2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Deep Active Inference in Physical Human-Robot Interaction: Balancing Exploration and Goal-Directed Behavior

Jefimija Borojevic[1†], Gabriel W. Haddon-Hill[1†], Juan Sandoval[2] and Shingo Murata[1]

***Abstract*— In physical Human-Robot Interaction (pHRI), a significant challenge lies in the unpredictable nature of human behavior, which can introduce a high level of uncertainty during an interaction with a robot. While traditional force-based control laws lack high-level reasoning, learning-based methods do not treat perception and action uniformly to decrease uncertainties about human intention. This paper presents a pHRI framework based on Active Inference (AIF) for planning and decision-making, which guides a robot to balance goal-directed behavior and the exploration of the human intention. The framework integrates a 1D-CNN based Conditional-Variational Autoencoder (CVAE) architecture for Expected free energy (EFE) computation and policy selection with a Cartesian impedance controller to allow the robot to adapt its motion according to the selected policy. Additionally, this paper investigates the influence of preference precision on the robot's behavior. The pHRI experiment includes pushing, pulling and no-applied interaction scenarios. The results show that the robot naturally favors goal-directed, high-stiffness behavior when it is undisturbed, while it prefers exploration behavior for lower preference precision values and goal-directed behavior for higher preference precision values during pushing and pulling interaction scenarios. The findings of this research demonstrate that the preference precision parameter significantly influences the process of minimizing uncertainty of human behavior, enabling the robot to adaptively balance exploration and goal-directed behavior in pHRI scenarios.**

I. Introduction

Recent advances in robotic technologies have enabled the development of robotic systems with enhanced safety and reliability, bringing them closer to direct interaction with humans [1]. As a result, developments in physical Human-Robot Interaction (pHRI), in which a robot aims to increase human capabilities in strength, speed and accuracy, while a human can guide interaction through their intention, have been demonstrated over diverse applications [2], ranging from robotic assistance in industrial and non-industrial settings [3], [4], [5], rehabilitation robots [6], [7], [8] to assistive suits such as exoskeletons [9], [10], [11].

Despite these advances, pHRI still faces challenges due to the high level of uncertainty inherent in interaction with humans whose behaviors can be unpredictable. Traditionally, classical force control laws, such as impedance and admittance control, have been dominantly used to enable robot motion compliance. However, since these are model-based approaches and reactive to external forces, they lack high-level reasoning of human intention during the interaction [12]. To address this limitation, learning-based methods have been increasingly adopted in order to infer human intention and adapt robot behavior, such as Bayesian inference, Markov models, supervised learning and reinforcement learning approaches [13].

Although these methods can capture complex data patterns related to the interaction between robot and human, they often treat the perception, prediction, and action as separate processes, possibly struggling in decreasing uncertainties across these stages. In contrast, a biological plausible theory called Active Inference (AIF), grounded in the free energy principle, has been introduced as a unified framework to integrate perception and action by minimizing a bound on surprise [14]. This process enables constant inference about the environment's states, adapts and updates the inner states accordingly, and acts on the environment in order to reduce the surprise.

In order to scale AIF to complex real-world scenarios, the key conditional distributions are often amortized using deep neural networks [15], [16], which is known as deep AIF. Advances in scalable AIF, including deep AIF, has improved its applicability to real-world robotics [17], [18]. Such works have been realized success in state estimation [19], adaptive control [20], [21], [22], planning [23], [24] and cognitive skills [25]. To the best of the author's knowledge, only the study by Sawada et al. [26] has applied free energy principle in a pHRI scenario to transfer one robot motion to another according to a human intention. However, this study does not consider the planning of future actions based on their predicted consequences, which is a key feature of AIF.

In this study we explicitly tackle a pHRI scenario using AIF, and explore the ensuing behavior of a robot during online interactions with a human partner. We implement a learned generative model using a 1D-CNN architecture to infer latent states from position and force data. Through the use of a dynamic prior model, the robot plans future actions by predicting their consequences. Our deep AIF model dynamically infers the human's intentions and adapts its behavior depending on both the interaction and a tunable preference precision parameter that modulates the exploration-exploitation trade-off. We demonstrate that the robot naturally balances exploration of human intention with its own goal-seeking desires. To the best of our knowledge, this is the first study to apply deep active inference with EFE policy selection in a pHRI setting.

