



# 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?

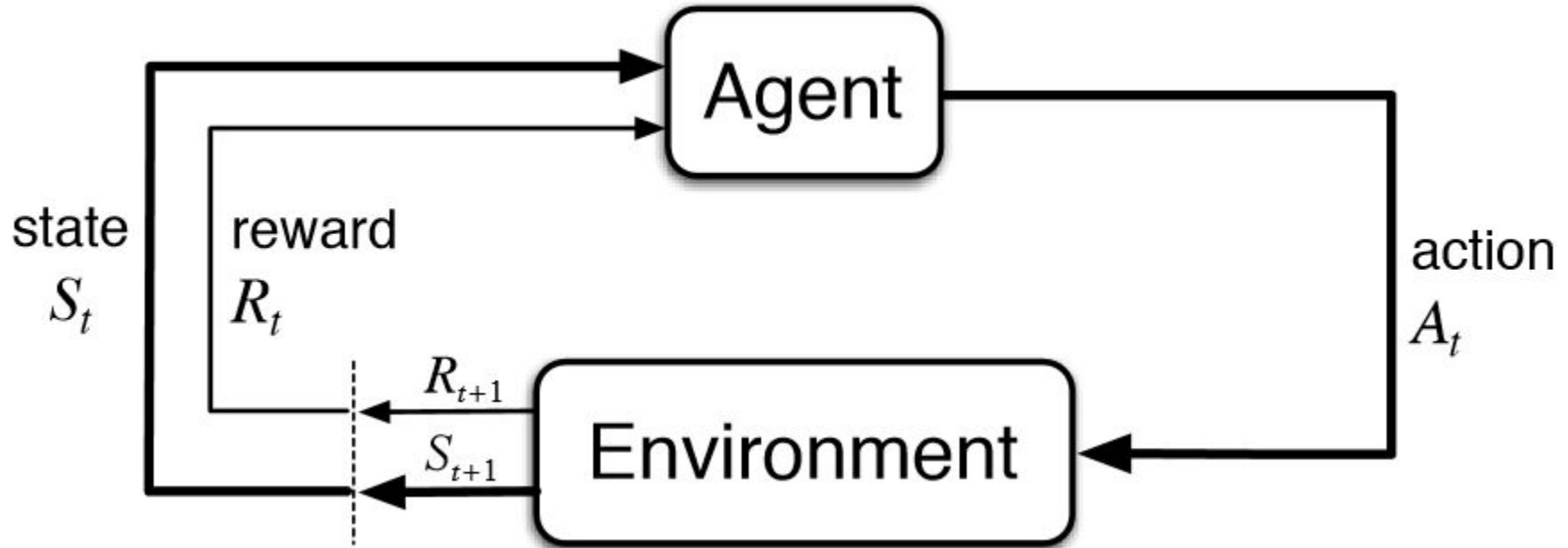
- **What?**

What should I actually represent from all this mess?

- **Why?**

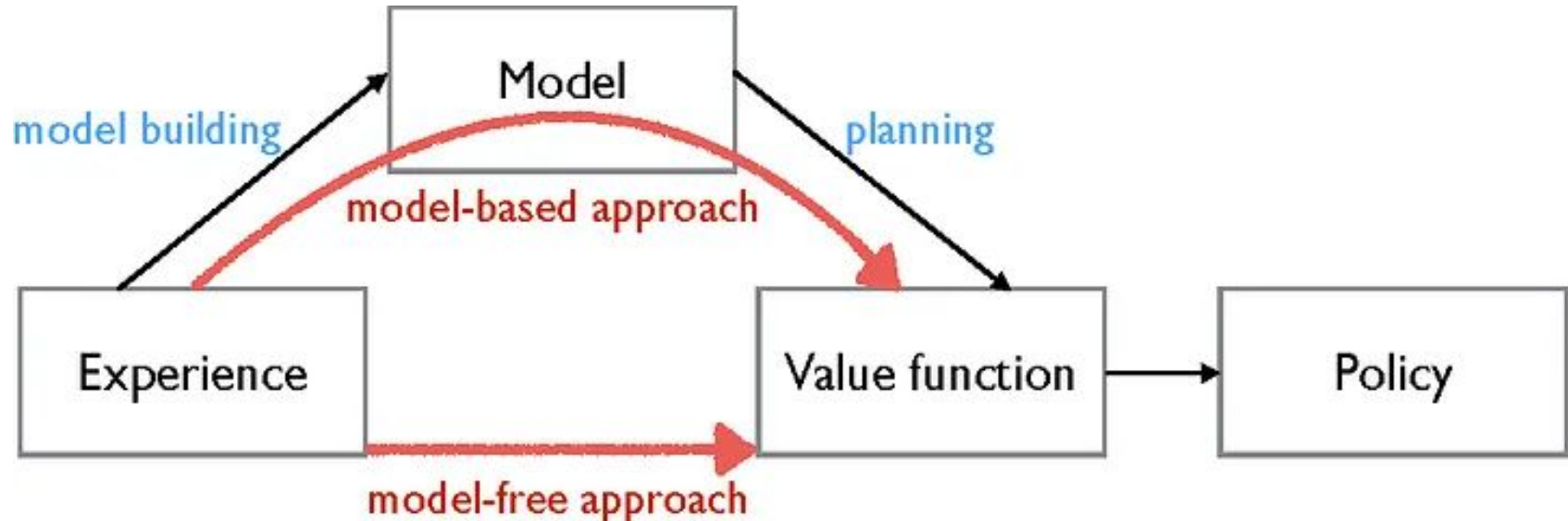
Why should I do anything at all? Is it worth doing anything if there are no rewards?

