

# Curiosity, Imagination and Information

Applied Machine and Deep Learning 190.015

**Vedant Dave** 

October 2023

Chair of Cyber-Physical-Systems



### Outline

- What are our Questions?
- Imagination-augmented agents
- Relevant feature selection in Conscious brain
- Curiosity-driven learning
- What future holds?



# What are our Questions?

#### How?

How can I represent all the data I see?

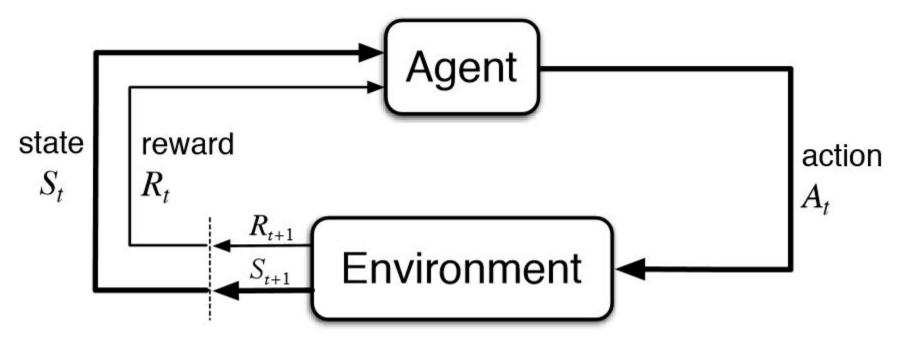
#### What?

What should I actually represent from all this mess?

### Why?

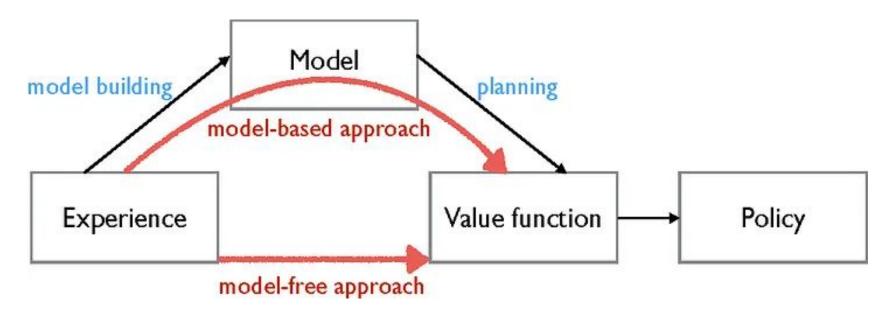
Why should I do anything at all? Is it worth doing anything if there are no rewards?

## Reinforcement Learning



Source: Sutton, R. S., Barto, A. G. (2018). Reinforcement Learning: An Introduction. The MIT Press.

## Model-free and Model-based learning



Source: https://medium.com/analytics-vidhya/model-based-offline-reinforcement-learning-morel-f5cd991d9fd5

Applications of Advances in Monlinear space of North N An On-Line Algorithm for Dynamic Reinforcement Learning and cations of a collection summarizes the minimum cost. The ar An Example of Self-Learning in Neural Networks Cations or a collection or algo optimization at minimum cost, social values of a cost. optimization at minimum cost. The at tric estimation on timization are in a cost. The at the cost of t Planning in Reactive Environments rensitivity analysis of models, new w following setting stimation, optimization, new w equesting "setting setting sett By Derrick Nguyen and Prof. Bernard Widrow a boat trailer will realize this. Normal driving in-Jürgen Schmidhuber\* Dept. Stanford University a poat traiter will realize tima. A great deal stincts lead to erroneous movements. A great deal Forecast Analysis and Evaluation : A, Room 7413, Department of Energ stincts lead to erroneous movements. A great deal of practice is required to develop the requisite skills. . Informatik practice is required to develop the required sound a  $PILCO:\ A\ Model-Based\ and\ Data-Efficient\ Approach\ to\ Policy\ Search$ -k, one often observes t Lacking again, g Reinforcement Learning in Markovian and the desire Data-Efficient Reinforcement Learning in Continuous-State POMDPs Non-Markovian Environments gineering, University of Washington, USA HINGTON.EDU Abs Rowan McAllister of Cambridge, UK Carl Rasmussen CAM.AC.UK Learning deep dynamical models from From Pixels to Torques: Policy Learning with Deep Dynamical Models Jürgen Schmidhuber RTM26@CAM.AC Institut für Informatik CER54@r Technische Universität M Niklas Wahlströn \* Thomas B. Schön \* Marc Peter Deisemroth \*\* Arcistr. 21, 8000 München NIKWA@ISY.LIU.SE ods that Department of Electrical Engineering, Linköping University, Sweden,
 (e-unuit: universities line.se) schmidhu@tumult.informat tre more (e-mail: mksou(hisp.lin.ar.)

