Accounting for the Minimal Self and the Narrative Self: Robotics Experiments Using Predictive Coding

Jun Tani

[1] Cognitive Neurorobotics Research Unit
Okinawa Institute of Science and Technology
Onna-son, Okinawa, Japan 904-0495
jun.tani@oist.jp

Abstract. This paper proposes that the mind comprises emergent phenomena that appear via intricate and often conflicting interactions between top-down intentional processes involved in proactively acting on the external world, and bottom-up recognition processes involved in inferring possible causes for the resultant perceptual reality. This view has been tested via a series of neurorobotics experiments employing predictive coding principles implemented in "deep" recurrent neural network (RNN) models. The current paper illuminates phenomenological accounts of the minimal self and the narrative self from the analysis of those synthetic neurorobotics experiments.

Keywords: Predictive Coding, Robot, RNN, Self, Consciousness.

1 Introduction

If we suppose that entangling interactions between top-down intention and bottom-up recognition of perceptual reality from the objective world is essential in development of embodied minds, what sorts of models could best account for such dynamic interactions? The developmental psychologists Gibson and Pick [1] suggest that learning an action is not just about learning a motor command sequence, but that it also involves learning possible perceptual structures extracted during intentional interactions with the environment. This view can be explained by predictive coding [2].

In predictive coding frameworks, the learning process involves extracting causal structure between the intention for action and resultant perceptual reality, by means of prediction error minimization. It has been suggested that brains utilize specific macroscopic constraints, such as timescale differences and connectivity among local regions in terms of downward causation for development of composition/decomposition mechanisms [3]; therefore, it is expected that the predictive coding framework implemented in neural network models provided with such spatio-temporal constraints could extract hidden causal structure in a compositional manner from accumulated sensory-motor experiences [4]. Our group conducted neurorobotics experiments following the aforementioned considerations [5-7]. The current review of these robotics experiments suggests that the phenomenology of self and consciousness can be best

explained by self-organizing phenomena that emerge through intricate interaction between top-down intentional processes and bottom-up perceptual reality.

2 Neurorobotics experiments

Yamashita and Tani [8] proposed a predictive coding recurrent neural network model, known as the multiple timescale RNN (MTRNN), which is composed of stacks of continuous-time RNNs (CTRNN) with different timescales assigned to each. This model has been extended to a dynamic visual processing predictive coding model, characterized by its multiscale property, both in temporal and spatial dimensions in terms of its range of local connectivity [6]. Furthermore, Hwang et al. [7] integrated this visual processing predictive coding model and MTRNN for the purpose of conducting human-robot interaction experiments. The following provides a review of this study.

2.1 P-VMDNN Model

Fig. 1 illustrates the integrated model, the predictive visuo-motor dynamic neural network (P-VMDNN) [7], or experiment setup using a simulated humanoid robot and part of the information flow. P-VMDNN consists of the visual pathway, the proprioceptive (movement) pathway, and the associative layers. Each pathway consists of stacks of CTRNNs, where CTRNNs in the lower layer comprise neural units with smaller time constants, while those in higher layers have larger time constants. In addition to this timescale constraint, the visual pathway also utilizes spatial-scale constraint in terms of the connectivity range. As in the convolutional neural network, retinotopically allocated neural units are connected only locally in the lower layers, while those units in higher layers are connected fully in the same layer. The associative layers also comprise CTRNNs with larger time constants for each unit.

The whole network functions as a generative model. The current internal state in the highest associative layer, which encodes the current intention, dynamically drives internal states in the next layer through top-down connectivity. The drive propagates downward through layers to both the visual and proprioceptive pathways, and finally generates a prediction of the pixel pattern and joint angles of next time step in the lowest layers in the visual pathway and the proprioceptive pathway, respectively.

