ActInf Livestream #051 ~ 'Canonical neural networks perform active inference'

A three-session series of discussions of the paper "Canonical neural networks perform active inference" by Takuya Isomura, Hideaki Shimazaki & Karl J. Friston.
https://www.nature.com/articles/s42003-021-02994-2

Presented by Active Inference Institute in 2022

Active Inference Institute information:
- Website: https://activeinference.org/
- Twitter: https://twitter.com/InferenceActive
- Discord: https://discord.gg/8VNKNp4jtx
- https://www.youtube.com/c/ActiveInference
- https://coda.io/@active-inference-institute/livestreams

LS #051.0: Background and context. https://www.youtube.com/watch?v=ZASG-rtkXDk
LS #051.1: First participatory group discussion. https://www.youtube.com/watch?v=IM_NlUzyq8M
LS #051.2: Second participatory group discussion. https://www.youtube.com/watch?v=hY_CajLpt9Q

Session 051.0, Oct. 26, 2022

https://www.youtube.com/watch?v=ZASG-rtkXDk

Background and context.

SPEAKER

Daniel Ari Friedman

CONTENTS

TRANSCRIPT

00:29 Daniel Friedman:

All right, welcome. It is ActInf Livestream Number 51 Dot 0. This is background and context, a first discussion, for “Canonical neural networks performance active inference” by Isomura et al.

It's October 26, 2022. Welcome to the active inference institute. We're a participatory online institute that is communicating, learning and plasticity applied active inference. You can find us at some of the links on the page. This is recorded in an archive livestream, so please provide us with feedback so we can improve our work.

All backgrounds and perspectives are welcome and will be following video etiquette for Solo Livestreams head to Active Inference.org to learn how to get involved with Active Inference Institute projects today. It's the first of several discussions that we'll have on the paper. Canonical neural networks perform active inference. 2022 by Tequia Isomura, Hijacki Shimazaki and Karl Friston The video is just an introduction to some of the ideas. It's not a review or final word.

01:43 There will be an overview of the structure of the paper and then we'll go through many of the set point. Also just to disclaim, there's many, many other and better resources to learn about neural networks. So I would very much welcome those with a technical understanding of neural networks and or some of the more applied computational or otherwise aspects of neural networks. It would be awesome to have them on for the dot one and the dot two, because it was not an area I was familiar with and so hope that I can hear more from the authors in our coming weeks and others. I'm Daniel, I'm a researcher in California, and this will just be a solo zero, which I guess hasn't happened in a while in the making of this.

Here are some of the generated art prompts area 51 Active Inference Brea. 51 neural network. Area 51 active inference. Ant Brea. 51 Active Inference Neural Network just some interesting images coming out of stable diffusion.

02:56 So as to some big questions that the paper is addressing and that one might be interested in to come to the paper. How can artificial neural networks be understood as generic optimization processes and what is the correspondence between hemodynamics and modern statistical inference methods? Other big questions are about the history and next steps of the enmeshment of natural intelligence, eg. Neuroscience and artificial intelligence, as well as, of course, whether to even play into this kind of distinction at all and have different integrated intelligence frameworks. And one paper where any of the authors or anyone who has kind of resonated with this work is recent by Zidore at all 2022 towards the Next Generation artificial Intelligence catalyzing the NeuroAI Revolution and so this is a bunch of authors.

And so it's interesting just to quote in terms of what some areas of discourse are saying right now, which is neuroscience has long been an important driver of progress in artificial intelligence AI. We propose that to accelerate progress in AI, we must invest in fundamental research in NeuroAI. So that's one way to lead some of the developments that are happening in the paper we'll discuss what does it mean to be particular but generic? That's a phrase used in the paper, so maybe that's kind of a jumping off point. And then how can active inference help us understand the past, present and future here of the interface with neural networks, statistics and neural AI?

04:38 Here's the abstract. This work considers a class of canonical neural networks comprising rate coding models where neural activity and plasticity minimize a common cost function and plasticity is modulated with a certain delay. We show that such neural networks implicitly perform active inference and learning to minimize the risk associated with future outcomes. Mathematical analyses demonstrate that this biological optimization can be cast as maximization of model evidence or equivalently minimization of variational free energy under the wellknown form of a partially observed Markov decision process model. This equivalence indicates that the delayed modularity of heavy and plasticity accompanied with adaptation of firing thresholds is a sufficient neuronal substrate to attain Bayes optimal inference and control.

We corroborated this proposition using numerical analyses of maze tasks. This theory offers a universal characterization of canonical neural networks in terms of Bayesian belief updating and provides insight into the neuronal mechanisms underlying planning and adaptive behavioral control.

05:53 Here is the roadmap I'll be driving. On the right side of the road, there's an introduction and a table with glossary of different expressions used. Then there's the Results sections, which has first an overview of equivalence between neural networks and variational base active inference formulated using a postdiction of past decisions. Canonical neural networks perform active inference and numerical simulations. So they work on the interface between neural networks and variational Bayesian methods and start with a more theoretical and mathematical background and then eventually present some maze simulations that are in MATLAB.

Then there's a discussion, and then the methods section is following the discussion and it has the subsections generative model variational free energy inference and learning, neural networks, simulations and data analysis. And I think that I have kept to the right color codes consistently from here on forward. The black quotes are the direct quotes from the paper Bleu quotes quotes are related to perception, orange for action, complete class theorems and neural networks together in purple. And then red is commentary and then the highlighting for red is related to the variational methods, but it's kind of redish.

07:29 Okay, the papers canonical neural networks perform active inference by the authors listed previously. It's in communications biology from 2022. And the key aims of the paper are laid out well in the first paragraph. So that's in black and other colors. And here the red is just kind of settling point initial conversation with the paper and connecting a little bit more to some previous streams before we go more into the specifics of the paper.

The sentient behavior of biological organisms is characterized by optimization biological organisms recognize the state of their environment by optimizing internal representations of the external environmental dynamics generating sensory inputs. So that's the sensory recognition model. In addition, they optimize their behavior for adaptation to the environment thereby increasing their probability of survival and reproduction. So that is bringing the entire active and inactive control theory. Formal theories of action.

08:35 Action selections. Planning and planning as inference all of these actionoriented pragmatic term type ideas come into play when this actionorange aspect is brought in and it is. Must be. Should be. Etc oriented towards survival and persistence otherwise we don't see that kind of thing for long over appropriate definitions for thing and long.

Et cetera. And this few sentences that the paper begins with summarizes a common theme in actinF which is that there's basically two pathways towards proficiency in the niche. There's changing the mind, which is the perceptual and learning and then there's changing the world through action. And that kind of integration of

inference, cognitive inference, whether it's perceptual or learning or other and action as enacted is part of the active formula and shared with other areas. This biological selforganization is typically formulated as the minimization of cost functions where in a gradient descent on a cost function furnishes neurodynamics and synaptic plasticity.

09:49 So they are working with a framework where the way that this perception action flow is tractable computationally either just for our current computers so that we can implement simulations and data fitting or whether more fundamentally like this is the computational context of action they're suggesting there's a way to think of it in terms of a cost function minimization and also note like minimization maximization. In a way they're interchangeable because there's just sometimes negative signs that can flip them. So it's the same energy and fitness landscape that is being navigated whether you're minimizing to the bottom of the bowl and thinking of action selection that way and inference or whether you're climbing to the top of the hill gradient descent. Gradient ascent they're both kind of two sides of the same .2 fundamental issues remain to be established. Namely so this is what the paper is saying they're filling the gap in literature the characterization of the dynamics of an arbitrary neural network as a generic optimization process and the correspondence between such neural dynamics and statistical inference found in applied mathematics and machine learning.

11:05 The present work addresses these issues by demonstrating that a class of canonical neural networks of rate coding models is functioning as and thus universally characterized in terms of variational bayesian inference under a particular but generic form of the generative model.

This is maybe the only slide that has resources from too far a field from paper. Well, almost a few more. It's just on neural networks. The search on a very common search tool resulted in five plus million results for what are neural networks? And the first one was three Bleu one brown in 2017 which is just a nice YouTube video as they very often make with these beautiful renditions, and it's a very great video.

And there's many, many others. A neural network is a network or circuit of biological neurons, and that is variously purely computational and or biological map territory stuff. And they can be thought of as nodes and edges, which are the firing rates and the connections between the biological neurons being modeled as weights between nodes. A positive weight reflects an excitatory connection, while negative values mean inhibitory connections. And here is data coming in and ending up activating different neurons in this final layer.

12:40 So this is kind of doing inference in a neural network, and then there's an action part. Okay, the paper says variational Bayesian inference offers a unified explanation for inference learning, prediction, decision making and the evolution of biological form, which can be considered over multiple timescales. So this returns to the earlier theme of like exteroception action and persistence within and among generations. And the citations that are provided here for this are Friston, Kilner Harrison 2006 and Friston 2010, both very classic acne FEP papers, just to show one figure from each here's from the first citation with a winged snowflake, where if the snowflake ends up being somewhere that's too warm, that's not able for it to maintain the structure, then it melts, melts into a dew, etc. And it must be acting as if in one way or another, it's staying above that phase boundary.

13:53 That's how it's statistically identified, it's how it's autogenetically identified, functionally identified. And then here is one of the kind of path setting figures in the first and 2010 paper with a tree of many different areas

to connect to, informalisms to represent as including probabilistic, neural coding, Bayesian brain, optimal control and value learning. These three and others can be really seen at play in this paper.

The kind of inference variational basing methods use rests upon a generative model that expresses a hypothesis about the generation of sensory input. Perception and behavior can then be read as optimizing the evidence for a generative model inherent in sensory exchanges for the environment, and that's the integration of perception and action within a single imperative in terms of if only computation. And that is described more on this slide, which one can pause and read, but the rest of the quotes are from the paper.

15:20 One eformalism that I had no familiarity with before this paper and associated readings and conversations with Ali was the Complete Class Theorem. So it would be great for anyone who's familiar with Complete Class Theorem to help a little bit here, but here's some interesting things that I came across crucially from the paper as a corollary of the Complete Class Theorem citations 19 through 21. Any neural network minimizing a class function can be viewed as performing variational Bayesian inference under some prior beliefs. Here are the citations 19 through 21 from 1940 719 81 2013, and also found some interesting resources. So quoting from this less wrong post linked here, dutch book defines belief as willingness to bet and Money Pump defines preference as willingness to pay.

16:28 This is in terms of the foundational arguments for a Bayesian Epistemological worldview. Hope that's correct. Again, it would be good to hear from anyone who knows more here, but in terms of different ways that one can consider the Bayesian proposition, which perhaps these are even better to use than referencing any person's last name. Because the interesting thing will be the different lenses that these different framings on Bayesian probabilities are interpreted as. Just like what a Pvalue is interpreted as in the frequentist worldlessly, like a beta factor interpretation.

So Dutch book defines belief as willingness to bet and Money pump defines preferences willingness to pay. In doing so, both arguments put the justification of decision theory into hypothetical exploitation scenarios which are not quite the same as the actual decisions we face. If these were the best justifications for consequentialism we could muster, I would be somewhat dissatisfied, but would likely leave it alone. Fortunately, a better alternative exists. The complete Class theorem.

17:46 So here's an image from that post possible world states observation likelihood maps to the possible observations hashtag a matrix, then decisionmaking rule may be probabilistic action selection as inference a possible actions affordances. And here we commonly see the loop being closed from actions to world states like through the B matrix changing how the world changes through time. And here these are both going to be mapped to a loss function l a real valued score from taking an action A when the world turned out to be theta realized theta. So that's quite an interesting framing and there were some other useful posts online. And from this Zhang blog the argument boils down to if you agree with expected utility as your objective, then you have to be Bayesian.

18:50 In a nutshell, a strategy is inadmissible if there exists another state that is as good in all situations and strictly better in at least one. If you want your strategy to be admissible, it should be equivalent to a Bayes estimator Complete Class Theorems. Only Bayes strategies are admissible and admissible strategies are Bayes. So there's probably a lot of interpretations of this, but it seems kind of related to optimal control or Bayes

optimal inference, perhaps a little bit more keenly. So these are some next two slides were contributed by Ali, so if you or if anyone familiar in this area wants to come on and the discussion would be great, but Ali pointed me to this book fundamentals of Bayesian Epistemology by title Bomb 2022 table of contents shown here, and some challenges and objections to Bayesian Epistemology are listed which can be read.

19:57 And then also there's some quite interesting logical structures which can be read here in terms of their premise, theorem and conclusion. In the areas of arguments for Bayesianism being representation theorem, argument for probabilism, the dutch book argument for probabilism and the Gradational accuracy argument for probabilism. So pretty interesting. Back to the paper they wrote we have previously introduced a reverse engineering approach that identifies a class of biologically plausible cost functions for neural networks. Citation 22 previous Paper of Isomura Karl Friston 2020 the paper was reverse engineering neural networks to characterize their cost functions, the abstracts shown here.

But this letter considers a class of biologically plausible cost functions for neural networks using generative models based on partially observable Markov decision processes. We show that neural activity and plasticity perform Bayesian inference and learning respectively by maximizing model evidence. So this is a precursor paper from 2020 that is cited and foundational for the 2022 paper. Here are some figures from that paper. They have comparisons among Markov decision process scheme and a neural network.

21:33 For example, a Markov decision process scheme expressed as a formy factor graph based on the formulation in Friston. So here is the Markov decision process as a formy factor graph and on the right side is the neural network with the hidden state sensory inputs and neural activity and some figures and concordances.

And they wrote in this context, the neural network can in principle be used as a dynamic causal model to estimate threshold, constants and implicit priors. This reverse engineering speaks to estimating the priors used by real neuronal systems under ideal Bayesian assumptions, sometimes referred to as metabasian inference. So they're laying out their research agenda and much of the work is going to reference this earlier paper and other work. And then there's also some new integrations and especially the maze simulation in this paper, and probably more so let's find out.

22:51 Referring to the earlier paper, this foundational work identified a class of cost functions for single layer field feed forward neural networks of state coding models with a sigmoid or logistic activation function. We subsequently demonstrated that mathematical equivalence between the class of cost functions for such neural networks and variational free energy under a particular form of the generative model, which is similar broadly to what the 22 paper dot. Two, this equivalent licenses variational Bayesian inference as a fundamental optimization process that underlies both the dynamics and function of such neural networks.

However, it remains to be established whether the active inference is an apt explanation for any given neural network that actively exchanges with its environment. In this paper, we address this inactive or control aspect to complete the formal equivalent of neural network optimization and the free energy principle. So this earlier work was not including action. And so this paper's key addition, as I understand it, hopefully, is that they bring action more formally into the model. And it would be interesting to know like what else is that change associated with?

24:17 Or what else happens or doesn't?

Here's some more text about the approach that they're going to take, and many of these details are going to be addressed later on.

Here's table one glossary of expressions, so these will also be broadly addressed later on.

Maybe it's useful though just to read the first one a canonical neural network in this work, the canonical neural network is defined by differential equations of neural activity derived as reduction of realistic neuron models through some approximations which give a network of rate coding neurons with a sigmoid activation function. In particular, we consider networks comprising a middle layer that involves recurrent connections and the output layer that provides feedback responses to the environment. So some category of neural network architecture or anatomy, physiology, whatever with sensory, cognitive and actionlike features in an environment or a generative process. So the results section, they write an overview of equivalence between neural networks and variational base. First we summarize the formal correspondence between neural networks and variational base.

25:56 A biological agent is formulated here as an autonomous system comprising a network of rate coding neurons. Figure one A so here's a small figure one A. We'll see it larger based upon the assumptions which will be brought up later on in a different fuller form. The update rule for the ith component of phi, which is the internal states so the cognitive states and the output layer y. So middle layer x and output layer y, these can be seen as the cognitive and the actinF selective aspects which are the autonomous states in contrast to the particular state of the active inference entity which is the whole blanket, and the cognitive states so including the sensory state.

But the sensory states can't be directly controlled, however action can influence them. And so that's what closes the causal loop and that set of the particular states comprising the autonomous states the update rule for the ith component of phi is derived as the gradient descent on the cost function. So the change in that by sub I is proportional to a derivative on the loss function. This determines the dynamics of neural networks including their activity and plasticity. So.