†These authors contributed equally to this work.
[1]School of Integrated Design Engineering, Keio University, Yokohama, Kanagawa 223-8522, Japan. Emails: jefimija_borojevic@keio.jp, gabe.haddon-hill@keio.jp, murata@elec.keio.ac.jp
[2]Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France. Email: juan.sandoval@ec-nantes.fr

978-1-6654-5784-2/26/$31.00 ©2026 IEEE

II. Theory

A. Active Inference

Active Inference (AIF) is a theory grounded in neuroscience that describes the processes of perception, behavior, and learning as the minimization of *Free Energy*—a variational bound on surprise ($-\log P(o)$) [27]. An AIF agent embodies a generative model $P(o, z)$ of its environment, which it uses to infer the hidden causes z of sensory observations o. Due to the intractability of exact Bayesian inference, it is assumed that the agent maintains an approximate posterior over hidden states $Q(z|o)$, which can be optimized using, for example, variational inference. Variational Free Energy (VFE) scores the mismatch between the agent's beliefs and the sensory evidence. The VFE for a given time-step t is given by:

$$F = D_{\mathrm{KL}}[Q(z_t|o_t) \| P(z_t)] - \log P(o_t|z_t). \tag{1}$$

A key aspect of AIF is the recognition that the agent has two methods of minimizing Free Energy at its disposal: updating its beliefs about the hidden causes (perception), and acting upon the world in order to elicit different sensory observations (action). Thus, the roles of perception and decision-making are united under a single variational objective.

B. Expected free energy

Whilst minimizing VFE allows an agent to perceive and react to its surroundings, the ensuing behavior is reflexive, since it does not permit planning sequences of actions (policies). Expected free energy (EFE) is an extension of VFE which incorporates the expected surprise of future states under a certain policy π. It therefore allows the agent to plan by selecting policies that minimize EFE. For a given future time-step τ, EFE can be expressed as:

$$G_\tau(\pi) = \underbrace{-\mathbb{E}_{Q(z_\tau, o_\tau|\pi)}\left[D_{\mathrm{KL}}[Q(z_\tau|o_\tau, \pi) \,\|\, Q(z_\tau|\pi)]\right]}_{\text{Information Gain}} \underbrace{-\mathbb{E}_{Q(z_\tau|\pi)}\left[\ln P(o_\tau|C)\right]}_{\text{Pragmatic Value}}. \tag{2}$$

This particular decomposition of EFE emphasizes the separation of epistemic and goal-seeking drives. The information gain (IG) term incentivizes states that the agent expects to learn a lot from, encouraging exploratory behaviors. The pragmatic value (PV) term, on the other hand, encodes the agent's goal-seeking drives by pushing it to maximize the likelihood of preferred sensations. The C parameter in the PV term refers to the agent's *prior preferences*, which result in a generative model that is biased towards certain desirable regions of the environment. Expressed in this way, EFE minimization can directly address the exploration-exploitation dilemma inherent in other decision-making schemes, and the relative balance between the two terms is modulated via the preference precision parameter (inverse variance) γ of the likelihood distribution $P(o_t|C)$ [28].

III. Methodology

A. Deep AIF framework and EFE computation

We present our deep AIF framework in Fig. 1. Each of the main distributions are parameterized by deep neural networks. These include prior $Q_\theta(z_t|o_{t-P:t}, \pi)$, posterior $Q_\phi(z_{t+F}|o_{t-P:t+F}, \pi)$ and likelihood $P_\theta(o_{t-P:t+F}|z_t, \pi)$ models with parameters θ and ϕ. Note that the output of all models is conditioned on the policy π. The prior and posterior receive observations comprising sequences (length $P+1$ for Q_θ and $P+F+1$ for Q_ϕ) of 3-dimensional end-effector positions and external forces, for a total of 6 channels, and output the means and log-variances of a Gaussian distribution over the 128-dimensional latent variable z_t. The likelihood model generates reconstructions of the full observation sequence $o_{t-P:t+F}$, and is modeled as a Gaussian with a fixed variance.

