Cognitive and motor compliance in intentional human-robot interaction

Hendry Ferreira Chame[*1] and Jun Tani[*2]

Abstract— Embodiment and subjective experience in human-robot interaction are important aspects to consider when studying both natural cognition and adaptive robotics to human environments. Although several researches have focused on nonverbal communication and collaboration, the study of autonomous physical interaction has obtained less attention. From the perspective of neurorobotics, we investigate the relation between intentionality, motor compliance, cognitive compliance, and behavior emergence. We propose a variational model inspired by the principles of predictive coding and active inference to study intentionality and cognitive compliance, and an intermittent control concept for motor deliberation and compliance based on torque feed-back. Our experiments with the humanoid Torobo portrait interesting perspectives for the bio-inspired study of developmental and social processes.

I. INTRODUCTION

Our society is probably changing into a world populated by natural and artificial beings, where we will coexist with robots at home and the office. However, at present, there is still an important gap to be overcome, in particular, when we pause to contemplate the amazing complexity of natural behavior. This sort of sophistication is related to important properties, among which are autonomy and intentionality.

When studying motor interaction between a human and a robot, some works have focused on goal-directed collaboration (e.g. mediated by physical objects [1], and behavior improvement [2]), leaving aside the social dimension of the interaction with the robot partner. With some exceptions, direct contact in autonomous behavior has been avoided [3]. Moreover, in social robotics, the non-verbal aspects of interaction has been studied from diverse modalities (e.g. facial expressions [4] and touch [5]), though motor clues in direct contact has attracted less attention.

In psychology, physical interaction is considered an intuitive means of communication during human early life, which underlies the acquisition and development of social and cognitive skills [6]. Hence, it is paramount for human development, and it is relevant for studying learning from the perspective of developmental robotics [7].

Intentional motor interaction is certainly a broad phenomenon that ought to be delimited. Thus, by taking inspiration on neuroscience research, we hold the assumption that intentionality involves an optimization process in hierarchical representation structures, in which a top-down information flow characterizes the intention to behave in a given context, whereas a bottom-up information flow accounts for

*Cognitive Neurorobotics Research Unit (CNRU), Okinawa Institute of Science and Technology (OIST), Okinawa, Japan 904-0495.
[1] Hendry.FerreiraChame@oist.jp
[2] tani1216jp@gmail.com

the consequences of action. We investigate this within the perspective of Friston's *free energy minimization* principle [8], according to which the brain attempts to resolve conflicts possibly appearing between the top-down intentional flow, developed by a generative model, and the bottom-up sensory flow, through minimizing free energy as a statistical quantity.

Our research is interested in the study of harmonious (or coherent) and disharmonious (or incoherent) interaction between the human and the robot. Thus, an important distinction is established between physical (or motor) and cognitive (or representational) compliance. The former is defined as the capacity to be driven by external physical action, which can be conforming or conflicting with the intended action; whereas the later involves the flexibility in modifying the generative process according to information from the ascending process, that is, the capacity to be driven by sensory evidence (the inference process).

After having established the previous considerations, we claim that the originality and main contribution of our research is to study the relation between physical and cognitive compliance in intentional motor interaction. Within the perspective of neurorobotics, we propose a variational model, inspired by the principles of predictive coding and active inference [9], to study intentionality and cognitive compliance. We propose an intermittent control concept for the study of motor deliberation and compliance that takes into account torque feed-back. From the analysis of interaction and behavior emergence in experiments with the human-sized Torobo platform, we illustrate the relevance of our work to the study of direct interaction, with interesting perspectives for the bio-inspired approach to developmental and social robotics.

II. RELATED WORK

Although independently developed, our research has been consistent with the *free energy minimization principle* [9]. In [10] a deterministic hierarchical neural network architecture operating on different time scales (i.e. a Multiple Timescale Recurrent Neural Network, or MTRNN) was proposed for learning temporal sequences. Inspired by these ideas, behavior imitation was studied from visual and proprioceptive representations (e.g. in [11]).

In order to improve generalization, stochastic modeling was adopted where uncertainty in training data is learned as a Gaussian distribution in the output layer [12]. However, a limitation of this approach is the fact that the context layers in the hierarchy remained deterministic.

The emergence of the variational Bayes auto-encoder (VAE) framework [13] paved the way for optimizing in-

ference and learning probabilistic distributions in latent variables, by re-parameterizing the variational lower bound. This framework has been extended for studying on-line interaction, as for example anticipating human behavior [14].

The Predictive-coding-inspired Variational Recurrent Neural Network (PV-RNN) [15] has been proposed recently in our research group. It is a variational model capable of learning probabilistic structures in fluctuating temporal patterns, by modifying dynamically the stochasticity of the represented latent states. Different from VAEs, inputs are not propagated in the network during forward computations. Instead, conforming to the framework of predicting coding [16], in PV-RNN prediction errors are back propagated through time (BPTT).

Unlike [11], where the robot imitates the human behavior sensed by visual input, our work is focused on direct physical interaction based exclusively on the proprioceptive source of information. Since our interest is to investigate autonomous and possibly conflicting interaction, different from [14], our study does not focus on predicting the human behavior, but on the inference process over proprioceptive representations. Thus, the agent receives the human intentions through how its body posture is changed within the interaction, which characterizes an egocentric experience.

