Leading or Following? Dyadic Robot Imitative Interaction Using the Active Inference Framework

Nadine Wirkuttis and Jun Tani

Abstract—This study investigated how social interaction among robotic agents changes dynamically depending on the individual belief of action intention. In a set of simulation studies, we examine dyadic imitative interactions of robots using a variational recurrent neural network model. The model is based on the free energy principle such that a pair of interacting robots find themselves in a loop, attempting to predict and infer each other's actions using active inference. We examined how regulating the complexity term to minimize free energy determines the dynamic characteristics of networks and interactions. When one robot trained with tighter regulation and another trained with looser regulation interact, the latter tends to lead the interaction by exerting stronger action intention, while the former tends to follow by adapting to its observations. The study confirms that the dyadic imitative interaction becomes successful by achieving a high synchronization rate when a leader and a follower are determined by developing action intentions with strong belief and weak belief, respectively.

Index Terms—Predictive coding, Cognitive robotics, humanoid robots, activity recognition, adaptive control, inference algorithms.

I. INTRODUCTION

SOCIAL interaction is considered an essential cognitive behavior. In both empirical studies and synthetic modeling, researchers have investigated underlying cognitive, psychological, and neuronal mechanisms accounting for various aspects of social cognitive behaviors. This study investigates mechanisms underlying synchronized imitation as a representative social cognitive act, by formulating the problem using the free energy principle (FEP) [1], [2]. In simulation experiments of dyadic robot imitative interaction, we examine how a leader and follower can be determined in conflicting situations by investigating the underlying network dynamics.

Numerous robotic studies have investigated imitative interaction. In the 90 s, imitation was identified as an indispensable human competency required in early development of cognitive behaviors [3]–[6]. Rizzolatti and colleagues [7] showed that the mirror neuron system uses observations of an action to generate the same action. Arbib and Oztop [8], [9] indicated that mirror neurons may participate in imitative behaviors. Upon this development, several research groups proposed computational mirror neuron models for imitation using Hidden Markov Models [10] and neural network models [11]–[14].

An essential unsolved question in modeling of imitative interaction is how a leader, who initiates an action, and a follower, who imitates this action, can be determined when multiple choices of actions are possible among a set of well-habituated ones.

Recent theories on predictive coding (PC) and active inference (AIF) based on the free energy principle (FEP) [2], [15] show that "action intention" and its "belief" can be formulated as a predictive model. "Action intention" is considered a top-down prediction of action outcomes and "belief" as an estimated precision of this prediction or the strength of intention (as described in [2], [15]). Analogous to PC and AIF, Ito and Tani showed that imitative interaction can be performed using an RNN model by minimizing the prediction error instead of free energy in order to update deterministic latent variables [13]. However, this deterministic model does not account for the belief of action intention because the precision of prediction cannot be estimated. On a related topic, Ahmadi and Tani developed predictive-coding inspired variational RNN (PV-RNN) [16]. Their model was used to investigate how the strength of top-down intention in predicting fluctuating temporal patterns was modulated, depending on learning conditions in the model. In the learning process, free energy represented by the weighted sum of the accuracy term and the complexity term is minimized. Ahmadi and Tani found that softer regulation of the complexity term during network training develops strong top-down intention. Predictions are more deterministic by self-organizing deterministic dynamics with the initial sensitivity characteristics in the network. Likewise, tighter regulation of the complexity term results in weaker intention and increased stochasticity. Compared to other neural network models based on the FEP [17]–[20], PV-RNN has advantages when applied to problems in robotics. It can cope with temporal structure by using recurrence associated with stochastic latent variables and by hierarchical abstraction through a multiple timescale structure [21].

Our research group further investigated human-robot imitative interaction using PV-RNN. Chame and Tani [22] showed that a humanoid robot with force feedback control tends to lead or follow the human counterpart in imitative interaction when its PV-RNN is set to softer or tighter regulation, respectively. However, the result is preliminary, merely showing a one-shot

experimental result without any quantitative analysis. In a similar experimental setup, Ohata and Tani [23] showed that this tendency can be also observed when regulation of the complexity term is modulated during the interaction phase, rather than during the prior learning phase. The study investigated pseudo-imitative interaction between a simulated robot and a human counterpart. This study, however, lacks genuine interaction between the simulated robot and the human counterpart because the outputs of the counterpart were replaced with static output sequences prepared in advance.

The main contribution of the current study is to clarify the underlying mechanism of how a leader and a follower can be determined in dyadic synchronized imitative interaction using the framework of AIF. This study is distinct from the author's aforementioned past studies because genuine interaction between two robots using the same model is examined and results are analysed both quantitatively and qualitatively. An advantage of performing a robot-robot interaction experiment is that the internal dynamics can be analyzed in a comparative way between the two robots.

The interaction experiment considers two robotic agents that are trained to generate a set of movement primitive sequences. When movements are generated by following a probabilistic finite state machine, the transition probability differs, depending on each of the two robots. After each robot learns the given probabilistic transition structure for a sequence, the experimental design allows us to investigate how two robots generate movement primitives in the synchronised imitative interaction. In particular, we examine conflicting situations in which each robot prefers to generate different movement patterns, depending on its learned experience. Do they synchronize to generate the same movement pattern with one robot following the other or leading by adapting the intention? Or do they desynchronize by generating different movement patterns, ignoring their counterparts by following their own action intentions? The current study hypothesizes that these dyadic interaction outcomes depend on the relative strength of the intention between the robots as a result of regulating FEP complexity.

II. MODEL

A. Predictive Coding and Active Inference

The current study applies the concepts of PC and AIF based on FEP [1]. PC considers perception as the interplay between a *prior* expectation of a sensation and a *posterior* inference about a sensory outcome. Expectation of the sensation can be modeled by a generative model that maps the prior of the latent state to the expectation of sensation. The posterior inference of the observed sensation can be achieved by jointly minimizing the error computed between the expected sensation and its outcome, i.e. the *accuracy*, plus the Kullback-Leibler (KL) Divergence between the posterior and the prior distributions, i.e. the *complexity*. Posterior and prior are both represented by Gaussian probability distributions using means and variances. This is to minimize free energy or to maximize the lower bound of the

logarithm of marginal likelihood:

$$
\begin{aligned}
\log p_\theta(\bar{\boldsymbol{X}}) &\geq \underbrace{\int q_\phi(\boldsymbol{z}|\bar{\boldsymbol{X}}) \log \frac{p_\theta(\bar{\boldsymbol{X}}, \boldsymbol{z})}{q_\phi(\boldsymbol{z}|\bar{\boldsymbol{X}})} d\boldsymbol{z}}_{\text{Evidence lower bound}} \\
&= \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{z}|\bar{\boldsymbol{X}})}[\log p_\theta(\bar{\boldsymbol{X}}|\boldsymbol{z})]}_{\text{Accuracy}} - \underbrace{D_{\mathrm{KL}}[q_\phi(\boldsymbol{z}|\bar{\boldsymbol{X}})\|p_\theta(\boldsymbol{z})]}_{\text{Complexity}}
\end{aligned}
\tag{1}
$$

$\boldsymbol{z}$, $\bar{\boldsymbol{X}}$, p_θ, and q_ϕ denote the latent state, the observation, the prior distribution, and the approximate posterior, respectively. θ and ϕ are the parameters of the generative and inference model. In maximizing the lower bound, the interplay between *accuracy* and *complexity* characterizes the model performance in learning, prediction, and inference.