The policies in our model are represented by a choice between two sets of stiffness parameters (high or low) used in the impedance controller. This is described in more detail in the discussion of our experimental framework. As there are two policies π is a flat binary vector with a length of either $P+1$ or $P+F+1$ depending on which network is being used. When predicting future outcomes, we concatenate the vector of the last $P+1$ actions with a vector of length F containing only a single action (high or low SP). We take inspiration from [29], which also uses action conditioning to imagine the results of different policies.

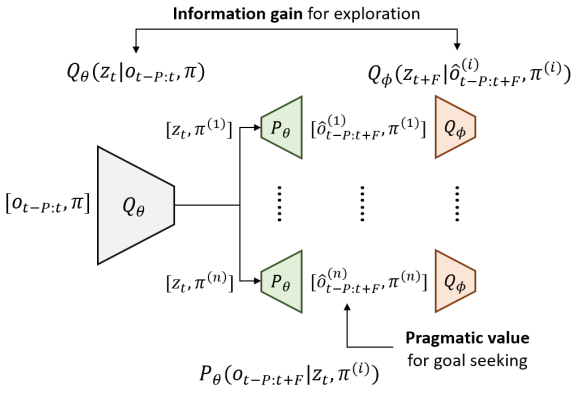


Fig. 1: Deep AIF framework for computing EFE

EFE is computed as follows. First, a prior is formed given the last $P+1$ observations and actions, representing the agent's belief about both the current situation and the next F time-steps. Next, a reconstruction $\hat{o}^{(i)}_{t-P:t+F} \sim P(o_{t-P:t+F}|z_t, \pi^{(i)})$ is generated for each policy $\pi^{(i)}$. Each is then provided to the posterior model giving $Q_\phi(z_{t+F}|\hat{o}_{t-P:t+F}, \pi)$, representing the agent's prediction of its future belief state under a $\pi^{(i)}$. The expectations in (2) are estimated via ancestral sampling from $Q_\theta(z_t|o_{t-P:t}, \pi)$. 10 samples of z_t are drawn, resulting in 10 reconstructions and predicted posteriors.

Both EFE terms can then be computed for each policy. Since our likelihood is assumed to be a homoscedastic

Gaussian, the pragmatic value term is equivalent to a γ-weighted MSE (ignoring constants):

$$PV^{(i)} = -\gamma \text{MSE}(\hat{o}_{t-P:t+F}, o_{goal}), \tag{3}$$

and the information gain term measures how much new information the agent expects to gain for a particular policy:

$$IG^{(i)} = D_{\text{KL}}[Q_\phi(z_t|\hat{o}^{(i)}_{t-P:t+F}, \pi^{(i)})||Q_\theta(z_t|o_{t-P:t}, \pi)]. \tag{4}$$

Once the EFE for each policy has been computed, the policy that minimizes EFE is selected deterministically:

$$\pi = \text{argmin}(G(\pi^{(n)})). \tag{5}$$

B. 1D-CNN CVAE model

Our model is implemented as a 1D-CNN based Conditional-Variational Autoencoder (CVAE) architecture with separate prior and posterior encoder networks. Using a CVAE means that we can condition the model on the policy and thus predict outcomes contingent on specific policies. Furthermore, the use of 1D convolutions instead of dense layers allows our model to better learn temporal patterns in the flat time-series input data.

Unlike the classical CVAE, our model uses a dynamic prior model conditioned on a reduced sequence of observations $o_{t-P:t}$. This choice allows for the prior to be used in a predictive capacity by estimating the probability of policy-conditioned futures based on currently available information. By contrast, the posterior model only seeks to encode information contained in observations up to the present moment. This is a crucial detail in our EFE computation. The overall model parameters, including kernel sizes and layer depths, are summarized in Table I.