27:24 L common loss function O observations sensory states by internal state comprising the middle and output layer neural activity that's the rate coding part and synaptic strength w the parameterization and other free parameters including that that characterize L. Then there's the output layer activity y is determining the network's actions or decisions delta o x and y y is outputting action environment generative process hidden state passing observations to the sensory cognitive active layers the neural network and that is analogous topologically to the variational Bayesian formulation in figure one B. And so left side gradient descent on a neural network cost function L determines the dynamics of neural activity and plasticity. Thus L is sufficient to characterize the neural network and then being shown next to on the right side variational free energy minimization allows an agent to self organize, to encode the hidden states of the external milieu and to make decisions minimizing future risk. Here, variational free energy F is sufficient to characterize active inference and behaviors of the agent.

28:47 So sentence structure parallelism, and these two schemes or approaches being Linsker and applied is a focus of this area.

So neural networks are minimizing the cost function L. In contrast, variational Bayesian inference depicts a process of updating the prior distribution of external states P of theta, so the corresponding posterior distribution Q of theta based upon the sequence of observations. And we see other familiar terms from discussions on variational DAGs like minimization of surprise. And there's some more model details shown here. A few points

are about choosing the family of distributions that one is doing variational inference with, and that allows for the gradient update rules in that family of distributions to be maze simpler.

29:54 For example, choosing a linear regression with an L two norm. At the very least, you can say that it has a simple decision rule for being fit. I'm sure there's like also counter arguments and other algorithms that are better and so on. But just to say that a simple decision rule in a known family of distributions can often be good enough as an approximation, if not more crucially. According to the Complete Class theorem from earlier, a dynamical system that minimizes its cost function can be viewed as performing Bayesian inference under some generative model and prior beliefs.

The Complete Class theorem goes on to describe it ensures the presence of a generative model that formally corresponds to the above defined neural network characterized by L. Hence, this speaks to the equivalence between the class of neural network cost functions and variational free energy under such a generative model equation one loss function on the time series of observations comma parameters. Three ines is defined as f the free energy function on the same O through time comma parameters. So there's a parallel being shown defined wherein the internal states of a network by encode or parameterize the posterior expectation theta. This mathematical equivalence means that an arbitrary neural network in the class under consideration is implicitly performing active inference through variational free energy minimization and more is written.

31:41 But this is one of many statements that will be made corresponding to correspondences between neural network loss function, generative model, active free energy minimization, and associated formalisms.

Note that being able to characterize the neural network in terms of maximizing model evidence lends it in explainability in the sense that internal neural network states and parameters encode Bayesian beliefs or expectations about the causes of observations. In other words, the generative model explains how outcomes were generated. However, the Complete Class theorem does not specify the form of a generative model for any given neural network. To address this issue in the remainder of the paper, we formulate active inference using a particular form of partially observable Markov decision processes PMDP models whose states take binary values. So this is one of several simplifications is one way to see it or areas for later generalization.

32:52 Figure Two in this section, we define a generative model and ensuing variational free energy that corresponds to a class of canonical neural networks that will be considered in the subsequent section. The internal model is expressed as a discrete state space in the form of a POMDP figure two. So to make figure two larger, there's a lot to probably say about this bigger. I'll just note that the caption is informative and some of the highlighted lines. It'll be awesome to have the author and other people who's familiar with some of the differences and symmetries between the, for example, bottom and top, above and below the observations, and also about the role of time and what these two risks are.

33:56 What is fictive causality?

Here's more details on that POMDP then equation two. This is in this section active inference formulated using a postdiction of past decisions. So if it was a prediction of future decisions, it's the opposite, a postdiction of past decisions. Hence, we define the generative model as follows PO delta s gamma theta, which is the model of observations decisions, observations decisions, hidden states risk and theta theta equals A, b and C constitutes a

set of parameters and more details are provided. This is the familiar notation with some slight differences that will be described.

34:56 And equation two reflects the factorizability and the dimensions and the kind of computational tractability expression. So that would be interesting to also learn about like under what conditions can this joint distribution be factorized? Under what simplifications or constructions the agent is making decisions to minimize a risk function capital gamma that's on the bottom of figure two, victim causality. Leading to this gamma. Coming from that gamma, equation three is shown.

And to characterize the optimal decisions as minimizing expected risk in our POMDP model, we use effective mapping from the current risk gamma to past decisions that's that retro addictive fictive causality. Although this is not the true causality in the real generative process that generates sensory data. Here we intend to model the manner that an agent subjectively evaluates its previous decisions after experiencing their consequences. This fictive causality is expressed in the form of a canonical distribution. So it could be other families hypothetically.

36:21 But this equation three is describing fictive causality and this interesting sign indicates the element noise division operator. Also note throughout the manuscript the overline variable indicates one minus that variable. So gamma bar equals one minus gamma. Or it can be understood as the complement of a statistical probability probability of a happening and the probability of not a happening in that one way.

Importantly, the agent needs to keep selecting good decisions while avoiding bad decisions. To this end, equation three supposes that the agent learns from the failure of decisions by assuming that the bad decisions were sampled from the opposite of the optimal policy. Mapping some more details and then by convention, active inference uses C to denote the prior preference. That's how we've seen the C variable in many models as preferences. Prior Preference this work uses C to denote a mapping to determine a decision depending on the previous state.

37:37 Herein the prior preference is implicit in the risk function due to construction. C does not explicitly appear in the inference. Thus it is omitted in the following formulations. So that's a key point about the notation of the variable C and something kind of maybe interesting to explore. Equation Four variational Bayesian inference refers to the process that optimizes the posterior belief q of theta based on the mean field approximation q of theta is expressed as and here is the factorization representation of the q of theta variational and another notation.

Note throughout the manuscript, bold case variables such as bold s subtow denote the posterior expectations of the corresponding italic case random variables and some more details about the model. For example, for simplicity, here we suppose that state and decision prior D and E are fixed, so one could imagine they don't have to be, but for simplicity they will be. Here, under the abovedefined generative model and posterior beliefs, the ensuing variational free energy is analytically expressed as follows equation Five this is a variational free energy F and recall equation one that it is being connected to, juxtapose with, etc. The loss function the gradient descent on variational free energy updates the posterior beliefs about hidden states s decisions, delta and parameters theta. The optimal posterior beliefs that minimize variational free energy are obtained as the fixed point of the implicit gradient descent, which ensures that change in F with respect to the change in hidden state through time equals zero.

39:45 And some more definitions. All of them are zero, that one might be an O, but they're all zero. And this is saying the ball rolls to the bottom of the hill in this gradient descent, and when the rate of change is zero, then that is a fixed point attractor like dynamic, and that can be used as a way to incrementally fit statistical models like the loss function is used to incrementally fit neural networks. To explicitly demonstrate the formal correspondence with the cost function for neural networks considered in the next section, we further transformed the variational free energy as follows some details Using these relationships, equation five is transformed into the form shown in Figure three. See the Methods section variational Free Energy for further details.

In what follows, we demonstrate that the internal states of canonical neural networks encode posterior beliefs. Here's figure three on the top from the caption figure three is the mathematical equivalence between variational free energy and neural network cost functions depicted by one to one correspondence of their components. Top variational free energy transformed from equation five using the Bayes theorem and foreshadowing. Equation seven. Using this relationship, equation seven is transformed into the form presented at the bottom of the figure.

41:20 So here's f variational free energy and L the neural network cost function and different elements as they're represented here with some resonances and concordance and beyond which we can explore, they're being shown to be equivalent. And it was in the conversation preparing for this dot zero with Dean, where we saw this as some chromosomes.

In this section, canonical Neural Networks Perform Active Inference in this section we identify the neuronal substrates that correspond to components of the active inference scheme defined above. We consider a class of two layered neural networks with reafferent connections in the middle layer. Figure one a those are those connections with loops recurrent in the middle layer. The modeling of the networks in this section, referred to as canonical neural networks, is based on the following three assumptions that reflect physiological knowledge. One gradient descent on a cost function l, determines the updates of neural activity and synaptic weights.

42:34 Method Section neural Networks Neural Two assumption two neural activity is updated by the weighted sum of inputs and its fixed point is expressed in a form of a sigmoid or logistic function. And assumption three, a modulatory factor mediate synaptic plasticity in a post hoc manner.

They write based upon assumption two, which is neural activity is updated by the weighted sum of inputs and its fixed point is expressed in the form of a sigmoid or logistic function. Based on assumption two, we formulate neural activity in the middle layer and output layer the autonomous states as follows equation six without loss of generality, equation six can be cast as the gradient descent on cost function l. Such a cost function can be identified by simply integrating the righthand side of equation six with respect to x and y, consistent with previous treatments. Citations because neural activity and synaptic plasticity minimize the same cost function l, the derivatives of l must generate the modulated synaptic plasticity under these constraints, reflecting assumptions one through three, a class of cost functions is identified as follows equation seven loss function. Awesome to hear somebody read this directly.

44:10 So there's a firing rate aspect that's related to neural firing in the short term more perceptual aspect of function. And there's a slower neurotransmitter neuroglial factormediated learning over perhaps a different time scale and with some different features, and they're being included as a joint model of inference. And here that is being connected to doing inference on action. Synaptic Plasticity in Neural Networks so synaptic plasticity rules conjugate to the above rate coding model can now be expressed as a gradient descent on the same cost function l,

according to assumption one, equations eight and nine showing that neural networks can integrate those modes of firing rate like and synaptic modulatory like. Neural networks.

45:13 Cost Functions and Variational Free Energy Based on the above considerations, we now establish the formal correspondence between the neural network cost function and variational free energy. Using under the aforementioned three minimal assumptions, we identified the neural network cost function as equation seven. Equation seven can be transformed. Hinton the form shown in figure three bottom using sigmoid functions of synaptic strengths. So equation seven loss function transformed into the form shown in figure three bottom.

Hence, this class of cost functions for canonical neural networks is formally homologous to variational free energy under the particular form of the POMDP generative model defined in the previous section. We obtain this result based on analytic derivations without reference to the complete class theorem, thereby confirming the proposition of equation one loss function and free energy. This in turn suggests that any canonical neural network in this class is implicitly performing active inference. Table two summarizes the correspondence between the quantities of the neural network and their homologues in variational Bayes. So two, a great concordance table.

46:51 On the left side, neural network formation. On the right side, variational Bayes formulation. So just like in figure three, we had the two long lines. Then table two, we have rotated 90 degrees and the variables for different parts of these models are laid next to each other.

So what papers of neural networks can we make active models for? Is it already done or is there just one script that needs to be done? And then conversely, what interesting, variational Bayesian models have interesting applications or some other value or information gain from being cast as neural networks or integrating neural networks into what had previously been only analytical variational Bayesian models as well as the empirical data fitting aspect which is related that's highlighted by the authors. So, in summary, when a neural network minimizes the cost function with respect to its activity and plasticity the network selforganizes to furnish responses that minimize the risk implicit in the cost function, this biological optimization is identical to variational free energy minimization under a particular form of the POMDP model. Hence, this equivalence indicates that minimizing the expected risk through variational free energy minimization is an inherent property of canonical neural networks featuring delayed modulation of hebion plasticity.

48:49 Okay, brief seconds to view some image memes and take a breath in a stretch.

On the left side, the top panel says Epineural network is implicitly performing active inference. That's awesome. In the middle image meme, one simply makes a neural network and it is implicitly performing active inference. Also question mark. And then in the right image meme, there are two pieces of paper neural network and implicit performance of active inference.

Corporate needs you to find the differences between this picture and this picture. And here is the paper. They're the same picture. So the previous sections have successfully extended the variational Bayes meets inference. Neural network results from the 2020 paper with an action loop and a loss function only on the autonomous states.

50:11 So that was the conceptual and technical feature that this paper brought in. Again, it'll be awesome to hear the author describe it more and differently. And then in this part of the paper, they turn from that kind of analytical theoretical towards some numerical simulations in MATLAB. Here we demonstrate the performance of canonical neural networks using maze tasks. As an example of a delayed reward task, the agent comprised the aforementioned canonical neural networks.

Figure four a thus it implicitly performs active inference by minimizing variational free energy. So now some entity or agent is going to be constructed and numerically simulated to support some of the aspects of their model and point towards utility in other settings. So here's figure four simulation of neural networks solving maze tasks. In part A, there is the architecture of the agent sensory input layer O of T synaptic weights into the middle layer x internal states and the output layer y decision action with risk gamma of T.

51:57 The sensation comes in a neural network then outputs a decision. There's some task that the entity must perform which is to move towards the goal from the start across this lateral maze. So one could imagine that if it were just a hallway with no maze features, simple strategies would be very overfit but effective like always move simply towards the goal. Whereas in more complex settings, which this is an example of where there's uncertainty as well as many local optima. So one has to sometimes take one or two or three or four or more steps or however many to get closer to the goal and sometimes may not know, for example how long those backtracking situations are going to be and all these other complexities for which this maze example is symbolic of.

53:18 So maybe there's some interesting mythos maze connections and computational and here's some performance measures on the maze task with the neural network entity. This way, before training, the agent moved to a random direction, each step resulting in a failure to reach the goal position right end within the time limit. During training, the neural network updated synaptic strengths depending on its neural activity and ensuing outcomes I e. Risk. This training comprised a cycle of action and learning phases.

This treatment renders neural activity and adaptive behaviors of the agent highly explainable and manipulatable in terms of the appropriate prior beliefs implicit in firing thresholds for a given task or environment. In other words, these results suggest that firing thresholds are the neuronal substrates that encode state and decision priors as predicted. Mathematically big if true.

54:24 Furthermore, when the updating of parameters is slow across these two now linked domains parameters of the variational Bayesian autonomous state inference and the neural network loss function parameters when the updating of these parameters are slow in relation to the experimental observations, the parameters can be estimated through Bayesian inference based on empirically observed neuronal response curve method. Section data Analysis for details using this approach, we estimated implicit prior E, which is encoded by PSI from sequences of neural activity generated from the synthetic neural networks used in the simulations reported in figure four. We confirmed that this estimator was a good approximation to the true e. So that's also pretty interesting. This is showing they don't just lay out this architecture and show that it's possible to fulfill this maze task with the best whatever.

55:37 It's not a classification accuracy imperative alone. They're describing that from empirically observed neuronal responses, which is to say the experimenters observation the sequences of neural activity generated

from the synthetic neural networks used in that figure numerically. So statistically, that estimator was a good approximation to the true e. Figure five a so here's figure five. Estimation of implicit priors enables the prediction of subsequent learning.

So that's pretty interesting and will be great to hear what each of the axes are and what all the variables mean.

With this setup in place, they did some numerical validation and talked a little bit more about fitting data from this simulation model.

56:51 Here's the discussion section. So the first paragraph of the discussion biological organisms formulate plans to minimize future risks. In this work, we captured this characteristic in biologically plausible terms under minimal assumptions. We Deneve simple differential equations that can be plausibly interpreted in terms of a neural network architecture that entails degrees of freedom with respect to certain free parameters, e. G firing threshold.

These three parameters play the role of prior beliefs in variational Bayesian formulation. Thus, the accuracies of inferences and decisions depend upon prior beliefs implicit in neural networks. And here's some more stable diffusion ant neural network ant gain neural network so some more summarization of what they did based on the view of the brain as an agent that performs Bayesian inference. Internal representation of Bayesian belief updating have been proposed which enables neural networks to store and recall spiking sequences eight learn temporal dynamics and causal hierarchy nine. Extract hidden causes ten, solve maze tasks eleven, and make plans to control robots twelve.

58:11 So citations eight through twelve listed here. In these approaches, the update rules are generally derived from Bayesian cost functions e. G variational free energy. However, the precise relationship between these update rules and the neural activity and plasticity of canonical neural networks has yet to be fully established.

We identified a onetoone correspondence between neural network architecture and a specific POMDP implicit in that network. Equation two speaks to a unique POMDP model consistent with the neural network architecture defined in equation six, where their correspondences are summarized in table two and the figures. This means that our scheme can be used to identify the form of the POMDP given an observable circuit structure. Moreover, the free parameters zhat parameterize equation six can be estimated using equation 24. This means that the generative model and ensuring variational free energy can in principle be reconstructed from empirical data.

59:24 This offers a formal characterization of implicit Bayesian models entailed by neural circuits, thereby enabling a prediction of subsequent learning. So what can be done with this? What is this new? What is new here? Does this second selection fully establish the precise relationship between these update rules and the neural activity and plasticity of canonical neural networks.