# 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>

## The Truck Backer-Upper: Self-Learning in Neural Networks

By

...Bernard  
Engineering  
University

a boat trailer will realize this. Normal driving in-  
structions lead to erroneous movements. A great deal  
of practice is required to develop the requisite skills  
of watching a truck driver backing toward a  
backing.

...k, one often observes  
backing again, g  
the desire  
...wa

\_\_\_\_\_

Data-F

# Reinforcement Learning in Markovian and Non-Markovian Environments

**Jürgen Schmidhuber**  
Institut für Informatik  
Technische Universität München  
Arcistr. 21, 8000 München  
schmidhu@tumult.informatik.tu-muenchen.de

## Abstract

**Abstract**

This work addresses three problems with reinforcement learning in a dynamic environment: 2. On-line learning based on system realizations and on two interacting fully recurrent continuous-time models/controllers. It is also described in terms of 'adaptive randomness'. Problems with parallel model/controller systems can be combined with vector-valued adaptive critics (previous critics have been scalar).

# 1 INTRODUCTION

At a given time, an agent with a *non-Markovian interface* to its environment derive an optimal next action by considering its current input only. The described below differs from previous reinforcement algorithms in the following issues: It has a potential for on-line learning and non-

# Data-Efficient Reinforcement Learning in Continuous-State POMDPs

Rowan McAllister  
Carl Rasmussen  
Department of Eng'

From Pixels to Torques: Policy Learning with Deep Dynamical Models

**Niklas Wahlström**  
f. Autom

Niklas Wahlström  
Division of Automatic Control, Lund University, Sweden

Thomas B. Schön

Department of Information  
Deisenroth

## Abstract

[illegible]

mation, (3) take new information into account for learning and adaptation. Effectively, any fully autonomous system has to close this perception-action-learning loop without relying on specific human expert knowledge. The *pixels to torques problem* (Brock, 2011) identifies key aspects of an autonomous system: autonomous thinking and decision-making using sensor measurements only, intelligent exploration from mistakes.

We consider the problem of learning closed-loop policies ("torques") from pixel information end-to-end. A possible scenario is a scene in which a robot is moving about, by a camera, only available sensor information is provided by a camera, i.e., no direct information of the robot's joint configuration is available. The objective is to learn a continuous-valued policy that allows the robotic agent to solve a task in this continuous environment in a data-efficient way, i.e., we want to keep the number of trials small. To date, there are no fully autonomous system that solves the pixels in the perception-action-learning loop and solves the pixels in the torque problem in continuous state-action spaces, the torques problem in robotics.

A promising approach toward solving the pixels to torques

Applications of Advances in Nonlinear S-  
Paul J. Werbos, U.S. Depart-  
Forecast Analysis and  
The following paper sum-  
s of a collection  
ation at  
vity.

The following paper summarizes the results of a collection of algorithms for optimization at minimum cost. The article presents a sensitivity analysis of models, a heuristic estimation of models, new modeling. The details, references, and requests for "Sensitivity Analysis and Forecast Analysis and Evaluation" should be sent to Dr. J. Werbos, Room 7413, Department of Energy Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.