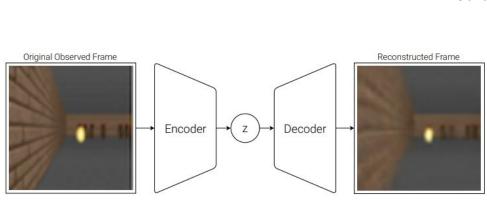
\*\*Department of Information Technology (typnale University, Sgreden, e-mail: thomas-achon grid.un.se.)

\*\*\*Department of Commutation. Immersal College London.TiK. THOMAS.SCHON@IT.UU.SE model lea mation This work addresses three problems with reinforceme Division of Automatic Control, Linköping University, Linköping, Sweden vironment. 2. On-line learning has interfaces Linköping of Automatic Control, Linköping University, Linköping, Sweden realization. Abstract Any single uch as tive neuro-control: 1. Non-Markovian interfaces betw. Thomas B. Schön vice. 2. On-line learning based on system the control of information of the control of (e-mail: thomas schonfiri un.se). London, UK, the Department of Computing, tmperial Coalege (e-mail: m.deisenroth@imperial.oc.uk) iust one of M.DEISENROTH@IMPERIAL.AC.UK odelwironment. 2. On-line earning based on system realization and on two interacting fully described which may learn.