Here is a discussion section on hebion plasticity, neurotransmitters and glia with a lot of citations listed. And here is just one interesting followon discussion from the computational aspects. Neurocognitive and computational aspects of heavy and plasticity is these modulations have been observed empirically with various neuromodulators and neurotransmitters such as Dopamine, Noradrenaline, Lescreen, and GABA, as well as glial factors. So here's a picture by Alexandra Michaelova Cultured astrocytes release, signaling and growth factors that regulate proper neuronal development so Cool, Glial, Pick, Dopamine and reinforcement learning. In

particular, a delayed modulation of synaptic plasticity is well known with Dopamine neurons citations 35 through 37.

1:01:07 Those citations are listed here. This speaks to a learning scheme that is conceptually distinct from standard reinforcement learning algorithms, such as the temporal difference learning with actorcritic model based on state action value objective function. Please see the previous work citation 50 for a detailed comparison between active inference and reinforcement learning that is state ball par Karl Friston active inference demystified and compared from 2021 and that's also active model stream number two. One, we mathematically demonstrated that such plasticity enhances the association between the pre post mapping and the future value of the modulatory factor, where the latter is cast as a risk function. This means that postsynaptic neurons selforganized to react in a manner that minimizes future risk.

So the neural network had three layers. It's the second and the third layer, not the initial sensory layer, but the second and the third layer, the cognitive and the active states, which are the autonomous states of the particular state. So that's quite interesting the self organization of synaptic neurons to react in a manner that minimizes future risk. Crucially, this computation corresponds formally to variational Bayesian inference under a particular form of Pom DP generative model, suggesting that the delayed modulation of Hebbian plasticity is a realization of active inference and regionally specific projections of neuromodulators may allow each brain region to optimize activity to minimize risk and leverage a hierarchical generative model implicit in cortical and subcortical hierarchies. This is reminiscent of theories of neuromodulator and meta learning developed previously.

1:03:16 Citation 52 Doya 2002 metal learning and neuromodulator cool. Our work may be potentially useful when casting these theories in terms of generative model and variational free energy minimization.

They then return the discussion to the complete class theorem and neural networks. So the Complete Class Theorem same citations from before ensures that any neural network whose activity and plasticity minimize the same cost function can be cast as performing Bayesian inference. However, identifying the implicit generative model that underwrites any canonical neural network is a more delicate problem, because the theorem does not specify a form of the generative model for a given canonical neural network. Which is pretty interesting. Is that to say that the form of the generative model as modeled is different in what ways?

1:04:22 From the given canonical neural network, the posterior beliefs are largely shaped by prior beliefs, making it challenging to identify the generative model by simply observing systemic dynamics. To this end, it is necessary to commit to a particular form of the generative model and elucidate how posterior beliefs are encoded or parameterized by the neural network states. This work addresses these issues by establishing a reverse engineering approach to identify a generative model implicit in a canonical neural network, thereby establishing onetoone correspondences between their components. Remarkably, a network of rate coding models with sigmoid activation function formally corresponds to a class of POMDP models, which provide an analytically trackable example of the present equivalents. Please refer to the previous paper citation 22 for further discussion.

So some of the analytical details, especially on the inferential side cognitive side burr captured in the earlier paper. This paper goes further into mapping the potentially necessary and sufficient aspects of the particular entity, which is to say the risk minimizing features of the autonomous states with respect to the entire particular states, including sensory states. Connecting that back of the envelope verbally expressible formulation to some

complete class of neural networks. So what's outside of the complete class, and why? And then what groups within the class have special or different features?

1:06:33 It is remarkable that the proposed equivalence can be leveraged to identify a generative model zhat an arbitrary neural network implicitly employs this contrast with naive neural network models that address only the dynamics of neural activity and plasticity. If the generative model differs from the true generative process that generates the sensory input, inferences and decisions are biased, ie. Suboptimal relative to base optimal inferences and decisions based upon the right sort of beliefs. Prior beliefs in general, the implicit priors may or may not be equal to the true priors. Thus, a generic neural network is typically suboptimal.

Nevertheless, these implicit priors can be optimized by updating free parameters e. G threshold factors, phi PSI based on the gradient descent on cost function l. By updating the free parameters, the network will eventually in principle become Bayes optimal for any given task. In essence, when the cost function is minimized with respect to neural activity, synaptic strengths, and any other constants that characterize the cost function, the cost function becomes equivalent to variational free energy with the optimal prior beliefs.

1:08:00 So the cost function for the neural network activity and synaptic strengths underlying the loss function are equivalent to the kind of gradient descent enabled variational free energy minimization under the Bayes optimality scenario from a risk perspective. Simultaneously, the expected risk is minimized because variational free energy is minimized only when the precision of the risk gamma is maximized. C method section generative model for further details. Very interesting.

They then say the class of neural networks we consider can be viewed as a class of reservoir networks. Citation 54 citation 55 here, the proposed equivalents could render such reservoir networks explainable and may provide the optimal plasticity rules for these networks to minimize future risk by using the formal analogy to variational free energy minimization under the particular form of PMDP models. A clear interpretation of reservoir networks remains an important open issue in computational neuroscience and machine learning.

1:09:38 So from Wikipedia reciprocal computing is a framework for computation derived from recurrent neural network theory that maps input signals into higher dimensional computational spaces through the dynamics of a fixed nonlinear system called a reservoir. After the input signal is fed into the reservoir, which is treated as a black box, a simple readout mechanism is trained to read the state of the reservoir and map it to the desired output. Then there's two key benefits of this approach. The first key benefit of this framework is that training is performed only at the readout stage as the reservoir dynamics are fixed. The second is that the computational power of naturally available systems, both classical and quantum mechanical, can be used to reduce the effective computational cost.

Here some stable diffusion reservoir computing, active inference neural network so this would be interesting to discuss. And I remember some Octave institute participants who are specifically interested empirical analysis and hypotheses. They write the equivalent between neural network dynamics and gradient flows on variational free energy is empirically testable using electrophysiological recordings or functional imaging of brain activity. So then another summarization crucially the proposed equivalence guarantees that an arbitrary neural network that minimizes its cost function, possibly implemented in biological organisms or neuromorphic hardware, can be cast as performing variational Bayesian inference. So to state it a few more times in the final paragraph, in

summary, a class of biologically plausible cost functions for canonical neural networks can be cast as variational free energy.

1:11:45 Formal correspondences exist between priors posteriors and cost functions. This means that canonical neural networks that optimize their cost functions implicitly perform active inference. This approach enables identification of the implicit generative model and reconstruction of variational free energy that neural networks employ. This means that in principle, neural activity, behavior and learning through plasticity can be predicted under Bayes optimality assumptions.

There's a code availability statement. The MATLAB scripts are available at GitHub in the repo of the first author, and it would be awesome for the author or anyone to bring this working MATLAB script up and see if we could do some realtime active inference. Then, as mentioned earlier, from the roadmap the methods are following the discussion. I'll just show the equations but not go into any details because there is not time nor familiarity. So those who have more of one or the other would be welcome to fill in some dots because this is a really awesome and important paper.

1:13:18 So I hope that it can be interpreted and critiqued and elaborated on and so on by those with familiarity in both sides of that free energy loss function.

Equation one generative model section, many details are provided. Equation ten is shown. So larger unpacking of the generative model section variational free energy many details are provided, equation 1112, 1314 and 15 section inference and learning details are provided. Equations 16 1718 then section on neural networks so there's just some interesting writing here. So I wanted to highlight that in this section we elaborate the one to one correspondences between components of the canonical neural networks and variational Bayes via an analytically tractable model.

1:14:26 So that's the figure three that we've been returning to. Neurons respond quickly to a continuous stimulus stream with a time scale faster than typical changes in sensory input. For instance, a peak of spiking neurons in the visual cortex of V one appears within 50 and 80 milliseconds after a visual stimulation citation 62 63, which is substantially faster than the temporal autocorrelation timescale of natural image sequences about 500 milliseconds. Citation 64 65. So that's pretty interesting.

What is the temporal autocorrelation timescale of natural sequences?

What timescale do neural firing and neuroplasticity etc processes actually occur at? And when might some type of function at a given time scale be understood to be functional or not?

1:15:41 Thus, in other words, downstream of the fact that neurons respond quickly at a time scale faster than typical changes in sensory input, we consider the case where the neural activity converges to a fixed point given a sensory stimulus. We note that the present equivalence is derived from the differential equations equation six, but not from its fixed point. Thus, the equivalence holds true irrespective of the time constant of neurons to rephrase neural networks with a large time constant formally correspond to Bayesian belief updating with a large time constant, which implies an insufficient coverage of the posterior beliefs. Pretty interesting related to learning rates and Bayesian updating rates and all of the nooks and crannies of Bayesian inference and the

challenges associated with dynamic uncertain, rugged fitness and free energy landscapes. These optimization challenges, which are addressed differently methodologically, culturally, etc.

1:17:05 In the variational Bayesian and in the neural cases, they have to deal with time. And so all of those different nooks and crannies mentioned, like catastrophic learning, catastrophic forgetting, simply memory, anticipation, attention, local maxima, choosing when to play all these higher order questions are connected here does that maze it? What kind of a problem space now or just what kind of space? Pretty interesting. And equation 19 202-021-2223 they have some more details on the simulation.

Maybe we could see the simulation go and in the last section data analysis. So this is kind of returning to that point about the time constants in Bayesian and neural network systems when the belief updating of implicit priors D and E is slow in relation to experimental observations, d and E, which are encoded by phi and PSI, can be viewed as being fixed over a short period of time as an analogy to homeostatic plasticity over longer time scales. 66 homeostatic Plasticity in the developing nervous System 2004 very interesting in light of our recent discussions on allostasis and other topics. Thus by and PSI can be statistically estimated by a conventional Bayesian inference or maximum likelihood estimation given a flat prior based on empirically observed neuronal responses. In this case, the estimators of phi and PSI are obtained as follows Nice equation number 24 mentioned earlier, so that will be definitely one to look into more.

1:19:27 Well, I hope you found this a useful and interesting zero. I'm looking forward to the discussion with the author and again, any other institute participants or those with knowledge or strong feelings to express about neural networks, active inference, applied, active inference, computational modeling of perception, cognition and action, neural networks in the wild, AI ethics, all these different areas can hopefully have a nice discussion in 51.1 and .2. That'll be the last paper streams for 2022. And yeah, if you want to be more involved with live streams whenever you're listening to this, join or recommend someone to join or help in any number of other ways. Just listening and learning is awesome.

1:20:55 And we also really look forward to those who want to help make some of these connections that are latent and sometimes exposed in these papers and conversations, and with a few motivated people who want to connect, for example, to the neural network communities and those who can facilitate discussions on these topics, that would be awesome. Just using my final thoughts on this dot zero, because it's always great to have others also join in the preparation for the dot zero. So just want to add that note on this rare solo stream. So thanks again, looking forward to talking or seeing you later. Bye.

Session 051.1, November 9, 2022

https://www.youtube.com/watch?v=IM_NlUzyq8M

First participatory discussion on the 2022 paper "Canonical neural networks perform active inference" by Takuya Isomura, Hideaki Shimazaki & Karl J. Friston.

SPEAKERS

Daniel Ari Friedman, Takuya Isomura

CONTENTS

TRANSCRIPT

00:39 DANIEL FRIEDMAN:
All right, hello everyone. Welcome. This is ActInf livestream number 51 one. We are in the second discussion of this paper, “Canonical Neural Networks Perform Active Inference. Welcome to the Active Inference Institute.
00:55 Daniel:
We're a participatory online institute that is communicating, learning and practicing applied Active Inference. You can find us on this slide and this is recorded in an archived livestream. So please provide us feedback so we can improve our work. All backgrounds and perspectives are welcome and we'll follow good video etiquette for live streams, head over ActiveInference.org to learn more about the institute and how to participate in projects and learning groups. All right, we're in ActInf

Livestream number 51 Dot One, and having our first nonsolo discussion on this paper, “Canonical Neural Networks Perform Active Inference, and really appreciative that you've joined today.
01:44 It's going to be a great discussion. We'll begin with introductions. I'll say hello and then please just jump in however you'd like. And we can start by setting some context. So I'm Daniel, I'm a researcher in California, and I was interested in this paper because we've been talking a lot about active inference from a variety of different perspectives, from the more fundamental math and physics to some applications, philosophy, embodiment, all these really interesting threads.
02:22 And this paper seems to make a really clear meaningful contribution and connection by connecting active inference entities and this approach of modeling to neural networks which are in daily use globally. So thought it was a fascinating connection and really appreciate that we can talk about this today. So to you and welcome. Go forward, Takuya, however you'd like to introduce and say hello.
02:54 Yeah. Hi. I'm Takua Isomura, neuroscientist in RIKEN Brain Science Institute in Japan. I'm particularly interested in universal characterization of neural network and brain using mathematical techniques.
03:16 TAKUYA ISOMURA:
So this work is I believe important as a link between active brain formal aspects, Bayesian aspect of the brain, and the dynamics system aspect of the neural network. So I'm very happy to join this discussion session. Thank you for invitation. Nice to meet you.
03:46 Daniel:
Nice to meet you as well. The first thing you added, the universal characterization of neural networks. What is the universal characterization of neural networks? Why is it being pursued in this area of research? So, as a narrow sense, my gain aim of this paper is that so, you know, people active inference lab communication to characterize brain activity, behavior, so on, so on, but which would be different from conventional neural network. So there is a crossover program which is associated with conventional neural network and it is not very clear whether all characterization of computational neural network can be explained by activity infrastructure principle or not.
04:50 Takuya:
So here universal characterization means that characterization of every aspect of conventional neural network which is a kind of dynamics system derived as association between biological phenomena and simple mathematics. Car formula using gift card, using differential equations as the broad sense. I think universal characterization means that well, it is a characterization of brain intelligence, but it's a big picture and the paper particular address is only one aspect of the big picture.
05:46 Daniel:
All right? So it'll be great to pull back to really understand what synthesis is happening. So I'm going to ask what makes a neural network model a neural network model and what makes active inference lab model an active inference model? Is this synthesis and connection you've made true? Because of what?
06:14 Takuya:
Because basically what we show is the mathematical equivalence between the formulation of canonical neural networks and the formulation active inference lab in the sense that we show that as possible neural networks can be characterized by minimization of some biological plausible cost function. And we show that that cost function can be least as variational based on inference and a particular cross of

gentlemen model in terms of well known partially observable position process.
07:06 Daniel:
Alright, shall we perhaps walk through some of the sections of the paper? It would be awesome. Just for each of these sections, maybe the numbered and the lettered sections. What does the section aim to show and why was it there in the paper?
07:37 Takuya:
Briefly it's over, right? So briefly. So first we introduce so the gain issue main program, our interest, which is relationship.
08:00 We try to make a formal ring between neural network and active reinforcements that gain program background. And then we first formulate the equivalence, mathematical equivalence, in a very Brea manner. So in the first section in results, we formulate the relationship using complete craft serum, which is well known statistical theorem proposed very long time ago. And using that we link a general form of neural network with a general form of variational data impress. But a problem is that this characterization does not address a specific generative model which is crucial to characterize a specific model, specific neural network dynamics.
09:13 So in the following sections, we characterize the problem using Pomodb or partially observable Markosition process and link that model with a particular class force canonical neural network. And then we simulated we use the simulation to propagate that property in terms of some major tasks.
09:55 Daniel:
All right, thank you for this. Could we talk about the complete class theorem? So what is the scope of the complete class theorem and why was it the relevant set of the neural networks to pursue or the right way to frame it? Thank you for asking that. So I like the slide you showed last week's video.
10:25 Takuya:
So computer cross theorem basically indicates the relationship between some crossover decision rule and vision in France. Here a crucial keyword is admissible decision rule, which is a rule which is as good as other decision rules or at least at one point better than other decision rules. So simply speaking, adomissibility indicates in some sense it is the best rule for some aspect. And usually we characterize such a goodness using cost function, loss function or risk function. And here what we did is we established some association with this type of loss function or risk function with canonical neural network which is we call cost function or biotic roles Costa function or neural network.
11:48 So our assumption is that neural network minimize cost function. So if it active the inclination and it is virusly active some sort of optimality so we can say it is adommissible with respect to that cost function. So the beauty of complete cross theorem is that if we find some admissible decision rule then automatically we can say that it is based on inference in terms of some Bayesian Costa function with gentlemen model a priori beliefs. So this computer chaos theorem is crucial as abstract characterization of the relationship between conventional neural network architecture, dynamics and variational. Beijing influence.
12:51 Daniel:
All right, thank you. What does it mean when you said it was biologically plausible of a loss function? The term is a little bit arbitrary because in this paper we mean by probability in the sense that this neural network model can be derived from realistic neural model through some approximation. And so here barricade probability, suggest means probability as a neural model or synaptic processing model.