Consistent with the AIF, actions are generated so that the error between the expected action outcome and the actual outcome is minimized. In robotic applications, this is equivalent to determining how expected proprioception in terms of robot joint angles can be achieved by generating adequate motor torque. A simple solution is to use a PID controller, in which adequate motor torque to minimize errors between expected joint angles and actual angles can be obtained by means of error feedback schemes. Finally, perception by predictive coding and action generation by active inference are deployed simultaneously, thereby closing the loop of action and perception.

B. Overview of PV-RNN

The PV-RNN model is designed to predict future sensation by means of prior generation, while reflecting the past by means of posterior inference based on learning (see Fig. 1). One essential element of the model is the introduction of a parameter **w**, the so-called meta-prior, which regulates the complexity term in free energy. Different **w** settings results in alternation of the estimated precision in predicting the sensation, as described later as *prior generation* (see section III-C). The model is also characterised by employing an architecture of multiple timescale RNN (MTRNN) [21]. The whole network comprises multiple layers of RNNs wherein the dynamics of each layer are governed by different time constant parameters τ. This scheme supports development of hierarchical information processing by adequately setting the timescale of each layer [14], [21]. This approach is considered as analogous to [24], [25].

The following briefly describes the two essential parts, a *generative model* which is used for prior generation to make future predictions, and an *inference model*, which is used for posterior inference about the past. For further details, refer to [16], [23].

1) Generative Model: The stochastic generative model is used for prior generation, as illustrated in the future prediction part (after time step 4) in Fig. 1.

PV-RNN is comprised of deterministic variables **d** and random variables **z**. An approximate posterior distribution q_ϕ is inferred based on the prior distribution p_θ by means of error minimization on the generated prediction **X**. The generative
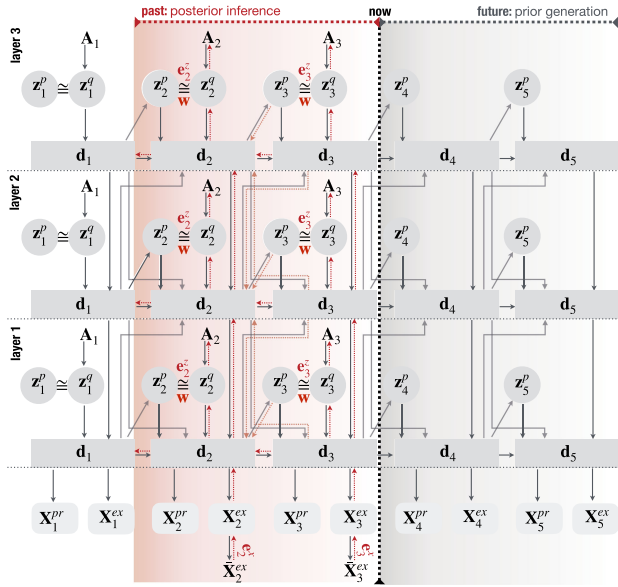
Fig. 1. **Graphical representation of a hierarchical 3-layer PV-RNN architecture.** Layers are indicated on the left. Time increases from left to right and is indicated as a subscript. The representation shows the network in $t = 3$ with a two-time-step posterior inference window [2,3] and prior generation for $t = [4, 5]$. In the posterior inference window, prediction error $\mathbf{e}^x$ and the $\mathbf{w}$ weighted KL Divergence $\mathbf{e}^z$ are minimized.

model can be factorized as:

$$p_\theta(\bar{\boldsymbol{X}}_{1:T}, \mathbf{d}_{1:T}, \boldsymbol{z}_{1:T}|\boldsymbol{d}_0)$$

$$= \prod_{t=1}^{T} p_{\theta_{\bar{\boldsymbol{X}}}}(\bar{\boldsymbol{X}}_t|\mathbf{d}_t) p_{\theta_d}(\mathbf{d}_t|\mathbf{d}_{t-1}, \boldsymbol{z}_t) p_{\theta_z}(\boldsymbol{z}_t|\mathbf{d}_{t-1}) \quad (2)$$

Although $\mathbf{d}$ is a deterministic variable, it can be considered to have a Dirac delta distribution centered on $\tilde{d}$ as $\sigma(d - \tilde{d})$. $\bar{X}$ is conditioned directly on z through $\tilde{d}$. At the initial time step, $\tilde{d}$ is set to 0. Otherwise, $\tilde{d}$ is recursively computed, for which the internal state before activation is denoted by h. This internal state h is a vector, calculated as the sum of the internal states of the current level l and its connecting layers of the previous time step $t-1$ plus the latent z in the same layer of the current time step t:

$$\tilde{\boldsymbol{d}}_t^l = \tanh(\boldsymbol{h}_t^l)$$

$$\boldsymbol{h}_t^l = \left(1 - \frac{1}{\tau^l}\right)\boldsymbol{h}_{t-1}^l +$$

$$\frac{1}{\tau^l}\left(\boldsymbol{W}_{dd}^{ll}\tilde{\boldsymbol{d}}_{t-1}^l + \boldsymbol{W}_{zd}^{ll}\boldsymbol{z}_t^l + \boldsymbol{W}_{dd}^{ll+1}\tilde{\boldsymbol{d}}_{t-1}^{l+1} + \boldsymbol{W}_{dd}^{ll-1}\tilde{\boldsymbol{d}}_{t-1}^{l-1}\right)$$

$$(3)$$

τ^l denotes the layer-specific time constant. With larger value for τ^l, slower timescale dynamics develop, whereas with a smaller value set, faster timescale dynamics dominate. $\boldsymbol{W}$ represents connectivity weight matrices between layers and their deterministic and stochastic units. The output with size N_x is computed as mapping from $\tilde{d}^1$ as:

$$\mathbf{X}_t = \boldsymbol{W}_{Xd}^{ll}\tilde{\boldsymbol{d}}_t^1 + \boldsymbol{b}_X \quad (4)$$

