Predictive Motor Control Based on a Generative Adversarial Network

Movitz Lenninger [1†], **Whan-Hyuk Choi** [2†] **and Hansol Choi** [3*]

[1] *Division of Information Science and Engineering, KTH Royal Institute of Technology*
[2] *Department of Mathematics, Kangwon National University, 1 Gangwondaehakgil, Chuncheon, 24341, Republic of Korea*
[3] *Department of Biomedical Engineering, College of Information-Bio Convergence Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea*
[†] *These authors contributed equally to this work and share first authorship*

Correspondence*:
Hansol Choi
capine@gmail.com

ABSTRACT

Predictive processing models suggest that a brain decides actions through inference using its internal generative model over the world's states and their transitions. Most of the predictive processing models have been formalized using explicit representations of the probability distributions, with explicit structures and parameters. They are difficult to learn in general and needs explanation about representation of structure and parameters of distributions, and the method for statistical arithmetic on them, each of which are questions not easy to answer. In this study, we explore an alternative representation for predictive processing which is based on an implicit model known as generative adversarial networks, which has been widely explored recently in machine learning studies as they can learn a distribution directly from data. We demonstrate how a generative adversarial network can be trained to learn an implicit generative model of motor dynamics. And then, we show that such a model can perform approximate inference using the trained model, providing the necessary computations for both the forward and inverse model of motor control. Our framework may provide another formalization for brain's inference model, especially for learning process. Additionally, we suggest that the functional architecture of the cortical-basal ganglia circuit may modeled as the generator and discriminator in the generative adversarial network model.

Keywords: implicit generative model, Generative Adversarial Network, generative model, predictive processing, action inference

1 INTRODUCTION

In Neuroscience, there exists a widely influencing idea that brains are inference machines. They use internal generative models to explain the cause of the observations and decide actions to maximize adaptability. The models under the idea are collectively known as predictive processing (Clark, 2012). In the core of the models are the generative model (Pickering and Clark, 2014; Botvinick and Toussaint, 2012; Solway and Botvinick, 2012; Friston et al., 2009), which can be understood as "stochastic recipes for generating observed data" (Gershman, 2019). For example, a probabilistic distribution $P(\tilde{o}, \tilde{s}, \pi|m)$ is a generative model to give the probability of observing a sequence of observation $\tilde{o}$, when the sequence of hidden states and the policy are $\tilde{s}$ and π. The distribution function is parameterized with m. The perception is

formalized to infer $P(\tilde{s}|\tilde{o}, \pi, m)$, which is the posterior probability of $\tilde{s}$ when the probability of a sequence of observation $P(\tilde{o})$ and the agent's policy $P(\pi)$ is given. The action is to find $P(\tilde{s}, \pi|\tilde{o}, m)$.

The predictive processing needs to explain how to use the generative models for inference mechanisms. Although the exact posterior can be computed using Bayes rule $P(s|o) = P(o|s)P(s)/\sum_s P(o|s)P(s)$, it is intractable for many of realistic conditions when the state s is high dimensional, or its functional form is not easy to integrate. The approximate inference methods alleviate the problems. For example, the active inference framework uses a variational Bayes method (Friston, 2010; Smith et al., 2022). It tries to find an approximate distribution $Q(s)$ of $P(s|o)$, of which the KL-divergence $KL(Q \parallel P) = -\mathbb{E}_Q \log(Q/P)$ is minimized. The method can keep the inference procedure tractable by carefully choosing the function Q. Active inference demonstrated the methods to approximate $Q(\tilde{s}) \sim P(\tilde{s}|\tilde{o}, \pi, m)$ for state perception on the observations and policy, and $Q(\tilde{s}, \pi) \sim P(\tilde{s}, pi|\tilde{o}, m)$ for action generation on observation. The method also provided an update rule for m to explain the (parametric) learning of the $P(\tilde{s}, \tilde{o}, \pi)$ to update the generative model. Studies have demonstrated that predictive processing can explain a wide variety of brain functions, including perception problems such as visual perception or binocular rivalry, or decision-making (Hohwy et al., 2008; Friston et al., 2009). One question is how to implement the statistical arithmetic in explicit model from neural network.

The previous studies use explicit generative models that the probability of samples are explicitly represented in probabilistic distribution functions. (Gershman and Beck, 2017; Gershman, 2019; Bogacz, 2017; Friston et al., 2009). The (approximate) inference is to compute the parameters of the distribution functions using Bayesian or statistical techniques. One question is the representation of explicit generative models in the brains and its inference mechanisms. Although several hypotheses, such as population coding, explain the representation of the probabilities in brains, it is not an easy-to-answer question (Pouget et al., 2013). Moreover, the result of inference is a probability of states or actions. A state or action of the agent must be finally sampled from the distribution. It means that the agent's state and the generative models should exist independently from each other.

The last important question is learning from the environment. Most predictive processing studies is that the learning are about inference with the model given a prior or to parametric learning. Structural learning of the generative model, for example, learning the functional family of the distribution, is an important task for the brain but mostly overlooked in predictive processing model's studies.

