Investigation of the Sense of Agency in Social Cognition, Based on Frameworks of Predictive Coding and Active Inference: A simulation Study on Multimodal Imitative Interaction

Wataru Ohata and Jun Tani
Cognitive Neurorobotics Research Unit (contact: wataru.ohata@oist.jp)

ABSTRACT

When agents interact socially, conflicts are inevitable. Although how social agents can resolve such problems autonomously has not been determined, dynamic characteristics of agency may cast light on the underlying mechanism. In this work, we employed predictive coding and active inference as theoretical frameworks and studied the sense of agency (SoA) in social interaction. Hypothesizing that complexity regulation of agent's model should affect the strength of the sense of agency and the interaction, we evaluated it by conducting simulation experiments of multimodal imitative interaction between a humanoid robot and a human. The results showed that with tight regulation, the agent showed a weak SoA, and with loose regulation, the agent exerted a strong SoA[1].

Keywords: sense of agency, predictive coding, active inference, multimodal perception, human-robot interaction, recurrent neural network, variational Bayes

MOTIVATION

In social interactions,
- humans sometimes cooperate while at other times conflic. What determines the type of interaction and how?
- how their *agency* is characterized?

APPROACH AND HYPOTHESIS:

Theoretical framework
- **Predictive Coding** for perception
- Active Inference for action generation

Perception is achieved by maximizing **the evidence lower bound (ELBO)** of **generative model** (a.k.a. *free-energy* minimization)

$$\ln p(X) \geq \underbrace{\int q(z|X) \ln \frac{p(X,z)}{q(z|X)} dz}_{\text{Evidence Lower Bound}} = \underbrace{\mathbb{E}_{q(z|X)}[\ln p(X|z)]}_{\text{Accuracy}} - \underbrace{D_{\mathrm{KL}}[q(z|X) \| p(z)]}_{\textit{Complexity}}$$

($q(\cdot)$ is approximate posterior)

HYPOTHESIS:

Regulation of the complexity should
- affect the strength of the agent's **sense of agency (SoA)**
- have a significant impact on social interaction

The feeling that *"I am the one gererating this action"* based on the match between the agent's intention in acting and the outcome

MODEL

Imitative interaction between a humanoid robot and a human

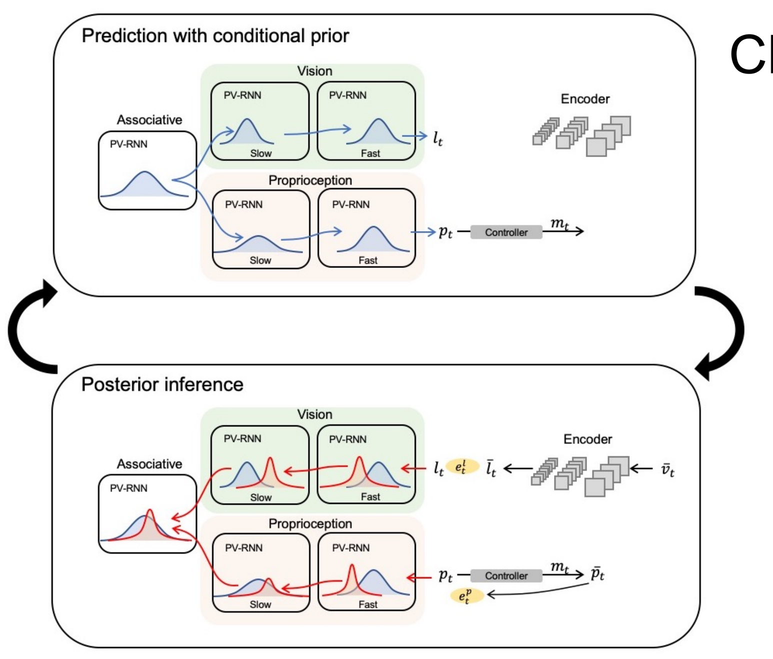


Figure 1. A schematic of the cycle of prediction with conditional prior and posterior inference