III. COGNITIVE COMPLIANCE

As a variational framework, PV-RNN includes a generative and an inference model (see Fig. 1). The generative model implies the prior distributions. It does hierarchical predictions on the output layer following a top-down flow. That is, the prediction is generated from the information represented in the latent state. The inference model involves posterior distributions, so the information flow goes in the bottom-up sense. Predictions are done starting from the evidence (i.e., a given observation), and the latent representation is modified in order to approximate the evidence.

Let[1] the generative model P_ϕ be defined from the parameters ϕ, distributed among the components: generated prediction $\mathbf{x}$, stochastic $\mathbf{z}$ and the deterministic $\mathbf{d}$ latent states. For a prediction $\mathbf{x}_{1:T} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)$, and considering the parameters ϕ_x, ϕ_z, and ϕ_d, P_ϕ factorizes such that:

$$P_\phi\left(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \mathbf{d}_{1:T} | \mathbf{z}_0, \mathbf{d}_0\right) = \prod_{t=1}^{T} P_{\phi_\mathrm{x}}\left(\mathbf{x}_t | \mathbf{d}_t\right) P_{\phi_\mathrm{z}}\left(\mathbf{z}_t | \mathbf{d}_{t-1}\right) P_{\phi_\mathrm{d}}\left(\mathbf{d}_t | \mathbf{d}_{t-1}, \mathbf{z}_t\right) \quad (1)$$

Let the deterministic states be defined according to a MTRNN structure. For the kth context layer at time t, with timescale ι^k, the internal dynamics are represented such that

[1]Notation: layer's latent states are denoted bold low-case, biases are denoted $\mathbf{b}$, weight connexion are denoted $\mathbf{W}$ with subscripts indicating the origin and destination of the connection (e.g., $\mathbf{W}_\mathrm{zd}$ are the weights connecting $\mathbf{z}$ to $\mathbf{d}$ units). Superscripts k indicate the level in the MTRNN hierarchy. Finally, the superscripts p and q are used to distinguish between variables that belong to the prior and posterior distributions, respectively.
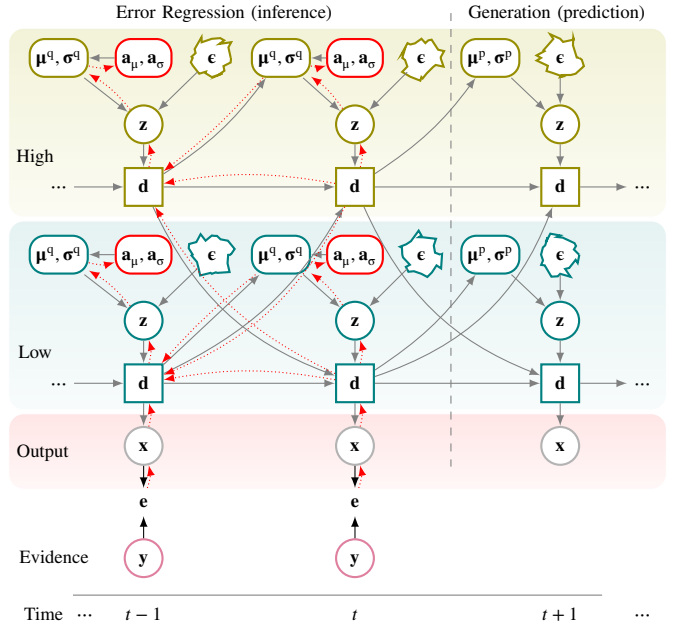

Fig. 1: A two-layer PV-RNN architecture [15]. The notation has been simplified for clarity, so the k and t indexes have been dropped. The time constant ι^k (see Eq. (2)) in greater in the High layer than in the Low layer. Since the High layer is the top on the hierarchy, the term $\mathbf{W}_\mathrm{dd}^{kk+1}\mathbf{d}_{t-1}^{k+1}$ is removed, analogously, the term $\mathbf{W}_\mathrm{dd}^{kk-1}\mathbf{d}_{t-1}^{k-1}$ is removed for the Low layer. The top-down process flow is represented by gray arrows. Red arrows illustrate the bottom-up process flow, where error is back propagated through time. The on-line inference process computed in a sliding time window is named Error Regression [10].

$$\mathbf{u}_t^k = \mathbf{W}_\mathrm{dd}^{kk}\mathbf{d}_{t-1}^k + \mathbf{W}_\mathrm{dd}^{kk-1}\mathbf{d}_{t-1}^{k-1} + \mathbf{W}_\mathrm{dd}^{kk+1}\mathbf{d}_{t-1}^{k+1} + \mathbf{W}_\mathrm{dd}^{kk}\mathbf{z}_t^k$$
$$\mathbf{h}_t^k = \left(1 - \frac{1}{\iota_k}\right)\mathbf{h}_{t-1}^k + \frac{1}{\iota_k}\mathbf{u}_t^k \quad .$$
$$\mathbf{d}_t^k = \tanh\left(\mathbf{h}_t^k\right)$$
$$(2)$$