Implicit generative models are worth investigating as an alternative formalization for predictive processing. The implicit models have a function $G(z')$, which transforms a random sample z' from an arbitrary distribution $z' \sim P_z(z)$ to vector x in a sample space. The density of the sample $x = G(z')$ implicitly represents the x's distribution density. A recently developed machine learning technique called generative adversarial network (GAN) has been successful in training implicit generative models for diverse domains of high dimensional data.(Goodfellow et al., 2014; Radford et al., 2015) Moreover, GAN can learn representation of the data to encode the semantic information. For example,when GAN is trained for face images latent states were found to to encode face directions.(Yeh et al., 2016)

Here, we will demonstrate predictive processing with an implicit generative model based on GAN. First, using a simple toy environment, we show how GAN can learn a generative model of motor system dynamics. Second, we demonstrate how approximate inference using the trained GAN can behave as two of the most important models for motor control: the forward and inverse models. The forward model predicts the system's future state from the current state and action, and the inverse model decides the actions required to achieve a target state (Pickering and Clark, 2014; Jordan and Rumelhart, 1992; Desmurget and

Grafton, 2000; Franklin and Wolpert, 2011). Finally, we will discuss the neural implications of the GAN models. We argue that the functional architecture of the Cortico-Basal Ganglial system can be interpreted in terms of the generator-discriminator system of our GAN predictive processing model.

2 MATERIALS AND METHODS

2.1 Implicit Generative Model and predictive processing

2.1.1 Planning as inference

Planning is to select a policy π which maximizes an objective function such as return. An agent with planning as inference uses an internal generative model $P(\tilde{o}, \tilde{s}, \pi|m)$ for planning. (Smith et al., 2022). Active inference agents use a prior preference distribution $p(o|C)$, which denotes the agent's preference for observing a certain observation o. The agent infers current hidden state and then infer action with under the current observation and the preferred observation as evidences. The action is biased toward achieving states by the preferred observation distribution.

One of the simplest examples is a single-step state-action-state condition. A state s and action u is followed by a s' in a discrete-time system. The generative model for a predictive processing agent is $P(o, s, u, o', s')$. When we set a pooled hidden state $z = (s, s')$ the generative model is $P(o, u, o', z)$. Planning is to infer the $P(u|o, o')$, when the current observation is o, and the preferred observation is o'. Bayes rule and marginalization gives $P(u|o, o') \sim \sum_z P(o, o'|z)P(u|z)P(z)$ as the exact inference result (Under POMDP setup, o, o', u are independent each other given z.). An agent can approximately decide latent states by finding optimal $\hat{z}$ to maximize $P(o, o'|z)P(z)$, then, use $P(u|\hat{z})$ to decide action.

Another important problem in motor control called the forward problem can be also formalized as inference. It is to infer $P(o'|o, u)$, the future state from the current state in action. Similarly, it is to decide $\hat{z}$ maximizing $P(o, u|z)P(z)$ and $P(o'|\hat{z})$ to infer the future observation.

2.1.2 Learning an implicit generative model by GAN

The agents in the real world cannot directly access the generative process of the world, which determines the states of the sample space. Instead, they should learn a generative model from the observations $x \sim p_{data}$, where p_{data} is the real distribution density from the generative process. GAN is developed to learn directly from the data. Two functions are learned simultaneously. One is a generator, $G(z; \theta_z)$, which maps a random noise z with prior distribution $p_z(z)$ to the sample space and parameterized by θ_z (Goodfellow et al., 2014; Arjovsky et al., 2017; Yeh et al., 2016). Another is $D(x; \theta_D)$ called discriminator, which maps a value in sample space to a scalar value. During the training, the purpose of G is to generate realistic samples. while D plays an adversarial role in discriminating between the motor sequence generated from G and the real data. G and D are trained by optimizing the loss function:

$$\max_D \min_G V(G, D) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z\ p_z(z)}[\log(1 - \log D(G(z)))] \tag{1}$$

2.1.3 Planning as inference using an implicit generative model

Inference using an implicit generative model is not easy and has not been studied much. For example, Monte Carlo sampling can be used on the implicit generative model. Assume we want to decide an action u when current and target observation $oando'$ are given. We can sample $G(z)$ using $z \sim P_z(z)$ and filter them by $O = o, O' = o'$. The U from the filtered samples may provide the target distribution. The method will

require to generate a lot of samples, memory to store them, and a mechanism to filter samples and count the samples for density. Moreover, we will need a decision process for action from the filtered samples. We do not want agent's states independently exist its distribution.

We will investigate a method from machine learning studies. Yeh et al. ((Yeh et al., 2016)) demonstrated how to infer using the GAN generative model. Their problem was to in-paint an omitted part $x_{omitted}$ of an image when the other part, $x_{evidence}$ is given. The problem is to find $x_{omitted}$ maximizing $P(x_{ommited}|x_{evidence})$. When z is the latent state of the generative model $P(x, z)$, the problem is to find z which maximizes $P(x_{evidence}|z)P(z)$ similar to the previous section. ($x_{evicence}$ is independent from $x_{ommited}$ given z).