And if this cost function loss function can derive such a plausible algorithm, then we can say that this cost function is barely plausible.
13:59 So what is the distinction between those neural and synaptic components in the loss function or what equation to look at? You mean distinction between dynamics and synaptic? Yeah. What is the distinction between them and how is it represented in the equations? Okay, basically neuropathivity equation means differentiate equation about a variable that represents firing intensity or some sort of variables associated with the firing.
14:43 Takuya:
On the other hand, dusty equation means an update rule about the synaptic weight or synaptic strengths which is a connection between two neurons. And beauty of this formulation proposed in this paper is that we characterize both heuristic equations synaptic procedure equations in terms of gradient descent on a same cost function, common cost function. So we can say that if we consider the partial derivative of some cost function with respect to new activity, then it's derived by gradient descent rule about if we consider a partial derivative of chaos function errors with respect to synaptic weights, then we derive a prosthesis rule.
16:10 Daniel:
Are those the only two aspects of a neural network or why are those the two key aspects?
16:20 Takuya:
It is a main, I think it's the main body of the neural activity. If we consider some inference running or action exhibit by neural networks in the sense that neural activity correspond to fast dynamics, fast gradient dynamics mix and scientific processes indicate through dynamics that minimize least function and cost function. But in general, we can consider any aspects, any variables associated with your method. For example, at least what we show in the paper is any free parameter which may be associated with firing threshold or although we don't discuss in this paper it would be possible to add other variables related to neural network. For example, here we ignored contribution of Griad factor but it would be possible to add the Griar factor in this correlation

or any other aspect of virus corporate neural network.
17:44 Daniel:
That's very interesting and it speaks also to a general separation of time scales. For example in different multi scale systems or in the renormalization group where it's describing some minimal multi time scale system where the faster time scale can be seen as perception like a slower time scale can be seen as more learning like. And then in some hierarchical model what's learning of one time scale can be perceptual for a slower time scale? So it's a very nice generalization.
18:32 Are there any examples of decision rules that will help us think about the action components of what the neural network is doing? Because it may be more familiar to think about digit characterization and image classification, some kind of classical tasks for neural networks. But how does the decision rule play out in the context of neural networks?
19:04 Takuya:
Okay, so in this paper we basically assume a closed loop so comprising a neural network part and environmental part. So Neuron receives sensor input from environment and provide some feedback to the environment.
19:31 Even with the example of classification, we can say that output correspond to classification

output, which is kind of generative model relevant. Example would be, for example, controlling agent like a robot control or any kind of control errors. Decision making tasks. For example, when we encounter some choice tasks, we need to advertise, for example, left or right or something. Any kind of such a decision can be associated with the admissibility or admissible decision.
20:27 Daniel:
So what would an example of an inadmissible or admissible strategy be in the decision making task?
20:40 Takuya:
Admissibility usually characterized by loss function or risk function.
20:52 Here admissivity indicates that there is another decision rule which is at least one point better than the forecast decision rule.
21:12 Simply speaking in Adobe CBD indicates that decision rule is not good relatively. Let's just say our decision rule is we always turn right. Is that an example of a decision rule? Because there might be settings where that is strictly effective and the simplest rule whereas there's other settings where that's going to be tragic. So what does it mean to be admissible for an agent in light of different environmental contexts?
21:52 That's an interesting point. So even with such a too much simplified rule it can be admissible under some particular situation, particular loss function. For example, the rulers that always turn right maybe the best under some situation, right? So the relationship of admissibility or enough adommissibility depends on both agent characteristics and environmental characteristics.
22:39 Daniel:
What aspects of the environment.
22:44 Takuya:
For example? For example, if that decision group matches the structure architecture of environment then maybe that decision always downright active the shortest past under some situation, some environment.
23:11 Daniel:
How does this admissibility help us think about like overfitting and how does it help us think about the way that different practices are used for neural networks to prevent them from being over fit in practice?
23:30 Takuya:
Well, strictly admissivity is characterized with the Bayesian risk.
23:50 We cannot observe a hidden states of the environment, only we can observe is a part of the entire universe. So the question is an important question is what is the best choice under such a limited information? Limited information? So this Bayesian list adoissibility or computer credit theorem tell us that well known, only the well known Bayesian framework achieved the adommissible decision. Which means that in this aspects Bayesian optimization give us a least choice strategy, otherwise we overfit or find the suboptima evolution.
25:13 So it's a nice association, nice linkage between the decision, but is a good decision about the decision and more established statistical inference. Freedom work.
25:31 Daniel:
Thank you, that's very helpful. So we're reducing our uncertainty and risk about hidden states in the environment. So in the special case where the entire environment is observable without errors like a

chess game, then there's an equivalence between correlation of risk or loss on observables or on hidden states. But they're not really hidden, but they are environmental states. Whereas any amount of uncertainty in the mapping between observations and hidden states, which is usually shown as a in the partially observable Markov decision process, any amount of uncertainty about unobserved or partially observed environmental states enables you to fit your uncertainty optimally about that hidden state and fit that uncertainty simply with the gradient descent.
26:45 And by doing so, you don't overfit a model of observables, which might be the fallacy or the issue with simply doing descriptive statistics you might get an infinitely small variance with a frequentist estimate because you have 1000 data points. So the variance from a descriptive statistics perspective might be very small.
27:21 I think it speaks very much to why neural networks are useful in practice from training with limited data sets because that's an empirical observation that they don't entirely over fit. But also I'm sure there's ways to construct them that are overfit. Yeah, overfit will occur if we select some optimal priorities. For example.
27:53 Takuya:
Well, I'm not sure if it is overfit in the sense what you mentioned because if we select some priorities then the Bayesian function itself changes and the neural networks that try to fit to that Costa function. So cost function minimization will be achieved agent such a situation. But that solution is not good for our original help us. That's the tricky part. Yeah, that is reminiscent of some discussions we've had discussing like driving off a cliff or blowing up is also reducing free energy.
28:46 Daniel:
Like dropping up a building reduces your potential energy. And so there are potentially decisionmaking or strategic trajectories that do for some time horizon minimize free energy, perhaps even or maybe even guaranteed better than some longer time horizon. Because if the shortterm strategies were somehow better than the longterm horizon. It would be difficult to imagine because the long term horizon would be at least as good as a shortterm strategy. So that speaks to the challenges of planning in action.
29:28 So how is planning addressed in modern neural networks and how does this work help us think about that?
29:39 Takuya:
That's another very important aspect.
29:45 I have to say that this framework addresses planning aspect, but that planning is not necessarily the optimal solution in the sense that what we interested in is optimization or learning under limited structure. The structure is characterized by here Prosperia neural networks. So yeah, planning occurred by association between risk in the future and our decision in the past. Here we model that aspects using delayed moderation of scientific activity mediated by some neuromodurator or neurotransmitters. This is the model.
30:58 This is model as the risk factor and the heavy product holding the neural network.
31:21 Daniel:
All right, I'm going to ask a great question from the chat and then we'll look at the figures a little closer. So ML Don wrote a question stuck in my mind for a long time. Could you please put it to rest? Do we need to have knowledge about all states possible actions and sensory inputs for active inference?

31:50 Takuya:
Well, you mean if you seek the exact solution, exact optimal solution, then maybe more information would help you to find that. But under some ideal assumptions then the is not necessary to achieve the optimal solution. I'm not sure if I correctly answer your point. So just to restate it. Of course, knowing all the state's possible actions and sensory inputs, it's not a bad thing.
32:44 Daniel:
Worst case, there's some computational complexity, trade offs, but the problem becomes fully stateable. But I think ML Dawn is asking about cases where you don't know all of the state spaces or potentially even the dimension or the semantics of hidden states, active states, sensory inputs and why not even add cognitive states? So in not just partially observed but partially known state spaces, how are these address in neural networks and how does active inference help us think about it?
33:37 Takuya:
Okay, I think the question is about how can we separate those states? Like sensory function interface entorhinal, how can. We separate not just in principle have these states be separated, but deal with the fact that some of these states we might have good knowledge on and some states like the hidden states we might not even know, like we don't know the dimension of the cause vector in the world. I see.
34:22 In terms of dimension, there is a statistical technique to estimate the dimensionality, for example via information criteria like I agent information criteria, based information criteria, all them try to info estimate plausible dimension about the environmental hidden states. There is an analogy with those information criteria and version of free energy minimization. So with version of free energy inclination we can identify the plausible model structure which in principle involves the dimension aspect. But in terms of Neural network in this paper we don't carefully consider about the dimensionality optimization because we first define the number of neurons and don't change during the training. But in principle we can consider the change in the number of neurons which is associated with the neurogenesis adult neurogenesis or development during the developmental stage.
35:57 That would be an important expansion of

this direction.
36:13 Daniel:
That's very interesting. Here's a remark. Well, one note is equation one summarizes a lot of what you've been describing. There's a parallelism or a concordance being drawn between the loss function of Neural networks and the variational free energy of the parameterized model there. So to come back to these processes that influence learning which we could think of as the Neural network becoming more fit from a loss function perspective or the variational Bayesian partially Observable Markov decision process entity generative model encoding better at doing what it does.
37:05 So there's the firing rate on the Neural network side, the synaptic plasticity at a slower time scale which we discussed a little earlier. And then now there's a third time scale with the birth and death of new cells and maybe even new layers. And that kind of multiscale temporal structuring is not intrinsic to the Bayes graph to represent multiple nested timescales in a Bayesian graph in the act of inference literature it's more common to make a hierarchically nested model, right? And just say that the time handling on one level is happening more rapidly with respect to clock time than deeper nested, slower models. Whereas the Neural formulation allows us to deal with multiple ongoing active states without appealing to hierarchical nesting, which is a very important feature.

38:31 Takuya:
Well, both distinctions will be possible. So without hierarchical or with higher car modeling so even with hierarchical modeling, the optimization of dimensionality should be possible. It would be possible. But in other distinctions we can consider that a population of Neural models so one has a single layer, another has two layers, three layers, four layers. And consider the probability of network architectures associated with Costa minimization and a particular environment which is in principle have the same computational architectures with the hierarchy model.
39:46 Daniel:
Very interesting. Yes, perhaps I over generalized or speculated because I thought about how one could have a 100 timestep POMDP that also performs multiscale behavior potentially extremely wastefully, but at least it could in principle. And similarly, within a neuron there could be another Neural network or some other structure approximated by that. So they almost both enable hierarchical and non hierarchical model modeling as you described, but in very different ways that lead to very different implementations.
40:44 Takuya:
Yes. I think this brings us to the topic of forward and reverse engineering. So you talked a lot about reverse engineering. What is reverse engineering and what is forward engineering and what has been done in these areas of engineering? Okay, I'm not an expert in this process, but I believe that liver here means your characterization of the blueprint of some device or machine from data observable information like activity or action behavior of some agent.
41:44 Goal is identification of blueprint and the crucially here blueprint correspond to generative models because once we define generative model, we can Deneve evolution, anthropology algorithm, running inference algorithm and any behavior of the agent. So here reverse means that we first observe some activity of agent and its mechanism is still unknown for us, but we can estimate its mechanism using that activity by identifying the most plausible guarantee model which can minimize some Costa function or risk function when we feed the data to the model. So, on the other hand, for the engineering would be more mainstream, way fast defined model blueprint gently model then drive everything including parasite functional running algorithms and behavior action prediction algorithm.
43:19 Daniel:
So, by reverse engineering neural networks, we're observing some already parameterized neural network and then fitting a POMDP to it. To what extent is it possible to take a given POMDP and create a neural network that performs that inference?
43:52 Takuya:
Okay, in this paper or in the following paper, what we consider is a strategy that we first feed empirical data whether force neural response data to BioScale prosper neural network model which is similar to a conventional model fitting approach where we have differential equation data and differential equation to explain the behavior with the minimum prediction. So now, a virtue of this framework we established is that we can naturally transform such neural network architecture with the very known partially observable markup action process architecture. Because for any kind of canonical neural network there is a cost function. So we Deneve cost function through neuroactive decision which is opposite with the conventional way we define cost function derived algorithm and then we use the formal equivalence between neural network Costa function and variant queen energy. So now

transform the journal architecture to Beijing model architectures and once we characterize vital energy, there should be some general that define that informational energy functional.
46:01 So in particular, in this example, canon network nicely correspond to well known across macquarlin process. So, by using this procedure, we identify a plausible home DP architecture which correspond to observed activity data.
46:49 Daniel:
Well, let's stay on this last point. So, after all those transformations, first the measurements of neurons using that data to fit the neural network and then by virtue of the relationships unpacked in the paper, transforming the neural network in the left side of figure one into a particular form of the P-O-M DP. So first, what are the constraints on that form of the P-O-M DP? Is this a little corner of model space or what are the space of acceptable P-O-M DPS?
47:38 Takuya:
That totally depends on what kind of neural network model you are considering. So for example, in this paper we discussed about a particular crossover from DP in which each state takes either zero or one. So it's very restricted compared to the general form of homedp. But we consider a factorization so in the sense that although each but we consider a vector of observation, a vector of hidden states where each element correspond to one single one hot vector but as an entire state it can represent high dimension discrete state space. And this architectures nicely correspond to neural network architectures because usually each neuron takes either zero one or some value continuous variable between zero and one.
49:04 So we use this association to characterize a particular OMDP which correspond to neural networks, and this follows a particular mini field approximation, approximation or approximation in generative model because we associate posterior belief in this particular homo DP with the neural activity, which means that posterior of action also has a factorization architecture in the sense that we don't fully consider about the second order statistics between neurons activity and activity, which is outside of this poem. RASM. So each neuron activity correspond to posterior expectation about a particular element of the state and we don't consider the joint posterior property of all state.
50:40 So although this is a implication, we see this Asia impress, but otherwise, for example, we can consider any recurrent network architectures which correspond to state to transition metrics and it would be possible to extend this architecture to higher call structure in the sense that it is straightforward. Consider a tree structure or any kind of higher car structure by assumptions that some neurons connect to other neuron but not connect to other neurons. So this is Lamme as considering the higher car structure in general.
51:43 Daniel:
That's very interesting. It's commonly remarked in the base graphs that they represent the connections amongst random variables and there's a relationship between their computability and their sparsity. The sparsity structure as in which variables do or do not influence each other makes the problem tractable through factorization and just kind of conceptually like if every one of a thousand variables or an unknown number of large variables if it was all by all the number of parameters to fit on that connectivity matrix would be very high. So statistical power would be very low for any given edge. Whereas the more and more constrained you make the connectivity of the variables, the more statistical power you have to resolve or kind of spend on fitting those edges like in a structural equation.

