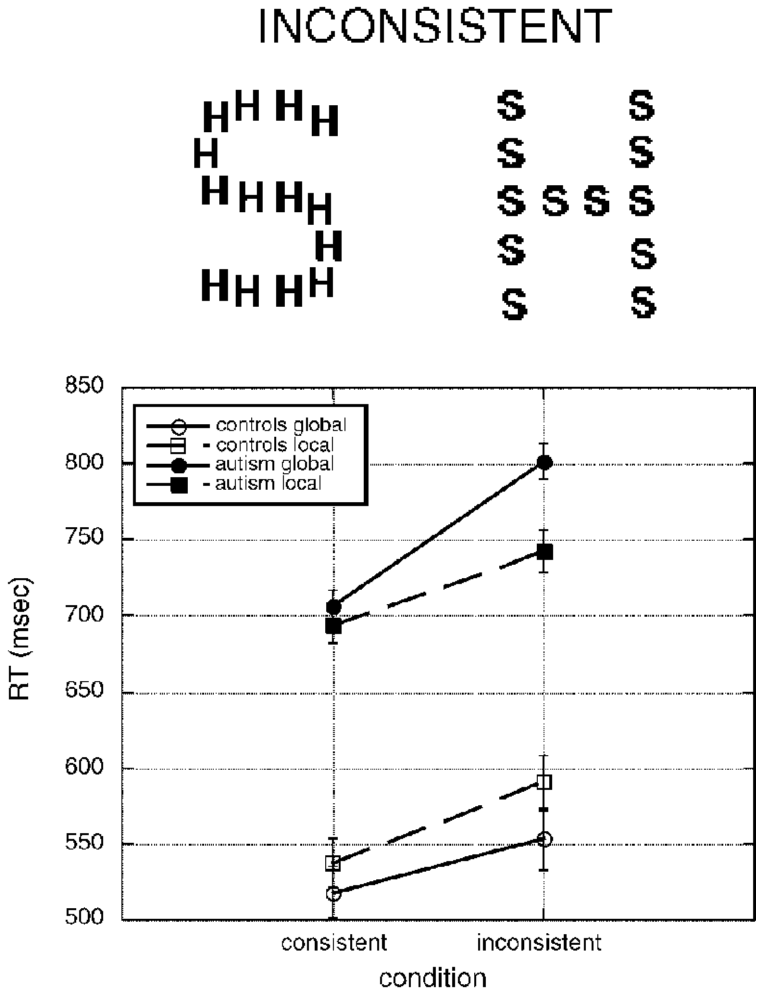
Autism Spectrum Disorder (ASD)

- Difficulties in **social interaction** [Baron-Cohen, 1995; Charman et al., 1997; Mundy et al., 1986]
  - Less eye contact
  - Difficulties in reading emotion
  - Lack of theory of mind, etc.

- Atypical **perception** and **information processing** [O'Neill & Jones,1997; Happé & Frith, 2006; Ayaya & Kumagaya, 2008]
  - Hyperesthesia/hypoesthesia
  - Local processing bias, etc.

Examples of Atypical Perception in ASD

[Ayaya & Kumagaya, 2008]



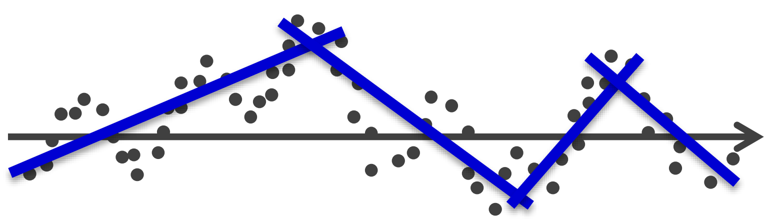
[Behrmann et al., 2006]

Our Hypothesis about Mechanism of ASD

- ASD might be caused by **an atypical tolerance for prediction error** in predictive learning.
  [Ayaya & Kumagaya, 2008; Nagai, in press]

Typically developing people

Proper tolerance for prediction error

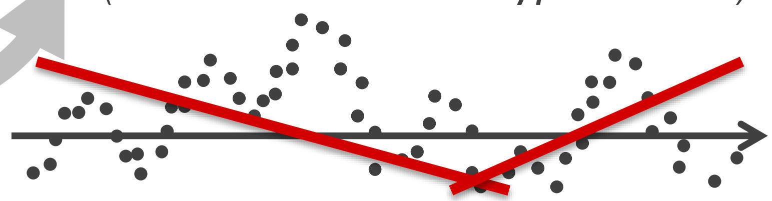


Sensorimotor information

People with ASD

Atypical tolerance for prediction error

(smaller tolerance → hyperesthesia)



(larger tolerance → hypoesthesia)

Simulator of Atypical Perception in ASD

[Qin et al., ICDL-EpiRob 2014; Nagai et al., in prep.]