Yeh et al. used GAN, the implicit generative model. Their main idea was to find an optimal input $\hat{z}$ which minimizes a combination of prior loss, L_p, and context loss, L_c by gradient-descent.

$$L_p(z) = D(G(z)) \tag{2}$$

$$L_c(z; x_{evidence}) = \|G(z) \odot M - \mathbf{x} \odot M\| \tag{3}$$

$$L_{total}(z) = L_p(z) + \lambda L_c(z) \tag{4}$$

G is the generator and D is the discriminator. They were pre-trained for the same domain data with the in-paint problems. x is a sample with an omitted part. M is mask vector for the $x_{evidence}$ that, $M_i = 1$ if x_i is given as evidence, $M_i = 0$, otherwise.

The context loss evaluates the difference between a generated sample $G(z)$ and the given part of a sample $x_{evidence}$. As the likelihood $P(x_{evidence}|z)$ in an explicit form evaluates how well a latent state z explains a piece of evidence $x_{evidence}$, $L_p(z)$ evaluates related property to the likelihood. (Further studies are needed for the exact relationship). The discriminator evaluates z whether a sample $G(z)$ exists in the trained data in another loss function L_p. It measures quantity related to the prior $P(G(z))$in the explicit Bayesian method, which provides the probability of $G(z)$ in the overall distribution. Together, the Bayesian posterior maximization for z where $P(z|x_{evidence}) \sim P(x_{evidence}|z)P(z)$ is related to Yeh's method, even though further studies are needed for finding the exact relationship between them.

2.2 Simulation

2.2.1 A Toy Environment

We will use an arm environment to demonstrate our idea in the following. We will use a simple two-joint model with a single time step (Fig1a), which 14 dimensions. And also, we additionally tested four joint and two-two joint arms models to scale up the model. (See supplementary material) Each has 24 and 28 dimensions.

The variables of our toy arm environment are as below.

- $Q = (q_1, q_2) \in \mathbb{R}^2$: The angle of the first and second joint.
- $\dot{Q} = (\dot{q_1}, \dot{q_2}) \in \mathbb{R}^2$: the angular velocities of the joints.
- $U = (u_1, u_2) \in \mathbb{R}^2$: u_i the action given on the i-th joint.
- $(x, y) \in \mathbb{R}^2$: The position of the cursor at the end of the arm.
- $S = (Q, \dot{Q}, x, y)$: The state of the arm.

- symbols with $'$: The state after action.

The dynamics of the system are as follows:

- $\dot{Q}' = \dot{Q} + U_t$
- $Q' = Q + \dot{Q}'$

In total, the toy environment has six degrees of freedom. We will use the notation $a \underset{given}{|} b$ to indicate a, which is fixed from b. For example,

$$S' \underset{given}{|} S, U$$

represents the future state, S' of our toy environment when a set of current states and actions, S and U, are given.

2.2.2 Training GAN and inference

We uniformly sampled current joint states and the current actions. The joint angles were between $[0, \pi/2]$, angular velocities between $[-\pi/6, \pi/6]$ and the actions between $[-\pi/6, \pi/6]$. Each dimension was independently normalized between -1 and 1 by first subtracting each dimension's average and then dividing by the maximum deviation from the average. Two fully connected multilayer perceptrons with leaky ReLU activation functions were used as G (6-8-14 nodes) and D(14-8-1 nodes; Fig. 1b and trained for the samples from toy environment. The forward problem is to infer S' when the current state S and action U are given. The inverse problem is to infer an action, U, which brings the cursor to a desired position (x, y) from the current state S is given. The approximate inference method using GAN explained above was used to infer the target values.

3 RESULTS

3.1 Training Generative Model for a Two-Arm System

We trained a GAN generative model on data generated from the toy environment. Fig. 2a shows representative samples from the naive and trained model, respectively. During the training process, we evaluated the GAN by letting it generate 1000 random inputs **z** and tested whether the S' in $G(z)$ is close to $S' \underset{given}{|} S, U$ where S and U from G. Comparing the joint states and cursor positions between the actual and generated future states shows that as the number of iterations increases, the model-generated samples get closer to obeying the dynamics of the actual toy model environment. Both the error between the actual and generated future joint states and the error between the actual and generated cursor position decrease during training (Fig. 2b). This demonstrates that the adversarial training process is capable of producing a generative model which reasonably reflects the actual dynamics of the environment solely based on training data.

3.2 The Forward and Inverse Model by Approximate Inference with GAN

To investigate whether the trained GAN model can be used as a forward model, we generated 100 test samples from the toy environment and masked the future states. Based on the knowledge of the current state and action(Fig. 3, black arms), the generative model is tasked with completing the sample for the prior and context loss. Compared to the naive model, the trained model is much better at predicting the future state of the arm (compare blue and red arms in Fig. 3a). The error for the future joint states between the

actual and predicted future state (joint angle and angular velocities) decreases with increasing iterations (Fig. 3b). The same holds for the error between the predicted position of the cursors from the forward model and the actual future cursor position (Fig.3b).

The same approximate inference method is used as an inverse model (Fig. 4). In the inverse model, the task is to predict the action when the agent's current state and target cursor positions are known (black arms and dots in Fig. 4a, respectively). As the generative model trains, the inferred action by the inverse model becomes more accurate compared to the untrained model, that it moves the cursor closer to the target position (red crosses in Fig. 4a).