52:49 But you might be losing sight of the unknown unknowns by constraining yourself to a very limited or fallacious topology of the variables. So there's this kind of structure learning statistical inference question in the Bayes graphs then on the neural side from the biological much of neuroscience is about understanding how the firing rate, connectivity patterns and other factors Hohwy the structure of those neural systems and their function like form and function enable adequate inference and inference on action. So it's like in both of those areas or really like in neural network artificial and neural networks and in variational. DAGs the discussion is about how the structure and the fine tuning work together to generate function and about some of the statistical or biological challenges of balancing different needs while also constraining the cost in terms of materials and biometabolism. So it's a very rich interoception that is being explored here.
54:29 If these models can really be moving back and forth.
54:38 Takuya:
In the sense that back and forth.
54:45 Daniel:
Moving back and forth, like there's some imprints of the model that is implementation independent or like some interlingua or some semantics or compatibility, I don't really know. I mean, that's

something we can explore is like what is it that is such that one could forward engineer and then reverse engineer and have like kind of an expectation maximization between these two areas. So what is it that's being solved?
55:25 Takuya:
Yes, important point, for example, about you is that we can use the knowledge of Bayesian inference to explain your activity dynamics, which is crucial because people often say that characterizing neurodynamics is no straightforward, we may obtain some solution on your net dynamics, but the meaning of that dynamics in terms of the functional aspect is very unclear. We don't know the meaning of connectivity strength matrices and what is the learning of the threshold factor, so on and so on those de Vries from the modern physiological phenomena. But it is not necessary to have clear linkage to functional exploration. So explanation of function of the brain. But once we transform translate this dynamics into Bayesian inference, then we can explain every functional aspect of the neural network diagrams architecture in terms of where established Bayesian inference under a particular crossover Bayesian model, in this case palm DP model.
57:08 So now it turns out that synaptic strength correspond to a matrix B matrix, which are very established culture meaning. So yeah, this is useful to explain neuronsynatic property in terms of established statistics.
57:44 Also, for the people in active inference lab site, it would be helpful to understand the neuronounce master straight about particular active interface model model. So I think it related to forward modeling. But finally to discuss with discuss about the border service rate of that forward model, we need to address the neural network architecture service property. So in that case, we can transform a particular force DP invasion modeling to a neural network architecture using this relationship and then get prediction about the substrate. So if we have this based on model, this particular quantity in this model should be it would be possible using.
59:14 Daniel:
Oh, it's all good. Can you just repeat the last 20 seconds? Yes. So in the last part I mentioned about first

we define the Bayesian model and then can predict what is the neural net substrates that correspond to that particular Beijing model. So this will be useful to identify the biological quantities that correspond to a quantity in Beijing.
59:53 Takuya:
Chaos.
1:00:03 Daniel:
There's a lot there. It makes me think about the inference of implementation and. Heuristics in the computational setting, which is often in the extreme disembodied, and the biological setting, which is in the extreme entirely embodied. And for a given generative model, the kinds of computational heuristics that can be applied include a whole host of different strategies ranging from sampling to tree exploration and branching to paralyzing the data architectures and all these other kinds of disparate strategies and software packages and implementations.
1:01:04 But on the biological side, what is needed is something that's very simple but also very inscrutable, which is a given pattern of interactions must embody that calculation. So that might mean that it can add three digit numbers, but it can't add two digit numbers under some constraints. But what isn't accessible to that kind of morphological, biological or like form and functional computing, what's not accessible are the tree branching, the database decentralization, like they're a different set of heuristics. Right. But they're both very useful when we're thinking about making sentience artifacts or benefiting simply from the explainability across both sides of this figure.
1:02:18 Takuya:
Yeah. So you now address an important point. So Homistry, it is very nontrivial whether there is a corresponding valve car architecture force any given Bayesian architectures. I believe it is impossible to design biography architectures to respond to arbitrary Bayesian architectures. So only a limited aspect of region model can be implemented in a vertical plausible manner.
1:02:55 And that point is crucial as capitalization of biological network. Biological brain.
1:03:09 Yeah. Wow. Well, just to kind of touch again on this forward in reverse engineering.
1:03:18 For. A given POMDP if we're willing to compose it within a certain class, which might be quite general still, but some class of PMDP, as written. On the paper. We may be able to have a neural network architecture that would be very amenable to deep learning, low energy computing, pretraining various features. And then on the other side, for a given artificial neural network that we come across in the wild or a model of neural dynamics that we fit using a neural network model.
1:04:08 Daniel:
So something in a neuroscience laboratory that model can have interpretability corresponding to the variables of a given POMDP. And just to kind of give one more point on how that's going deeper than, for example, statistical parametric mapping SPM. So let's just assume that the neural network we're dealing with is fit from brain data from some lucky Kant, right? Now, what would be possible or prior to this line of work or without this line of work, one could fit a neural dynamics model and then do all kinds of analyses, like power analyses on the different frequency spectra and say, look at the average firing rate or the correlation coefficients of firing rate. So fit the firing rates and the synaptic Plasticities and store all that data.
1:05:14 And then we could just pick a POMDP that we've seen in the literature without any reference to the neural network and optimize the POMDP. And then we could say well, it turns out that when the

POMDP o is high there's increased theta power in this firing pattern. So it's like comparing the descriptive statistics from the neural model to the descriptive summary statistics of the POMDP decision making model. However, with this formal connection there is actually an interpretability to the unobserved neural states which are what are being inferred from the fMRI measurement, from the EEG measurements and so on. Those underlying variables have a specific interpretability in relationship to the structure of the P-O-M DP.
1:06:21 Takuya:
Right? So yeah, that's also very interesting important aspect. So what you said is I think more conventional strategy and it is also formally related to model comparison aspect. So we usually think various modeling and identify or select what is the best model to explain a given data. And this reverse engineering idea involves such a model comparison aspect in the sense that we try to find the model with the best expandability which should we have the identical functionality, right directory address, the exact same Costa function architectures using the information natural transformation.
1:07:31 So it should be up to explain the neural data in the Bayesian sense.
1:07:42 Daniel:
Yeah, one can imagine how that would transform the way that current neuroimaging studies and technologies describe what it is about the measurement that provides information about the cognition model. So, to give another related example, let's just say a person was wearing an EEG headset and a previous study had shown that increased alphaband activity was associated with this behavior. That's comparing a descriptive statistic of the observations of the sensor and correlating the summarized observable to some other variable like anxiety or performance on a behavior.
1:08:47 In contrast, an unobserved variable in this setting the actual underlying neural state is being correlated to some semantic generative models component. So it's no longer necessarily that any single frequency band would be associated more or less with a given outcome, but it's actually some hidden state variability which gains the interpretability across this transformation. Which is a subtle point, but it speaks to how broadly the equivalents would reinterpret empirical neuroimaging results as well as a variety of artificial neural network experiments and diagnostics where people do lesion studies and double knockouts on artificial neural networks.
1:10:11 So anywhere where somebody with awareness sees that a neural network, artificial or biological, is having summary features described and correlated to something that's more semantic in a quest for meaning may now have a different approach that involves formalizing. The model explicitly in terms of unobserved hidden states with a cost function akin to a variational free energy minimizing risk bounding surprise on the Unobservables. So even though the unobservables were modeled in a sense in the other conventional strategy like neural activity is a variable in fMRI experiments, it's underlying the bold signal. Yet this formalism concordance is a more coherent and powerful connection.
1:11:35 Takuya:
Lib sold. So you now address this very important point. So first to address that so we need to clarify about what is a program, consider here. So this is a program Socalled metabasian problem in the sense that researchers try to infer or estimate neuro activity or brain activity which infer the external world dynamics. Right.
1:12:15 So neuron or brain environment and we research brain activity. So there are two step processes.

So this sort of meta Bay is quite tricky intractable because sometimes London variable becomes posterior about other aspects. So I think there is some established approach about metabolism. But this paper provides some alternative in the sense that we separate two programs by saying that here what we import is simply neural network dynamics which is shown in the left hand side of this figure.
1:13:25 So we feed data to conventional neural network model which is a simple differential creation. But thanks to this formal recovery

between neural network dynamics and home VP behavior, then we can transform the resulting neural network architectures or dynamics into the page and in force in some sense post Hog mana. So we nicely avoid the directory addressing the meta agent program but obtain the same kind of solution in that sense. Yes, with combining with brain activity recording Lieke de Boer imaging. Yeah, we can estimate a plausible neural network model in the right hand side and we can transform that to home DB in the right hand side.
1:14:40 Daniel:
Awesome. I'm going to show an image and ask a question from Dave in the chat. So, Dave made this image, it's the right side of figure one that we've just been looking at with the variational Bayesian information and he wrote the arc shown as impinging on the S self arc. Is this intentional? If so, it could represent tuning or modulation of the feedback of S into itself.
1:15:17 Do you have a thought on this? It's attention? Yes. I think it's related to the usual formulation of home DP architecture and active inference concept in the sense that our decision or policy in the usual setting modify the state transition matrix b matrix.
1:15:53 Takuya:
Here, delta is an alternative of policy of agent. So basically the director indicates stated transition metrics under a particular decision which agent made. In that sense, what the agent changes is state transition metrics, not state itself directly. That's why we use this illustration.
1:16:29 Daniel:
Awesome. Very subtle but important point, which is when we look at the classical POMDP formulation. So here we'll look at a version shown in figure two. I'll just bring just figure two in.
1:16:49 Could you describe what you just did about the role of the B matrix in influencing how hidden states change and how that is where our policies have impact? And also please, how do the top and the bottom of figure two differ?
1:17:13 Takuya:
Okay, so in the usual correlation under active inference with palm DB structure. So we for us to consider the prior inference and depending on the prior preference, we compute the expect free energy and its minimization provide the policy and the policy moderate state transition. So now in the upper Brea, we instead use the builder which is the option of the agent. So here option or decision was made for each timestep so that unlike the conventional formation, we have a sequence of delta and for each time step delta moderates active states cognition matrix B. So B is a matrix that transformed hidden state in the previous step to the current time step and its moderation indicate that under a specific decision rule.
1:18:51 For example, if this F indicates our cognition in the virtual environment with the Gold decision move forward. But if we choose the no go decision, then it unchanged. So such a moderation of state transition was made by choosing debuta and the lower part correspond to Beijing inference made by

Bayesian agent. So basically there is a symmetry between a third part and roar part because we assume that this Beijing agent has a plausible guarantee model which nicely correspond to given environment defined in the above upper part in this figure. But one interesting thing asymmetry is that to model this particular canonical neural network, we don't consider an arrow or link from delta posteria to S posteria which is in the environment data moderate S in the next step through P matrix moderation.
1:20:45 In this particular Bay jets which formally correspond to a canonical neural network, we don't consider that it is correspond to an absence of the projection from output layer to the middle layer.
1:21:19 Okay.
1:21:25 Daniel:
This is from the 2020 paper, but it shows the neural network architectures, the two layer architectures. So could you restate the top and the bottom of figure two in the 22 paper and connect it to why it's important that you're studying two layer neural network models? I miss you. Yeah, can you just connect the asymmetry between the top and the bottom on figure two with the two layer neural network architectures? You said that the asymmetry, there's no direct link between.
1:22:27 Takuya:
This is another story. So in the previous paper there is only output or concept layer because we basically consider a single layer feedforward network. So my apologies for some confusion about the network architectures in the 2020 paper. So now upper part of this network architectures correspond to environmental generative process and only a lower part correspond to single foot feed for neural network architectures. So now this part is identical to a map from O to S.
1:23:33 OSS area in the 2022 papers.
1:23:43 Daniel:
Okay, so on the top of figure two is the actual generative process. It's the true structure of causation in the environment, which is to say that actions delta actually influence how states change through time via B delta. The generative process through the A matrix emits observations, sequences of observations. And here on the bottom with a mirrored structure is the generative model of the entity. So what's the relevance of the arrows and the more force factor graph structure on the bottom?
1:24:40 Takuya:
The arrow indicates active inference.
1:24:50 So it's a flow of the information in the sense that to calculate in the step two, we use the information of step two conversation and step one's posterior expectation about hidden states. So those two determine the s two's expectation. Usually in the following graph, we consider retrospective arrow so in the sense that s three also affect the s two inference. But this corresponds to Bayesian smoother in the sense that we update every time step simultaneously to better inference. However, what we consider here is more filtering approach in the sense that for each step we compute the latest hidden states and then we don't change any other states in the past.
1:26:11 So that's why we don't consider the arrow from future to the past.
1:26:22 Daniel:
Awesome. Yeah. Just to highlight that in the Bayesian smoothing approach, it's kind of like fitting a spline because it takes the whole time series and it fits the smoothest possible line or the line whose smoothness is on the AIC BIC frontier. But here on the bottom with the almost pseudocode implementation provided by the Force Factor graph, which was demonstrated to be equivalent with the

Bayesian graph in the 2017 work with Friston, Par and de Vries. This architectures is reflecting a filtering scheme like a common filter or just generalized Bayesian filtering through time, where estimates are being carried forward and changed time point to time point, such that the decision rules, or the updates perhaps more accurately, are defined between time points.
1:27:33 And the total time series does not have to be loaded into memory or remembered at once. And then the Bayesian filtering approach has the asymmetry with a different consideration of action. So why again is it that action is considered differently in the Bayesian filtering approach on the bottom of the generative models than the consideration of action in the generative process.
1:28:11 Takuya:
That correspond to lack of cognition from Y to X in the figure one? Or probably a figure four is helpful to that relationship.
1:28:37 Daniel:
Four? Yeah. This is an example network architecture comprising input Brea, output Brea. What we consider is information flow from sensory to middle Brea and middle area have a self connection, recurrent connection and middle area project to output layer. So there is no connection from all output layer to middle layer.
1:29:08 Takuya:
Right. So that's why we don't consider the link from data in the bottom layer of the figure to posterior. So this is different from true generative process in the environment.
1:29:37 This is a kind of simplification. So because our purpose is identifying the plausible Bayesian model which correspond to this type of neural network canonical network. So in other words, this neural network uses approximation about that point or use limited form of palm DP scheme.
1:30:23 Daniel:
Thanks. So could you describe W-V-K and Gamma? Just what is the biological or functional interpretation of those variables? What brain regions or what processes or pathologies do they map to? Okay, so basically, WVK, synaptic strength is in the form of matrix and active inference.
1:31:04 Takuya:
They represent connection in the different layer or different architectures in the sense that W means forward connectivity from sensory Laje to middle Laje, k correspond to recurrent network recurrent connectivity and V correspond to projection from middle Brea to output layer. So in this paper, we don't discuss the relation to brain anatomy in detail, but what one can consider analogy, for example, say x corresponds to several cortex activity and Y, for example correspond to cerebral wrong in the sense that it determines the action. So it is considered that in the cerebrum there is a signal that represents choice. This is joined for examples goal no go decision made in cergram.
1:32:31 It's analogous to this particular architectures. On the other hand, in the several cortex we compute the sensory information to generate some inference, prediction and planning the way it is computer by this recurrent network. In this particular modeling, although we don't separate brain region in detail, but this recurrent network is sufficient this graph of recurrent network is sufficient to characterize any type brain architecture in the sense that we can design any higher car or mutually connected architecture using a generic crossover recurrent network by changing weight.
1:33:54 Daniel:

Awesome. So the middle layer we can think of as like the cognition stuff. It's the internal states when we talk about perception, cognition,

action in the active scheme or even in the sandwich model of cognition, perception, cognition action. So W is describing how those sensory inputs either in one step or composably in multiple steps become processed to these internal representation of hidden external causes inferred external states. And so these are the states that have that sigma relationship and a generalized synchrony with external states.
1:34:48 The sigma and the generalized synchrony are not discussed in your paper, but it connects to other work and the recurrent connections are facilitating attention or waiting of the stimuli. This is the recurrent learning loop and the relationship of the A between observations and hidden state estimates. And then a different kind of modulation comes Hinton play between the hidden state estimate of the internal states state and the action selection. So what is gamma corresponding to? And why is the gamma modulation between layers two and three differing functionally from the k synaptic modulation of one and two?
1:35:47 Takuya:
Yeah, so K matrix basically formally correspond to B matrix in the Bayesian information. So we rotate the information about the prediction, right, our narrator or our expectation about the next state based on the previous state. On the other hand, Laurel gamma is quite different from such a computation. Gamma basically means risk function, which is in principle can we use arbitrary risk function. So this is a part of generative models we designed and the rule of risk function in generative model formulation is attention form of generative models depending on that value of gamma which examples retrospective moderation of evaluation of task decisions given an outcome in the future.
1:37:15 In terms of neural network, of course it corresponds to some neural modulation. For example, Dopaminergic moderation is famous in the literature which moderates the activity and fluxicity of various brain vision. But we particularly focus on Dopaminergic or any kind of neuromoduration of cyanogic prosthesis in the output trigger which may correspond to Cergram. So in the Serbram neural activity or processes moderated by Dopaminergic, input from is used as the optimization action rule, decision rule or sometimes attention help us.
1:38:29 Daniel:
Awesome. Very interesting because in some previous papers and models that we've looked at, attention is dealt with as policy selection on mental states. So internal action selection, it's an action like variable describing attention and awareness and even metacognition. And so that connects the role of Dopamine in motor decision making seen in many Dyskinesias but also with the role of Dopamine in seemingly nonmotor based decisionmaking like gambling or investing where it doesn't seem to immediately translate to a given motor sequence. Yet it has analogous computational characteristics and the comorbidities and the side effects of different drugs that affect the Dopamine neurophysiology are known to have carryover in terms of like the rigidity or excessivity of motor and decision making aspects.
1:39:50 So it's like interesting that Dopamine has long been understood to have that parallel role in attention as cognitive action and motor action and that was established empirically through modifications of Dopamine signaling and also had been modeled analogously with computational neuroscience. And this is providing again a slightly different interpretation of that very well studied Dofaminergic modulation of attention and policy.