Niklas Wahlström

Nikla Abstract: Modeling dynamical systems is important in many disciplines, such as control, noted for a neurolecturology. Commonly the state of these systems is tool directly observed. ed the obse Abstract: Modeling dynamical systems is important in many disciplines, such as control, and the state of these systems is not directly observed, the said of these systems is not directly observed, but only available through noisy and potentially high-dimensional observations. In these cases, valued adaptive critics. An algorithm is described which have performed in parallel. Problems with the parameter of computer that the parameter of Computer of the parameter of the param from er the numb roboles, or neumotechnology. Commonly the size of these systems is not directly observed, but only analishic through using and potentially high-dimensional observations. In these cases, the control of the size of the second control of the size of but only available through noisy and potentially high-dimensional observations. In these cases are selected through noisy and potentially high-dimensional observations. In these cases, and the selection of the control of the contro inising data-e works which mad on two interacting fully securised which made attacked by a plearn in parallel. Problems with part continuation of Computing, Imperial College London, United Kingdom attacked by any learn in parallel. Problems continuated which made plearing and advantage of the continuation of Computing, Imperial College London, United Kingdom attacked by approximation of Computing, Imperial College London, United Kingdom interacting fully rectarged which made provided the continuation of Computing, Imperial College London, United Kingdom interactions of Computing College London, United Kingdom interactions of College London interactions ffiidentification i.t., fluding the measurement mapping and the transition mapping (symmetry) in latest space can be challenging. For linear system dynamics and measurement of the symmetry of t mation, (3) take new information into account for learning the low data works which may learn in parallel. Problems with para attacked by 'adaptive randomness'. It is also described by igent and adaptation. Effectively, any fully autonomous system mappings efficient solutions for system identification are available. However, in practical applications the linearity assumptions those not hably requiring nomlinear system [dentification are light-dimensional [e.g., mages], undilized system for additionally the observations are high-dimensional [e.g., mages], undilized system. attacked by adaptive randomness. It is also described it model/controller systems can be combined with vector-va and adaptation. Effectively, any ruly autonomous and adaptation. Effectively, any ruly autonomous and adaptation. Effectively, any ruly autonomous without traditional in this to close this perception-action-learning loop without traditional in the conditional in the conditional in the condition of the condition als or applications, the linearity assumptions does not hold, requiring nonlinear system identification techniques. If additionally, the observations are they followed the problem of nonlinear system indentification indicated the problem of nonlinear system indentification is inherently hard. To suffere the problem of nonlinear system indentification is inherently hard. To suffere the problem of nonlinear system in deep learning and several linear tension of the problem of nonlinear system in the problem of nonlinear system in the problem of nonlinear system in the problem of nonlinear system is not several advances. has to close this perception-action-learning non-modifical a relying on specific human expert knowledge. The pixels relying on specific human expert knowledge appears of House 2011) identifies key aspects of House 2011 identifies key aspects of identification is inherently hart. To address the problem of nonlinear system themification from high intramsional observations, we combine econt advances in deep learning and system to high intramsional observations, we combine econt advances in deep learning and system high intramsional observations, we combine econt advances in deep learning and system is the problem of nonlinear positions of the observation in the problem of the problem of nonlinear positions of the problem of the problem of nonlinear positions of the problem of the pro to forques problem (Brock, 2011) identifies key aspects of from high-dimensional observations, we combine recent advances in deep jearning and system is identification. In particular, we found a few dimensional embeddings of the observations are identification. In particular, we found the seminance of the observations are interested in the seminance of the observations and a predictive transition model in this localization and the seminance of deep auto-encoders and a predictive transition model in this localization and the seminance of the observations and a predictive transition model in this localization and the seminance of the observations are seminanced in the seminance of the seminance of the observations are seminanced in the seminance of the observation are seminanced in the seminance of the observations are seminanced in the seminance of the observation are seminanced in the seminance of the obse an autonomous system: autonomous thinking and decision blentification. In particular, we jointly learn a low-dimensional embedding of the observation by some of deep auto-encoders and a predictive transition model in this low-dimensional specimension of degraciant transitions and the forest predictive transition model on this low-dimensional specimens of degraciant transitions and the forest predictive models of dynamical spaces. Abstract Data-efficient learning in continuous state-action **L** algorithms making using sensor measurements only, intelligent explo-INTRODUCTION spaces using very high-dimensional observations cs, they will sa remains a key challenge in developing fully fficiency bene making using sensor measurements ration and learning from mistakes. At a given time, an agent with a non-Markovian interface to its environ autonomous systems. In this paper, we con-We consider the problem of learning closed-loop policies rords: Deep neural networks, system identification, nonlinear systems, low-dimensional dilutes auto-encoder. we consider the promises or searning crosed-noop poincies ("torques") from pixel information end-to-end. A possible At a given time, an agent with a non-Markonian interface to its environ derive an optimal next action by considering its current input only. The autonomous systems. In tims paper, we con-sider one instance of this challenge, the pixwe cussion of the land of the derive an optimal next action by considering its current input only. In described below differs from previous reinforcement algorithms in at els to torques problem, where an agent must only available sensor information is provided by a camera. learn a closed-loop control policy from pixel in-Learning nonlinear dynamical models is an inherently dif-ficult problem, and it has been one of the most active fould problem, and it has been one of the most active arms in nonleast identification for the last devades (I must arms in nonleast identification for the last devades (I must described below unters from previous reinforcement algorithms in at of the following issues: It has a potential for on-line learning and non-A formation only. We introduce a data-efficient, i.e., no direct information of the robot's joint configuraficult problem, and it has been one of the most perfect of the most perfect perfect of the most perfect perfec re- area in system identification for the last decades (Linux, and Montal Control of the last decades (Linux, and Montal Control of the Linux, and Linux, We demonstrate that our mix from pixel information only. model-based reinforcement learning algorithm tion is available. The objective is to learn a continuoustion is available. The objective is to tearn a community with the control of the that learns such a closed-loop policy directly from pixel information. The key ingredient is in this continuous environment in a data-efficient way, i.e., High-dimensional time series include video streams, elec-tropy of the series include a deep dynamical model that uses deep autour uns communus environment in a suns-emerien way, i.e., we want to keep the number of trials small. To date, there 2011) see also the recent survey (Kantas et al., 2015).
While methods based on SMC are powerful, they are also computationally expensive. Learning unitnear dynamic also computationally expensive. High-dimensional time series include video atreams, educations of the series received to the series of the series to the series of the series encoders to learn a low-dimensional embedding is no fully autonomous system that convincingly closes troosee-platography (EEG) and sensor network data. Di-naminal models describing such data are desired for for-ter than the sensor of the sensor of the sensor of the naminal models describing such data are desired for forof images jointly with a predictive model in this the perception-action-learning loop and solves the pixels nomical models describing such data are desired for forther the controlled design, both of which the controlled design, but the controlled design, but the controlled design des low-dimensional feature space. Joint learning to torques problem in continuous state-action spaces, the casting (prediction) and controller design, both of which play an important role, e.g., in antenuments system complete translations, colored and attractions explications. All the colored translations role of the colored translations of the colored translations of the colored translations is governed the colored translations. ensures that not only static but also dynamic chine translation, robotics and surreillance applications, for challenge is system identification. In the challenge is system identification in malescented model of the dynamical season based on malescenters. properties of the data are accounted for. This A promising approach toward solving the pixels to torques A loy challenge is system identification, i.e., finding, a manufacture of the dynamical system based on the dynamical system based on the information provided by unsummental from the uniis crucial for long-term predictions, which lie at natural domains in robotics. unabranatical model of the dynamical system based on the information provided by measurements from the unthe core of the adaptive model predictive control strategy that we use for closed-loop control. Compared to state-of-the-art reinforcement MONTANUNIVERSITÄT LEOBEN CYBER-PHYSICAL-SYSTEMS