The prior distribution $p_\theta(\boldsymbol{z}_t)$ is a Gaussian distribution represented with mean $\boldsymbol{\mu}_t^p$ and standard deviation $\boldsymbol{\sigma}_t^p$. The prior depends on $\tilde{\boldsymbol{d}}_{t-1}$ by following the idea of a sequence prior [26], except at $t = 1$ where it follows a unit Gaussian distribution.

$$p_\theta(\boldsymbol{z}_1) = \mathcal{N}(0, I)$$

$$p_\theta(\boldsymbol{z}_t|\tilde{\boldsymbol{d}}_{t-1}) = \mathcal{N}(\boldsymbol{\mu}_t^p, (\boldsymbol{\sigma}_t^p)^2) \text{ where } t > 1$$

$$\boldsymbol{\mu}_t^p = \tanh(\boldsymbol{W}_{d\mu^p}^{ll}\tilde{\boldsymbol{d}}_{t-1}) \quad (5)$$

$$\boldsymbol{\sigma}_t^p = \exp(\boldsymbol{W}_{d\sigma^p}^{ll}\tilde{\boldsymbol{d}}_{t-1})$$

Based on the work on variational autoencoders, we use the *reparameterization trick* to formulate the latent prior of $\boldsymbol{z}_t$ as mean $\boldsymbol{\mu}_t^p$ and standard deviation $\boldsymbol{\sigma}_t^p$. The *reparameterization trick* was introduced by Kingma and Welling [27] to make random variables differentiable for backpropagating errors through the network for learning. The same consideration is taken for the posterior of $\boldsymbol{z}_t$ in the inference model as well (cf. below 7).

2) Inference Model: Posterior inference is performed during learning and afterward, during action and perception. Fig. 1 illustrates information flow in the posterior inference in a time window from time step 2 to time step 3. The inference model for the posterior is described as:

$$Q_\phi(\boldsymbol{z}_t|\tilde{\boldsymbol{d}}_{t-1}, \boldsymbol{e}_{t:T}) = \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{\mu}_t^q, \boldsymbol{\sigma}_t^q) \quad (6)$$

where $\boldsymbol{e}_t$ denotes the error between the target $\bar{\mathbf{X}}_t$ and the predicted output $\mathbf{X}_t$. Like the prior p_θ, the posterior q_ϕ is also a Gaussian distribution with mean $\boldsymbol{\mu}_t^q$ and standard deviation $\boldsymbol{\sigma}_t^q$. For $\boldsymbol{z}_{1:T}$ it is defined as:

$$q_\phi(\boldsymbol{z}_t|\boldsymbol{e}_{t:T}) = \mathcal{N}(\boldsymbol{\mu}_t^q, \boldsymbol{\sigma}_t^q)$$

$$\boldsymbol{\mu}_t^q = \tanh(\boldsymbol{A}_t^\mu) \quad (7)$$

$$\boldsymbol{\sigma}_t^q = \exp(\boldsymbol{A}_t^\sigma)$$

Since computing the true posterior is intractable, an approximate posterior q_ϕ is inferred by maximizing the lower bound, analogous to Eq. (1). Here, the adaptation variable $\mathbf{A}_{1:T}$ forces the parameters ϕ of the inference model to represent meaningful information. The lower bound of PV-RNN can be derived as:

$$L(\theta, \phi) = \sum_{t=1}^{T} \left(\frac{1}{N_X} E_{q_\phi(\boldsymbol{z}_t|\tilde{\boldsymbol{d}}_{t-1}, \boldsymbol{e}_{t:T})} \left[\log p_{\theta_{\bar{X}}}(\bar{\boldsymbol{X}}_t|\tilde{\boldsymbol{d}}_t) \right] - \right.$$

$$\left. \sum_{l}^{L} \frac{\mathbf{w}^l}{N_z^l} D_{KL} \left[q_\phi(\boldsymbol{z}_t|\tilde{\boldsymbol{d}}_{t-1}, \boldsymbol{e}_{t:T}) || p_{\theta_z}(\boldsymbol{z}_t|\tilde{\boldsymbol{d}}_{t-1}) \right] \right)$$

$$(8)$$

where the first term is the *accuracy* and the second term is the *complexity* (for details referred to [16]). N_x and N_z^l are the number of sensory dimensions and the number of the latent random variables at the lth layer, respectively. $\mathbf{w}^l$ serves as a weighting parameter for the complexity term in layer l and is referred to as the meta-prior [16]. The meta-prior represents the strength for regulating the closeness between the posterior and the prior distributions. In $t = 1$, $\mathbf{w}_1^l$ is set with 1.0. $\mathbf{w}_{2:T}^l$ is set to a specific value when the sequence prior [26] is used after time step 1. In the posterior inference, all learning-related network parameters of θ, ϕ, and the adaptive variable $\mathbf{A}$ are updated to
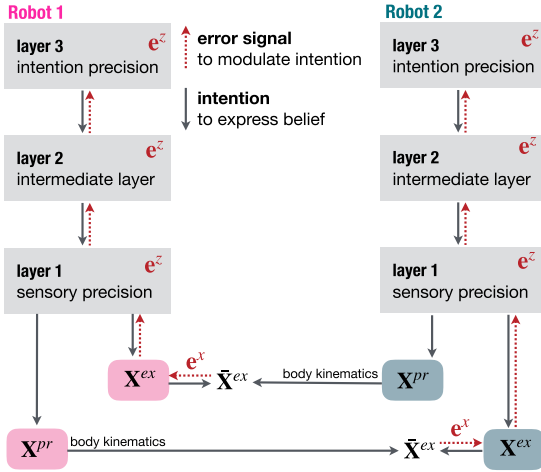
Fig. 2. **Schematic of dyadic robot interaction** where robots are equipped with the PV-RNN model.

(a) robot movements

(b) movement trajectories

(c) probabilistic transition

Fig. 3. **Task design.** Robot movement primitives **A**, **B**, and **C** of the training dataset (a). Proprioceptive trajectories $\bar{\mathbf{X}}^{pr}$ plus exteroceptive trajectories $\bar{\mathbf{X}}_r^{ex}$ and $\bar{\mathbf{X}}_l^{ex}$. Colors represent six dimensions of joint angles for $\bar{\mathbf{X}}^{pr}$ and xy-coordinate positions of the right and left end effector $\bar{\mathbf{X}}_r^{ex}$ and $\bar{\mathbf{X}}_l^{ex}$ (b). Two P-FSMs representing different movement primitive transition patterns of **A**20%**B**80%**C** and **A**80%**B**20%**C** (c).

maximize the lower bound by back-propagating the error from time step T back to t_1 [28].