We tested our system on two more conditions. One with a case with non-task related dimensions that two two joints arms exist and only one arm is related with task (supplementary figure 2). The other was with four joints arm (supplementary figure 2). They had more dimensions to learn the generative models (28 and 24 dimensions correspondingly). For the both of the cases, our system could learn the generative model and be used to infer the better action to achieve target condition compared to the naive system.

4 DISCUSSION

4.1 Implication in neuroscience

We hypothesize that this functional architecture of GAN for implicit generating model and inference method can be related to the functional architecture of the Cortico-Basal Ganglial (BG) system: the neocortex as the generator and the BG as the discriminator.

Firstly, the discriminator requires (1) to receive signals from different sources of inputs to discriminate actual motion from internally generated, (2) to be capable of reducing the high-dimensional inputs to a lower-dimensional evaluation value, and (3) to send the evaluation signal to both the generator and the discriminator to teach and instruct.

The Basal Ganglia (BG) system may have those desired characteristics. (1) The dense input neurons of BG receive inputs from the diverse areas of the brain (Gurney et al., 2001; Bar-Gad and Bergman, 2001; Bar-Gad et al., 2003; Watabe-Uchida et al., 2012), especially from the neocortex and BG reduces the dimensions of the inputs to the activity of the smaller dimension of DAergic neurons (Bar-Gad et al., 2003; Watabe-Uchida et al., 2012). (2) We suggest that the phasic DAergic signal may act as the discriminator signal. Phasic DAergic activity has been suggested to function as a temporal difference signal (i.e., the difference between the predicted and arrived state value) for reinforcement learning (Montague et al., 2004; Schultz et al., 1997; Houk et al., 1995). Importantly, recent studies suggest alternative roles of DA neurons to represent non-reward information (Tian et al., 2016; Noudoost and Moore, 2011) and state prediction errors (Wei et al., 2015; Engelhard et al., 2019; Schwartenbeck et al., 2016; Suarez et al., 2019; Takahashi et al., 2017; Nakahara et al., 2004). Gardner et al. (Gardner et al., 2018) suggested that the phasic DA signal may not only encode the reward prediction error but perhaps act as a general prediction error signal in a diverse feature space. The prediction error can be interpreted as an evaluation signal of the generated samples, as suggested by Pfau and Vinaylas (Pfau and Vinyals, 2016). (3) The DA projects back to the basal ganglia as well as to diverse brain areas, including the neocortex (see (Puig et al., 2014) for a review). Thus, the discriminatory signal can be accessible to both the discriminator and generator networks. These functional structures of dimensional reduction, encoding of low-dimensional evaluation signals about how realistic a sample is, and signal projections of the evaluation signals are all necessary (even though not sufficient) properties for a discriminator.

The generator requires (1) to receive evaluation input from the discriminator and to be trained from the signal, (2) to generate potential samples for the control sequences, and (3) to map abstract latent input into a specific sample. We propose that the neocortex may have the necessary properties.

(1) The neocortex receives DAergic projections from the BG, which produce synaptic triad (Yao et al., 2008) and modifies the synapse by the prediction error signal. The DAergic prediction error signal may teach the neocortex area. (2) The motor cortex activity is correlated with the current state and actions (Fried et al., 2011; Cisek and Kalaska, 2005; Li et al., 2016; Churchland et al., 2010, 2012) as well as imaginary motor movement (Schnitzler et al., 1997; Doud et al., 2011). The imagination of motor movement is one way of motor sample generation, which is a possible behavior by a generator. (3) The cortical area is suggested to have a functional hierarchy ranging from abstract higher causes to more concrete sensory-motor states for visual perception (DiCarlo et al., 2012), motor control (DeWolf and Eliasmith, 2011; Tani, 2016; Dumoulin et al., 2017), and decision making (Ribas-Fernandes et al., 2011, 2018).

The place for the forward and inverse model inference process is also needed. First of all, the generator needs to generate samples matching the context cues. The affordance competition (Pezzulo and Cisek, 2016; Cisek and Kalaska, 2005; Cisek, 2007) suggests a process of contexts limit the sample space. A signal of target states can activate the neurons correlated with the potential actions to achieve the target (Cisek and Kalaska, 2005). When multiple actions are possible for the target signals, the neurons correlated with competing actions are activated together, and the neural correlates for the irrelevant actions for the given context are reduced.

Secondly, another criterion to generate actions in our model is prior loss, in which the discriminator evaluates the samples generated by the generator by a trained model. We hypothesize that the thalamic output from BG may have the necessary properties. The BG projects to the thalamus as one of the circuit's outputs. Interestingly, the thalamic output of BG has been suggested to be correlated with decision-making among multiple choices (Dunovan and Verstynen, 2016; Wei et al., 2015; Gurney et al., 2015; Keuken et al., 2015). This evaluation is not only limited to the value of the decision but also related to prior decision-making(Ding and Gold, 2013; Lak et al., 2017). With the function of DA as a general prediction error signal explained above, we suggest that the BG might function as a general evaluator for generated samples based on the reward size and the probability of the sample.