# An On-Line Algorithm for Dynamic Reinforcement Learning and Planning in Reactive Environments

Jürgen Schmidhuber\*  
Informatik

---

# PILCO: A Model-Based and Data-Efficient Approach to Policy Search

---

gineering, University of Washington, US

of Cambridge, UK

Learning deep dynamical models from  
image pixels  
Thomas B. Schön\*\*

Niklas Wahlström\* Thomas B. ...  
Marc Peter Deisenroth\*\*\*

\* Department of Electrical Engineering, Linköping University,  
(e-mail: nikwa@isy.liu.se)  
Information Technology, Uppsala University, Sweden  
(e-mail: thomas.schon@it.uu.se).  
Serial College London, UK,  
(e-mail: ...@ac.uk)

\*\*\* Department of Computing, Imperial College  
(e-mail: m.deisenroth@imperial.ac.uk)

[illegible]

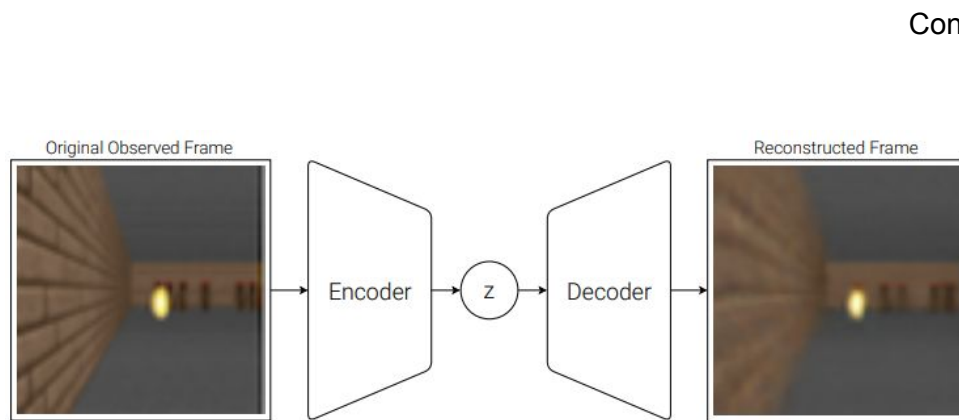
**Keywords:** Deep neural network, word embedding, auto-encoder

**Keywords:** Deep neural network, auto-encoder, embedding.

High-dimensional time series include video streams, electroencephalography (EEG) and sensor network data. Dynamical models describing such data are desired for forecasting (prediction) and controller design, both of which play an important role, e.g., in autonomous applications such as machine translation, robotics and surveillance. Finding a mathematical model of the dynamical system based on the information provided by measurements from the

Learning nonlinear dynamical models is an inherently difficult problem, and it has been one of the most active areas in system identification for the last decades (Ljung, 2010; Sjöberg et al., 1995). In recent years, sequential Monte Carlo (SMC) methods have received attention for identifying nonlinear state-space models (Schön et al., 2011); see also the recent survey (Kantas et al., 2015). While methods based on SMC are powerful, they are also computationally expensive. Learning nonlinear dy-