3) PV-RNN in Dyadic Robot Interaction: Two robots equipped with the PV-RNN model interact during synchronized imitation. In the interaction, the robots predict proprioception $\mathbf{X}_{t+1}^{pr}$ and exteroception $\mathbf{X}_{t+1}^{ex}$ for the next time step. The predicted $\mathbf{X}_{t+1}^{pr}$ regulates joint angle movements of a robot by considering a PID controller. This movement $\mathbf{X}_{t+1}^{pr}$ can then be sensed by the other robot in terms of exteroception $\bar{\mathbf{X}}_{t+1}^{ex}$. This is provided through the kinematic transformation of joint angles $\mathbf{X}_{t+1}^{pr}$ (cf. Fig. 2). While in the training phase, the error signal is taken from the proprioceptive $\bar{\mathbf{X}}^{pr}$ as well as the exteroceptive $\bar{\mathbf{X}}^{ex}$ target sequences, in the interaction phase the error signal for each robot is taken only from $\bar{\mathbf{X}}^{ex}$[1]. During interactions, prediction errors $\mathbf{e}^x$ are generated and propagated bottom-up throughout all layers, as well as time steps in the posterior inference window, in terms of the latent error $\mathbf{e}^z$. This updates posterior distributions in the network and minimizes the variational free energy. In this phase, only $\mathbf{A}_{1:T}$ is updated without updating network parameters θ and ϕ.

III. ROBOT EXPERIMENTS

To investigate how the interaction of two robots changes with tighter and looser regulation of complexity, each robot was trained and tested individually, as described in III-B and in III-C, respectively. Finally, two robots were examined during a dyadic interaction (III-D).

A. Task Design

Robotic agents are trained with three movement primitives **A**, **B**, and **C** (Fig. 3 (a)). Each primitive is 40 time steps in length. A human experimenter generated the primitive data via a master control of a humanoid OP2[2]. The experimenter controlled six

joints in the upper body of one humanoid $\bar{\mathbf{X}}^{pr}$. The exteroceptive trajectory $\bar{\mathbf{X}}^{ex}$ is generated by mirroring its own movement $\bar{\mathbf{X}}^{pr}$ and transformed into $\bar{\mathbf{X}}^{ex}$ xy-coordinate positions of the left hand and right hand tips of the robot (Fig. 3(b)). $\bar{\mathbf{X}}^{pr}$ and $\bar{\mathbf{X}}^{ex}$ are six and four dimensions, respectively. Individual movement primitives are sampled and combined to form a continuous pattern of 400 time steps that follows a probabilistic sequence (analogous to [23]). Two probabilistic patterns were generated, **A**20%**B**80%**C** and **A**80%**B**20%**C** as shown in the form of a probabilistic finite state machine (P-FSM) (Fig. 3 (c)). The difference between these two probabilistic patterns is that **C** is biased and comes more often (80%) than **B** (20%) after **A** in the former, and vice versa for the latter.

A point of interest is the interaction phase after the learning phase. It is expected that both robots can generate **A** synchronously, since it is a deterministic state. This could be different from generating **B** or **C** as two robots learned different preferences in terms of transition probabilities. One robot may lead so as to generate **B** or **C** while the other may just follow it. However, both robots may generate their own biased movements and, thus, desynchronize their behavior. The current study hypothesizes that whether **B** or **C** is generated synchronized or desynchronized between the two robots depends on the complexity regulation of each robot.

---

[1]Considering only $\bar{\mathbf{X}}^{ex}$ for the error term in interaction settings assumes that the PID controller generates only negligible position errors for the joints.

[2]Humanoid OP2 and its master controller are developed by Robotis: www.robotis.us/robotis-op2-us/ www.robotis.us/robotis-op2-us/.
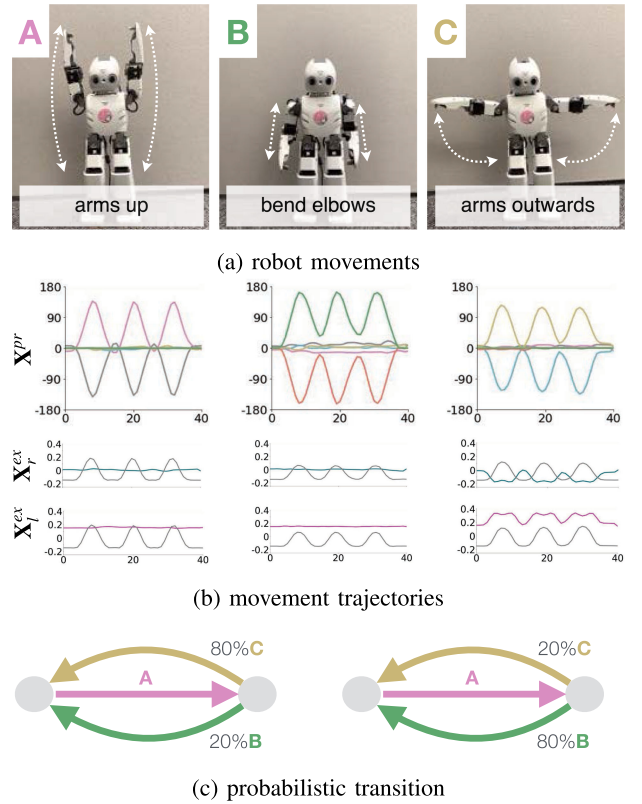
TABLE I
NETWORK PARAMETERS FOR TRAINING PV-RNNS

	$\mathbf{d}$	$\mathbf{z}$	τ	$\mathbf{w}_1$	$\mathbf{w}_{2:T}$
layer 1	40	4	2	1	$\mathbf{w}^1 = [0.0, 0.001, ...4.999, 5.0]$
layer 2	20	2	4	1	$\mathbf{w}^2 = w^1 \times 10$
layer 3	10	1	8	1	$\mathbf{w}^3 = w^1 \times 100$

B. Robot Training

The PV-RNN was trained with 20 data samples on a set of different parameters (TABLE I). All network specific parameters were fixed during training. To explore the influence of the meta-prior, w, only this parameter changed for different networks and was repeated with different random seeds to ensure reproducibility.

Networks were trained for 80 000 epochs, using Adam Optimizing and back-propagation through time (BPTT) [28] with learning rate 0.001. After training, network performance was first analysed in stand-alone robot experiments (subsection III-C). Thereafter, dyadic robot interaction was studied using networks trained with **w** set for the two representatives of tight and loose regulation of FEP complexity (subsection III-D).

C. Preparatory Analysis of Training Results

To investigate how the model learns the probabilistic structure of the training data, we conducted a first analysis in the form of *prior regeneration*. For prior regeneration we choose one training sample and use two time steps of the adaption variable $\mathbf{A}_{1:2}^{\bar{\mathbf{X}}}$ to initialize the prior distribution $p(\mathbf{z}_{1:2})$ in the PV-RNN. Thereafter the future prediction $\mathbf{X}_{3:400}$ for the remaining training sample length can be calculated (cf. *prior generation* in Fig. 1). Using this scheme, we generated 20 sequences for each meta-prior **w**. This was repeated for each network that was trained for that parameter for all random seeds. For brevity, training analysis is reported only for the network that was trained on the probabilistic sequence **A**20%**B**80%**C**. Training of **A**80%**B**20%**C** showed comparable results. An Echo State Network for multivariate time series classification [29] with reservoir size $N = 45$, 25% connectivity and leakage 60% was used for classification of movement primitives. Movement patterns were identified as *not classified* below a 55% threshold.