Characteristics:
- Variational Bayes Recurrent Neural Network (PV-RNN[2])
- Multimodal
  - Vision (CNN-based)
  - Proprioception (joint angle)
- Perception and action generation through the cycle of
  - Prediction with prior
  - Posterior inference

Online posteior inference by iterative ELBO maximization using back-propagation through time

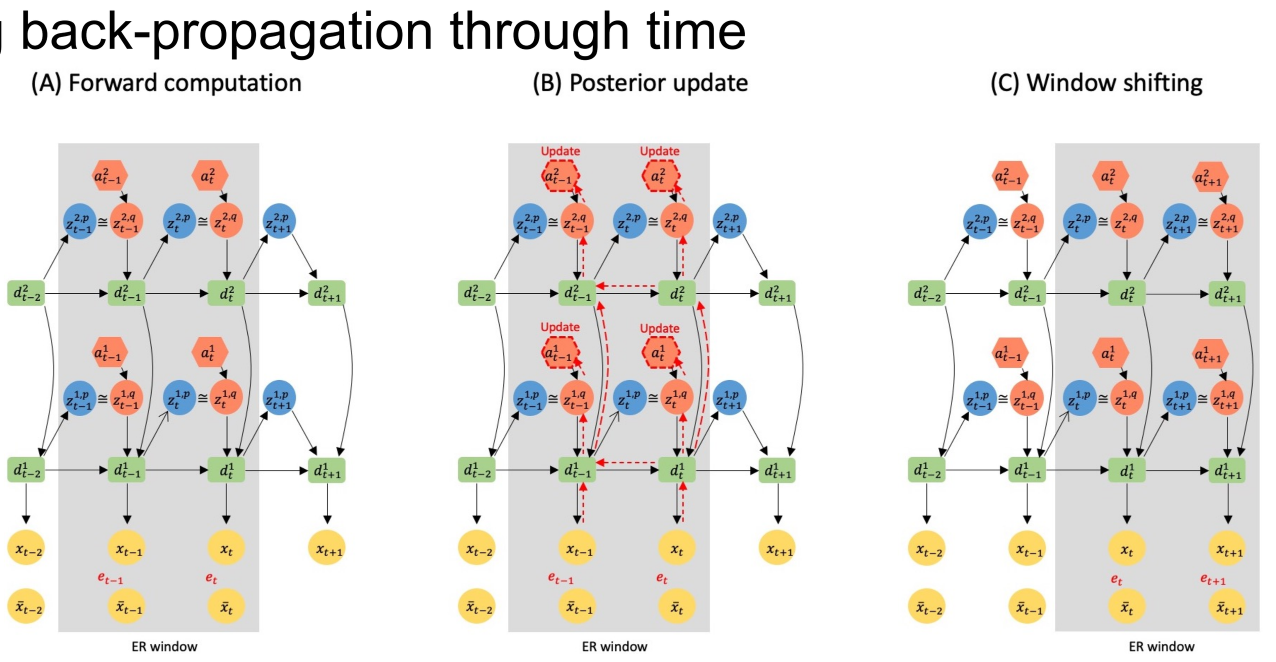


Figure 2. A schematic of posterior inference. At each sensory time step the network iterates (A) forward computation and (B) posterior update inside a window for the immediate past using back-propgation through time such that the ELBO is maximazed. At next time step, the window shifts (C) and the above iteration repeats.

EXPERIMENTAL DESIGN

The robot and the human imitate each other following a probabilistic finite state machine (PFSM) in simulation.