The loss function is the standard β-VAE ELBO loss, containing reconstruction and regularization terms:

$$\begin{aligned}\mathcal{L}_{\text{CVAE}} = \beta\, D_{\text{KL}}[Q_\phi(z_{t+F}|\hat{o}_{t-P:t+F}, c) \,\|\, Q_\theta(z_t|o_{t-P:t}, c)] \\ + \text{MSE}(\hat{o}_{t-P:t+F}, o_{t-P:t+F})\,.\end{aligned} \tag{6}$$

Note that this loss function is equivalent to the VFE shown in (1). The β parameter scales the influence of regularization term relative to the reconstruction term. The regularization loss represents the difference between the approximate posterior distribution $Q_\phi(z_{t+F}|\hat{o}_{t-P:t+F}, c)$ and the prior distribution $Q_\theta(z_t|o_{t-P:t}, c)$. Since our framework uses a moving prior, the regularization term acts both to regularize the posterior representation and to improve the predictive capabilities of the prior model. The reconstruction loss is the mean squared error between the observation data (without conditional vector) and likelihood model's reconstruction. This term encourages accuracy of the model's latent representation.

TABLE I: 1D-CNN based CVAE architecture layers (stride and pooling are always 2 and 1 respectively)

Posterior Encoder	Prior Encoder	Likelihood Decoder
Conv1d(32, k=7)	Conv1d(32, k=7)	Linear 3200
LeakyReLU	LeakyReLU	LeakyReLU
Conv1d(64, k=5)	Conv1d(64, k=5)	Unflatten
LeakyReLU	LeakyReLU	ConvTrans1d(64, k=3)
Conv1d(128, k=3)	Conv1d(128, k=5)	LeakyReLU
LeakyReLU	LeakyReLU	ConvTrans1d(32, k=5)
Flatten	Conv1d(128, k=3)	LeakyReLU
Linear 256	LeakyReLU	ConvTrans1d(6, k=7)
/	Flatten	Tanh
/	Linear 256	/

C. Experimental framework

The overall framework developed in ROS Noetic distribution is presented in Fig. 2. It consists of two ROS nodes: one for the robot control and another for planning future actions by computing EFE: *cartesian_impedance_controller* and *deep_active_inference_framework* respectively.

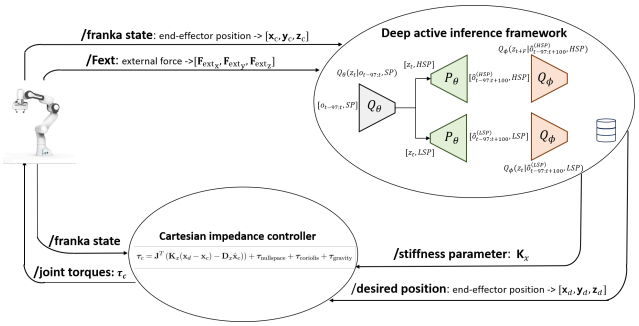


Fig. 2: ROS implementation of the overall framework on the Franka robot, featuring deep AIF and impedance control nodes.

The robot control node is a modification of the official ROS node from *franka_ros* package, which implements Cartesian impedance control. We modified the node to dynamically receive different stiffness parameters $\mathbf{K}_x$ during robot motion execution. It receives desired end-effector position data at 200 Hz, which is transferred into robot joint torques through the impedance control law:

$$\boldsymbol{\tau}_c = \mathbf{J}^T\left(\mathbf{K}_x(\mathbf{x}_d - \mathbf{x}_c) - \mathbf{D}_x\dot{\mathbf{x}}_c)\right) + \boldsymbol{\tau}_{\text{nullspace}} + \boldsymbol{\tau}_{\text{coriolis}} + \boldsymbol{\tau}_{\text{gravity}}, \tag{7}$$

where $\boldsymbol{\tau}_c$ is the commanded joint torque vector, $\mathbf{x}_c$, $\dot{\mathbf{x}}_c$ and $\mathbf{x}_d$ are the current end-effector pose and velocity, and the desired end-effectors pose, $\mathbf{K}_x$ and $\mathbf{D}_x$ are the Cartesian stiffness and damping matrices, $\boldsymbol{\tau}_{\text{nullspace}}$ is the torque component for nullspace control, $\boldsymbol{\tau}_{\text{coriolis}}$ compensates for Coriolis forces, and $\boldsymbol{\tau}_{\text{gravity}}$ compensates for gravity.