1) Analysis of Probabilistic Transition: A robot that is trained with **A**20%**B**80%**C** will first generate an **A** movement, and then transition to **B** with 20 percent probability and to **C** with 80 percentage probability. We found that smaller **w** settings are less stable in reproducing the probabilistic structure of the training data. The **BC**-ratio was either greater or less than 20% for **B** or greater or less than 80% for **C**. Networks trained with larger meta-priors become more reliable in regenerating the probabilistic training sequence (**BC**-ratio in Table II). In addition to the capacity of learning the probability distribution of the training data, we found that smaller meta-priors show noisier pattern generation. Non-classified movements were as high as $22\% \pm 4$ with $\mathbf{w} = 0.01$ and decreased to $6\% \pm 0.6$ with $\mathbf{w} = 3.4$.

2) Divergence Analysis: Repeatability in generating sequences in prior generation was examined by conducting a divergence analysis. Sequences are considered diverged when a comparison per time step of $\mathbf{X}^{pr}$ exceeds a threshold[3]. Out of 20 regeneration sequences, we randomly select one as a reference and calculate the average divergence step of the other sequences to this the reference. Out of 400 time steps of prior generation, sequences diverged from the reference around time step 43 for networks trained with smaller **w**. With increasing **w**, repeatability of the trajectories increased. Here the divergence step was around 139 (cf. divergence step t in Table II).

3) Summary of Preparatory Analysis: Loose regulation of the complexity term results in noisier, less repeatable prior generation performance. Also the learned probability for transition to either **B** or **C** is not accurate. This observation changes with increasing meta-prior. The larger **w**, the more accurate the learned transition probability becomes. Also, prior generation becomes more repeatable by developing more deterministic dynamics with the initial sensitivity characteristics (i.e., the sequence is generated solely depending on the latent state in the initial time step). For subsequent dyadic robot interaction experiments, we empirically select the meta-prior setting $\mathbf{w} = 0.005$ and $\mathbf{w} = 3.4$ as two representatives of tight and loose regulation of the FEP complexity.

D. Dyadic Robot Interaction Experiments

In the following experiments, robots are either trained with $\mathbf{w} = 0.005$ or $\mathbf{w} = 3.4$. For readability, we will consider $\mathbf{R}_w^1$ and $\mathbf{R}_w^2$ with subscripts of the respective meta-priors **w**. In the dyadic interaction, we present the network of each robot with observations of movements of the counterpart robot $\bar{\mathbf{X}}^{ex}$ as the target and perform posterior inference in a regression window with size $win_{size} = 70$. Inference is performed from the current time step t back to $t - win_{size}$, or t_1 in case $t - win_{size} \leq 1$. After 200 epochs of iteration to maximize the lower bound, the time window is shifted one time step forward. Note, all experiments were conducted in simulation due to the difficulty of real-time posterior inference computation.

We investigated how two robots interact in three different dyadic conditions (TABLE III). We then analysed whether the robots trained with **A**80%**B**20%**C** maintained the learned preference between **B** and **C** or adapted to their counterparts that were trained with **A**20%**B**80%**C**. We also calculated the so-called **BC**-synchronization rate during the interaction. If at any time step t, one of the robots generated **B** or **C** and the other robot generated the same movement primitive, the interaction was considered synchronized. Note that time steps in which movement patterns were identified as *not classified* by the Echo State Network (cf. subsection III-C) were excluded from the computation.

TABLE III shows the summary of the analysis for all three experiments. To better understand effects of loose and tight regulation of FEP complexity, exemplar plots of robot movement patterns, as well as corresponding network dynamics, are shown (cf. Fig. 4 and Fig. 5). We provide supplementary movies of the experiments showing humanoid robot interaction and network

---

[3]As the threshold for the divergence analysis, we use the mean squared error of joint angle data $[-180, 180]$ of $\mathbf{X}^{pr}$. The threshold is set to 55.

TABLE II
TRAINING PERFORMANCE OF REPRESENTATIVE META-PRIOR **w**. THE MEAN $\pm$ STANDARD DEVIATION REPRESENT THREE RANDOM SEEDS AND 20 REPETITIONS OF PRIOR GENERATION FOR EACH **w**

w	training sequence A20%B80%C				BC-ratio	divergence step t
	A	**B**	**C**	not classified		
0.005	34 ± 1	11 ± 3	40 ± 2	15 ± 1	22 ± 6**B** 78 ± 5**C**	43
0.01	35 ± 2	13 ± 0.2	30 ± 2	22 ± 4	30 ± 5**B** 70 ± 5**C**	50
1.0	36 ± 1	11 ± 2	40 ± 2	13 ± 0.2	22 ± 4**B** 78 ± 4**C**	91
2	41 ± 0.7	10 ± 0.5	39 ± 0.7	10 ± 0.3	21 ± 8**B** 79 ± 8**C**	120
3.4	45 ± 1	11 ± 1	38 ± 2	6 ± 0.5	24 ± 2**B** 76 ± 2**C**	139

TABLE III
INTERACTION PERFORMANCE OF THREE EXPERIMENTAL SETTINGS

Experiment		BC-ratio				BC-sync
ID	robots	stand-alone		interaction		
		B	**C**	**B**	**C**	
1	$\mathbf{R}^1_{0.005}$ $\mathbf{R}^2_{3.4}$	22 ± 6 75 ± 2	78 ± 5 25 ± 3	70 ± 11 73 ± 10	30 ± 10 27 ± 10	56 ± 23
2	$\mathbf{R}^1_{3.4}$ $\mathbf{R}^2_{3.4}$	24 ± 2 75 ± 2	76 ± 2 25 ± 3	17 ± 12 61 ± 12	83 ± 18 39 ± 12	31 ± 24
3	$\mathbf{R}^1_{0.005}$ $\mathbf{R}^2_{0.005}$	22 ± 6 44 ± 11	78 ± 5 56 ± 11	52 ± 13 20 ± 7	48 ± 9 80 ± 8	42 ± 20

dynamics here: https://doi.org/10.6084/m9.figshare.14 099 537 https://doi.org/10.6084/m9.figshare.14 099 537.