4.2 Conclusion

In this study, we suggested a predictive processing model using an implicit generative model. GAN was capable of training a generative model for our toy environment motor control sequences. The training lowered the error between the generated samples and the real constraints of the toy environment. Importantly, although minimal structural assumptions were imposed beforehand, the system could work adequately after training in the environment. The forward and inverse models of motor control were constructed as approximate inference processes as data in-painting processes. Interestingly, the same approximate inference architecture was used for both models. The forward and inverse models only differed in what information is considered the evidence and what information is omitted. We suggested the potential relationship between the explicit Bayesian statistics and the context and prior loss that the context loss is related to the likelihood of $P(o|z)$ latent state for the observation and the prior loss as the prior for $P(G(z))$ or $P(z)$. Clearly, our suggestion remains conceptual and further mathematical and experimental studies are required.

We focused on the functional architecture of the system. How can the generated samples' evaluation signal teach the generator and discriminator? Our GAN framework for modeling brain functions needs a

lot more clarifications. For example, it uses back-propagation in the learning and inference process, which is still a topic of debate and may not be feasible in real neural systems. Another algorithm, such as the recently suggested forward-forward algorithm (Hinton, 2022), may provide alternative options but still needs further studies.

Another limitation is from the simple one-step generative model; the generative cannot explain the complex behavior requiring multiple steps of choice. Further studies are needed, such as a hierarchical generative model or successor representation, which represents the multiple steps of data as the density of states.

4.2.1 Permission to Reuse and Copyright

Figures, tables, and images will be published under a Creative Commons CC-BY license, and permission must be obtained for the use of copyrighted material from other sources (including re-published/adapted/modified/partial figures and images from the internet). It is the responsibility of the authors to acquire the licenses, follow any citation instructions requested by third-party rights holders, and cover any supplementary charges.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

ML designed, implemented code, analyzed data, and wrote the manuscript. WC analyzed the result and wrote the manuscript. HC designed, analyzed data, wrote the manuscript, and supervised the project.

FUNDING

Whan-Hyuk Choi is supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government (NRF-2022R1C1C2011689).

ACKNOWLEDGMENTS

We thank Prof. Dr. Carsten Mehring, University of Freiburg, for the comments that greatly improved the manuscript.

REFERENCES

Arjovsky, M., Chintala, S., and Bottou, L. (2017). enWasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (PMLR), 214–223

Bar-Gad, I. and Bergman, H. (2001). Stepping out of the box: information processing in the neural networks of the basal ganglia. *Current Opinion in Neurobiology* 11, 689–695. doi:10.1016/S0959-4388(01)00270-7

Bar-Gad, I., Morris, G., and Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in Neurobiology* 71, 439–473. doi:10.1016/j.pneurobio.2003.12.001

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology* 76, 198–211. doi:10.1016/j.jmp.2015.11.003

Botvinick, M. and Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences* 16, 485–488. doi:10.1016/j.tics.2012.08.006. 00020

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., et al. (2012). Neural population dynamics during reaching. *Nature* 487, 51–56. doi:10.1038/nature11129. 00000

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., and Shenoy, K. V. (2010). Cortical preparatory activity: Representation of movement or first cog in a dynamical machine? *Neuron* 68, 387–400. doi:10.1016/j.neuron.2010.09.015

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, 1585–1599. doi:10.1098/rstb.2007.2054

Cisek, P. and Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron* 45, 801–814. doi:10.1016/j.neuron.2005.01.027

Clark, A. (2012). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci*

Desmurget, M. and Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences* 4, 423–431. doi:10.1016/S1364-6613(00)01537-0

DeWolf, T. and Eliasmith, C. (2011). The neural optimal control hierarchy for motor control. *Journal of Neural Engineering* 8, 065009. doi:10.1088/1741-2560/8/6/065009. 00000

DiCarlo, J., Zoccolan, D., and Rust, N. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi:10.1016/j.neuron.2012.01.010

Ding, L. and Gold, J. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron* 79, 640–649. doi:10.1016/j.neuron.2013.07.042

Doud, A. J., Lucas, J. P., Pisansky, M. T., and He, B. (2011). Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface. *PLoS ONE* 6. doi:10.1371/journal.pone.0026322

Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., et al. (2017). Adversarially learned inference. *arXiv:1606.00704 [cs, stat]* ArXiv: 1606.00704

Dunovan, K. and Verstynen, T. (2016). Believer-skeptic meets actor-critic: Rethinking the role of basal ganglia pathways during decision-making and reinforcement learning. *Frontiers in Neuroscience* 10. doi:10.3389/fnins.2016.00106

Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H. J., Ornelas, S., et al. (2019). Specialized coding of sensory, motor and cognitive variables in vta dopamine neurons. *Nature* 570, 509–513. doi:10.1038/s41586-019-1261-9

Franklin, D. and Wolpert, D. (2011). Computational mechanisms of sensorimotor control. *Neuron* 72, 425–442. doi:10.1016/j.neuron.2011.10.006. 00126

Fried, I., Mukamel, R., and Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69, 548–562. doi:10.1016/j.neuron.2010.11.045

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11, 127–138. doi:10.1038/nrn2787

Friston, K. J., Daunizeau, J., and Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE* 4, e6421. doi:10.1371/journal.pone.0006421