Two Challenges in Developing ASD Simulator

1. **Objective evaluation**

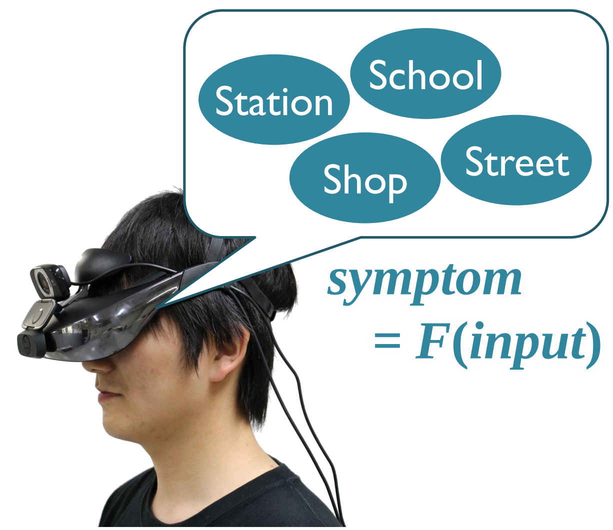
   Atypical perception is subjective experiences.

   → How to **objectively evaluate** the experiences?

2. **Quantitative evaluation**

   Atypical perception is associated with social contexts.

   → How to **quantitatively evaluate** the social contexts?

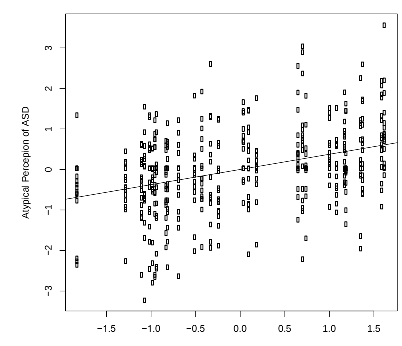


Our Approach Employing Computational Techniques

1. Prepare multiple patterns of atypical perception using visual processing techniques

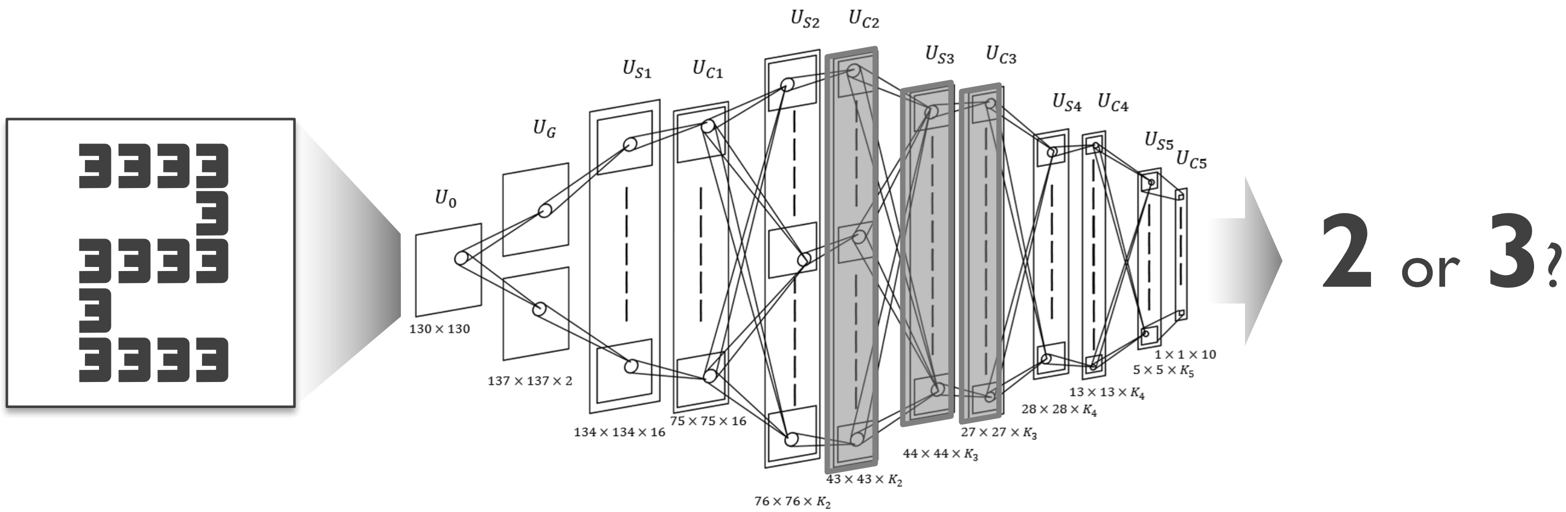
2. Ask ASD participants to **reproduce their experiences** using the prepared patters
   - Select experienced pattern
   - Adjust its strength