The prior distribution $P_{\phi_z}\left(\mathbf{z}_t | \mathbf{d}_{t-1}\right)$ is modeled as a Gaussian with diagonal covariance matrix, such that

$$P_{\phi_z}\left(\mathbf{z}_t | \mathbf{d}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t; \boldsymbol{\mu}_t^\mathrm{p}, \boldsymbol{\sigma}_t^\mathrm{p}\right), \quad (3)$$

where $\boldsymbol{\mu}_t^\mathrm{p}$ and $\boldsymbol{\sigma}_t^\mathrm{p}$ are, respectively, the mean and standard deviation of $\mathbf{z}_\mathbf{t} = \boldsymbol{\mu}_t^\mathrm{p} + \boldsymbol{\sigma}_t^\mathrm{p} * \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon}$ sampled from $\mathcal{N}(0, 1)$. The variables $[\boldsymbol{\mu}_t^\mathrm{p}, \log(\boldsymbol{\sigma}_t^\mathrm{p})] = f_{\phi_z}(\mathbf{d}_{t-1})$ are obtained with $f_{\phi_z}(.)$ the one layer feed-forward neural network, such that

$$\boldsymbol{\mu}_t^{\mathrm{p},k} = \tanh\left(\mathbf{W}_{\mu\mathrm{d}}^{kk}\mathbf{d}_{t-1}^k + \mathbf{b}_\mu^{\mathrm{p},k}\right)$$
$$\log\left(\boldsymbol{\sigma}_t^{\mathrm{p},k}\right) = \mathbf{W}_{\sigma\mathrm{d}}^{kk}\mathbf{d}_{t-1}^k + \mathbf{b}_\sigma^{\mathrm{p},k} \quad . \quad (4)$$

Let the inference model Q_π (the approximate posterior) be defined from the parameters π, such that

$$Q_\pi(\mathbf{z}_t | \mathbf{d}_{t-1}, \mathbf{e}_{t:T}) = \mathcal{N}\left(\mathbf{z}_t; \boldsymbol{\mu}_t^\mathrm{q}, \boldsymbol{\sigma}_t^\mathrm{q}\right), \quad (5)$$

where $\boldsymbol{\mu}_t^{\mathrm{q}}$ and $\boldsymbol{\sigma}_t^{\mathrm{q}}$ are, respectively, the mean and standard deviation of $\mathbf{z_t} = \boldsymbol{\mu}_t^{\mathrm{q}} + \boldsymbol{\sigma}_t^{\mathrm{q}} * \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon}$ sampled from $\mathcal{N}(0,1)$. The variables $[\boldsymbol{\mu}_t^{\mathrm{q}}, \log(\boldsymbol{\sigma}_t^{\mathrm{q}})] = f_{\pi_z}(\mathbf{d}_{t-1}, \mathbf{a}^{\bar{\mathbf{x}}})$ are obtained with $f_{\pi_z}(.)$ the one layer feed-forward neural network, such that

$$
\begin{aligned}
\boldsymbol{\mu}_t^{\mathrm{q},k} &= \tanh\left(\mathbf{W}_{\mu\mathrm{d}}^{kk}\mathbf{d}_{t-1}^k + \mathbf{a}_{\mu,t}^{\bar{\mathbf{x}},k} + \mathbf{b}_\mu^{\mathrm{q},k}\right) \\
\log\left(\boldsymbol{\sigma}_t^{\mathrm{q},k}\right) &= \mathbf{W}_{\sigma\mathrm{d}}^{kk}\mathbf{d}_{t-1}^k + \mathbf{a}_{\sigma,t}^{\bar{\mathbf{x}},k} + \mathbf{b}_\sigma^{\mathrm{q},k}
\end{aligned} \quad .
\tag{6}
$$

The parameters $\mathbf{a}_{1:T}^{\bar{\mathbf{x}}}$ are introduced to provide the network with information about the prediction error in relation to a given pattern $\bar{\mathbf{x}}$. Thus, $\mathbf{a}_{1:T}^{\bar{\mathbf{x}}}$ is changed back propagating through time the prediction error $\mathbf{e}_{t:T}$, so information about the future steps of $\bar{\mathbf{x}}_{t:T}$, and existing dependencies with the current time step t, are captured. These terms are defined by

$$
\begin{aligned}
\mathbf{a}_{\mu,t}^{\bar{\mathbf{x}},k} &= \mathbf{a}_{\mu,t}^{\bar{\mathbf{x}},k} + \alpha\frac{\partial L}{\partial \mathbf{a}_{\mu,t}^{\bar{\mathbf{x}},k}} \\
\mathbf{a}_{\sigma,t}^{\bar{\mathbf{x}},k} &= \mathbf{a}_{\sigma,t}^{\bar{\mathbf{x}},k} + \alpha\frac{\partial L}{\partial \mathbf{a}_{\sigma,t}^{\bar{\mathbf{x}},k}}
\end{aligned} \quad .
\tag{7}
$$

with α denoting the learning rate.