Training of the P-VMDNN is conducted in an end-to-end supervised manner. During the robot tutoring phase, a set of visuo-proprioceptive sequences consisting of video frame sequence and arm joint trajectories is sampled. This training enables the network to regenerate exemplar sequence patterns by adapting connectivity weights from the whole network. This is performed using a backpropagation-through-time (BPTT) algorithm [9]. Actual training of multiple sequence patterns uses initial sensitivity characteristics of nonlinear dynamics of the network. More specifically, training by BPTT infers optimal values for initial internal states in all layers, for each sequence, as well as all connectivity weights. After training converges, each training sequence can be regenerated with top-down prediction by setting values of the initial

internal states to those inferred for this sequence through training. By this means, it can be said that initial states represent the intention for generating the corresponding sequence. Then, recognition of a particular sequence pattern after learning can be conducted by inferring the corresponding intention in terms of initial states that can reconstruct the sequence pattern while preserving connectivity weights as fixed

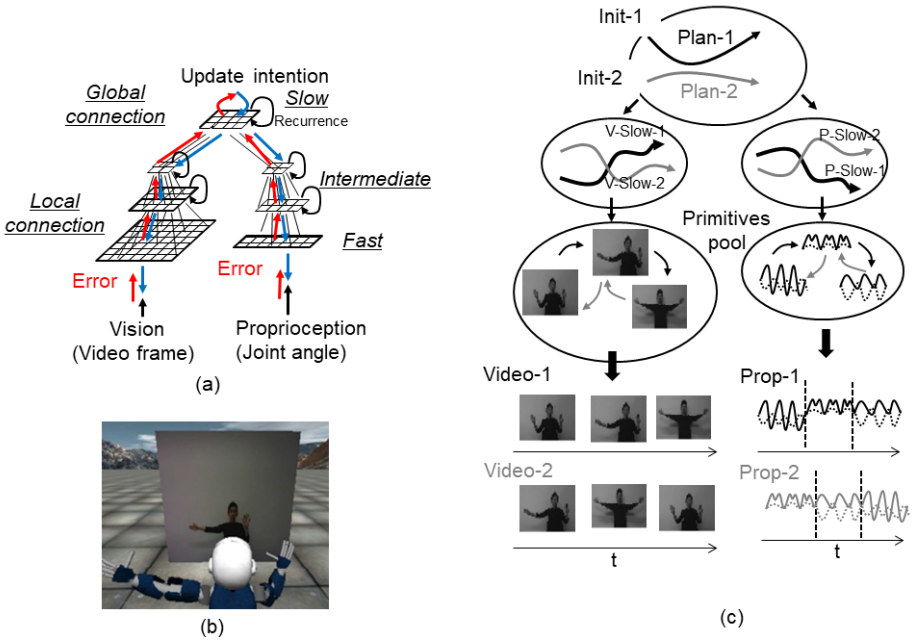

Fig. 1. P-VMDNN model (a), a simulated iCub imitates visually perceived human movement patterns (b), and the underlying mechanism of compositional movement generation (c).

2.2 Learning to imitate compositional movement patterns

In the current task of imitation learning shown in Fig. 1a, tutoring of the simulated robot was conducted as follows. Three human subjects volunteered to play the role of imitatees. Volunteers were instructed in 9 different cyclic two-arm movement patterns as movement primitives. Then, each subject generated 9 different ways of concatenating 3 movement primitives arbitrarily selected from 9 predefined movement primitives. All 27 3-primitive concatenated patterns were demonstrated by each subject. While the robot visually perceived each concatenated pattern demonstrated by the subjects, a tutor guided the movement trajectory of both arms of the robot by synchronously imitating the movement patterns of the subjects. In this manner, 27 visuoproprioceptive sequences were sampled for training the network. Network training was conducted for 50,000 epochs. Fig. 2 shows some examples of generation of imitated movement patterns after training.