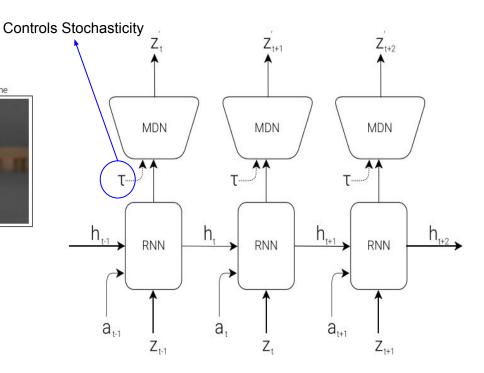
## World Models



Variational Autoencoder

### Policy

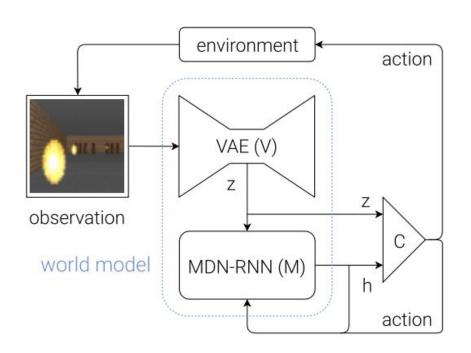
$$a_t = W_c \left[ z_t \ h_t \right] \ + b_c$$