Let the Variational Evidence Lower Bound (ELBO) $L(\phi, \pi)$ be defined by

$$
\begin{aligned}
L(\phi, \pi) = \sum_{t=1}^T \Big(&E_{\mathrm{q}_\pi}\left[\log\left(\mathbf{x}_t|\tilde{\mathbf{d}}_t, \mathbf{z}_t\right)\right] - \\
&w\mathrm{KL}\left[Q_\pi\left(\mathbf{z}_t|\tilde{\mathbf{d}}_{t-1}, \mathbf{e}_{1:T}\right) \| P_{\phi_z}\left(\mathbf{z}_t|\tilde{\mathbf{d}}_{t-1}\right)\right]\Big).
\end{aligned}
\tag{8}
$$

Since $\mathbf{d}_t$ is deterministic given $\mathbf{d}_{t-1}$ and $\mathbf{z}_t$, $\tilde{\mathbf{d}}_t$ denotes the center of a Dirac distribution. The first term at the right of the equation is a reconstruction component, it corresponds to the expected log-likelihood under the posterior distribution Q_π. The second therm is a regulation component, it corresponds to the Kullback-Leibler (KL) divergence between the prior and the posterior distributions of the latent variables. The meta-parameter w adjusts the optimization weight in learning the posterior and the prior distributions. After dropping the random variable notation to improve readability, the KL component can be expressed as

$$
\mathrm{KL}\left[Q_\pi \| P_{\phi_z}\right] = \log\left(\frac{\boldsymbol{\sigma}^{\mathrm{p}}}{\boldsymbol{\sigma}^{\mathrm{q}}}\right) + \frac{(\boldsymbol{\mu}^{\mathrm{p}} - \boldsymbol{\mu}^{\mathrm{q}})^2 + (\boldsymbol{\sigma}^{\mathrm{q}})^2}{2\,(\boldsymbol{\sigma}^{\mathrm{p}})^2} - \frac{1}{2}
\tag{9}
$$

Finally, the variable $\mathbf{x}_{i,t}$, related to the ith dimension of the output space, is defined such that

$$
\mathbf{x}_{i,t} = \mathrm{softmax}\left(\mathbf{W}_{\mathrm{dx}_i}\mathbf{d}_t^0 + \mathbf{b}_{\mathrm{x}_i}\right).
\tag{10}
$$

Unlike [15], we do not include in $\mathbf{x}_{i,t}$ connections from the stochastic latent distributions at the Low level.

IV. MOTOR COMPLIANCE

In conformity with the principles of ethics in robotics experiments, inspired by the famous Assimov's three laws of robotics [17], motor compliance is studied within the context of social ethical conventions. Thus, the human and the robot actions must preserve each other's integrity and safety.

Computational theories of human control have pointed out the plausibility of continuous and intermittent control systems in the brain, with the latter being possibly driven by events [18]. Intermittent control is related to the prefrontal cortex, the premotor cortex, and the basal ganglia areas, it has been studied in the context of postural regulation [19].

We model motor control as a hybrid intermittent process driven by intentionality and social interaction. The rationale behind this is the following. When contact between the human and the robot is established, their intended motion may complement or be to some extent incongruent. Thus, in both the human and the robot cases, the body posture must be adapted to the other's influence to preserve a safe interaction. Expressed in other terms, once external forces acting on the body induce significant joint torques, the joint should exhibit viscoelasticity, so it becomes compliant to the external force. Once the external constraints cease to be relevant, active control should be resumed from a gradual transition to the desired state. Hence, in this hybrid view, each robot articulation is controllable instantaneously by one of two possible schemes: a compliant and an active scheme.

Let the switching between the compliant and the active modes of joint j at time t rely on a continuous observation process of the torque $\hat{\tau}_{j,t}$, and the prediction on the body dynamics $\tau_{j,t}^{\mathrm{act}}$. The estimation of the external torque induced in the interaction is such that

$$
\hat{\tau}_{j,t}^{\mathrm{ext}} = \hat{\tau}_{j,t} - \tau_{j,t}^{\mathrm{act}}.
\tag{11}
$$

Thus, switching to the compliant mode occurs once $|\hat{\tau}_{j,t}^{\mathrm{ext}}| > \tau_j^{\mathrm{th}}$ exceeds a threshold τ_j^{th}.

Let the joint position $\theta_{j,t}$ be regulated by the active control scheme in charge of tracking a reference position $\theta_{j,t}^{\mathrm{net}}$, intended by the agent (i.e. it is generated by the neural network). The target position $\theta_{j,t+1}^{act}$, in relation to the observation $\hat{\theta}_{j,t}$, is obtained from a discrete proportional control law, such that

$$
\theta_{j,t+1}^{act} = \eta_{j,t}^{\mathrm{a}}\left(\theta_{j,t}^{\mathrm{net}} - \hat{\theta}_{j,t}\right).
\tag{12}
$$

Since $\theta_{j,t}$ may differ considerably from $\theta_{j,t}^{\mathrm{net}}$ under the compliant scheme, in order to avoid abrupt motions, the proportional gain $\eta_{j,t}^{\mathrm{a}}$ is set as the cosine transition modulated by $\left|\theta_{j,t}^{\mathrm{net}} - \hat{\theta}_{j,t}\right|$, from a minimum to a maximum gain.