1) Experiment 1: $\mathbf{R}^1_{0.005}$ *Vs.* $\mathbf{R}^2_{3.4}$*:* In Experiment 1, $\mathbf{R}^1_{0.005}$ adapts to the probabilistic transition of $\mathbf{R}^2_{3.4}$ by increasing the probability of performing **B** from 22% in the stand-alone condition to 70% in the dyad (Table III Experiment 1). Both robots are performing more **B** than **C** with a **BC**-synchronization of $56 \pm 23\%$ which is significantly higher than the chance rate of 32%[4].

Fig. 5 shows an example of how prediction of the future and posterior inference of the past proceed as time passes from time step 199, 229, to 259 for both robots. We observe that the intended future behavior (the prior generation) of $\mathbf{R}^1_{0.005}$ is not consistent with the actually performed actions after posterior inference. On the other hand, in the case of $\mathbf{R}^2_{3.4}$, the performed action complies with its prediction.

This behavior can be explained by looking at exemplar priors μ^{lp}_i and posteriors μ^{lq}_i for layer l and neuron i between two robots. In layer 1, selected posterior network dynamics μ^{1q}_1 and μ^{1q}_2 are deviating from prior dynamics μ^{1p}_1 and μ^{1p}_2 for robot $\mathbf{R}^1_{0.005}$. Whereas the dynamics of $\mathbf{R}^2_{3.4}$ are mostly overlapping (cf. Fig. 4(a) and https://doi.org/10.6084/m9.figshare.14 099 537 supplementary movie). More specifically, the average KL Divergence $\mathbf{e}^z$ of $\mathbf{R}^1_{0.005}$ is larger for all layers $((\mathbf{e}^{z,1}, \mathbf{e}^{z,2}, \mathbf{e}^{z,3}) = (109.1, 1.4, 0.06))$ than for $\mathbf{R}^2_{3.4}$ $((\mathbf{e}^{z,1}, \mathbf{e}^{z,2}, \mathbf{e}^{z,3}) = (0.4, 0.0003, 0.00001))$. This means that $\mathbf{R}^2_{3.4}$ tends to behave as intended because the posterior is attracted by the prior. On the other hand, $\mathbf{R}^1_{0.005}$ tends to adapt

to $\mathbf{R}^2_{3.4}$ since the posterior is rather attracted by the observation than by the weaker prior belief.

Note that μ^{3q}_1 and μ^{3p}_1 in layer 3 change only slowly with time. This indicates that these latent variables represent how movement primitives transit from deterministic states to non-deterministic states using their slower timescale properties characterized by τ^3.

2) Experiment 2: $\mathbf{R}^1_{3.4}$ *Vs.* $\mathbf{R}^2_{3.4}$*:* When two robots with loose complexity regulation interact, both robots maintain their learned preferences in terms of probability in generating either **B** or **C**. $\mathbf{R}^1_{3.4}$, which learns a 76% transition to **C** in a stand-alone situation, shows its preference to **C** in the dyad with probability of 83%. $\mathbf{R}^2_{3.4}$, which in a stand-alone condition would maintain its preference to **B** with a probability of 75%, shows 61% percentage transition to **B** in the interaction. **BC**-synchronization rate turns out to be low as $31 \pm 24\%$, which is almost equal to the chance rate. Examining the network dynamics of the prior and posterior distributions shows that the robots executed movements based upon their prior action intention without adapting their posteriors to observations of the other robot's movement (cf. Fig. 4(b) and https://doi.org/10.6084/m9.figshare.14 099 537 supplementary movie).

3) Experiment 3: $\mathbf{R}^1_{0.005}$ *Vs.* $\mathbf{R}^2_{0.005}$*:* When two robots with tight regulation of complexity interact, both try to adapt their own action to the one demonstrated by the other. Indeed, Fig. 4(c) shows that the prior and posterior do not comply, but deviate. Whether trained with the probabilistic transition of **A**20%**B**80%**C** or **A**80%**B**20%**C**, both robots significantly reduce the tendency to perform their own intended behavior **C** or **B**, respectively. This is evidenced by changes of the **BC**-ratio from stand-alone compared to the dyadic setting (TABLE III Experiment 3). **BC**-synchronization rate is $42 \pm 20\%$ which is higher than the chance rate but not significantly. The interaction becomes noisier, compared to results of Experiments 1 and 2 (cf. Fig. 4(c) and https://doi.org/10.6084/m9.figshare.14 099 537 supplementary movie), which indicate that tight regulation makes robots more sensitive to temporal fluctuations in observations of their counterparts.

IV. DISCUSSION

The current study examined how social interaction in robotic agents dynamically changes depending on how the complexity in the free energy is regulated. For this purpose, we conducted simulation experiments on dyadic imitative interactions using humanoid robots equipped with PV-RNN architectures. PV-RNN is a hierarchically organized variational RNN model that employs a framework of predictive coding and active inference based on the

---

[4]We assume that **B** and **C** are independent probabilistic events. Then we can consider the probabilities for a robot R to perform either a **B** movement as $\mathbf{P}^R(B)$ or a **C** movement as $\mathbf{P}^R(C)$. The actual **BC**-synchronization chance level can then be calculated as: $\mathbf{P}^1(B) \times \mathbf{P}^2(B) + \mathbf{P}^1(C) \times \mathbf{P}^2(C) = 0.8 \times 0.2 + 0.8 \times 0.2 = 0.16 + 0.16 = 0.32$.