1:40:40 Takuya:
Yes. In addition to that, I believe another important aspects is correlation of scientific processing by document.
1:41:15 Daniel:
Do you want to show something or yeah. Can you see this paper?
1:41:27 Takuya:
I sent you a chat. If you can't, I'll send you a PDF. Okay, let me see. I'll look at it up now.
1:41:47 Daniel:
All right. The paper is a critical time window for Dopamine actions on the structural plasticity of dendritic spines from 2014 byagasha. So what is interesting about this paper. Yeah, it basically explained conversation of plasticity by Dopamine, which is common but crucial point of this paper is that it shows that it proved that Dopaminezic input can moderate after hebion prosthesis is established. So this paper showed that they add domain logic input for about 2 seconds after or several seconds after the Hebbian process is established.
1:42:50 Takuya:
But such a post hoc moderation, post hoc introduction of heterotopamagic Impetu is sufficient to change the past capacity which may be associated with the Costa hoc evolution of our past decisions. So by decision making we of course changes the changes the weight matrix by through trust 50. But to evaluate the goodness or badness of our decision, we need to see observe the future outcome which is propagated by for example, Dopamine. And this paper nicely show empirically that Dopamine actually can change the past evaluation, maybe after such a psychic level, very local level, ecoscopic level.
1:44:12 Daniel:
So there's a short term window, the critical time window that they're describing. But there's some window. Yeah, some window by which dopamine potentially unrelated to the initial heavy and plasticity events, right. Where secondary dopamine signaling or not secondary just after the initial fact, potentially of a different valence or the same valence can synergize or cancel the plasticity formed in the moment. Exactly.
1:45:05 Takuya:
And this is not limited to Dopamine, but other neuro moderator can also do this.
1:45:18 Daniel:
Well, on one hand, how does this change our understanding of animal neurophysiology? And then I guess, on the other hand, how does this influence how we would design sentient artifacts.
1:45:42 Takuya:
For both animals and artificial agent?
1:45:53 One important message I free with us. So this tells us possible simple architectures to make learning. This is association between past decision and future reward or any risk factors, which is otherwise computed by computing forward prediction by iterating some computational, this is a usual way to predict the future event and then select the option. But using this property, biological property, which is observed in experiment, we can design, we can imagine other simpler architecture to make a planning.
1:46:57 So for both animals and synthetic Bayesian agent, it provides an alternative explanation about the association between our past decision and the future risk and the optimization of our decision to

maximize, reward or minimize risk.
1:47:25 Daniel:
Well, one interesting note is we spoke earlier about the difference between the Bayesian smoothing all at once approach and the Bayesian filtering step by step approach. Now, if one had infinite knowledge and computational resources, the Bayesian smoothing approach is the way to go. Like, you don't want the decision rule for investment. You want to look at the whole time series past, present and future, and know the best moments to have made the trades. I mean, there's no comparison.
1:47:58 You're going to do better with the Bayesian smoothing. However, it's just implausible computationally and because it requires total memory of the past and knowledge of the future. So that's what motivates the development of Bayesian filtering approaches, which are tractable and calculable through time. Yet with this time delayed modulation. Part.
1:48:27 Of the Bayesian smoothing strength comma back into play. It doesn't enable true anticipation of future states, but that's what the expected free energy does. However, the delayed neuromodulation allows for reconsideration of a window of past states. And so in that way it corresponds to like a slightly deeper filter, not just a filter of a time step of one, but a filter of like a rolling window of five or with some decay. And you don't want that window to be too big because if the window were ten minutes, then too many contrasting stimuli would get piled together.
1:49:23 The Dopamine level would just converge to a mean field average. But there's some time decay or time constant on the post hoc modulation where that neuromodulatory signal is actually a parameter of interest. And that's not an infinitely long or infinitely short window, but it's some niche dependent amount of time. And that's a very interpretable and first principles interpretation of the computational role of neuromodulators in a way that is also consistent with all these other concordances we've been exploring. So it's quite an interesting connection back, I guess, in our final minutes of this discussion.
1:50:26 What are you? Well, maybe go to the beginning at the end, which I meant to ask earlier, but it's a good way that we can sort of close today and look forward, which is how did you come to this line of research specifically studying neural networks in this way with Karl Friston and your colleagues?
1:50:58 Takuya:
So, yes, so my interest was the characterization of Barricade network. So my first motivation is to make biologically plausible artificial intelligence. But to achieve that, we need to know about biological brain or biological neural networking.
1:51:41 In these several years, I collaborated with the doctor professor Californiston to study about his salary principal after doing forest.
1:52:01 My question during that period was the priority principle, is everything about the biological possible neural network or is there another aspect that can characterize the virus car neural network? So it is non trivial. It was non trivial. So that's why I tried to start from characterizing the neural network first. So our strategy is not considering the way of implementing any Bayesian algorithm as the brain architectures, but my interest is rather characterization of a given vertical network in terms of some other things.
1:53:17

One possible way is of course based on inference free energy transplant reinforce. So that's why I start from characterizing power's network. But just defining neural network architecture is insufficient.
1:53:41 It is not tractable, it is far beyond the computational tractability as the mathematical analysis.

And we need some assumptions or some trick to increase the tractability. One day I came up with an idea that in which we consider that both new activity and fastest follow the same cost function gradient. This is very much an allergy with physical system like Lagrangian information geometry, Hamiltonian formation. So usually we consider some energy landscape and design plausible trajectory as the evolution of some principle of minimum action or restruction.
1:54:59 So we imagine that what if we applied such idea to computational neural network or biological neural networks to characterize their dynamics in the first principle, that's the first computational step to come up with this framework. And finally we noticed that it is not easy to connect the Newtonian dynamics with this type of neural activity study because neural activification not necessary to be a second order differential equation, but rather it is fast order and considering many things. Then we finally use a Costa function proposal in the papers, which is not necessary to have a former identity with the so called lavalier in the Newtonian physics, but it is rather plausible as the rule or underlying mechanism of such type of network.
1:56:51 Daniel:
Awesome. Well, it has been quite an interesting dot one. I really appreciate everything you've shared today. Is there anything else you want to add at this point? Otherwise we'll talk again.
1:57:10 Takuya:
Yeah, I already speak a role. Thank you. Alright, talk to you later. Bye. Thank you very much for a nice discussion.

Session 051.2, November 9, 2022

https://www.youtube.com/watch?v=hY_CajLpt9Q

Second participatory discussion on the 2022 paper "Canonical neural networks perform active inference" by Takuya Isomura, Hideaki Shimazaki & Karl J. Friston.

SPEAKERS

Daniel Ari Friedman, Takuya Isomura

CONTENTS

TRANSCRIPT

00:33 DANIEL FRIEDMAN:
Hello and welcome everyone. This is ActInf livestream number 51. Two. It's November 9, 2022. Welcome to the the active inference institute.
00:45 We're a participatory online institute that is communication, learning and practicing applied active inference. This is a recorded and an archived livestream, so please provide us feedback so we can improve our work. All backgrounds and perspectives are welcome and we'll be following video etiquette for live streams, head over active inference.org to learn more about participating in different institute projects. Alright, well, we're in ActInf Stream number 51.2. We're in our third discussion on the paper
01:26 “Canonical neural networks perform active inference,” from 2022. We had a Dot Zero video with some background and context and overview. And then last week in 51.1 we had a great discussion,

went over many interesting details of the paper and related topics. So today we're going to jump in, cover some empirical details, some implications, connect some more dots, maybe look at some code. And thanks again to Takuya for joining these discussions.
02:07 I'm Daniel, I'm a researcher in California and thought a lot over the last week about what this kind of neural network synthesis or translation really means, and just want to learn more about what fundamentals or foundational aspects of these different kinds of models enable that synthesis or translation. And then again what that means for areas where one or the other kind of model is already in use. So thanks again for joining and I'll pass it to you if you want to say hi or give any a second interpretation.
03:01 TAKUYA ISOMURA:
Oh yeah, I'm at RIKEN Brain Science Institute, Japan. So I look forward to discuss another different aspect of this work.
03:23 Daniel:
Well, let's just remind ourselves of the fundamental parallel being made in the paper and then we'll get to these two questions about kind of the two directions that things can go. One representation is in equation one with loss function of a neural network and the free energy on a POMDP. And that's also seen visually in figure one, with a neural network being drawn a concordance against the variational base of the action perception loop. So maybe just let's begin by restating. What is this parallel that is in equation one and figure one and how was it reached in this paper?
04:19 Takuya:
So basically idea here is that we derived to characterize the dynamics and activity of canonical neural network in terms of Bayesian inference, because arbitrary dynamics of neural network is interoceptive in the sense that we don't know what is the function underlying such a dynamics and what is the coherence of the self organization or activity. So once we translate that dynamics in terms of Bayesian, we can assign quantities in Bayesian for any biological quantities, which enables us to lend the explainability to the neural network dynamics and architectures. So that's a basic idea. And what we have done in this paper is that we consider a biological plausible cost function for this particular canonical neural network. And show the equivalence between that Costa function and the variation navy energy and the particular partially observable cognition process model.
05:57 Daniel:
Awesome. So let's look at the parallel between the cost function for neural networks and the informational free energy. So one representation of that was in figure three. So maybe could you just describe what is the structure of the informational free energy expression and what is the structure of the loss function? Okay, so there is a clear parallel between the functional structure or those component in informational free energy and component in neural network Costa function.
06:43 Takuya:
So let's say the first time in F correspond to the it correspond to the expectation about hidden states is a hidden states Australia. So that part basically indicates the free energy with respect to the hidden state. Yeah, and the second part correspond to the free energy about the decision posterior. So the indicates the posterior belief about agent decision or action. And now in terms of the correspondence between the free energy and neural network function here the first time in the neural network function correspond to middle layer neural activity which has a recurrent connection and receive sensory input

from sensory layer and then project the output to the output layer and the second term correspond to output layer which receive signals from middle layer and send the feedback response to the environment.
08:25 Daniel:
So both of us expressions have the first term being more like a cognition perceptual sensory learning term and the second term is more like a control theoretic action selection. And how did you see this analogy or concordance because it looks like a zipper, like everything is totally lined up.
08:58 Takuya:
Well, this graph itself showed a clear correspondence because now we are considering a particular form of on DP in which each element of hidden states takes either zero or one. But there are many states so it is expressed in a form of factorization. So now we consider that in terms of the s fosteria Bordeaux. Upper part of Bordeaux correspond to the expectation about each element of s taking one and lower part of the bordese correspond to the expectation about s taking zero. So it is broke vector about the posterior expectation and this nicely correspond to the Brea vector shown in the bottom up this figure it is a vector of x and Bijan sorry it is a vector of x and by x and here by x indicate one minus x in the element y sense which is exactly correspond to block vector or expectation.
10:50 This correspondence also observed in the second tab. Here, log S correspond to log X and also log A correspond to the broke matrix of W log W. Here W hat indicates the sigmoidal function of W and its bar means sigmoidal function of W. So actually, because we now consider binary hidden state and binary observation it's like reviewed mapping. Mapping from hidden states to conversation is expressed as block matrix, which is exactly correspond to broke matrix shown in the bottom of this figure.
12:04 So like this, for every tab we have the exact correspondence between the upper expression and the lower expression.
12:17 So that's why we can say that this is a natural mapping from neural network formation to parishional vision formation.
12:31 So it speaks a sort of identity between those two different expressions. So although 1 may be able to consider another mapping from neural network to Bayesian inference, this is a sort of simplest mapping.
12:57 Daniel:
So how would it look different if it were three states categorical distribution or a continuous distribution? What aspects would change? Thank you for asking that. So that's in some sense outside of this paper because only when we consider a binary hidden state, this analogy is established nicely. Otherwise we need to consider some attention.
13:38 Takuya:
So because consider that each neuron code the probability or expectation of some value taking one, then the probability or expectation of taking zero can be simply computed by computing one Ines neural activity. So actually neural activity which is a single dimensional variable is sufficient to express the expectation. Right? But once we consider the three state hidden states program, this doesn't work. So we need to consider at least two variables but it's relation to neural network expression is not very clear in general.
14:51 Daniel:

That's very interesting why it would be so strong of a concordance in a binary case but immediately unclear for other distributions. Yeah, generally for poem DB expression we consider the one hot expression, one hot vector expression which means that we normalize the value in the sense that the summation of all variable to be one. Maybe there is some neural substrate that achieve that communication. But for classic type of neural network like canonical neural network, consider in this paper what is that neural substrate is not very clear. So that's why we selected the binary case because it's simple and have a clear analogy.
16:14 So what does it mean for an artificial or for a biological neuron to have activity dynamics or plasticity context? That justifies it being described as playing like a belief role in a Bayesian setting.
16:41 Higher firing means more belief, higher firing means lower belief. What does it really mean to have a connection between in this episode of street talk belief states. I see, so if you assign a mapping, a particular mapping, then its meaning is also determined. In this case, we assign that neural activity correspond to the posterior expectation about an element of hidden states taking one. So once we define this mapping, then higher neural activity indicates the higher probability of taking one.
17:32 So neural activity is the probability neural activity is on the x axis and the y axis of the sigmoid function is the probability of taking one. Well. Neural activity encodes the expectation. So neural activity is sigmoidal function of something itself.
18:08 Takuya:
This is because once we see a fixed point of neural activity equation which is derived from this cos function, it has a form of sigma WADA function so x equals sigmawilder function of graph, graph, graph. So this form is exactly correspond to softwaremax function of something which is seen in the solution of possibly expectation.
18:46 Daniel:
That'S what the neural activity encodes and what is the Bayesian interpretation or the update rules on the plasticity. Okay, that's another important point. So in terms of posture of parameters so in the case of Bayesian force us we consider update about deleted parameters of a matrix and B matrix which is usually

expressed by the small case variable. And its meaning is that if we compute the partial derivative of a partial derivative of F with respect to small A then its solution it's fixed point solution looks like an computer product of which is also known as Hebbian product because it has an errors drink to update depending on the precinaptic neuron activity and postsynaptic neuron activity. And according to this formal equivalent we revisit we can see again such analogy in a formal sense here if we computer the partial derivative of neural network function with respect to W then we can formally derive the Hebbian prosthesis which depends on the activity of prey and postsynaptic neuro activity.
21:08 Okay, so hebion plasticity often described as neurons that fired together, wired together. Here you're discussing it in terms of a matrix operation on the POMDP side between observables and hidden states. So there's a hebion plasticity happening between the perceptual layer and the cognitive layer, right? So the first half of the neural network is trained according to heavy and plasticity rules that optimize the A in terms of the perceptual and learning like relationship between hidden states and observables. And then the second half of the neural network has a slightly different structure.
22:13 It is optimizing based upon retroactive re analysis of consequences of action according to the fictive causality construction.