3. Analyze the correlation between social contexts and atypical perception

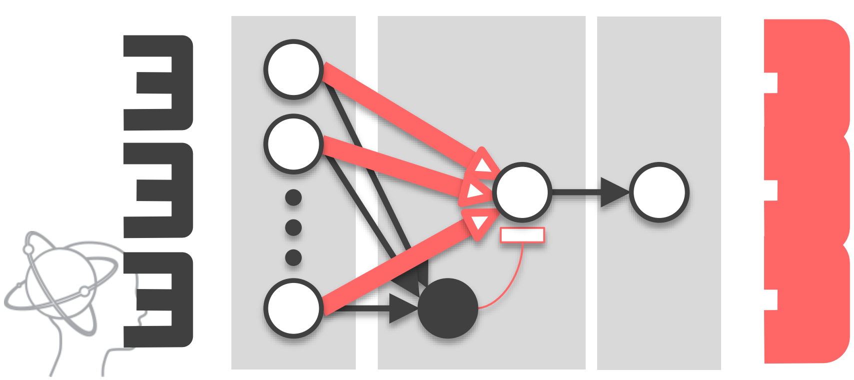


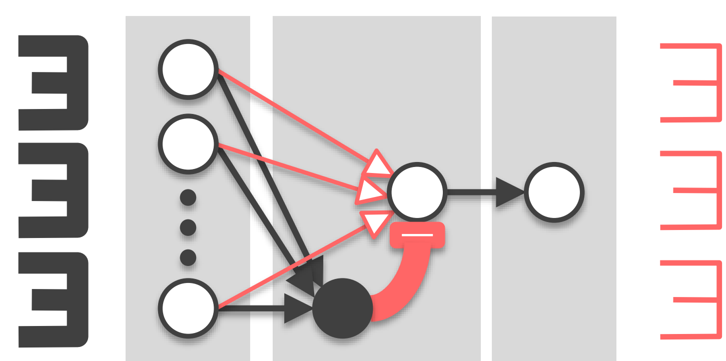
A Neural Network Model for Atypical Perception in ASD [Nagai et al., CogSci 2015]

- **Imbalance between excitatory and inhibitory** connections causes local processing bias in ASD.