(a) Experiment 1: $\mathbf{R}^1_{0.005}$ vs $\mathbf{R}^2_{3.4}$

(b) Experiment 2: $\mathbf{R}^1_{3.4}$ vs $\mathbf{R}^2_{3.4}$

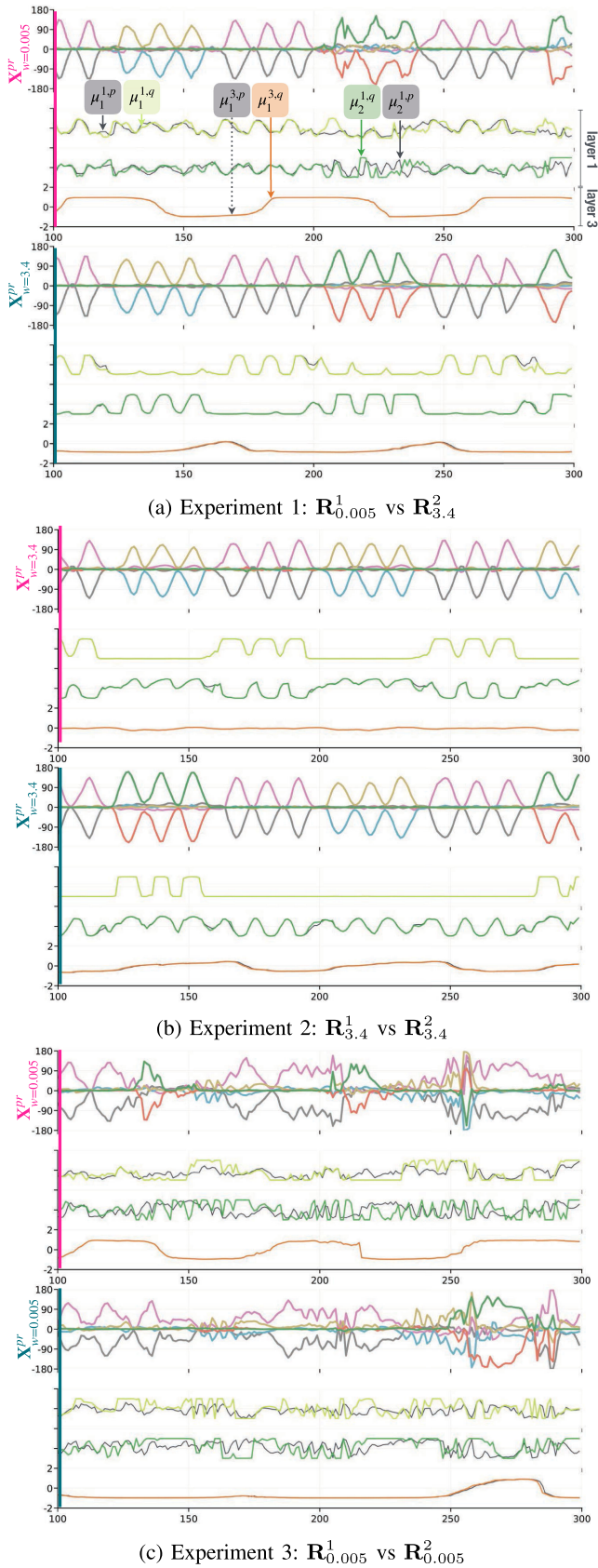(c) Experiment 3: $\mathbf{R}^1_{0.005}$ vs $\mathbf{R}^2_{0.005}$

Fig. 4. **Movement trajectories and network dynamics of dyadic robot interaction** for Experiment 1 (a), Experiment 2 (b) and Experiment 3 (c). Time steps $t = [100\,300]$ of movements and selected neurons in layer 1 and 3 are shown.
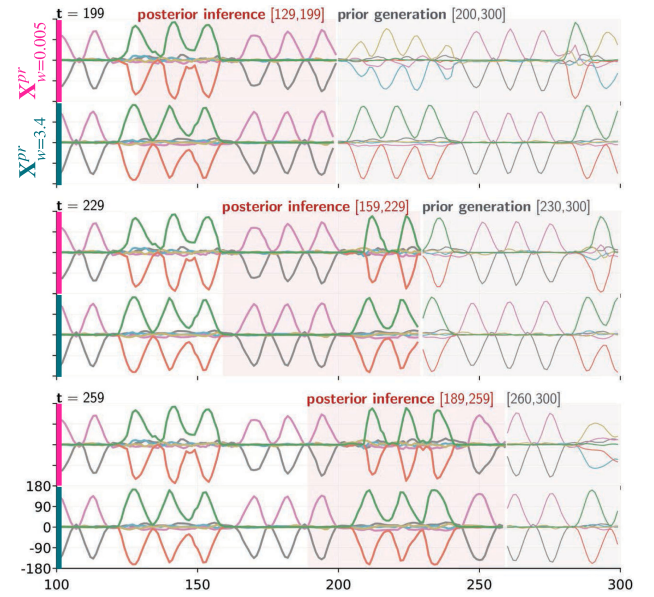

Fig. 5. **Posterior inference and prior generation in Experiment 1**. Interaction of $\mathbf{R}^1_{0.005}$ (upper) and $\mathbf{R}^2_{3.4}$ (lower) in terms of $\mathbf{X}^{pr}$. The first, the second, and the third row show $\mathbf{X}^{pr}$ after the posterior inference in the inference window with size $win_{size} = 70$, as well as its future prior generation with current time steps of 199, 229, and 259, respectively.

free energy principle. In a preparatory analysis we showed that PV-RNNs trained with looser regulation of complexity develop stronger action intentions by self-organizing more deterministic dynamics with strong initial sensitivity. Networks trained with tighter regulation develop weaker intentions by self-organizing more stochastic dynamics.

Our experiments revealed different types of interactions between robots. In the experiment where a robot having looser regulation interacts with a robot with tighter regulation, the former tends to lead the interaction by exerting action intention with stronger belief, while the latter tends to follow the other. The following robot adapts its posterior to its observations of the leading robot. In this setting, the synchronization of movement **B** and **C** (**BC**-synchronization rate) between the two robots was significantly higher than the chance rate. When two robots with looser regulation, i.e. intentions with stronger belief, interact, each robot tends to generate its own intended movements. Finally, in case both robots have tighter regulation, a fluctuating dyadic interaction develops where each robot attempts to adapt to the counterpart with an intention with weaker belief. In the last two cases, the **BC**-synchronization rate was not significantly higher than the chance rate. It can be summarized that the dyadic imitative interaction, including situations where the other's movements are unpredictable, tends to be synchronized successfully when a dedicated leader and follower are determined; a leader develops action intentions with strong belief whereas a follower develops action intentions with weak belief.

The readers may ask why tighter or looser regulation of the complexity term results in development of weaker or stronger belief of action intention for each robot. Let us consider a situation in which the PV-RNN learns to predict probabilistic sequences $\bar{\mathbf{X}}_{1:T}$ with meta-prior $\mathbf{w}$ set either with a large value (loose regulation) or a smaller one (tight regulation). The learning

process infers the posterior mean $\boldsymbol{\mu}_t^q$ and standard deviation $\boldsymbol{\sigma}_t^q$ at each time step t. In order to minimize the error $\mathbf{e}$ in the accuracy term, $\boldsymbol{\mu}_t^q$ is fitted with an arbitrary value, where $\boldsymbol{\sigma}_t^q$ will be minimized, in both cases. Notably, when the data $\overline{\boldsymbol{X}}_t$ is observed as random, the corresponding posterior $\boldsymbol{\mu}_t^q$ also becomes random.