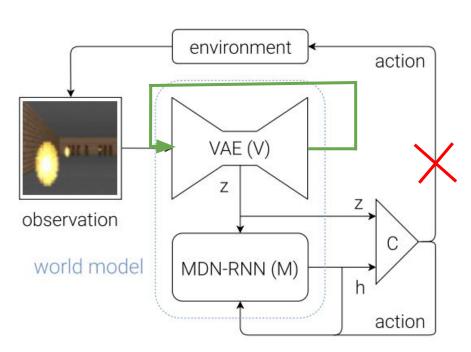


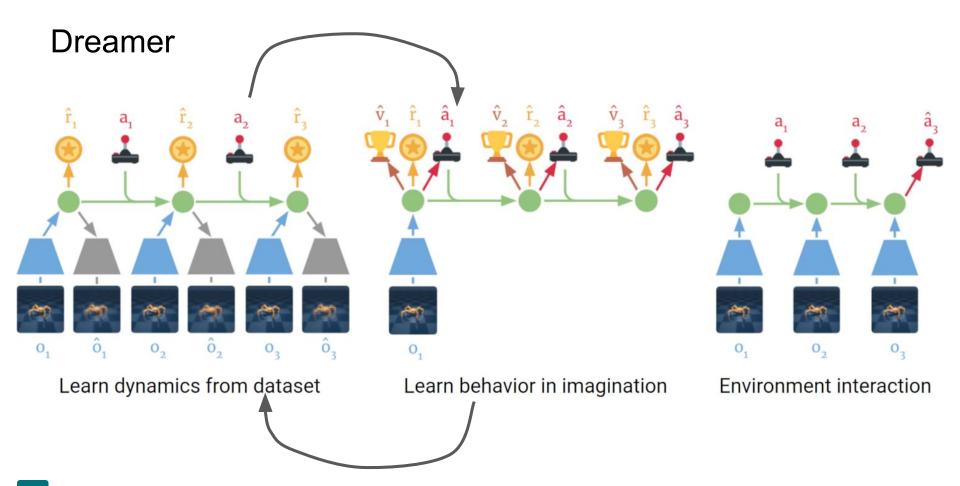
**Transition Model** 

### **World Models**

#### https://worldmodels.github.io/







## Dreamer



## Dreamer-v2











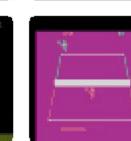


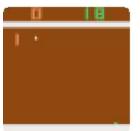










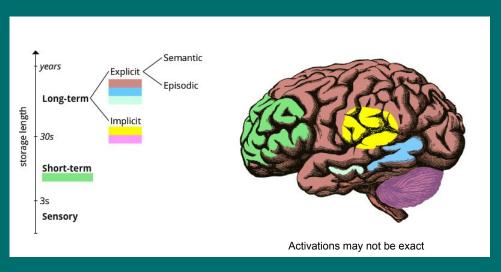








# Relevant Feature selection in Conscious Brain



https://gohighbrow.com/the-brain-and-memory/

• Short-term Memory: Holds seven, plus minus two, pieces of information [Miller]



Numbers are much better and Relevancy matters [Jacobs]

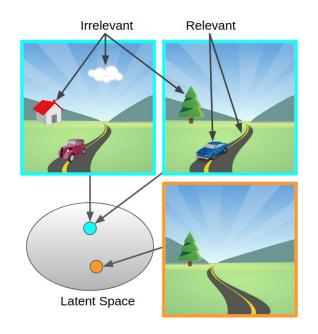
- Do we need so much information to solve a task? (<u>Selective Attention Test</u>)
- How to select relevant features from this noisy world? What is even meant by "relevant"?
- Can we imagine noise-free worlds from noisy observations?

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97. <a href="https://doi.org/10.1037/h0043158">https://doi.org/10.1037/h0043158</a>
Joseph Jacobs, Experiments on "Prehension", Mind, Volume os-12, Issue 45, 1 January 1887, Pages 75–79.

## • One of my inspiration



## Research in relevancy prediction



### **Deep Bisimulation for Control**

Bisimulation Metric: Measures similarity between similar states

Zhang, Amy, et al. "Learning invariant representations for reinforcement learning without reconstruction." arXiv preprint arXiv:2006.10742 (2020).













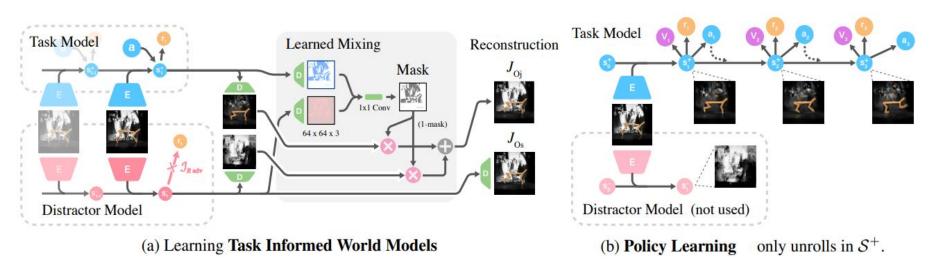


## **Some Assumptions**

- Used same video for test and train
- No RGB background
- Used Ground Plane

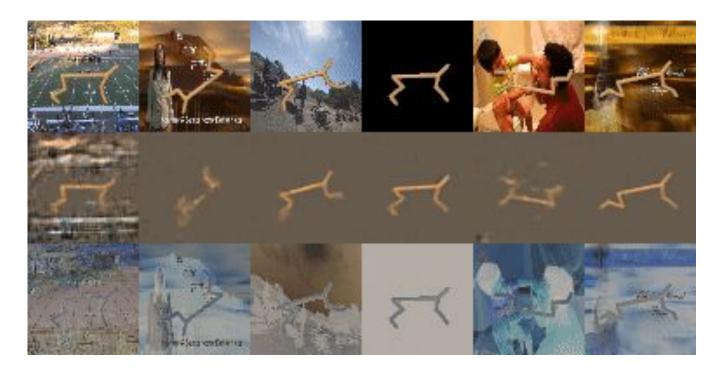
## Research in relevancy prediction

### **Learning Task Informed Abstractions**



Fu, Xiang, et al. "Learning task informed abstractions." International Conference on Machine Learning. PMLR, 2021.