Gardner, M. P. H., Schoenbaum, G., and Gershman, S. J. (2018). Rethinking dopamine as generalized prediction error. *Proceedings of the Royal Society B: Biological Sciences* 285, 20181645. doi:10.1098/rspb.2018.1645

Gershman, S. J. (2019). The generative adversarial brain. *Frontiers in Artificial Intelligence* 2. doi:10.3389/frai.2019.00018

Gershman, S. J. and Beck, J. M. (2017). *Complex Probabilistic Inference* (John Wiley and Sons Ltd). 453–466. doi:10.1002/9781119159193.ch33

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). *Generative Adversarial Nets* (Curran Associates, Inc.). 2672–2680

Gurney, K., Prescott, T. J., and Redgrave, P. (2001). A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological Cybernetics* 84, 401–410. doi:10.1007/PL00007984

Gurney, K. N., Humphries, M. D., and Redgrave, P. (2015). A new framework for cortico-striatal plasticity: Behavioural theory meets in vitro data at the reinforcement-action interface. *PLOS Biology* 13, e1002034. doi:10.1371/journal.pbio.1002034

Hinton, G. (2022). The forward-forward algorithm: Some preliminary investigations doi:10.48550/arXiv.2212.13345. ArXiv:2212.13345 [cs]

Hohwy, J., Roepstorff, A., and Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition* 108, 687–701. doi:10.1016/j.cognition.2008.05.010

Houk, J. C., Adams, J. L., and Barto, A. G. (1995). *A model of how the basal ganglia generate and use neural signals that predict reinforcement* (The MIT Press). Computational neuroscience. 249–270

Jordan, M. I. and Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science* 16, 307–354. doi:10.1207/s15516709cog1603_1

Keuken, M. C., Maanen, L. V., Bogacz, R., Schäfer, A., Neumann, J., Turner, R., et al. (2015). The subthalamic nucleus during decision-making with multiple alternatives. *Human Brain Mapping* 36, 4041–4052. doi:10.1002/hbm.22896

Lak, A., Nomoto, K., Keramati, M., Sakagami, M., and Kepecs, A. (2017). Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Current Biology* 27, 821–832. doi:10.1016/j.cub.2017.02.026

Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex during motor planning. *Nature* 532, 459–464. doi:10.1038/nature17643

Montague, P. R., Hyman, S. E., and Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature* 431, 760–767. doi:10.1038/nature03015

Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y., and Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. *Neuron* 41, 269–280. doi:10.1016/S0896-6273(03)00869-9

Noudoost, B. and Moore, T. (2011). Control of visual cortical signals by prefrontal dopamine. *Nature* 474, 372–375. doi:10.1038/nature09995

Pezzulo, G. and Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences* 20, 414–424. doi:10.1016/j.tics.2016.03.013

Pfau, D. and Vinyals, O. (2016). Connecting generative adversarial networks and actor-critic methods. *arXiv:1610.01945 [cs, stat]* ArXiv: 1610.01945

Pickering, M. J. and Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends in Cognitive Sciences* 18, 451–456. doi:10.1016/j.tics.2014.05.006

Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). enProbabilistic brains: knowns and unknowns. *Nature Neuroscience* 16, 1170–1178. doi:10.1038/nn.3495

Puig, M. V., Rose, J., Schmidt, R., and Freund, N. (2014). Dopamine modulation of learning and memory in the prefrontal cortex: insights from studies in primates, rodents, and birds. *Frontiers in Neural Circuits* 8. doi:10.3389/fncir.2014.00093

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434 [cs]* ArXiv: 1511.06434

Ribas-Fernandes, J., Solway, A., Diuk, C., McGuire, J., Barto, A., Niv, Y., et al. (2011). A neural signature of hierarchical reinforcement learning. *Neuron* 71, 370–379. doi:10.1016/j.neuron.2011.05.042

Ribas-Fernandes, J. J. F., Shahnazian, D., Holroyd, C. B., and Botvinick, M. M. (2018). Subgoal- and goal-related reward prediction errors in medial prefrontal cortex. *Journal of Cognitive Neuroscience* 31, 8–23. doi:10.1162/jocn_a_01341

Schnitzler, A., Salenius, S., Salmelin, R., Jousmäki, V., and Hari, R. (1997). Involvement of primary motor cortex in motor imagery: A neuromagnetic study. *NeuroImage* 6, 201–208. doi:10.1006/nimg.1997.0286

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)* 275, 1593–9

Schwartenbeck, P., FitzGerald, T. H. B., and Dolan, R. (2016). Neural signals encoding shifts in beliefs. *NeuroImage* 125, 578–586. doi:10.1016/j.neuroimage.2015.10.067

Smith, R., Friston, K. J., and Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology* 107, 102632. doi:10.1016/j.jmp.2021.102632

Solway, A. and Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychological Review* 119, 120–154. doi:10.1037/a0026435

Suarez, J. A., Howard, J. D., Schoenbaum, G., and Kahnt, T. (2019). Sensory prediction errors in the human midbrain signal identity violations independent of perceptual distance. *eLife* 8, e43962. doi:10.7554/eLife.43962