Let the joint position $\theta_{j,t}$ be regulated by the compliant control scheme in charge of following the torque induced by the human. A discrete time proportional integral (PI) feedback control law, with gains η_j^{p} and η_j^{i}, is adopted, so

$$
\theta_{j,t+1}^{\mathrm{ext}} = \hat{\theta}_{j,t} + \eta_j^{\mathrm{p}}\hat{\tau}_{j,t}^{\mathrm{ext}} + f\left(\eta_j^{\mathrm{i}}\sum_t \hat{\tau}_{j,t}^{\mathrm{ext}}\right).
\tag{13}
$$

In both the active and the compliant schemes the target correction is saturated to preserve safety in behavior. Thus, the function $f(.)$ acts as a reset windup to the integral term.

As explained before, due to the fact that the control is done in the joint space, an interesting situation emerges where the

joints may be set to different control schemes. Hence, the human provides feedback to the robot through the compliant joints, while receiving feedback about the robot's intention from the active joints. However, it is also interesting to include soft impedance in the compliant joints in order to enhance the interaction experience. Thus, the target position in the compliant scheme is obtained such that

$$\theta^{com}_{j,t+1} = \theta^{ext}_{j,t+1} + \eta^n_j s\left(\theta^{net}_{j,t} - \hat{\theta}_{j,t}\right), \qquad (14)$$

with η^n_j a gain parameter and $s(.)$ a saturation function.

V. METHODOLOGY

On the robot side, the variables studied are *intentionality* (desired behavior in relation to the human's actions) and *compliance* in the physical (or motor, M) and the representational (or cognitive, C) dimensions. As shown in Table I, several profiles were defined to study compliance. Regarding the motor dimension, the exclusively compliant mode M_{teach} was employed for kinesthetic demonstration of the behavior primitives, and the interaction mode M_{inter} was set based on the hybrid controller (including the active and the compliant schemes). Concerning the cognitive dimension, the profiles C_{rigid}, C_{mod}, and C_{flex} were defined corresponding, respectively, to a rigid, a moderate, and a flexible agent.

TABLE I: The cognitive and motor profiles.

Profile	Description
M_{teach}	Kinesthetic teaching (with the compliance scheme only)
M_{inter}	Interaction (with the active and compliant schemes)
C_{rigid}	Strong intentionality ($w = 0.01$)
C_{mod}	Moderate intentionality ($w = 0.001$)
C_{flex}	Low intentionality ($w = 0.0001$)

The implementation of the motor control in the Torobo platform is illustrated in Fig. 2. A dataset was constituted with three behavior primitives (see Fig. 3), sampled at 4 Hz, during 90 time steps. The cognitive profiles in Table I were modeled in the PV-RNN architecture (see Fig. 1). It included 40 **d** units, 4 **z** units, and $\iota = 2$ in the Low layer; and 10 **d** units, 1 **z** units, and $\iota = 10$ in the High layer. The models were trained with distinct values for the optimization parameter w (Eq. (8)), during 50K epochs. All the parameters were learned in training, whereas in on-line experiments only the parameters **a** (see Eqs. (6)(7)) were updated.

On the human side, the variables under study are *intentionality* (desired behavior in relation to the robot's actions) and *engagement* (physical efforts invested in the task). Finally, on the mutual interaction side, the variable *behavior emerge* (how similar the robot's posture is to the known primitives) is studied through a regression observer. Since the motion primitives were reasonably different from one another, a feed-forward model was defined (12 Input, 150 Hidden$_1$, 15 Hidden$_2$, 3 Output, with tanh activation for the hidden layers and sigmoid activation for the ouptut layer), and trained by supervised learning with the joint instantaneous positions. A success rate of 100% was achieved in the test set.
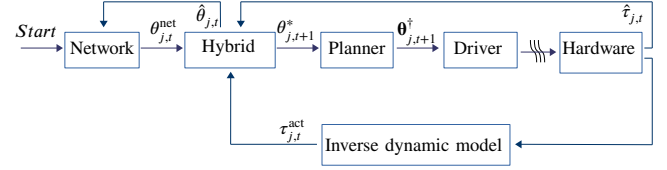

Fig. 2: The control scheme. The Network block contains the PV-RNN model that computes the desired position $\theta^{net}_{j,t}$ based on the current estimation $\hat{\theta}_{j,t}$. From the measured torque $\hat{\tau}_{j,t}$ and the dynamic torque $\tau^{act}_{j,t}$, calculated by the inverse dynamic model of the robot, the Hybrid controller computes the next target $\theta^*_{j,t+1}$. The Planner calculates a trajectory to achieve the target in open-loop via intermediate positions $\boldsymbol{\theta}^\dagger_{j,t+1}$. The robot Driver manages the Hardware plant.
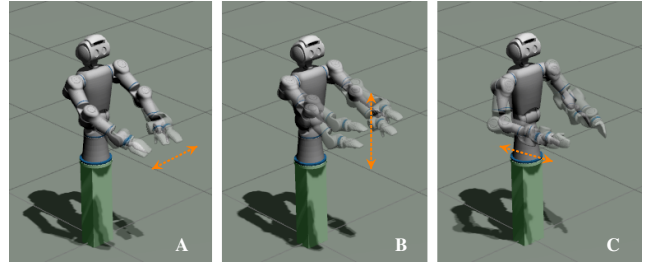

Fig. 3: On the top, the experimental scene. The primitives on the bottom were captured by kinesthetic demonstration with the real robot, and visualized in the Gazebo simulator.