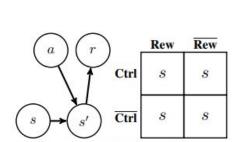
### **Learning Task Informed Abstractions**



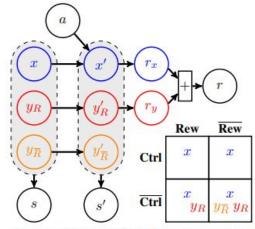
Fu, Xiang, et al. "Learning task informed abstractions." International Conference on Machine Learning. PMLR, 2021.

## Research in relevancy prediction

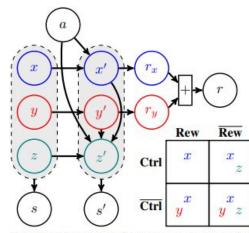
#### **Denoised MDPs**



(a) Transition without useful structure. s may contain any type of information.



(b) Transition that factorizes out uncontrollable information in  $y_{\overline{R}}$  and  $y_{\overline{R}}$ .



(c) Transition that factorizes out uncontrollable y and reward-irrelevant z.

Tongzhou Wang, Simon S. Du, Antonio Torralba, Phillip Isola, Amy Zhang, and Yuandong Tian. Denoised mdps: Learning world models better than the world itself. In International Conference on Machine Learning. PMLR, 2022.

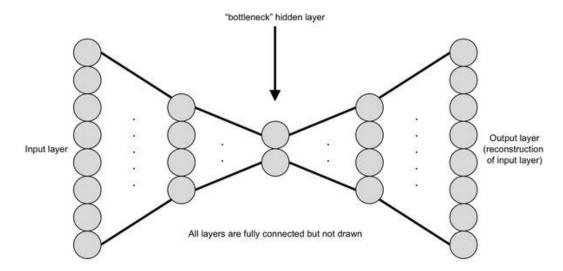
#### **Denoised MDPs**



Tongzhou Wang, Simon S. Du, Antonio Torralba, Phillip Isola, Amy Zhang, and Yuandong Tian. Denoised mdps: Learning world models better than the world itself. In International Conference on Machine Learning. PMLR, 2022.

### **Denoised Predictive Imagination**

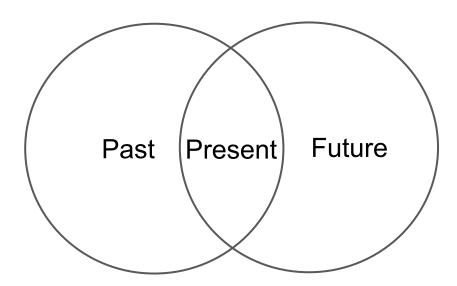
Information Bottleneck Principle



Tishby, Naftali, Fernando C. Pereira, and William Bialek. "The information bottleneck method." *arXiv preprint physics/0004057* (2000).

### **Denoised Predictive Imagination**

Predictive Information

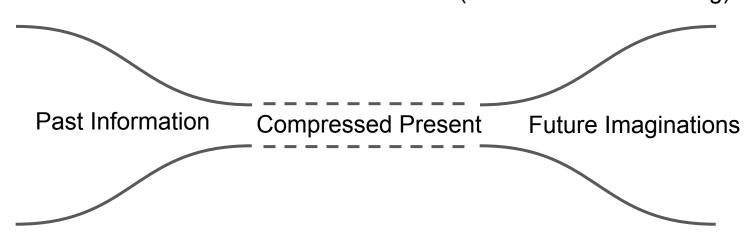


Bialek, William, and Naftali Tishby. "Predictive information." arXiv preprint cond-mat/9902341 (1999).



### **Denoised Predictive Imagination**

 Information Bottleneck + Predictive Information + Soft Actor-Critic (Reinforcement Learning)