At the same time, the deep AIF node uses the generative model to compute EFE and select the policy. It receives the current end-effector positions and applied external force data at 200 Hz, which is scaled and resampled to match the training data. A prior window is constantly created from the past 97 robot data samples, as well as the robot policy. After receiving the first prior window, the deep AIF framework computes EFE for both policies at 1 Hz, chosen to provide a smooth policy transition, and selects the policy that minimizes EFE for the current prior window. The selected

policy, stiffness parameter $\mathbf{K}_x$, is sent to the control node to adjust the robots behavior accordingly.

D. Experimental scenario

In this study, deep AIF has been examined in a pHRI experimental scenario with a collaborative robot Franka Research 3 (FR3). The robot's predefined motion is to go along the X-axis of its base coordinate system, moving forward-backward during 10 cycles. Y and Z coordinates should remain constant, where the goal observation o_{goal} is to maintain the same Y value as in the beginning of the motion, $Y = 0$. The duration of the motion is around 100 seconds.

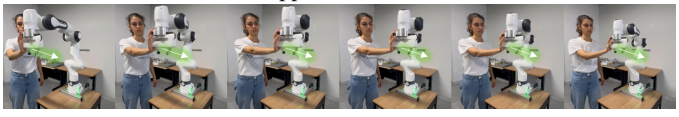
The robot's mechanical adaptability in response to external forces depends mainly on the choice of the stiffness parameter (SP) values $\mathbf{K}_x$ in the Cartesian impedance control scheme. For this experiment, two stiffness parameter values were chosen as actions. These values are selected empirically to represent different interaction modes: the high stiffness to provide sufficient robot rigidity, and the lower to allow human to easily guide the robot during the interaction.

- **High Stiffness Parameter (HSP):**
  $\mathbf{K}_x = \mathrm{diag}\big([300, 300, 300]\ \mathrm{N\,m^{-1}},\ [30, 30, 30]\ \mathrm{N\,m\,rad^{-1}}\big)$
- **Low Stiffness Parameter (LSP):**
  $\mathbf{K}_x = \mathrm{diag}\big([300, 50, 300]\ \mathrm{N\,m^{-1}},\ [30, 30, 30]\ \mathrm{N\,m\,rad^{-1}}\big)$

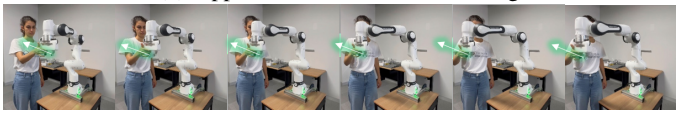
In our experimental setup, the pHRI is characterized by three different types of interaction, regarding the applied external force, as presented in Fig. 3:

- **No applied external force:** There is no physical interaction between the robot and the human; the robot continues following its defined motion along the X-axis.
- **Applied external force – Pushing:** User is pushing the robot along the Y-axis. The applied external force along the Y-axis is positive.
- **Applied external force – Pulling:** User is pulling the robot along the Y-axis. The applied external force along the Y-axis is negative.

(a) No applied external force



(b) Applied external force – Pushing



(c) Applied external force – Pulling

Fig. 3: Experimental scenarios: (a) no applied external force, (b) human pushing the robot, (c) human pulling the robot.

The interaction differs depending on the SP. In the HSP case the human applies short, regular bursts of pressure to the robot, whilst in the LSP case, the human moves the arm as far as the SP will allow.

In addition, the preference precision parameter γ is varied during the pHRI experimental scenario in order to investigate its effect on the trade-off between exploration and exploitation robot performance. The γ values used for the investigation are selected to cover a range of non-extreme robot behaviors for the chosen setup, varying from low to high. Specifically, the tested values are: $\gamma = \{0.1, 0.5, 0.75, 1, 1.5, 2, 5\}$.

E. Dataset and training process of 1D-CNN CVAE model

For our training dataset, we recorded 180 pHRI scenarios (60 for each scenario individually), each around 100 seconds long. These included robot end-effector positions and external forces along three axes and SP. During the recording process, SP were dynamically switched after random intervals between 1 and 15 s. This process ensured a rich variety of collected motion patterns that contain sufficient examples of action switching for the model to learn how to conditionally generate observations based on the policy.