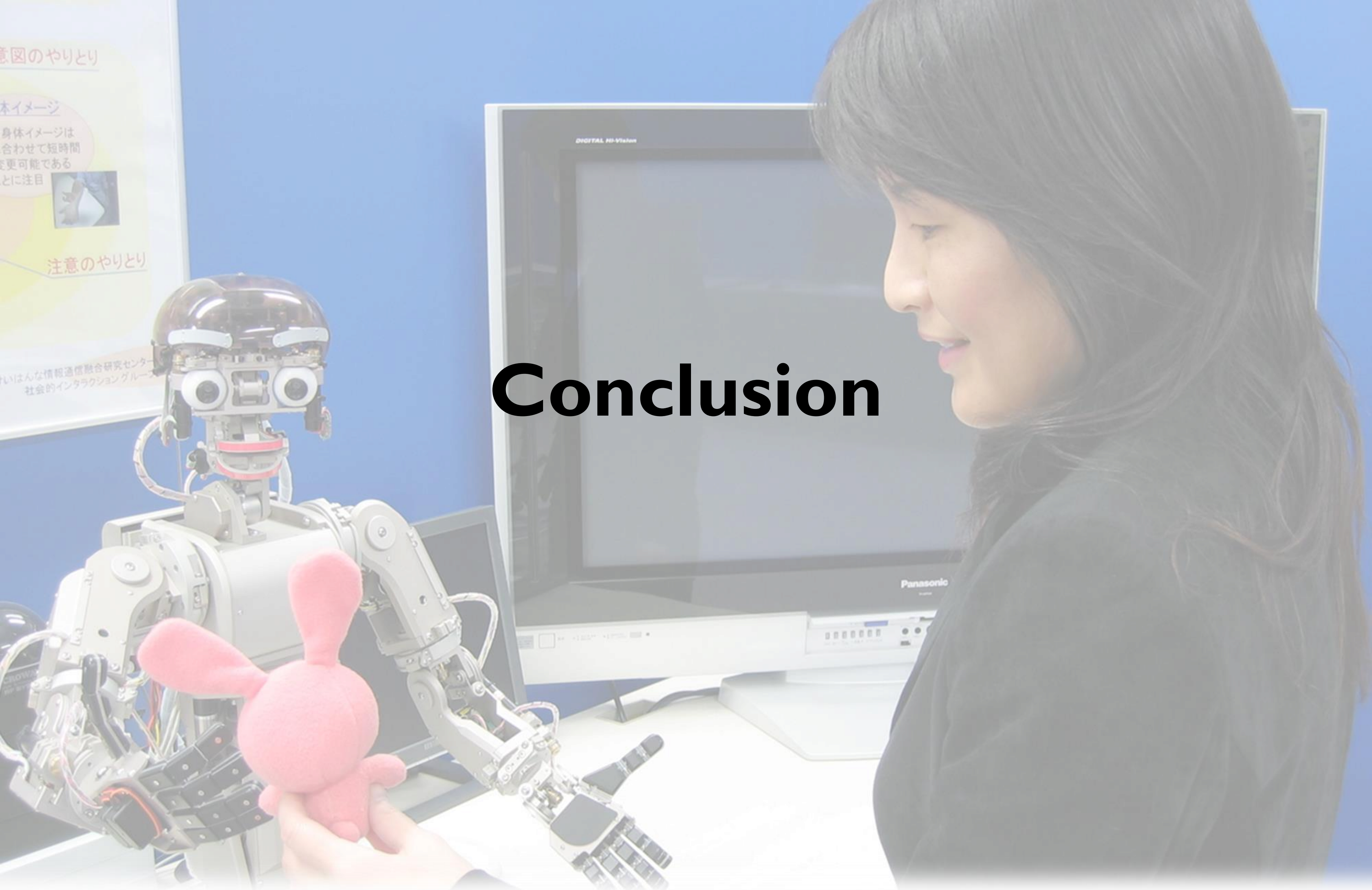


- Weaker inhib. → global ("2") bias
- Stronger inhib. → local ("3") bias





Conclusion



Our Theory about Cognitive Development

[Nagai, in press]

Predictive learning of sensorimotor information (i.e., minimizing prediction error $e_i(t+1)$) leads to cognitive development.