Submitted to International Conference of Learning Representations (ICLR) 2024



### **Denoised Predictive Imagination**

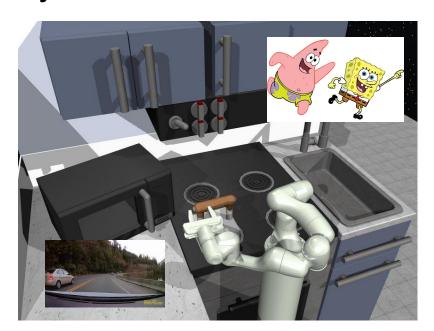


Submitted to International Conference of Learning Representations (ICLR) 2024



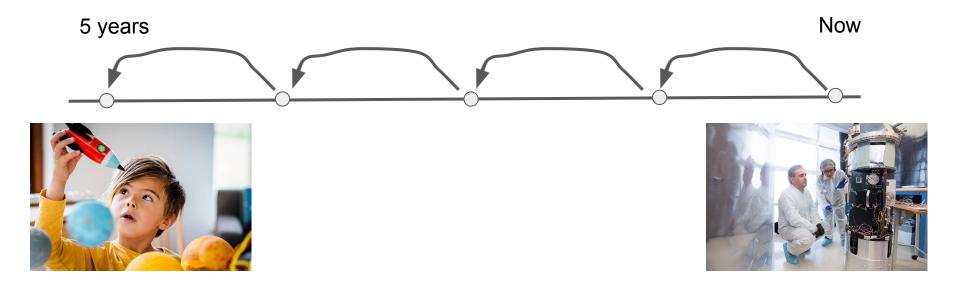
## New Ideas

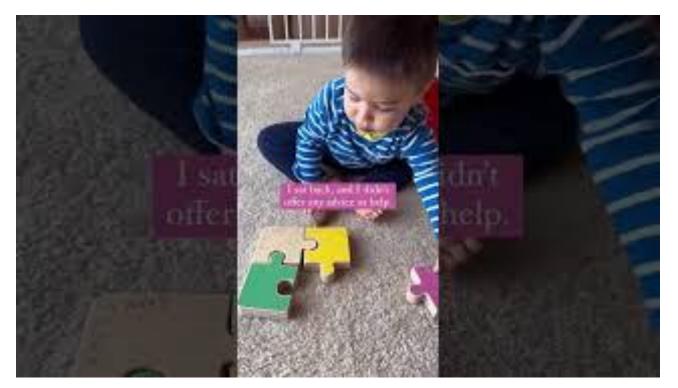
## Why not Attention?





Done by ChatGPT-4

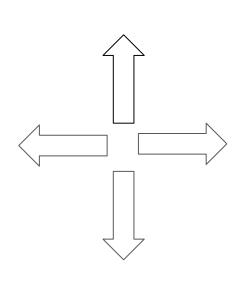


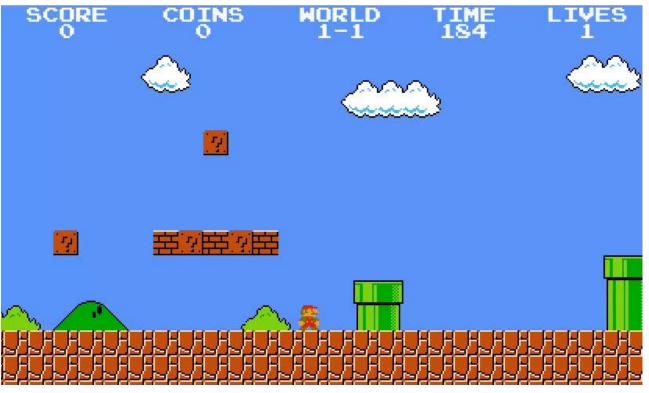


Silvia, P. J. (2012). Curiosity and motivation. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 157–166). Oxford

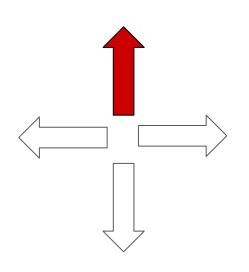


How to formulate Curiosity?

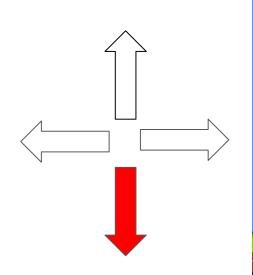




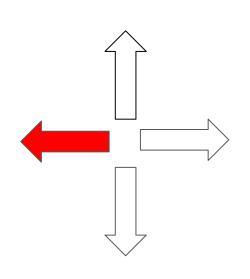
Source: <a href="https://supermario-game.com/de">https://supermario-game.com/de</a>