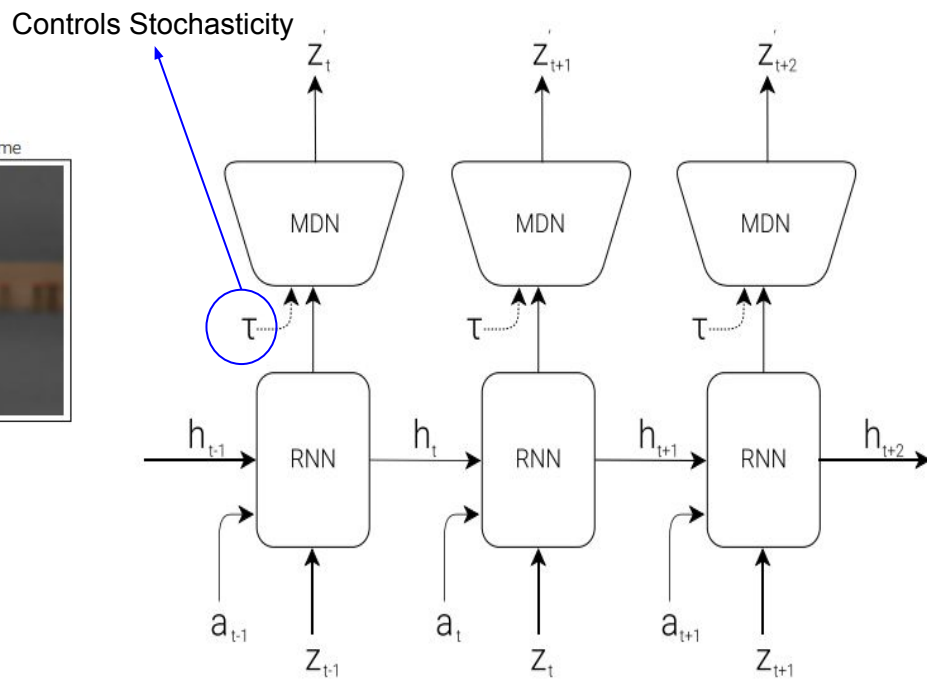
# World Models



Variational Autoencoder

Policy

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

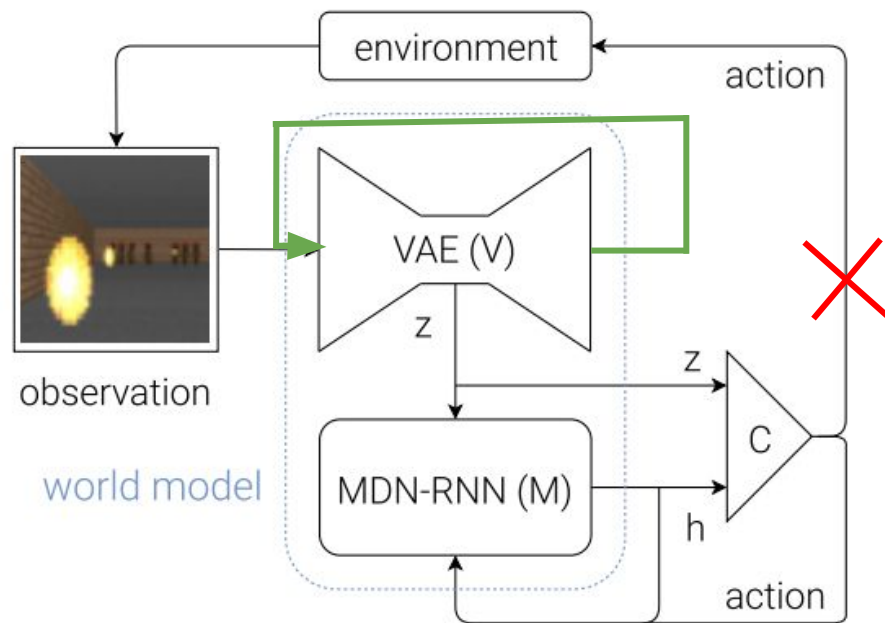
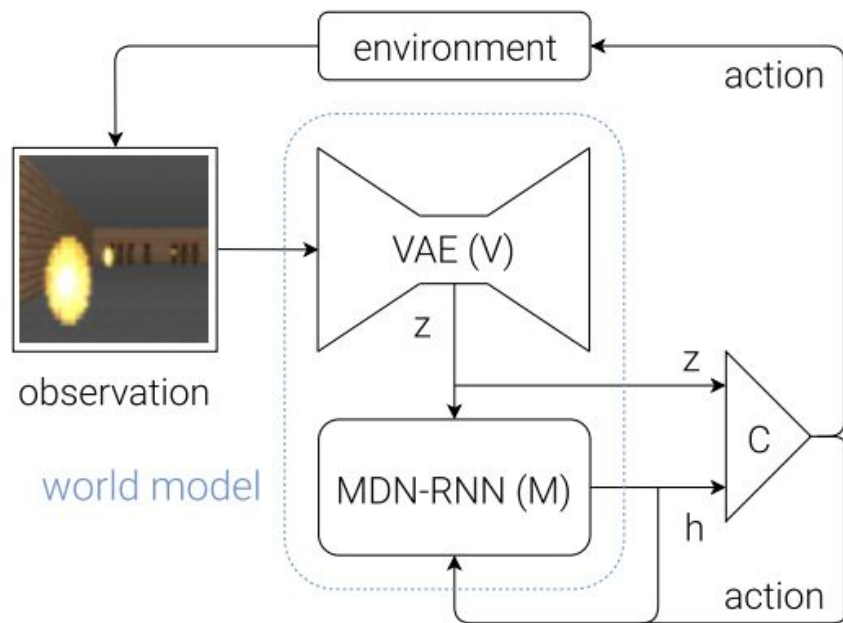