The experimental protocol: Six scenarios were portrayed conforming the pairs AB, AC, BA, BC, CA, CB; with the left term representing the robot's intended behavior, and the right term corresponding to the human intended behavior (the subject was instructed to induce a given primitive in the robot behavior). For example, AB means the robot wants to do A and the human wants to do B. A total of 54 trials were registered for the experimental subject (6 pairs x 3 cognitive models x 3 times), during 300 time steps each.

Implementation details: The software architecture run under the Robot Operative System (ROS) version Kinetic Kane. The Network block in Fig. 2 ran at 4 Hz, it was implemented in an Alienware Aurora R7 computer, equipped with 12 Intel® Core™ i7-8700K CPU at 3.70GHz, and 31.1 GiB RAM memory, running Ubuntu 16.04 LTS. The PV-RNN model was implemented in C++ language from scratch. The BPTT algorithm in the on-line inference process was computed in a 20 time steps sliding window during 28 iterations. The models learned 12 degrees of freedom (6 for each arm), constant desired references were given to the torso and the head joints.

VI. RESULTS

The training results are shown in Fig. 4. As noticed, although the reconstruction error component of the ELBO (see Eq. (8)) approached to zero in all cases, which indicates good predictions from the posterior distribution; the smaller the meta-parameter w was set (right plot), the more dissimilar the posterior and the prior distributions were, which implies more stochasticity in the generative process.
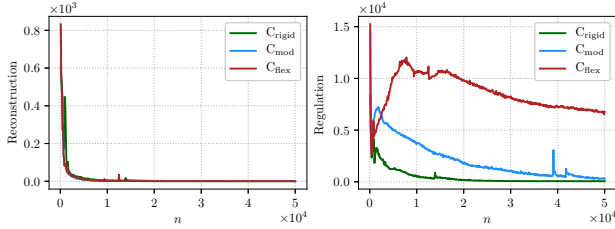

Fig. 4: Training during 50k epochs. The reconstruction (left) and the regulation (right) terms of the ELBO (see Eq. (8)).

A. Simulations

Two preliminary studies were conducted. The first one investigated the accuracy of the generative process. For this, the primitives were generated during 90 time steps (the length of the captured sequences in the dataset) for each cognitive model (see Table I). The comparison in done by calculating at each time step the mean squared error (MSE) between the reference $\bar{x}$ and the generated x sequences. The models C_{rigid} and C_{mod} performed similarly well (see Fig. 5), whereas C_{flex} had a more stochastic generative process.
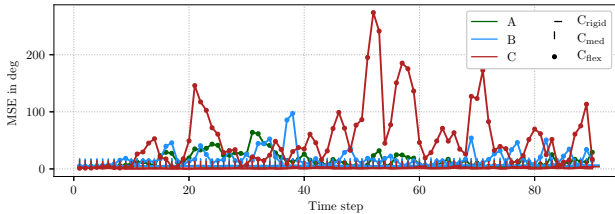

Fig. 5: Generative process accuracy.

The second study analyzed the cognitive compliance of the models, based on the values of the latent state d. Figure 6

presents a comparison on two principal component analysis (PCA) for the generation process of the primitives A, B, and C (left column), and the inference process (right column). Errors were BPTT by taking the difference between the model's prediction and recorded joint positions of A, B, and C. What is being analyzed is the possibility of transition from the generation of one primitive to another, given the evidence. That is, to what extend stochasticity from the hidden random distributions, received through the parameters a (see Eqs. (6)(7)), is allowed to affect the contextual representation during on-line inference. All transitions could be obtained, except from BC with C_{rigid}. As noticed, C_{flex} representations were simpler, and converged faster to conform the evidence.
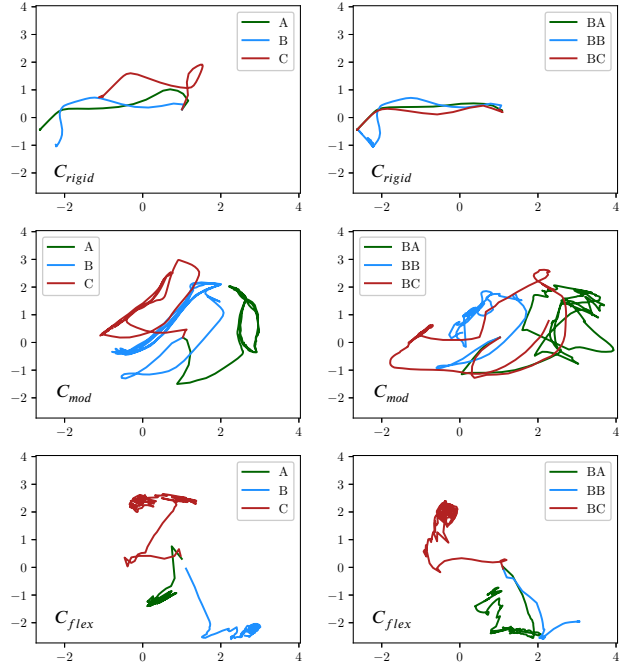

Fig. 6: Two PCA of the High layer's latent state d for the generation (left column) and the inference (right column) processes. The agents were set to generate B (intention), but received A, B, and C as evidence.