Let us consider the two cases when the meta-prior $\mathbf{w}$ is either set large or small. In case $\mathbf{w}$ is set large, the KL Divergence between the posterior and the prior is strongly minimized. Thus, $\boldsymbol{\mu}_t^p$ and $\boldsymbol{\sigma}_t^p$ of the prior latent state become close to $\boldsymbol{\mu}_t^q$ and $\boldsymbol{\sigma}_t^q$ of the posterior. By this, $\boldsymbol{\sigma}_t^p$ in the prior is forced to take a minimal value close to 0; therefore, the prior generation becomes deterministic. Since $\boldsymbol{\mu}_{1:T}^p$ should be reconstructed as close to the sequence $\boldsymbol{\mu}_{1:T}^q$ inferred with randomness, the prior generative model is forced to develop strongly nonlinear deterministic dynamics with the initial sensitivity through learning. On the other hand, if $\mathbf{w}$ is set with a small value, the KL Divergence is only weakly minimized. In this case, prior $\boldsymbol{\mu}_t^p$ and $\boldsymbol{\sigma}_t^p$ can diverge from the posterior ones; therefore, the learning becomes "relaxed". As a result, the prior generative model develops stochastic dynamics with only weak non-linearity, wherein $\boldsymbol{\mu}_t^p$ takes an average of $\boldsymbol{\mu}_t^q$ over all occurrences and $\boldsymbol{\sigma}_t^p$ takes their distribution at each time step. Consequently, with larger $\mathbf{w}$, the generative model develops action intention with stronger belief (i.e. smaller $\boldsymbol{\sigma}^p$) whereas in the case of tighter regulation using a smaller $\mathbf{w}$, the generative model develops action intention with weaker belief (i.e. larger $\boldsymbol{\sigma}^p$).

The current experiments consider a fixed meta-prior setting only. Since the meta-prior is the essential network parameter to guide the strength of action intention in the proposed framework, future studies should target meta-learning of the meta-prior in developmental processes or through autonomous adaption within dyadic contexts. This could provide further understanding of more complex social interaction phenomena, including turn-taking in the context of adaptive regulation of the complexity term in free energy.

ACKNOWLEDGMENT

REFERENCES

[1] K. Friston, "A theory of cortical responses," *Philos. Trans. Roy. Soc. B: Biol. Sci.*, vol. 360, no. 1456, pp. 815–836, 2005.

[2] K. Friston, J. Mattout, and J. Kilner, "Action understanding and active inference," *Biol. Cybern.*, vol. 104, pp. 137–160, 2011.

[3] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by watching: Extracting reusable task knowledge from visual observation of human performance," *IEEE Trans. Robot. Automat.*, vol. 10, no. 6, pp. 799–822, Dec. 1994.

[4] G. Hayes and Y. Demiris, "A robot controller using learning by imitation," *Citeseer*, vol. 676, 1995.

[5] K. Dautenhahn, "Getting to know each other - artificial social intelligence for autonomous robots.," *Robot. Auton. System*, vol. 16, no. 2-4, pp. 333–356, 1995.

[6] P. Gaussier, S. Moga, M. Quoy, and J. P. Banquet, "From perception-action loops to imitation processes: A bottom-up approach of learning by imitation," *Appl. Artif. Intell.*, vol. 12, no. 7-8, pp. 701–727, 1998.

[7] G. Di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti, "Understanding motor events: A neurophysiological study," *Exp. brain Res.*, vol. 91, no. 1, pp. 176–180, 1992.

[8] M. Arbib, "The Mirror System, Imitation, and the Evolution of Language.," *Imitation in Animals and Artefacts*, pp. 229–280, 2002.

[9] E. Oztop, M. Kawato, and M. Arbib, "Mirror neurons and imitation: A computationally guided review," *Neural Netw.*, vol. 19, no. 3, pp. 254–271, 2006.

[10] T. Inamura, Y. Nakamura, H. Ezaki, and I. Toshima, "Imitation and primitive symbol acquisition of humanoids by the integrated mimesis loop," in *Proc. IEEE Int. Conf. Robot. Automat.*, vol. 4, 2001, pp. 4208–4213.

[11] A. Billard and M. Mataric, "Learning human arm movements by imitation: Evaluation of a biologically-inspired connectionist architecture," *Robot. Auton. Syst.*, vol. 941, pp. 1–16, 2001.

[12] E. Oztop, T. Chaminade, G. Cheng, and M. Kawato, "Imitation bootstrapping: Experiments on a robotic hand," in *Proc. 5th IEEE-RAS Int. Conf. on Humanoid Robots*, 2005., 2005, pp. 189–195.

[13] M. Ito and J. Tani, "On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system," *Adaptive Behav.*, vol. 12, no. 2, pp. 93–115, 2004.

[14] J. Hwang, J. Kim, A. Ahmadi, M. Choi, and J. Tani, "Dealing with large-scale spatio-temporal patterns in imitative interaction between a robot and a human by using the predictive coding framework," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 5, pp. 1918–1931, May 2020.

[15] A. Clark, *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. London, U.K.: Oxford Univ. Press, 2015.

[16] A. Ahmadi and J. Tani, "A novel predictive-coding-inspired variational rnn model for online prediction and recognition," *Neural computation*, vol. 31, no. 11, pp. 2025–2074, 2019.

[17] S. Murata, Y. Yamashita, H. Arie, T. Ogata, S. Sugano, and J. Tani, "Learning to perceive the world as probabilistic or deterministic via interaction with others: A neuro-robotics experiment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 830–848, Apr. 2017.

[18] P. Lanillos and G. Cheng, "Adaptive Robot Body Learning and Estimation Through Predictive Coding," *CoRR*, vol. abs/1805.03104, 2018.

[19] C. Lang, G. Schillaci, and V. Hafner, "A deep convolutional neural network model for sense of agency and object permanence in robots," in *Proc. Joint IEEE 8th Int. Conf. Develop. Learn. Epigenetic Robot.* (ICDL-EpiRob), pp. 257–262, 2018.

[20] A. Philippsen and Y. Nagai, "Deficits in prediction ability trigger asymmetries in behavior and internal representation," *Front. Psychiatry*, vol. 11, p. 1253, 2020.

[21] Y. Yamashita and J. Tani, "Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment," *PLoS Comput. Biol.*, vol. 4, no. 11, 2008.

[22] H. F. Chame and J. Tani, "Cognitive and motor compliance in intentional human-robot interaction," *IEEE ICRA*, 2019, pp. 11291–11297, 2020.

[23] W. Ohata and J. Tani, "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," *Front. Neurorobot.*, vol. 14, p. 61, 2020.

[24] L. Pio-Lopez, A. Nizard, K. Friston, and G. Pezzulo, "Active inference and robot control: A case study," *J. Roy. Soc. Interface*, vol. 16, 2016.

[25] G. Schillaci, A. Ciria, and B. Lara, "Tracking emotions: Intrinsic motivation grounded on multi-level prediction error dynamics," *10th Joint IEEE ICDL-EPIROB*, pp. 1–8, 2020.

[26] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio, "A recurrent latent variable model for sequential data," in *Adv. neural Inf. Process. Syst.*, 2015, pp. 2980–2988.

[27] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," 2014.

[28] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.

[29] F. M. Bianchi, S. Scardapane, S. Løokse, and R. Jenssen, "Reservoir computing approaches for representation and classification of multivariate time series," *IEEE Trans. neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2169–2179, May 2021.