The data was collected at 200 Hz and then down-sampled to 20 Hz. We also scaled all values to be between -1 and 1. The processed data was then cropped to prior windows (4.8s, 97 samples) and posterior windows data (10.1s, 203 samples). The specific window lengths were chosen to provide enough information to the posterior and prior models without making reconstructions too challenging, and to maintain a posterior–prior length ratio that allows the prior to predict many future observations while still providing sufficient data to inform its representation.

We held out 20% of the data for testing and validation purposes and trained 600 epochs using the Adam optimizer with a learning rate 10^{-6}. During each epoch, corresponding prior and posterior windows are randomly sampled from the pre-recorded episodes. We set $\beta = 50$ to prioritize regularization of the prior and posterior distributions over achieving perfect reconstruction of the output data.

IV. Results and Discussion

The following results illustrate how γ affects policy selection and, consequently, the robot behavior under different interaction types. In the absence of external force provided by the human, the robot prefers to choose the high stiffness policy and exhibits goal-directed behavior for any chosen γ values, meaning that it does not have high effect on robot behavior when there is no human intention to change it. On the other hand, during applied external force interactions, both pushing and pulling, the results show that for low γ values, the robot tends toward exploratory behavior, deviating from its goal trajectory in order to acquire more information about human intention. At high γ values, the robot exhibits goal-directed behavior, not allowing human to disturb its target motion.

To provide a demonstration on the obtained results, specifically on the direct physical interaction scenarios between human and robot, Fig. 4 provides examples of robot data and

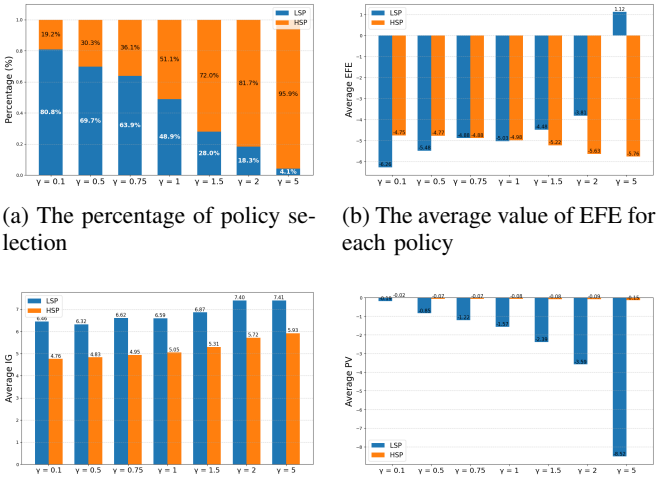
(a) $\gamma = 0.5$ (b) $\gamma = 1$ (c) $\gamma = 2$

Fig. 4: Examples of robot data results for pushing pHRI scenario

EFE computation results for pushing scenario on low ($\gamma = 0.5$), medium ($\gamma = 1$) and high ($\gamma = 2$) preference parameter values. The robot data is presented in the first three rows of Fig. 4: the first row shows robot's end-effector X position, the second row shows robot's end-effector Y position in red line, the third row presents robot's external force along Y-axis in green line. The second and third row also present used policy in a particular moment: high or low SP in the orange line. The last three rows demonstrate differences between LSP and HSP computed data from deep AIF framework: IG difference in the fourth row, PV difference in the fifth row and EFE difference in the sixth row.

Examining Fig. 4, it is revealed that the deviation of the end-effector position along Y-axis is larger for low γ value than for high γ value. For low γ, EFE is dominantly minimized by LSP, leading the robot to adapt more to the human applied force and to explore human's intention. As γ increases, HSP becomes more effective in minimizing EFE, making the robot maintain goal motion without adapting or showing an interest to discover what human aims. Medium γ is the one in which mixed robot behavior can be observed, both the stiff motion of not allowing human to make it deviate from its goal, such as shown in the part of Fig. 5b around 20 seconds, and more adaptive behavior where robot allows human to change its motion and push it more, such as the part around 70 seconds shows.