B. Experiments

Figure 7 shows a comparison on the performance of the intermittent controller for congruent and incongruent interaction, with a single joint (the right elbow), following a sinusoidal limit cycle reference. As noticed, smooth trajectories resulted from tracking the network signal while complying to the external torque induced by the human. Table II compares emergent behavior to the robot's intended behavior for the three cognitive agents, based on the regression observer's evaluation. Table III shows the mean and standard deviation of the estimated external torque during the experiment. In Fig. 8, emergent behavior for the pair BC is compared to the intended behavior of the robot[2].

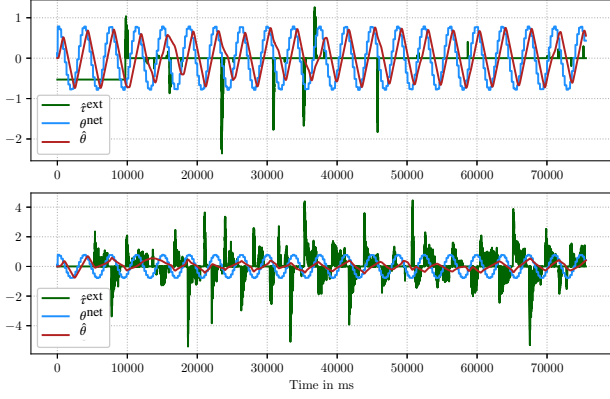[2]See the experiment video at: `https://youtu.be/MxbmIhstESE`

Fig. 7: Control performance (torque in Nm, angle in rad) for congruent (top) and incongruent (bottom) interaction.
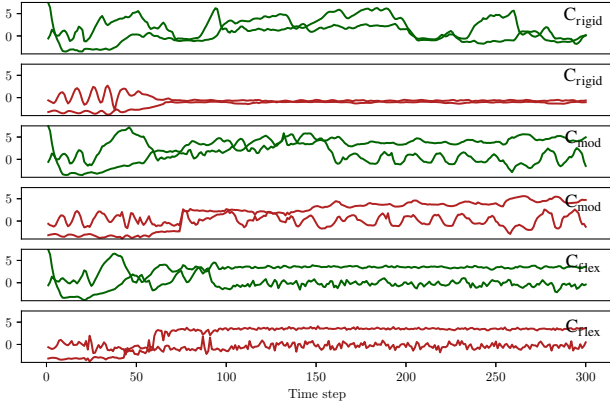

Fig. 8: Two PCA of emergent behavior (green) and intention (red) for the third BC trial.

C. Discussion

The results supported the plausibility to model distinct cognitive styles through the meta parameter w. Thus, by learning to strongly approximate the prior and posterior distributions (see Fig 4), the rigid agent was able to generate accurate behavior (see Fig 5), but it was less sensitive to evidence. Indeed, it obtained greater differences between the intended and observed behaviors (see Table II). Contrarily, the flexible agent was trained investing less efforts in learning to approximate these distributions, which resulted in stochastic or hesitating behavior. The moderate agent presented a good balance between accuracy and cognitive flexibility.

When analyzing the human engagement in the interaction, qualitative differences can be observed among the agents (see Table III, the lowest mean values per trial are highlighted in bold). Since the subject was instructed to induce as long as possible a certain behavior, with the rigid agent efforts were probably more invested in modifying the robot posture to encourage the generation of the desired primitive, given the agent's cognitive reluctance to change. Hence, the interactions were arduous and relied mostly on the compliant com-

TABLE II: Probability of the robot intending I and behaving B, according to the desired motions induced by the human.

Data	$\mathbf{C}_{\text{rigid}}$		$\mathbf{C}_{\text{mod}}$		$\mathbf{C}_{\text{flex}}$	
	$p(I)$	$p(B)$	$p(I)$	$p(B)$	$p(I)$	$p(B)$
Simulation	0.466	0.467	0.761	0.759	**0.843**	**0.841**
Trial 1	0.122	0.365	**0.562**	**0.588**	0.245	0.393
Trial 2	0.274	0.476	**0.705**	**0.686**	0.394	0.451
Trial 3	0.302	0.498	**0.590**	**0.619**	0.405	0.501

TABLE III: External torque $\sum_j |\hat{\tau}_{j,t}^{\text{ext}}|$ in Nm for trials.