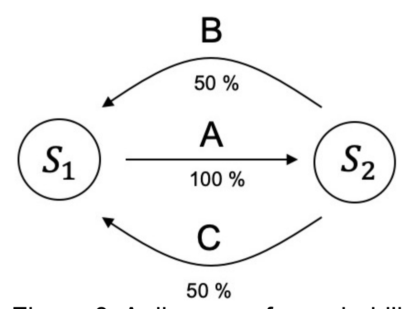


Figure 3. A diagram of a probabilistic finite state machine

- Composed of two states
- Transition with three primitive movements (A, B, and C)
- Deterministic and probabilistic transision

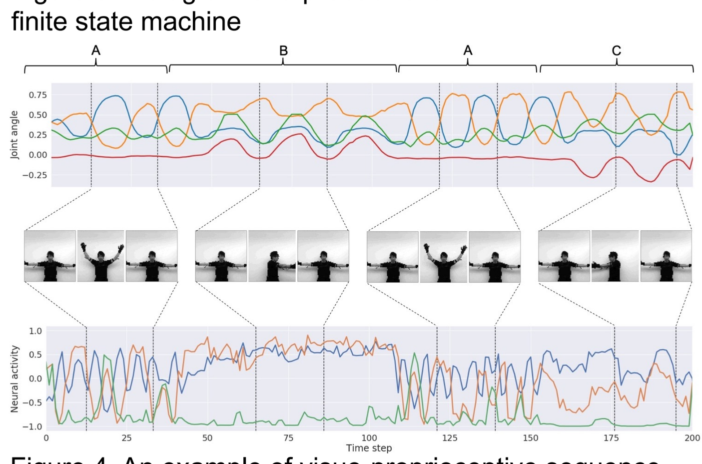
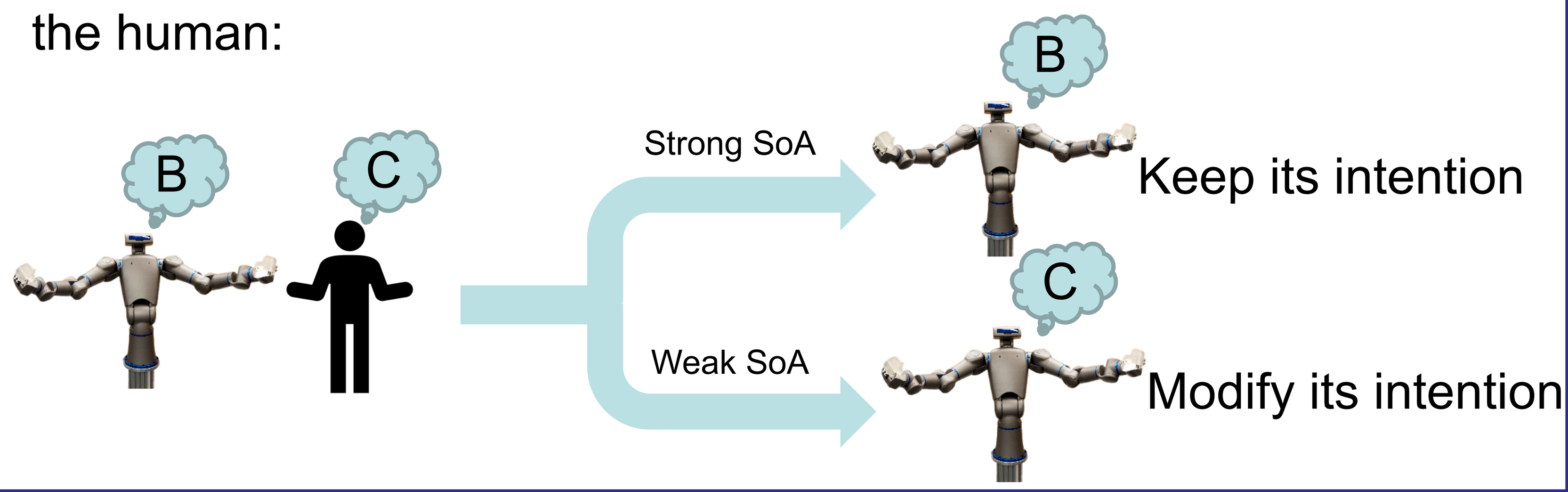


Figure 4. An example of visuo-proprioceptive sequence.