Since Fig. 4 shows only examples of pushing interaction scenario, Fig. 5 provides the overall trend for applied force scenarios, both pushing and pulling together, obtained from five recordings for each of the presented γ values. Fig. 5a and Fig. 5b reveal that the γ parameter crucially influences on robot behavior. For low γ values, LSP policy dominantly minimizes EFE, leading to an exploratory behavior, while as γ increases, robot behavior tends toward HSP policy, resulting in goal-directed behavior.



(a) The percentage of policy selection

(b) The average value of EFE for each policy

(c) The average value of IG for each policy

(d) The average of PV for each policy

Fig. 5: Deep AIF data analysis for both applied force scenarios

Fig. 5d of average PV and Fig. 5c of IG average can further explain the policy selection, as they are two main components of EFE. At low γ values, PV is small for both policies, making the policy choice more sensitive to IG value. Since the model expects to gain more knowledge about human intention if it adapts its motion and deviates more, the model selects LSP policy, as it provides higher IG value on average. On the other hand, higher γ values amplify PV term, especially for LSP policy. which will influence on producing higher EFE value and allowing HSP to become a dominant policy that minimizes EFE. This behavior results in choosing goal-directed behavior. Moreover, it is noticed that IG value increases as the γ increases as well, because the robot resistance and goal-directed choice for higher γ prevents the model from learning about human intention.

Overall, the results demonstrate that the robot behavior,

based on EFE minimization principle, is mainly influenced by preference precision parameter γ value during pHRI. They confirm that developed deep AIF framework enables the robot to manage the uncertainty of human intention across different pHRI scenarios by handling the trade-off between exploration and goal-directed behavior, and also show how the IG term can be framed in terms of exploring interactions with other agents in AIF.

V. Conclusion

This study examined the AIF principle to tackle a fundamental challenge of dealing with uncertainty of human behavior during interaction with a robot. A deep AIF framework was developed to enable a robot to infer human intention and make action choice based on EFE minimization principle. This framework integrated perception, prediction and action in the same planning process. The experimental results showed that the robot behavior choice was highly influenced by preference precision parameter γ value during applied external force interaction scenarios: low γ values favored exploratory behavior, while high γ encouraged goal-directed behavior. These findings demonstrate that the proposed deep AIF framework can support adaptive and goal-directed robot performance in the presence of differing human intentions. Future research will extend this work implementing more complex pHRI experimental scenarios and possibly the integration of multimodal sensory inputs, such as vision, to assist the inference process of human intention.

Acknowledgment

This work was supported by the Japan Science and Technology Agency (PRESTO Grant Number JPMJPR22C9 and SPRING Grant Number JPMJSP2123), and JSPS KAKENHI Grant Number JP24K03012. Gabriel W. Haddon-Hill is also supported by the Graduate School Doctoral Course Scholarship from the Amano Institute of Technology.

References

[1] S. Haddadin and E. Croft, "Physical human–robot interaction," in *Springer handbook of robotics*. Springer, 2016, pp. 1835–1874.

[2] M. Farajtabar and M. Charbonneau, "The path towards contact-based physical human–robot interaction," *Robotics and Autonomous Systems*, vol. 182, p. 104829, 2024.

[3] C. Y. Wong, L. Vergez, and W. Suleiman, "Vision-and tactile-based continuous multimodal intention and attention recognition for safer physical human–robot interaction," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 3205–3215, 2023.

[4] J. Luo, C. Zhang, W. Si, Y. Jiang, C. Yang, and C. Zeng, "A physical human–robot interaction framework for trajectory adaptation based on human motion prediction and adaptive impedance control," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 5072–5083, 2024.

[5] M. Charbonneau, F. J. A. Chavez, and K. Mombaur, "Slow waltzing with reem-c: a physical-social human-robot interaction study of robot-to-human communication," *arXiv preprint arXiv:2408.05301*, 2024.

[6] A. Mohebbi, "Human-robot interaction in rehabilitation and assistance: a review," *Current Robotics Reports*, vol. 1, no. 3, pp. 131–144, 2020.

[7] J. M. Prendergast, S. Balvert, T. Driessen, A. Seth, and L. Peternel, "Biomechanics aware collaborative robot system for delivery of safe physical therapy in shoulder rehabilitation," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7177–7184, 2021.