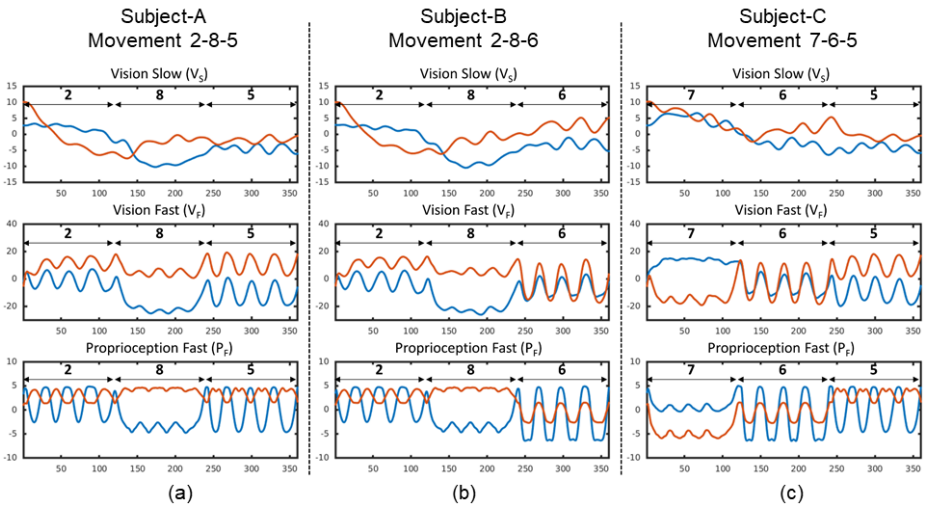

Fig. 2. Imitative generation of differently concatenated movement primitive sequences.

Three concatenated movements were generated as pantomime-like behaviors, while generating visual imagery by providing corresponding initial internal state values obtained via training. Each column shows an imitation of a movement pattern demonstrated by a different subject, in which labels above indicate the subject ID and the 3-primitive-concatenated pattern, with numbers indicating the corresponding primitive. For each column, the top, middle, and bottom rows show neural activity after PCA in the visual higher layer, visual lower layer, and proprioceptive lower layer, respectively. Each concatenation of cyclic movement primitives can be seen in the lower visual and proprioceptive layers. However, those sequences are abstractly represented with more slowly changing profiles in the higher layer.

Further analysis of network activities converged upon an understanding of the underlying mechanism for generation of compositional movements (Fig. 1c). Slowly changing profiles develop differently from different initial states in the higher layer (as illustrated by “plan-1 and plan-2” (Fig.1c). On the other hand, in both lower layers in the visual and proprioceptive pathways, a set of movement primitives is self-organized as limit cycle attractors (three movement primitives illustrate (Fig. 1c, middle). Eventually, the higher layer manipulates the lower layers by feeding bifurcation parameters that change slowly in a specific manner. This induces corresponding sequential transitions from one limit cycle to another. Consequently, movement patterns can be generated in a compositional, yet fluid manner, using nonlinear dynamic characteristics, including the initial sensitivity, bifurcation, and self-organization.

2.3 Online inference of intentions of others

This subsection describes an experiment involving online inference of intentions of human subjects during synchronous imitation [6]. In this experiment, only the visual pathway, consisting of 6 layers, was used. The network was trained to predict visually

perceived movement patterns, with 6 non-concatenated movement primitive patterns generated by 5 subjects. After learning, the network task was to continue to predict the next step visual image demonstrated by human subjects, while they occasionally switched movement patterns from one of the trained patterns to another. For the purpose of following such sudden switching, a scheme called online error regression [10] was used. In this scheme, a certain step length for the previous window is allocated. When a prediction error is generated, the error is back-propagated through time to the onset of the previous window for the purpose of updating initial internal states in the direction of error minimization. This corresponds to online inference of the intentions of movement demonstrators.

Fig. 3 shows an instance of online inference where the target visual image and its prediction after PCA, the prediction error, the internal state in the higher and lower layers is shown from top to bottom. The movement pattern in the target is switched near step 290, which is accompanied by a sharp rise in errors. However, error is reduced within 10 steps, as prediction outputs start to follow the switching. This recovery is accompanied by a shift in internal states in the higher layer. This immediate shift, enabled by means of error minimization, accounts for possible mechanisms for online inference of intentions of others, as well as segmentation of the perceptual flow.