22:33 Takuya:
So actually in this figure b up layer correspond to environment and lower part correspond to agent. So this structure corresponds to figure eight, this correspond to a simpler version of foam DP. So for version of POMDP, its corresponding neural network is showing figure four or this paper image task. This is the neural network architectures.
23:22 So as you say, there is a network connection from sensory layer to cognition layer which is expressed by W here and recurrent connection which corresponds to state transient matrix is expressed by K matrix which is recurrent Sinematic connectivity. And as you say the action generation through retrospective reward or risk evolution is done by output trigger through the synaptic connectivity expressed as V in this figure. So V is the synaptic connectivity between cognitive states in the middle layer and the action selection states in Y. Exactly. And so in that way V is exactly analogous to W.
24:25 Daniel:
But why and how does gamma come into play only in this second layer? I mean why not have gamma one in the first layer? Gamma two in the second layer.
24:38 Takuya:
Generally speaking, it is possible to moderate plasticity in first layer using another moderator gamma. But for complexity we focus only on neural modulation in the output layer. Analogy is that for example, as you said, the first rigor computer more perceptual things so perception of external world and instead on the other hand middle secondary which is mapping from cognition layer to action layer perform the optimization of its own action. So for example, in the story item in the brain action prediction is optimized by conversation of dopamine as a input. So usually that socket receives signal from ecological neural socket and send signal to another neuronal nucleus in meat grain.
26:21 But the point here is that neuron in storatum encodes some decision for examples goal or no goal. So such a decision is encoded. So now we consider analogy between pond DP expression in the Bayesian formation and neural socket in the brain that optimize action through some sort of moderation by another factor. Here that factor corresponds to gamma and gamma has variety of function. But in this paper we focus only on the moderation of activity.
27:26 So here the behavior activity is not determined by only a preposter relationship but determined by three factors relationship in the sense that the activity is updated by the product of gamma and prayer and postsynaptic activity. So there are three times in one.
27:58 This is why this comma can moderate prosthesity.
28:06 Daniel:
So how would a glial factor look different computationally? And where in the brain have people identified levels or other factors as relevant for learning? Yeah, that's an interesting point. I'm not really sure about the equation of the real moderation of neural activity or plasticity. There are many discussions and I'm sorry, I don't know the exact form, but one possible implementation is similar to this type of neuromodulation.
28:54 Takuya:
So it would be possible to model some real contribution or free factor to plasticity in the form of three factor learning room which is mathematically speaking Lamme as this type of neural moderation.
29:23 Daniel:
Here in table two we have another set of correspondences. It's like a sideways figure three, right? But a

little bit more like a dictionary.
29:40 Anything to add? Or any variables that we haven't really mentioned. What about the firing thresholds? Because these are common parameters in a neural model, however, we don't really hear about the interpretation of these constructs within the variational base POMDP. Yeah, there is an interesting story.
30:17 Takuya:
That's a very interesting point. So when we first tried to make analogy between neurons network and one program is the law of threshold factor because as you said, it is not absorbing POMDP structure. But there is another factor in Pompey which is prior expectation about hidden state which is usually expressed by D matrix. And what we consider is the relationship between d matrix and firing threshold. And finally, what we found is that firing threshold is not equal to the matrix itself, but it is a summation of B matrix under some function of synaptic strength or which is equal to a matrix b matrix in the POMDP formation.
31:46 In other words, what we found is that each which is a firing solution in neural network architecture, it is actually an adaptive threshold which is not a fixed value, but h is a function of W sinatic strengths and h changes depending on W's value. For example, if W is too Laje, then your activity can be unstable. So h behavior to reduce the activity to make neural activity more stable. So we can see an analogy of omeostatic mechanism here.
32:55 If we design A as the function of w and function of another factor which is all part of the term in this table, then we could make common analogy between this h and some variable in palm DP correlation which is shown in the right hand side of this table. Although its value is not simple because it chaos three different tasks. So all of them contribute to make h or M.
33:48 But anyway, once we map, so once we establish a mapping between h and this value, then everything works. So the cost function in different settings have Omar correspondence.
34:14 Daniel:
H and the M firing thresholds. So H correspond to middle rare M indicator output raider threshold which are different variables. And interestingly h correspond to prior expectation about hidden states because it corresponds to community rare and M correspond a priori belief about its own action because it is a bias in the action layer. Yeah, it's very interesting that the perceptual firing threshold h only includes prior beliefs on hidden states, beliefs about how observations map to hidden states A and beliefs about how hidden states change their time B. So that's like pure passive inference.
35:27 And then the firing thresholds for M correspond to only beliefs about preferences and beliefs about actions or habits with C and E. So there's like a complete division of labor or partitioning functionally between these structurally different parts of the neural network and structurally different and functionally different parts of the POMDP. Yet they're integrated in unified loss functions or unified imperatives.
36:10 And so it's like there's extreme separability of perception and action on both sides of the figure one divide, but also they're integrated, but they're separate. And that's what kind of grants it the best of both worlds because if they were any more integrated you couldn't really pull them apart. And if they were any less integrated then the imperative, the loss function or the variational free energy would be ad hoc and unprincipled. But there's kind of a middle ground where they have a principled integration but still a distinguishment.

36:58 Takuya:
Right?
37:01 This is caused by network structure defined or it is because the structure of Bayesian network defined in the MDV model. So both of them define the causal relationship between elements or quantities.
37:40 Its substrate is not important, so it's relationship is crucial to determine the cost function or it's a fixed point in this context. So that's why we can see the data analogy.
38:04 Daniel:
Well, there's a few technical points I think we can now go into and then there will be some more general points about applications and intelligence. So first the code availability statement. Awesome to see that the MATLAB scripts are available and also active on Zinodo. So here is the GitHub repo for reverse engineering. Do you want to give any overview descriptions of what people can expect to see in this repo?
38:41 And also what about using MATLAB?
38:50 Why did you use MATLAB? What advantages or limitations do you see in MATLAB?
38:58 Takuya:
So, because this is a very simple simulation, so Macrab is sufficient to encode the whole script.
39:14 We also try some implication in the material. See here, if you run the script, then you can see the process of an agent solving the maze task.
39:39 Daniel:
What did they do in the maze task? So here the aim of this agent is to reach the right hand side of this maze. Because this is a typical example of the rate moderation task. That's why we select the main task. So to achieve this next task, is it required

to make some plan to be able to select a good decision?
40:19 Takuya:
Because without planning, you may encounter the war and cannot go further and you may fail the image. But with learning, it is possible to see the path to reach the right hand side of this space.
40:46 Daniel:
So does it know its exploration?
40:53 Takuya:
Yeah, it received a state from neighboring eleven times eleven Jelle. So which is shown in the bottom part. Yeah, this left figure C indicates the observation. So eleven times eleven state around the agent position. Okay.
41:22 Now agent is on the right hand side of image at the goal position and it observes our neighboring state.
41:36 Daniel:
Well, a few interesting things here. It's looking off the right end. Yeah. And it has this kind of periodic belief in the key distribution. Why?
41:55 Takuya:
I think it is because when the agent is in the middle point of maze, then all neighboring state is in the maze. So there is a path and there is a wall. So this makes some ergodicity because mazes have some structures and actually have a periodic structure and only at the goal position, then right hand side

becomes war. But it is not common for this agent. This is because this agent show such a priori pattern.
43:04 Daniel:
Yeah, the streets are one wide and they tend to be separated by one. So we see this periodicity. What is the numbers in this middle bottom plot and what does the checker board represent? Yeah, hered correspond to possibility, expectation about active states and decision. So middle indicate decision posterior and decision.
43:39 Takuya:
Here we characterize decision as a secret of four step actions. So each action correspond to a movement to right or left or up or down. And we consider a four step sequence of that option which is expressed as D. So it has four power, four possibility. 256.
44:23 Yeah, 256. So this is a protein on XY coordinate because in the middle panel, middle point correspond to the current position of agent. And with four step movement, agent can go one of any current position and the current brightness corresponds to the expectation about the agent decision. Well this is very interesting. If we just were to think about you're at a point and you can go up, down, left, right, you have four moves.
45:16 Daniel:
Naively it sounds like, well it should look like a gaussian blur. Most of those should cancel out and then it should become rarer and rarer monotonically. But actually you start in the middle, you can't end up on these white squares because it's like one, two, three and then you have to leave. Right? So it's kind of like horses in chess or other pieces where actually their embodiment, it's very unexpected that you can't in four moves end up next to where you began when you can be so much further.
46:06 And then we see this kind of like embodied inferential prior with QS that embodies regularity beliefs about the width of the road and the separation of the roads. And then there's these like embodied action priors and real consequences that have to do with the structure of movement. So what it's doing? It's thinking about policies of length four. There's 256 policies of length four.
46:44 There's some degeneracy because there's obviously not 256 squares here. So while only one policy is going to take you up, up, down, down, other squares are reachable. Like the center square is probably the mode because it can be reached at least a handful of ways.
47:11 And then at each time point it's basically saying okay, I know where my X position is and given my local eleven by eleven view, I'm trying to plan to go right.
47:34 And then here through time in the simulation here it starts at 30 something, it quickly figures out how to get to about 40 and then it's kind of going up and down on 40. But it can't really break out because all of these bottom four routes are closed. It has a breakout and then very quickly it hits another plateau around 60. Right. Then it kind of has a very nice breakout and in just a few steps goes very far.
48:15 So what is dopamine doing? Or how is Dopamine helping it in the plateau and then to break out of the plateau? Yes. So this agent learned this particular structure through many trials. So before training it failed to reach the goal, but after training it achieved such a nice behavior.
48:52 Takuya:
So to active this, the role of domain is that we design gamma function such that if the agent can move rightward with some distance during some time limit, then risk becomes small like say comma equals zero nor risk situation in that sense. In that case, this agent updates synaptic weights through hebion

frostbust. But if the agent failed to go rightward with some distance during a limited time frame, then gamma becomes large like zero six which is larger than the average zero five. Then we design that drawing in attention antihebian prosthesis occurred instead of Hebbian. So antihabion indicates the works as the disassociation between the current state and current decisions.
50:20 Because the current decision does work, it's not good decision. So we try to make the agent who will get that particular decision rules through conversation of heavy and plasticity done by Dharma factory. So this can be an arrow to the Dopamine moderation heavy and activity. So if the policy is resulting in the expected outcome, gamma stays at .5, the policy is as risky or consequential as expected, and then the policy can either go better than expected, which facilitates learning to support that decision. To be made more or the outcome of the policy can be worse than expected, which disassociates previous conceptions to discourage that kind of behavior.
51:26 Exactly. Crucial point is that this association with different time frame in the sense that we consider multiplication of current risk and least decisions to average over past two present Hebbian product. This makes an association between past decision secrets and the current risk which enables to optimize decision to minimize the future risk. It is just a safe time frame.
52:11 Daniel:
So here risk is being used in a formal sense similar to how it's used in economics which is the associated uncertainty of outcomes with respect to a policy. Where does danger come into play? Like what if there was an adversary in the maze or something that was dangerous? How does this kind of model accommodate or hunger or different kinds of competitions? Because right now it's basically just trying to diffuse right word with a bias.
52:53 Takuya:
Right? But how do different kind of situational elements become interface into the generative model and generative process? Okay, any of those factors can be involved in risk factor, a single risk factor. So you can arbitrary design and risk factor because risk factor moderate generative model. So that's why agent try to minimize the risk through basic embryo updating.
53:40 But the risk itself is in some sense outside of such a Bayesian framework. So we can design arbitrary risk. So it may involve some danger factor, any other factor.
54:03 Daniel:
And this simulation, it is a POMDP or it is a neural network. And what scripts might we look at to understand the structure of the maze agents? Okay, it is basically expressed using the quantity in home depp for tractability. But for example, if you see the MDP learning probably okay, there is a variable Lamme type in the definition of SIM type correspond to the type of simulation. So if it's one or two it becomes homo DP or neural network to my understanding.
55:21 Takuya:
Well, in this particular example, Jelle, we use the let's say maybe it's not good example that DeForest is learning the deforesting this script as well. So maybe another as an example, let's see.
56:09 Daniel:
What is MDP init is initiating the markup decision process. Exactly. It's just determining the initial state of the computer generative model fe compute variational free energy or risk MDP computer risk function. So basically we use the neural network structure computation in this particular setup. So when you click maze m then in the line 31 line we determine that Lamme type is two.

57:08 Takuya:
This correspond to neural network architecture. So there is a very slight difference between home depicture and neural network architecture because assuming neural network architecture correspond to, you know, considering considering okay, well, if you choose the palm DP architectures, then we sometimes use the gamma function to computer the posterior expectation about parameters. But in the neural network modeling the gamma function doesn't appear but it is replaced with the logarithm of some function. And simply speaking, the difference between the gamma function of something and the logarithm function of something is asymmetry Lamme. So that's why we can transform home DP two neural network architectures.
58:48 When the number of samples is sufficiently large.
58:57 Daniel:
Which form do you expect performs better under small or large amounts of data?
59:07 Takuya:
Well, for large amount of data they work in the same manner. Same manner. For small amount of examples, I'm not truly sure but it corresponds to assumptions about the posterior belief distribution. So if you assume delicious distribution then your resulting function form is something that used the gamma function in terms of basic inference probably which is optimal.
1:00:00 Daniel:
All right, let's return to the earliest questions from today. So in your script, which people can reference, there's basically a toggle between having it in SIM type one or SIM type two corresponding to the POMDP in the neural network. What about if there's a published neural network or POMDP? How can we use this architecture to create a translation?
1:00:48 Is there any difference in this? Kind of like translating models

in the wild different than the full construction of a special script that can speak both languages?
1:01:10 Takuya:
Well, in terms of script there's no difference, sympathetic difference. Right. So they work in the same manner. So only a translation of variable the same source code in two different ways. So if you see that this is a neural network generation, then it is translated as a neural network.
1:01:55 Or if you see that this is a POMDP, then it's POMDP.
1:02:03 Daniel:
So for some neural network being used in an industrial setting, how would we get from the neural network to a POMDP? And where or how would that representation be valuable? Right? So when neural network in the different architectures the important point is that we consider a particular form of neural network which is called canonical neural network architecture. So only when we assume this crossover neural network then you can find the exact correspondence to a particular form of POMDP.
1:02:55 Takuya:
Otherwise you need to establish another equivalent between another form of neural network architectures and some sort of Bayesian model.
1:03:12 This may be expressed by POMDP, but maybe not so straightforward to be expressed as the computational AP architectures. So what is it about the canonical neural network architecture that facilitates its translation into the POMDP form? Yeah, first of all, it assumes sigmoid or activation function. It is nicely correspond to enthralpy time in the force DP equations from DP formulations. So

that's why we can clear marketing.
1:04:00 So yeah, in other words, simply speaking, they have the same nonlinearity. That's why this translation is very easy with another nonlinearity or neural network equation, then we need to find another type of entropy equation or another type of prior distribution. It is very nontrivial. How does one even go about doing that research?
1:04:49 If you want to go that direction, then I think the first step is to find the prior brief, which makes the prior brief and find the equivalence between a particular neural network architectures and particular Bayesian model.
1:05:25 Daniel:
This sigmoidal activation is interesting. It corresponds to general patterns seen in psychophysics, like two objects that are the same weight. You're going to have a chance of saying that one is heavier and then initially the difference has the most returns on that decision being made accurately. And then as it crosses some threshold where it just is beyond a noticeable difference, the decision becomes essentially probabilistic, like the firing curve becomes saturated, the neuron chaos, a very low belief about zero or very high belief about zero or one flip that.
1:06:26 So there is a nice grounding of that kind of a sigmoidal response curve with respect to stimuli differences and it has of course, tractable analytical properties, but it also just happens to be a good response summarizer. Yeah, you're right. So sigmoidar function is also known as a psychometric function, as you say. We observe that characteristic in many psychical experiments. And the previous work also said that even at the single neuron level, neuronal level, the same behavior were absorbed, which means that each heuristic we can reobserve the similar property, which is sometimes called as neurometric function, which is which have the form of sigmoida activation function.
1:07:43 Takuya:
So it is nice reason to design neural network architecture using a sigmoid or function.
1:07:58 Daniel:
All right, let's cover a few questions in the chat from Dave and then in the end turn to some general thoughts. Okay, this was when we were looking at figure three. So you described these vectors or matrices. What kind of matrix or vector did you describe? The mass block matrix.
1:08:34 Takuya:
Block matrix.
1:08:38 Daniel:
Block rock learning what? Okay, rock matrix of rock. Vector is a vector vector or matrix of matrix. So imagine that.
1:09:02 Sorry, just zoom, just glitch just repeat the last piece. Okay, well, broke matrix Dean that the element of matrix is a matrix. So let's say two by two matrix like matrix in the ear pointing. So this here W one hat is a matrix and W zero hat is another matrix. And combining four matrices, we define a single block matrix.
1:09:46 All right, thank you. So Dave then asks the hosts of Machine Learning Street Talk Number 67 with Karl Friston. Another podcast, Karl Friston has spoken. Pressed Karl Friston on why is it so important that most of the values in a generative model matrix assume values of exactly zero?
1:10:17 Why is it important that generative model matrices are sparse? Why?
1:10:30 Takuya:

I'm not Bull sure. I think there is some context before that point. I think on that particular situation, then, yeah, as you say, the many elements in the matrix or gentle model should be zero, but I'm not sure if it's a general statement at all. What do you think about compressed analyses on sparse matrices?
1:11:08 Daniel:
Is that a useful technique or direction?
1:11:25 Takuya:
You can use that knowledge to construct model, so you can use that knowledge to make more accurate inference. So in that sense, generally speaking, such assumptions should be useful. For example, yeah, as you said, it would be possible to use some sparse prior to restrict the value of parameter. Like, it is in principle same as assuming some L one norm to design the distribution. To design the prior distribution, which is mass Dutch speaking.
1:12:18 Exactly same as considering Lasso regression. Yes. So we've explored a little bit how from a canonical neural network to a particular form of a POMDP, gives us some semantics and interpretability around the dynamics and plasticity of the neural network. What do we gain by taking a stated POMDP generative model and deriving an analogous neural network?
1:13:03 Daniel:
Do we gain access to efficient computation, new software packages, different applications?
1:13:15 Takuya:
Well, if one use Home DP and one's goal is so design an efficient basic model, then I think your Home DP expression is sufficient. So you don't need to consider neural network architecture, probably because Homedippy architecture and Bayesian correlation is designed to achieve some sort of mathematically optimal inference and decision making. Right? So it itself is optimal scheme. But if one need to consider a link between Bayesian inference and biological substrates, then this mapping is crucial.
1:14:28 Simply speaking, we consider that we assume that a brain perform Beijing inference, but its substrate is still unclear. So we need to link the Bayesian quantity to biological quantities. So this mapping, this equivalent, helps us to its translation. Right? So when you start from on the model, then this translation facilitates the process of finding its neuronal substrate.
1:15:09 So once you translate that to a basic to neural network quantities, then it facilitates the experimental validation application to reality. So if its modeling is apt for a particular neural network neural circuit architecture, then it should provide some prediction about the architecture or dynamics of the empirical data. Right? So first we start from Bayesian model, which is not necessary to be equal to empirical data. So there is some mapping, but it's mapping is not straightforward.
1:16:10 We may have multiple mappings, but once you translate Bayesian model to a particular neural network architectures then mapping or relationship between applicable data to such a particular neural network model is straightforward. So it helps us to apply base to an explanation of MP card data.
1:16:46 Daniel:
So is it fair to say that neural networks have found wide recent application because they facilitate statistical learning in cases where the inference problem has not been a priori well specified? One can just have a folder of images and a list of labels and just say, here's the data. Run it through this architecture or this architecture Explorer. And so with this concordance we gain new interpretation into those settings that kind of arose from ill specified inference problems. And then on the other hand, for

problems that we already have well specified in terms of a POMDP generative model of a particular form, we gain the connection to actually implement it with empirical data and bring it into relevant industrial settings.
1:18:07 What systems or phenomena are promising to continue research on the image example obviously is a simple case, but are you continuing research into more advanced computational agents? Robotic animal.
1:18:36 Takuya:
Oh, hello can design a more sophisticated agent which performs some difficult tasks based on canonical network neural network but there is some limitation, clear limitation on that direction. Right? So yeah, I should emphasize that across the canonical neural network which correspond to a particular home DP is much smaller than general home DP framework. So there are some limitations, a least of limitations. So if one's goal is designed and sophisticated DAGs to perform some task or control robot, then one direction is just forget such limitations and take the mathematical optimality right and another direction is barrelscale probability.
1:19:53 So if one wants to image some agent which is barely possible, then this correspondence is crucial because it tells us biological limitations through the existence, no existence of such a mapping between POMDP and particular neural network architectures. So, yeah. So it would be useful to consider a vertical substrate to achieve exafferent task.
1:20:42 And that task would be related to high dimension image processing. Image recognition or sound recognition, such as multimodality, can be inbox or decision. Can be higher dimensional like in the mainstays, we just consider

the four directional movement, but it can be extended to higher dimensionality, like arm movement, body movement, so on, so on. So in principle it can be extended in that direction.
1:21:30 Daniel:
Which directions or questions are you excited about? Or what areas of studying the basis of biological and computational intelligence are relevant? Yes. So in terms of the importance of canonical network, as you said, virtue is a biological probability. So it would be nice that if we model some task which is conducted by learning and one already recorded some neural activity, then we design a task which is exactly same as the task which is done by the animal and then compare the simulated agent and Mpcar data to discuss about the similarity or difference between the simulated agent and the animals.
1:22:50 Takuya:
That would be very interesting direction of research.
1:22:56 Daniel:
Yeah, and if there could be some unexpected prediction or explanation in the computational agent, that would bolster the relationship. And then one other aspect is it would help with the reproducibility and the documentation around behavior studies if the computational agent were preregistered and someone said we've already done the statistical power analysis and we've already explored with parameter sweeps how many observations we need to make of the two armed bandit. How many observations of the three armed bandit should we do? Three mice 100 times or 100 mice three times? Those are the total substance of designing research programs.
1:24:02 And so having a formal representation of behavioral tasks that are being studied will help us design behavior observations and experiments that aren't simply ad hoc.

1:24:20 Takuya:
All right, that's an interesting application.
1:24:26 This framework helps to design the experimental setup itself. And what we often consider is the prediction ability of these modern canonical neural networks to predict the Jelle phoneization or dynamics of the VR neural network in the animal during the learning process. So in place for it is possible to predict the behavior after learning based only on data in the initial stage because once we obtain some empirical data, then we can fit that data to design a particular canonical network. And canonical neural network makes some correlation through a minimization cost function which is exactly same as the Bayesian belief updating under a particular guarantee model. So which means that its dynamics goes through the shortest path on the free energy landscape which means that we can make some quantitative prediction about the sinatic trajectory or neural activity or any kind of parameters.
1:26:00 So we demonstrated that using in virtual neural network and uploaded some footprint recently. So at the stage, at least at the level in vitro network, which is much simpler than VR brain, we could predict the self organization of in virtual network using this canonical network architecture and this support the probability of free energy principle because this canonical network predict the communication through the variational free energy minimization and its solution. Its results have a very tight correlation between correlation to Archer synchronization.
1:27:07 Daniel:
That's a very interesting experiment. So what animal were the neurons from and what was measured about these neurons? Yes, so that in vitro network is obtained from blood embryo, we use cortical cells to make that individual network and task is sort of causal inference task which can be designed in the form of OMDP. So imagine that we usually simulate agents that receive signals generated by OMDP generative process and process and Beijing task. So we just replaced that Bayesian agent to neural network.
1:28:07 Takuya:
So we stimulate neuron with some signal which is made by some hidden sources through like a human mapping and question is whether in viral network can info the hidden states through some communication and they can they could Ines the hidden culture.
1:28:35 Daniel:
What does it look like functionally when the neural network has succeeded at inferring the hidden causes? Yeah, the direct conversation is done by the response number response spikes to a particular pattern sensory input. So again, we can see a clear correspondence between activity level and posterior belief about hidden state. So here we see re a book response to an electropaste memory. We see the response from ten to 13 milliseconds after each stimulation and we compute the number of spikes and that spikes changes their preference in the sense that some neurons learn to preferentially respond to force one but not source two.
1:29:42 Takuya:
So which is not a response to input itself, but it looks like a response to particular source. So it is inference which is empirical evidence that neural network actually perform some sort of causal inference in a manner consistent with traditional Bayesian inference. And then we compute another quantity in Bayesian inference in the real vertical data. We show that firing special factor is consistent

with the prior belief about hidden states and we also compute the synaptic rate statistically through some connection strength estimation method and show that the estimated synaptic strength is consistent with something encoding posterior belief about parameters, as expected by the theory. Well, we looked at table two earlier and this is almost like the next step after the theoretical concordance is all right, well, let's measure the release of a neurotransmitter or the empirical synaptic strength or the firing threshold or all these different features in different empirical systems.
1:31:28 Daniel:
So what experimental systems does your group work in?
1:31:42 Takuya:
That in virtual system was made when I was a PhD student. So that is experiment we done in my previous route and now I COVID to the Bijan Institute and I'm a Princepal investigator of Celery unit. So now actually we don't use any experimental setup, so any experimental bidding is down with some collaboration. So although I cannot say detail about that exploration. But yeah, we learn some collaborating work about the implication of salary using various animals.
1:32:37 Yeah, so we hope we can show some interesting results following results using animal data.
1:32:49 Daniel:
Very interesting. Yeah, well, it speaks a lot to the stage that our field is in in certain ways where we've seen a lot of graphics that are suggestive. This paper and the building on the previous 2020 paper made a suggestive possibility much closer to an analytically demonstrated translation and then took the next step incrementally into the in silico agent. And so it's only natural to then explore different embodied systems as well.
1:33:50 Are there any other sections that you wanted to like, look at or highlight or any other topics about the paper or adjacencies or active inference that you think are interesting to go into?
1:34:10 Takuya:
Yeah, okay, I would like to mention about some implication of these papers, which is not the director discussing the papers. So for example, well, we focus on a discrete state space model. So we avoid to assume some subscript that encodes the coherence of the distribution. So once you assume homedp then it is categorical distinctions. So it is different from assuming Gaussian distribution characterized by me and variants.
1:34:56 So the neural substrate of variance is still unclear and we now try to figure out that. So this is one direction of limitation and another implication is that thanks to a simple ODP structure in this paper, we don't care about the hierarchical optimization. But generally it is crucial to update parameters through hierarchical optimization through some backpropagation like computation. Although it is unclear whether back propagation itself occurred in the real brain. But we still have some alternative that achieve such optimization and it's neural substrate still unclear and this paper doesn't address that direction.
1:36:25 Daniel:
Another area I'm wondering about is like where in neural structures is the learning reflected? Where is the function and learning reflected? Well, sometimes it has to do with not just structural Tweaking, but the presence or absence of synapses. So obviously this model does not expand into synaptogenesis synaptic pruning, let alone neurogenesis and neuro allostasis which we mentioned in the previous discussion. But understanding how these larger scale structural changes which are certainly important

in biological systems become reflected in artificial neural networks and then how that translates all the way back to P-O-M.
1:37:21 DP and then whether we could go the other way. What kinds of POMDP structures in their neural network realization would have structural modification. Like you COVID imagine a POMDP that does structure learning but the neural network doesn't have structural change. Or there's A-P-O-M DP that doesn't do structural learning but it's manifested by a neural network that does have a structural change element. So structure is doing something very different in these two different categories of model.
1:38:00 And then also even within neural firing, which is different amongst different species and so on, there's different aspects of what that firing is that would have different implications for the actual biological substrate of cognition niche. The simple connection is firing rate to posterior belief. Average firing rate, no change in posterior. Reduce the firing rate if the posterior should be going down. Increase it if it should be going up.
1:38:40 Or maybe there are neurons that have a flipped valence but the same type of relationship, but there's other firing patterns like spike time dependent plasticity synchronization amongst different brain regions. There's

a lot of things that don't change the rate overall. That again from the biological systems we know that those phenomena and mechanism are important for different cognitive processes.
1:39:17 So there will be many many years of a fruitful relationship.
1:39:25 I'm going to bring in this picture that Alexandra had taken. Maybe we need a third panel in figure one because these three systems moving between them is going to be the substance of the field for a long time and there may be other edges to build. But understanding how artificial neural networks intermediate between the empirical measurements and manipulations that we can make of real neural systems and the interpretability and factorizability of POMDPs it might be a bridge too far to go from the POMDP to the neuron. You could always use this technique but it would be a purely descriptive statistic type approach.
1:40:42 But it's so interesting that by intermediating through a formal connection established in figure one I Dean in equation one but also shown here, then we can kind of extend the chain of exploration, prediction, control, design all the way on through. And that just unlocks like an incredible amount of neuroscience that hasn't been formalized mathematically and an incredible amount of generative models that have been specified for different learning settings sometimes even by analogy to biological settings. But the metaphor remains just a metaphor until it's possible to intermediate with this type of neural network development. Yeah, that's a crucial point.
1:41:56 Takuya:
It is easy to imagine that phenomena can be modeled using very realistic neural model or free model synaptic model. Right.
1:42:14 We believe that it is possible and then Laje model is not necessarily tractable, not necessarily useful because it's too much complicated to analyze something. So we use some reduction, usually mathematically speaking, which correspond to topological transformation to make the model simpler. And then we need to consider the translation of that simplified neural network model because neural network model itself is not explainable which just represents some dynamics and its functional meaning is not clear. But thanks to the Bayesian framework we have a very nice event framework to

least the experiment ability. And this translation, this correspondence helped us to link such a phenomena base equation modeling and functional base equations.
1:43:32 Daniel:
Yeah, one paper from 2017 that was much discussed by some could a neuroscientist understand a microprocessor? And this group with Jonas and Courting, they had a simulation of a microprocessor from an earlier video game console, I believe. And then using the analogy of like the transistors and their connections as neural firing and structural connectivity they were able to simulate experimental settings, input and action and then make measurements from every neuron including doing lesions and loss of functions and so on. And it turns out that a lot of the techniques that are used to derive scientific explanation from analogous data collected from a biological system, those techniques which ostensibly should be isolating functional explanations in fact did not isolate effective exploration. You could have a deletion over here that induces some statistical change all the way over here and that may or may not be a useful clue towards the function of even subcircuits.
1:45:04 And so I think that was a wake up call with respect to the interpretability of simply this connection between the biological and the neural network. This connection alone is of limited applicability even when the neural network model becomes so complex as to recapitulate the biological phenomena, you're never under any guarantee that you're going to recover interpretability. You may have just created an atomic level simulation of the phenomena, but of course, a map that is the same scale as the phenomena isn't a map. It's just a copy that has no more interpretability.
1:46:06 And it's almost like what is now extended again, as we kind of just summarize this and think about how we move forward, is that connection can now be extended into the space of interpretable causal models and the generalized Bayesian graphical computational frameworks and all the heuristic that we can then use like variational, Bayes and all these other methods. So it'd be interesting to look back at different data sets of in vitro and in vivo and in silico neural activation, especially if the task was of this constrained set of POMDPs and it was already amenable. Because, as you brought up, other settings would require a little bit more theory development before we understand what POMDP family would be applicable.
1:47:19 Takuya:
Cool. Well, do you have any final thoughts or questions?
1:47:27 Well, do you hop.
1:47:34 Daniel:
I can download the MATLAB scripts and generate the figures, play around with a few of these parameters? Like, I see that you can change how far the entity can see. And then with these models, I'm also always curious about the computational complexity. Like if you extended the planning horizon from four to five or you dropped it down to three, what is the runtime consequences and what is the performance consequences? And where might we be able to use single or swarms of really simple agent, maybe even making binary decisions and achieve high performance?
1:48:32 And where do we really need to move into these large combinatoric spaces in order to solve problems and the kinds of complex planning problems that we solve, whether it's planning our day or planning our week, are those force like true Deep Harrison planning problems with extensive consideration of counterfactuals and calculation of alternatives? Or are those actually composite decisions that are made up of smaller, simpler, sub decisions that we may or may not have flexibility to

restructure?
1:49:31 So that a decision, a complex chess maneuver, a sacrifice in chess or another game. It may be possible to model that as a Deep Horizon scan or a kind of intuitive heuristic for an appropriate skilled entity.
1:50:01 Takuya:
For this particular two Brea structure, there is a clear limitation about the horizon forward because it doesn't use forward prediction, it used post action approach. So it's a clear limitation, but still, it may have some performance ability, active, some levels of affordance. That provides even another way to look at planning. The two ways I was describing planning, as it's often described in the PMDP literature is again, is it a true Deep Horizon consideration or is it just short term heuristics or nested models that are short? And I think that this paper says maybe neither.
1:51:07 Daniel:
Maybe it's purely the active causality on the past that leads to the emergence of sentient and maybe even teleological planninglike behavior through the ongoing reconsideration of the consequences of least action. But it's neither a short nor a long term planning challenge. It's actually like a memory and learning challenge. And no planning occurs. Right.
1:51:42 Takuya:
Indirect planning planning element is involving C matrix. Planning as a phenomenon occurs and derisking through time occurs. But it says something quite interesting and deep that that phenotype or function could be enacted by a system that explicitly looks a long way ahead, explicitly looks a short way ahead or moves forward and looks backwards only, right, which is what they sometimes say about the past and the future. So that may be a very biologically plausible form of learning. And it's already intimately connected with the dynamics and the activity in terms of an integrated loss function.
1:52:46 Daniel:
So these are all excellent directions to keep learning on. Right? And I'm also interested in the barrel implementation of such a short term or long term for the prediction and running. And I hope to find some nice connectivity to such implementation of Bayesian motor and implementation in VR brain network. Cool.
1:53:21 And also I'm always curious about the invertebrate brain as an ant researcher. And so many of the brain architectures as well as the brain architecture that people discuss are mammalian centric, which makes sense. The mammalian cortical column and the relationship with Dopaminergic midbrain and the cortical regions and the spinal reflex arc those are all important systems of interest. Yet the micro and meso anatomy of the invertebrate nervous system is pretty distinct. So our model should be able to describe neural and cognitive systems, of course, across invertebrates and vertebrates.
1:54:19 So I look forward to also seeing what those models of the invertebrate nervous system and collective behavior where you could have some type of backwards looking risk inference of the swarm. Who knows?
1:54:43 Takuya:
Well.
1:54:46 Daniel:
We really appreciate the time that you took for these discussions. I think they are immensely important. And we wish you the best of luck in these continued directions.

1:55:03 Takuya:
Yes.
1:55:06 Daniel:
Okay, that's it. Thank you. Thank you very much. See you next time. Bye.