Case	$\mathbf{C}_{\text{rigid}}$			$\mathbf{C}_{\text{mod}}$			$\mathbf{C}_{\text{flex}}$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
				Mean					
AB	2.0	**1.4**	1.9	**1.9**	1.8	**1.8**	2.0	2.0	1.9
AC	2.4	2.0	**1.1**	**2.2**	1.8	1.7	2.4	2.0	2.2
BA	1.9	1.6	**1.4**	**1.6**	**1.4**	1.6	1.7	**1.4**	**1.4**
BC	2.2	2.0	1.4	**2.1**	1.9	2.0	2.3	**1.8**	2.2
CA	**1.9**	2.2	**1.9**	2.0	**1.5**	2.1	1.8	1.9	2.2
CB	**1.9**	2.1	**1.8**	**1.9**	**1.9**	**1.8**	2.0	2.2	2.1
				Standard Deviation					
AB	0.9	0.8	0.9	1.0	1.0	0.9	1.1	1.0	1.0
AC	1.2	0.9	0.7	1.0	0.9	0.8	1.1	1.0	1.1
BA	1.0	0.9	0.7	0.9	0.9	0.8	0.9	0.8	0.8
BC	1.0	0.9	0.8	1.0	0.9	1.0	1.2	0.9	1.0
CA	0.9	0.9	0.8	1.0	1.0	1.3	0.9	0.9	1.0
CB	0.8	1.0	0.8	0.9	1.0	0.8	1.0	1.1	0.9

ponent of the hybrid control scheme. Considering the flexible agent, it likely adopted the desired posture shape, but efforts were required to keep the interaction going consistently. The subject faced difficulties when inducing gradual changes, due to erratic behavior from the robot, which would explain the difference between simulated and experimental results. In relation to the moderate agent, smoother interactions were obtained and less efforts appeared to be invested, since the agent was able to both change the posture and generating the behavior consistently. Finally, in agreement with [7], learning probably occurred on the human side, since both $p(I)$ and $p(B)$ were generally larger from the first to the third trials.

VII. CONCLUSIONS AND FUTURE WORK

This work focused on the study of physical interaction between a human and a robot, and considered both co-herent and incoherent scenarios. An important distinction was established between motor and cognitive compliance. A variational model, inspired by the principles of predictive coding and active inference, was proposed to model cognitive compliance as the capacity to be driven by sensory evidence. An intermittent control concept was proposed to study motor deliberation while adapting to the human interaction, based on torque feed-back. The experiments results pointed out a trade-off between cognitive compliance and refinement in autonomous motion. We believe that this trade-off can be explored in developmental robotics to investigate on-line learning [7]. When considering the social dimension of the interaction, our results open interesting perspectives for the research of motivational aspects inherent to the interaction, and how behavior is determined from the self [20].

REFERENCES

[1] S. Pellegrinelli, H. Admoni, S. Javdani, and S. Srinivasa, "Human-robot shared workspace collaboration via hindsight optimization," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 831–838.

[2] C. T. Landi, F. Ferraguti, C. Fantuzzi, and C. Secchi, "A passivity-based strategy for coaching in human-robot interaction," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–6.

[3] M. Kollmitz, D. Büscher, T. Schubert, and W. Burgard, "Whole-body sensory concept for compliant mobile robots," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–5.

[4] C. T. Ishi, T. Minato, and H. Ishiguro, "Analysis and generation of laughter motions, and evaluation in an android robot," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019.

[5] Y. Hu and G. Hoffman, "Using skin texture change to design emotion expression in social robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 2–10.

[6] L. S. Vygotsky, *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.

[7] S. Ikemoto, H. B. Amor, T. Minato, B. Jung, and H. Ishiguro, "Physical human-robot interaction: Mutual learning and adaptation," *IEEE Robotics Automation Magazine*, vol. 19, no. 4, pp. 24–35, Dec 2012.

[8] K. Friston, "The free-energy principle: a unified brain theory?" *Nature reviews neuroscience*, vol. 11, no. 2, p. 127, 2010.

[9] K. Friston, J. Mattout, and J. Kilner, "Action understanding and active inference," *Biological cybernetics*, vol. 104, no. 1-2, pp. 137–160, 2011.

[10] J. Tani, "Learning to generate articulated behavior through the bottom-up and the top-down interaction processes," *Neural Networks*, vol. 16, no. 1, pp. 11–23, 2003.

[11] 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 Transactions on Systems, Man, and Cybernetics: Systems*, 2018.

[12] S. Murata, J. Namikawa, H. Arie, S. Sugano, and J. Tani, "Learning to reproduce fluctuating time series by inferring their time-dependent stochastic properties: Application in robot learning via tutoring," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 4, pp. 298–310, 2013.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[14] J. Bütepage, H. Kjellström, and D. Kragic, "Anticipating many futures: Online human motion prediction and generation for human-robot interaction," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–9.

[15] A. Ahmadi and J. Tani, "A novel predictive-coding-inspired variational rnn model for online prediction and recognition," *Neural Computation*, in press.

[16] K. Friston, "A theory of cortical responses," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 360, pp. 815–36, 05 2005.

[17] I. Asimov, *I, robot*. Gnome Press, 1950, vol. 1.

[18] P. Gawthrop, I. Loram, M. Lakie, and H. Gollee, "Intermittent control: a computational theory of human control," *Biological cybernetics*, vol. 104, no. 1-2, pp. 31–51, 2011.

[19] H. Tanabe, K. Fujii, Y. Suzuki, and M. Kouzaki, "Effect of intermittent feedback control on robustness of human-like postural control system," *Scientific reports*, vol. 6, p. 22446, 2016.

[20] H. F. Chame, F. P. Mota, and S. S. da Costa Botelho, "A dynamic computational model of motivation based on self-determination theory and CANN," *Information Sciences*, vol. 476, pp. 319–336, 2019.