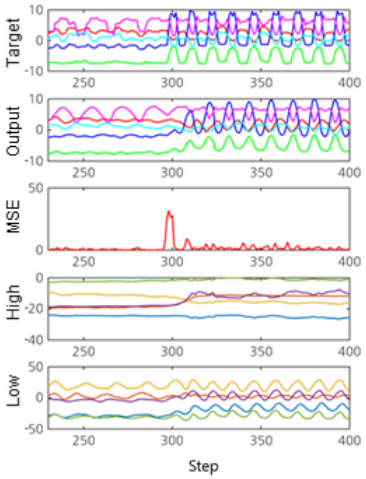

Fig. 3. Online inference of intention.

3 Discussion: Phenomenology of self from synthesis

Here, I discuss possible correspondence of this synthetic robotics study to phenomenological understanding of self and consciousness. The second experiment showed that shifts in intentions of others can be inferred by means of the error regression in the previous window. When a robot's own prediction went well, adequately synchronizing with the other, everything went smoothly and automatically, where the distinction between self and other was submerged under coherently coupled dynamics. However, when synchrony broke down due to a sudden intention change by the other, the robot should become aware of the gap between the two. This should entail consciousness, because the process of inferring the newly changed intentional state requires considerable effort to minimize the prediction error. Here, self, as defined via the gap, could consciously represent the minimal self [11, 12] in a pre-reflective form.

Gallagher [11] accounts that the minimal self should entail the sense of agency by considering possible neural mechanisms underlying the pathology of delusion of control in schizophrenia. Our prior study [13] on synthetic modeling of schizophrenia shows that the proposed predictive coding model can support this account. It was

shown that mild perturbation in the connectivity from the higher level to sensory-motor level possibly caused by this disease can generate fictive error and resultant maladaptation of the intention state. This could generate the sense of fictive agency.

Next, let us consider an account of the narrative self [11] in a reflective form. The first experiment showed that the robot was able to extract compositional structure from the experience of continuous perceptual flow by learning iterative interaction with the other. This process includes segmentation of continuous visuo-proprioceptive stream into a set of primitives, accompanied by minimal self-consciousness. These primitives can be reused later by the higher layer to account for different experiences in different contexts by recombining them. The experience re-represented in the higher layer in this manner is no longer a pure experience, but an objectified experience. Finally, we see the development of a narrative self that can represent its own experience objectively, as episodes.

References

1. Gibson, E.J., & Pick, A.D. (2000). *An ecological approach to perceptual learning and development*. New York: Oxford University Press.
2. Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1), 79.
3. Bassett, D. S., & Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5), 200– 209.
4. Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS computational biology*, 4(11), e1000209.
5. Tani, J. (2016). *Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena*. New York: Oxford University Press.
6. Choi, M., & Tani, J. (2018). Predictive coding for dynamic visual processing: development of functional hierarchy in a multiple spatio-temporal scales RNN model. *Neural Computation*, 30, 237–270.
7. Hwang, J., Kim, J., Ahmadi, A., Choi, M., & Tani, J. (2018). Dealing with large-scale spatio-temporal patterns in imitative interaction between a robot and a human by using the predictive coding framework. *IEEE Trans. on SMC: Systems*, (99), 1-14.
8. Yamashita, Y., & Tani, J. (2008). Emergence of functional hierarchy in a multiple time-scale neural network model: a humanoid robot experiment. *PLoS Computational Biology*, Vol.4, Issue.11, e1000220.
9. Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
10. Tani, J., Ito, M., & Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Networks*, 17(8-9), 1273-1289.
11. Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14-21.
12. Tani, J. (1998). An interpretation of the 'self' from the dynamical systems perspective: A constructivist approach. *Journal of Consciousness Studies*, 5(5-6), 516-542.
13. Yamashita, Y., & Tani, J. (2012). Spontaneous prediction error generation in schizophrenia. *PLoS* One, 7(5), e37843.