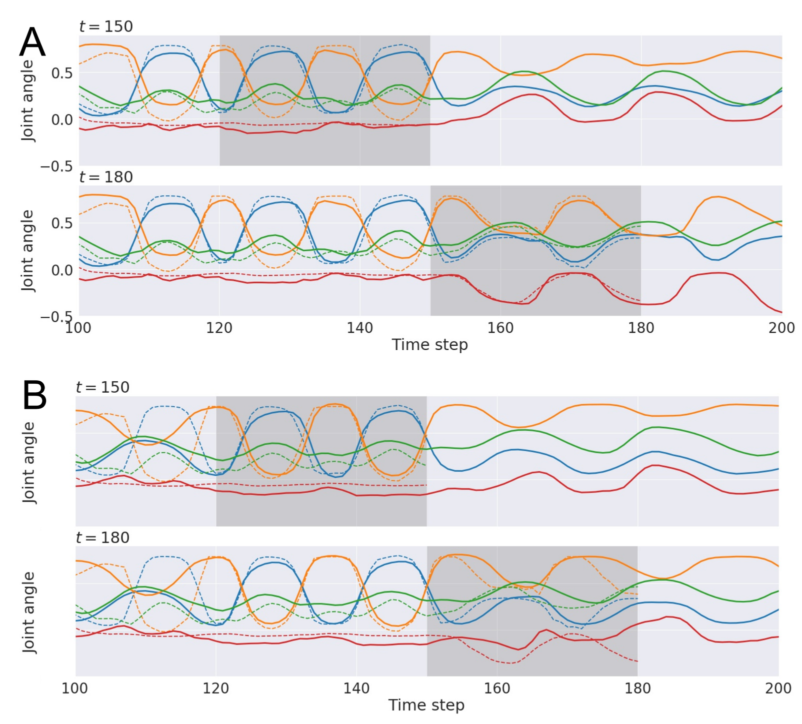
An example of visuo-proprioceptive sequence based on the PFSM
- Top row: joint angle trajectories
- Bottom row: latent representation of visual images

When the robot and the human make different predictions at the probabilistic state transition, for example, B for the robot and C for the human:



RESULTS

Simulation results with different conditions: tight and loose regulation



TIGHT regulation:
- Roconstruction error was minimized well
- Prediction was modified accordingly

LOOSE regulation:
- Roconstruction error was **NOT** minimized well
- Prediction was **NOT** modified much

Figure 5. A comparison of tight regulation (A) and loose regulation (B). Reconstruction of the past observation and the future prediction at time step 150 (top) and at time step 180 (bottom). Solid lines represent prediction outputs, and dashed lines represent observation. The shadowed area indicates the window for the posterior inference.

CONCLUSION AND FUTURE WORKS

- **TIGHT** complexity regulation → **STRONG** sense of agency
- **LOOSE** complexity regulation → **WEAK** sense of agency
- Next step is scaling up to physical experiments.
- Dynamic change of SoA leads to turn-takings in social interaction?

REFERENCES

[1] Ohata, W., & Tani, J. (2020). Investigation of the Sense of Agency in Social Cognition, based on frameworks of Predictive Coding and Active Inference: A simulation study on multimodal imitative interaction. Frontiers in Neurorobotics.
[2] Ahmadi, A., & Tani, J. (2019). A novel predictive-coding-inspired variational rnn model for online prediction and recognition. *Neural computation*, *31*(11), 2025-2074.

OKINAWA INSTITUTE OF SCIENCE AND TECHNOLOGY GRADUATE UNIVERSITY
沖縄科学技術大学院大学
1919-1 Tancha, Onna-son, Okinawa, 904-0495 Japan Phone. +81-98-966-8711
〒904-0495 沖縄県国頭郡恩納村字谷茶 1919-1 Phone. 098-966-8711
http ://www.oist.jp