(1) *Learn the predictor* through sensorimotor experiences

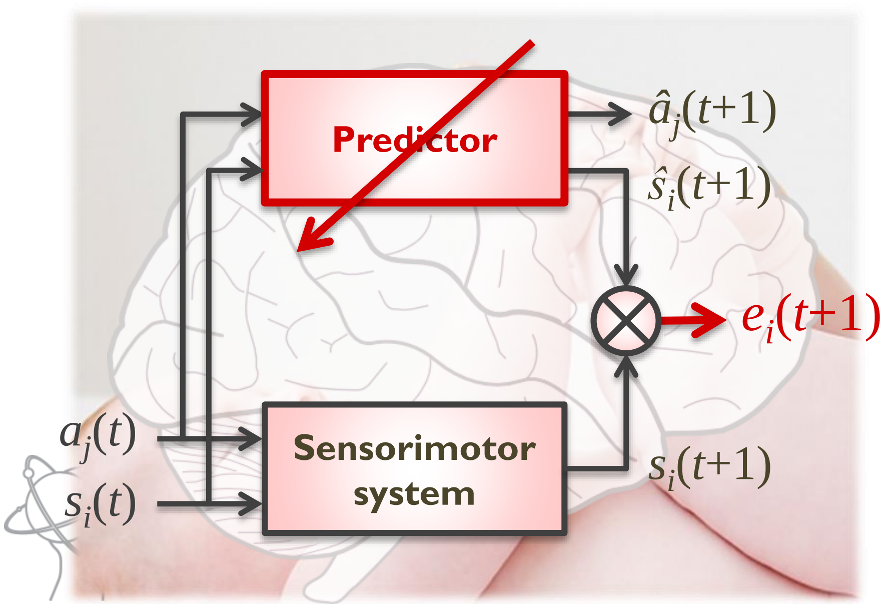
→ Self-other cognition

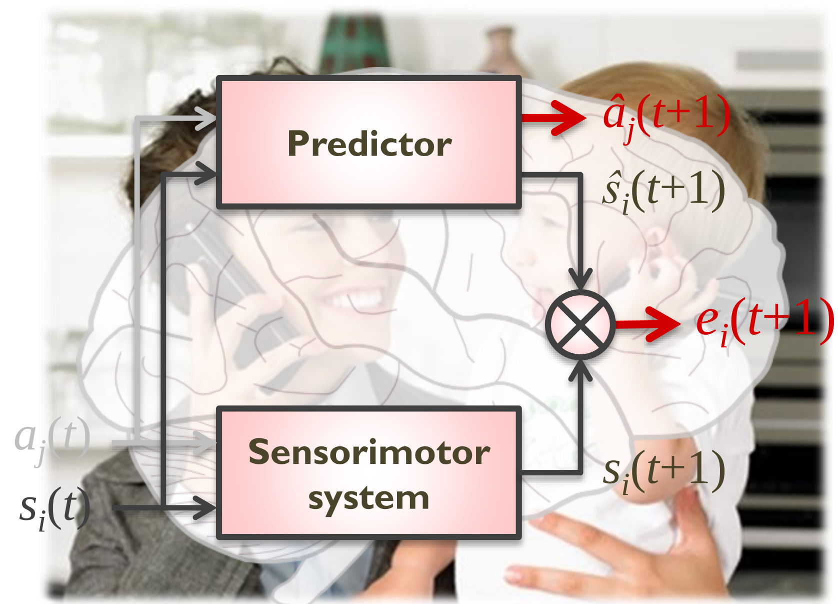
→ Goal-directed action, etc.

(2) *Produce an action* in response to other's action

→ Imitation

→ Altruistic behavior, etc.





Our Theory about Cognitive Development

[Nagai, in press]

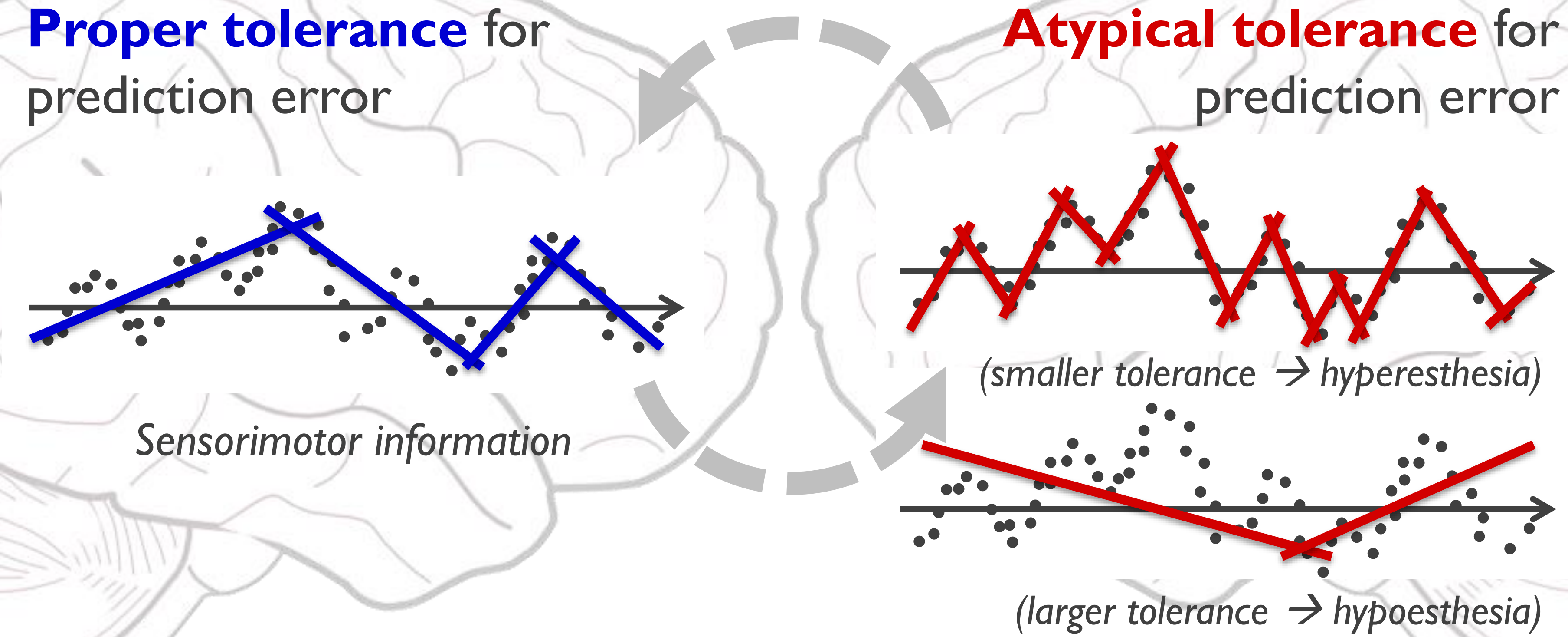
- ASD might be caused by **an atypical tolerance for prediction error** in predictive learning.
  [Ayaya & Kumagaya, 2008; Nagai, in press]