[8] J. Zhou, Y. Sun, R. Song, and Z. Wei, "Dynamic movement primitives modulation-based compliance control for a new sitting/lying lower limb rehabilitation robot," *IEEE Access*, vol. 12, pp. 44 125–44 134, 2024.

[9] M. P. De Looze, T. Bosch, F. Krause, K. S. Stadler, and L. W. O'sullivan, "Exoskeletons for industrial application and their potential effects on physical work load," *Ergonomics*, vol. 59, no. 5, pp. 671–681, 2016.

[10] M. Gunasekara, R. Gopura, and S. Jayawardena, "6-rexos: Upper limb exoskeleton robot with improved phri," *International Journal of Advanced Robotic Systems*, vol. 12, no. 4, p. 47, 2015.

[11] H. Lee, B. Lee, W. Kim, M. Gil, J. Han, and C. Han, "Human-robot cooperative control based on phri (physical human-robot interaction) of exoskeleton robot for a human upper extremity," *International Journal of Precision Engineering and Manufacturing*, vol. 13, no. 6, pp. 985–992, 2012.

[12] A. Kalinowska, P. M. Pilarski, and T. D. Murphey, "Embodied communication: How robots and people communicate through physical interaction," *Annual review of control, robotics, and autonomous systems*, vol. 6, no. 1, pp. 205–232, 2023.

[13] G. Hoffman, T. Bhattacharjee, and S. Nikolaidis, "Inferring human intent and predicting human action in human–robot collaboration," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 7, 2024.

[14] K. Friston, "The free-energy principle: a unified brain theory?" *Nature reviews neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[15] K. Ueltzhöffer, "Deep active inference," *Biological cybernetics*, vol. 112, no. 6, pp. 547–573, 2018.

[16] T. Champion, M. Grześ, L. Bonheme, and H. Bowman, "Deconstructing deep active inference," *arXiv preprint arXiv:2303.01618*, 2023.

[17] L. Da Costa, P. Lanillos, N. Sajid, K. Friston, and S. Khan, "How active inference could help revolutionise robotics," *Entropy*, vol. 24, no. 3, p. 361, 2022.

[18] P. Lanillos, C. Meo, C. Pezzato, A. A. Meera, M. Baioumy, W. Ohata, A. Tschantz, B. Millidge, M. Wisse, C. L. Buckley *et al.*, "Active inference in robotics and artificial agents: Survey and challenges," *arXiv preprint arXiv:2112.01871*, 2021.

[19] P. Lanillos and G. Cheng, "Adaptive robot body learning and estimation through predictive coding," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4083–4090.

[20] G. Oliver, P. Lanillos, and G. Cheng, "An empirical study of active inference on a humanoid robot," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 462–471, 2021.

[21] C. Pezzato, R. Ferrari, and C. H. Corbato, "A novel adaptive controller for robot manipulators based on active inference," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2973–2980, 2020.

[22] K. Fujii, T. Isomura, and S. Murata, "Real-world robot control based on contrastive deep active inference with demonstrations," *IEEE Access*, 2024.

[23] K. J. Friston, T. Parr, and B. de Vries, "The graphical brain: Belief propagation and active inference," *Network neuroscience*, vol. 1, no. 4, pp. 381–414, 2017.

[24] G. W. Haddon-Hill and S. Murata, "Active vision for physical robots using the free energy principle," in *International Conference on Artificial Neural Networks*. Springer, 2024, pp. 270–284.

[25] 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," *Frontiers in Neurorobotics*, vol. 14, p. 61, 2020.

[26] H. Sawada, W. Ohata, and J. Tani, "Human–robot kinaesthetic interactions based on the free-energy principle," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024.

[27] T. Parr, G. Pezzulo, and K. J. Friston, *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press, 2022.

[28] K. Igari, K. Fujii, G. W. Haddon-Hill, and S. Murata, "Selection of exploratory or goal-directed behavior by a physical robot implementing deep active inference," in *International Workshop on Active Inference*. Springer, 2024, pp. 165–178.

[29] T. Van de Maele, T. Verbelen, O. Çatal, C. De Boom, and B. Dhoedt, "Active vision for robot manipulators using the free energy principle," *Frontiers in neurorobotics*, vol. 15, p. 642780, 2021.