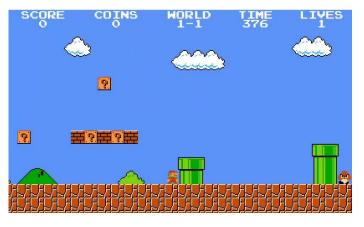


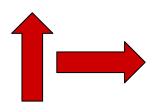




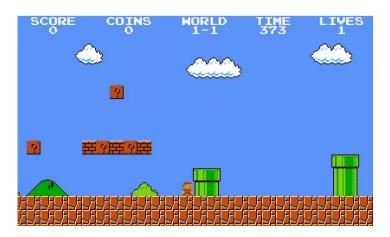
#### **Expected Behaviour**

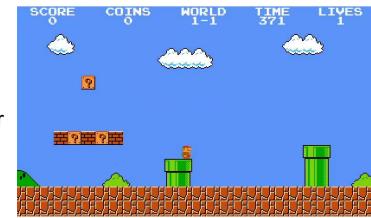
#### Curiosity = Prediction Error





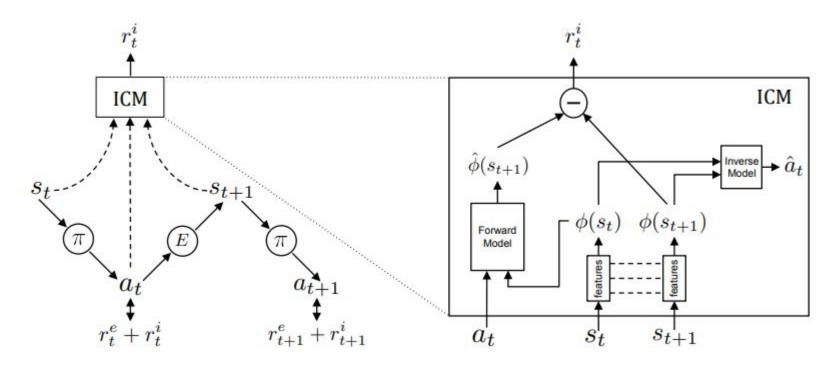
**Actual Behaviour** 



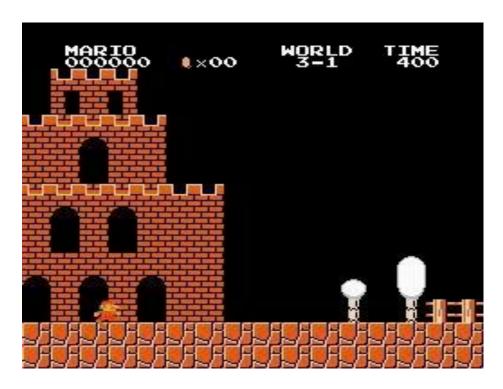




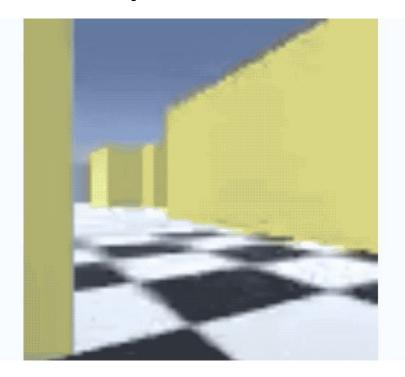
Pathak, Deepak, et al. "Curiosity-driven exploration by self-supervised prediction." *International conference on machine learning*. PMLR, 2017.



Pathak, Deepak, et al. "Curiosity-driven exploration by self-supervised prediction." International conference on machine learning. PMLR, 2017.



Pathak, Deepak, et al. "Curiosity-driven exploration by self-supervised prediction." International conference on machine learning. PMLR, 2017.





Burda, Yuri, et al. "Large-scale study of curiosity-driven learning." *arXiv preprint arXiv:1808.04355* (2018). Pong, Vitchyr H., et al. "Skew-fit: State-covering self-supervised reinforcement learning." *arXiv preprint arXiv:1903.03698* (2019).



### What future holds?

- Better representation in Reinforcement Learning
- Learning with Causal Models
- Long-term Planning
- New Formulations on Curiosity

### **Credits**

From Deep Learning of Disentangled Representations to Higher-level Cognition (<a href="https://www.youtube.com/watch?v=Yr1mOzC93xs&t=1529s">https://www.youtube.com/watch?v=Yr1mOzC93xs&t=1529s</a>)

Curiosity: Autonomy Talks - Deepak Pathak: Robotics In The Wild: Continual Improvement by Watching & Practicing (https://www.youtube.com/watch?v=mBkbnQoknr8)