Typically developing people

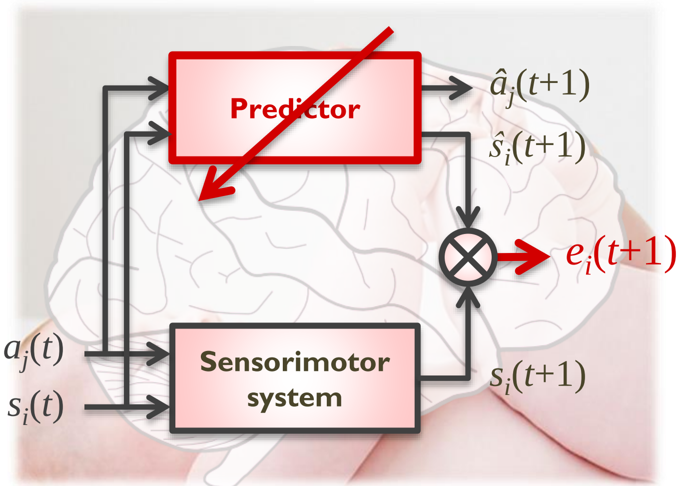
Proper tolerance for prediction error

People with ASD

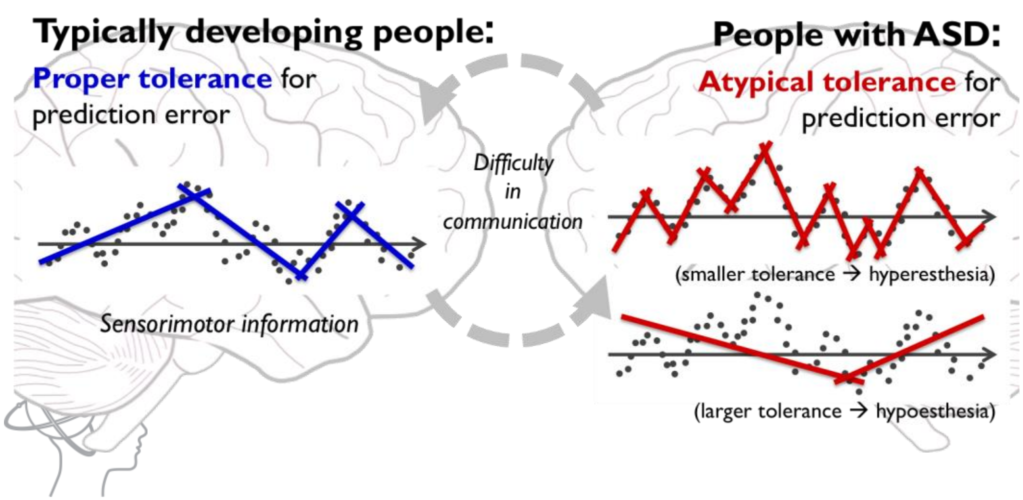
Atypical tolerance for prediction error



What is Consciousness?
Relationship to Predictive Learning?



- Consciousness ∝ prediction error
  - Learning *new* actions (e.g., walking for babies)→ *conscious*
  - Executing *acquired* action (e.g., walking for adults) → *unconscious*



- Individuals with ASD
  - Often *producing prediction error* due to smaller tolerance
  - Difficulty in developing unconscious process

Thank You!

Osaka University

- Minoru Asada
- Jimmy Baraglia
- Yuji Kawai
- Takakazu Moriwaki
- Shibo Qin
- Many students

University of Tokyo

- Shinichiro Kumagaya
- Satsuki Ayaya

KAIST

- Jun-Cheol Park

devsci
Constructive Developmental Science

Constructive Developmental Science
Based on Understanding the Process
from Neuro-Dynamics to Social Interaction

yukie@ams.eng.osaka-u.ac.jp
http://cnr.ams.eng.osaka-u.a.jp/~yukie/