Takahashi, Y. K., Batchelor, H. M., Liu, B., Khanna, A., Morales, M., and Schoenbaum, G. (2017). Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron* 95, 1395–1405.e3. doi:10.1016/j.neuron.2017.08.025

Tani, J. (2016). *Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena* (Oxford University Press). Google-Books-ID: NtooDQAAQBAJ

Tian, J., Huang, R., Cohen, J., Osakada, F., Kobak, D., Machens, C., et al. (2016). Distributed and mixed information in monosynaptic inputs to dopamine neurons. *Neuron* 91, 1374–1389. doi:10.1016/j.neuron.2016.08.018

Watabe-Uchida, M., Zhu, L., Ogawa, S., Vamanrao, A., and Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74, 858–873. doi:10.1016/j.neuron.2012.03.017

Wei, W., Rubin, J. E., and Wang, X.-J. (2015). Role of the indirect pathway of the basal ganglia in perceptual decision making. *Journal of Neuroscience* 35, 4052–4064. doi:10.1523/JNEUROSCI.3611-14.2015

Yao, W.-D., Spealman, R. D., and Zhang, J. (2008). Dopaminergic signaling in dendritic spines. *Biochemical Pharmacology* 75, 2055–2069. doi:10.1016/j.bcp.2008.01.018

Yeh, R., Chen, C., Lim, T. Y., Hasegawa-Johnson, M., and Do, M. N. (2016). Semantic image inpainting with perceptual and contextual losses. *arXiv:1607.07539 [cs]* ArXiv: 1607.07539

FIGURE CAPTIONS

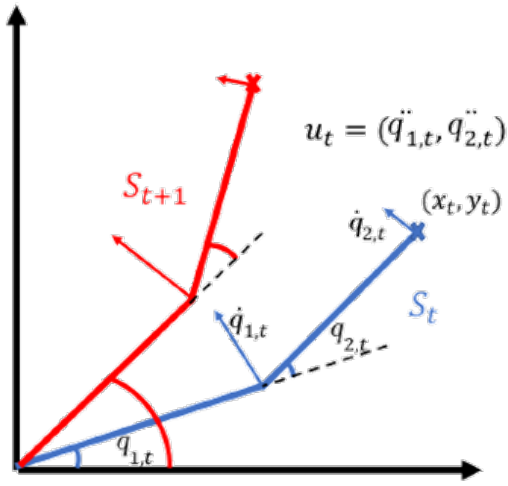

Figure 1a. The representative toy model: the two joint model

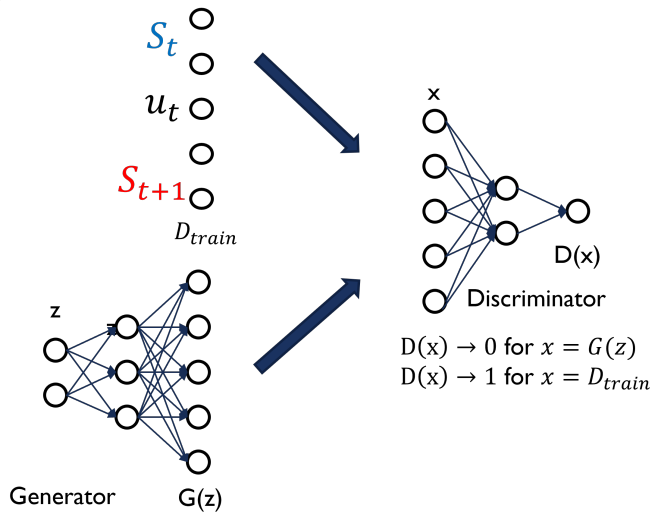

Figure 1b. Training GAN to generate the possible samples in our toy environment.

Figure 1. Overall architecture of our study Our GAN model learns the possible samples in our toy environment. The learnt information is used for action by inference

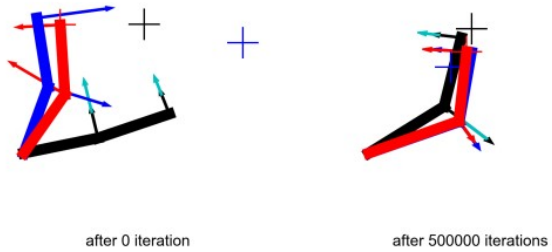

Figure 2a. Representative samples generated from (**left**) the untrained GAN and (**right**) the GAN after 500,000 training iterations. Both samples show the corresponding median error sample (in joint states) for the respective networks. **Black**: The current state generated from the generator. The black cross is the generated cursor position. **Blue**: The future state generated from the generator. **Red**: The actual future state from the generated current state and action, $S' \underset{given}{|} Q, \dot{Q}, U$.

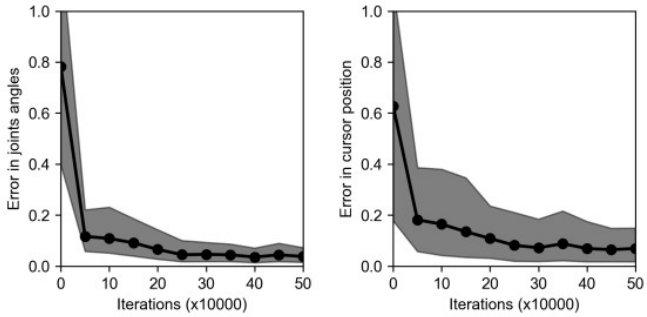

Figure 2b. The error between generated future states and actual future states for the generated current state and action. The error reduces by training The bold lines show the median error. The shade indicates the 5% and 95% quantile.

Figure 2. Training a Generative Model. GAN could internalize the dynamics of our toy model that, it could generate the samples, which are close to realistic samples after training

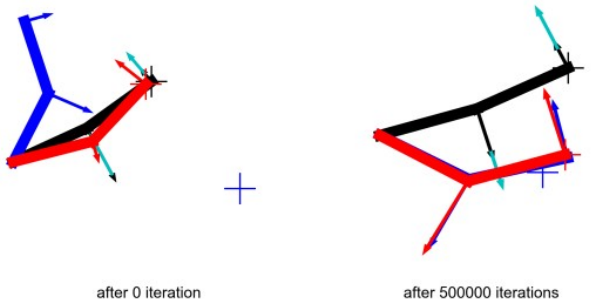

Figure 3a. Representative samples with a forward model using (**left**) the untrained generative model and (**right**) the model after 500,000 training iterations. The samples with the median error in joint states are shown. Black arm: The current state of the arm. Blue: The future state predicted by our forward model, S'^{g}. Blue arm: the future arm state predicted by our forward model from the current state and action, S'^{f} from the forward model. Red arm: the actual future state of the arm, $S' \underset{given}{|} S^{f}, U^{f}$.

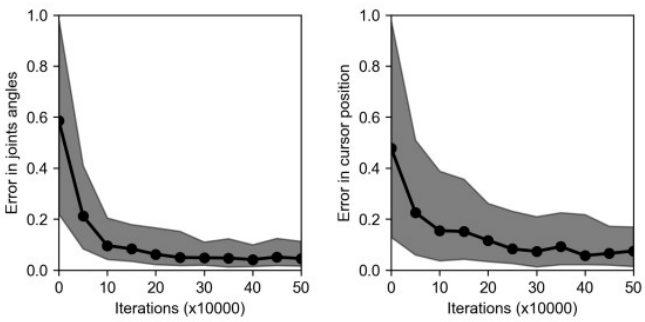

Figure 3b. Change of the distribution of errors by training iteration between the actual and forward model predicted future state. The bold lines show the median error. The shade indicates the 5% and 95% quantiles.

Figure 3. The forward model. Using the trained GAN network for the toy model samples, the future state of an arm could be inferred from the current state and action. The error between inferred and real future states were reduced by the training.

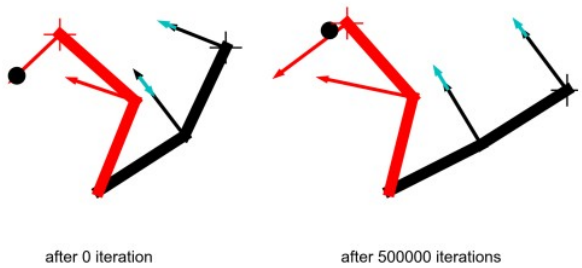

Figure 4a. Representative arm movement using our inverse model with a generative model after (**left**) 0 iteration and (**right**) 500,000 iterations. The samples with the median error between the target cursor position and actual cursor positions are shown. Black arm: The current state of the arm. Black dot: the target cursor position in the future state. Red arm: the future arm state when the arm moved with the action from our inverse model, $S'|S, U^i$.

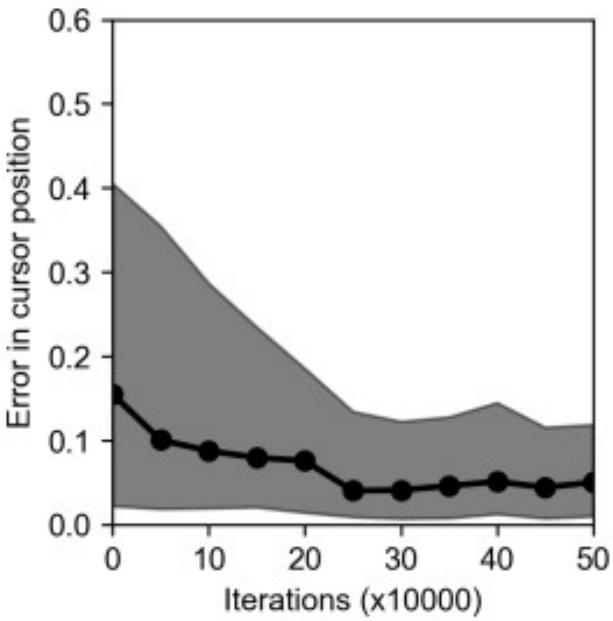

Figure 4b. The error between the target cursor position and the actual cursor position. The bold lines show the median error. The shade indicates the 5% and 95% quantiles.

Figure 4. The inverse model Using the trained GAN network for the toy model samples, the action could be inferred from the current state and target cursor state. The inferred action could move the cursor closer to the target as the iteration increase.