Transition Model

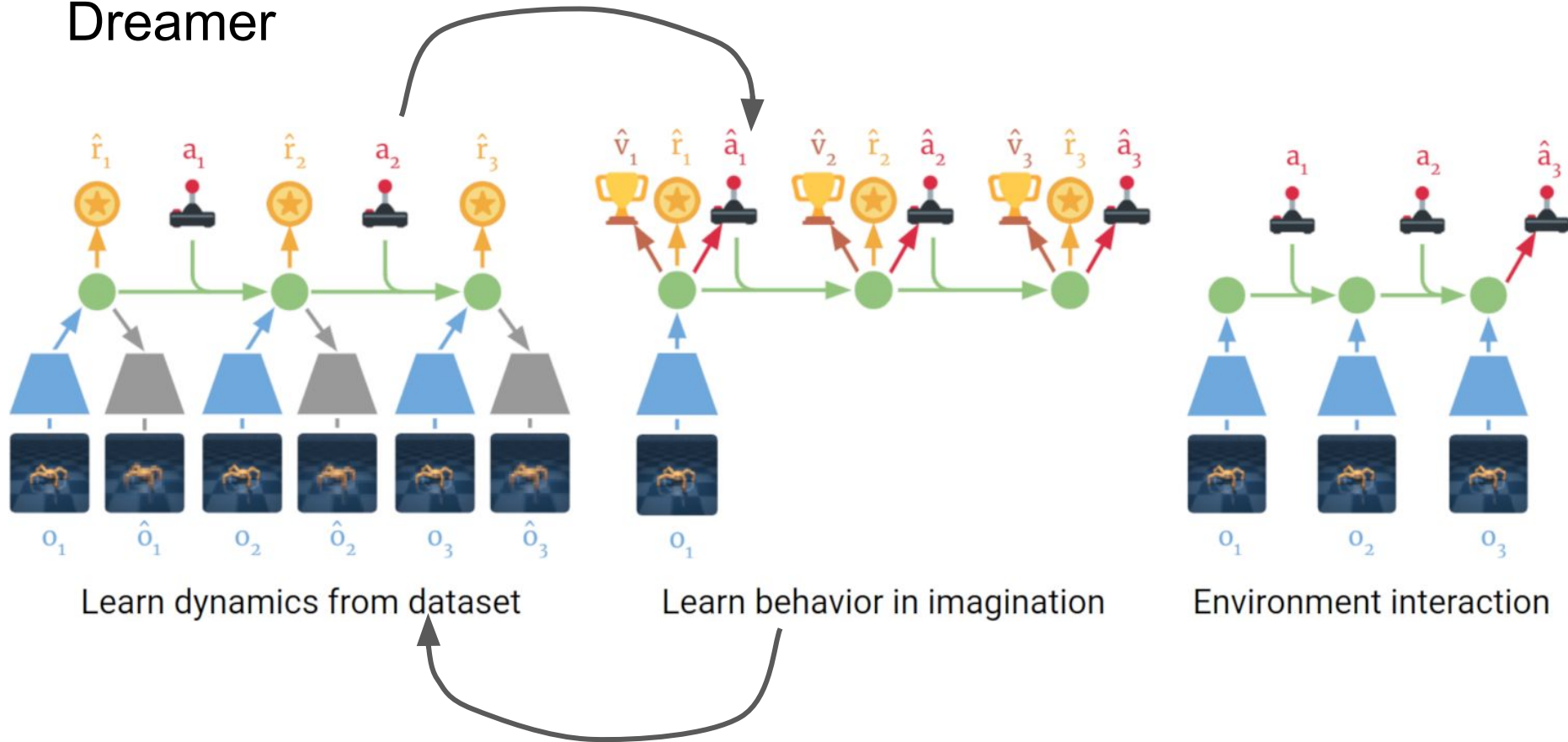


# World Models

<https://worldmodels.github.io/>



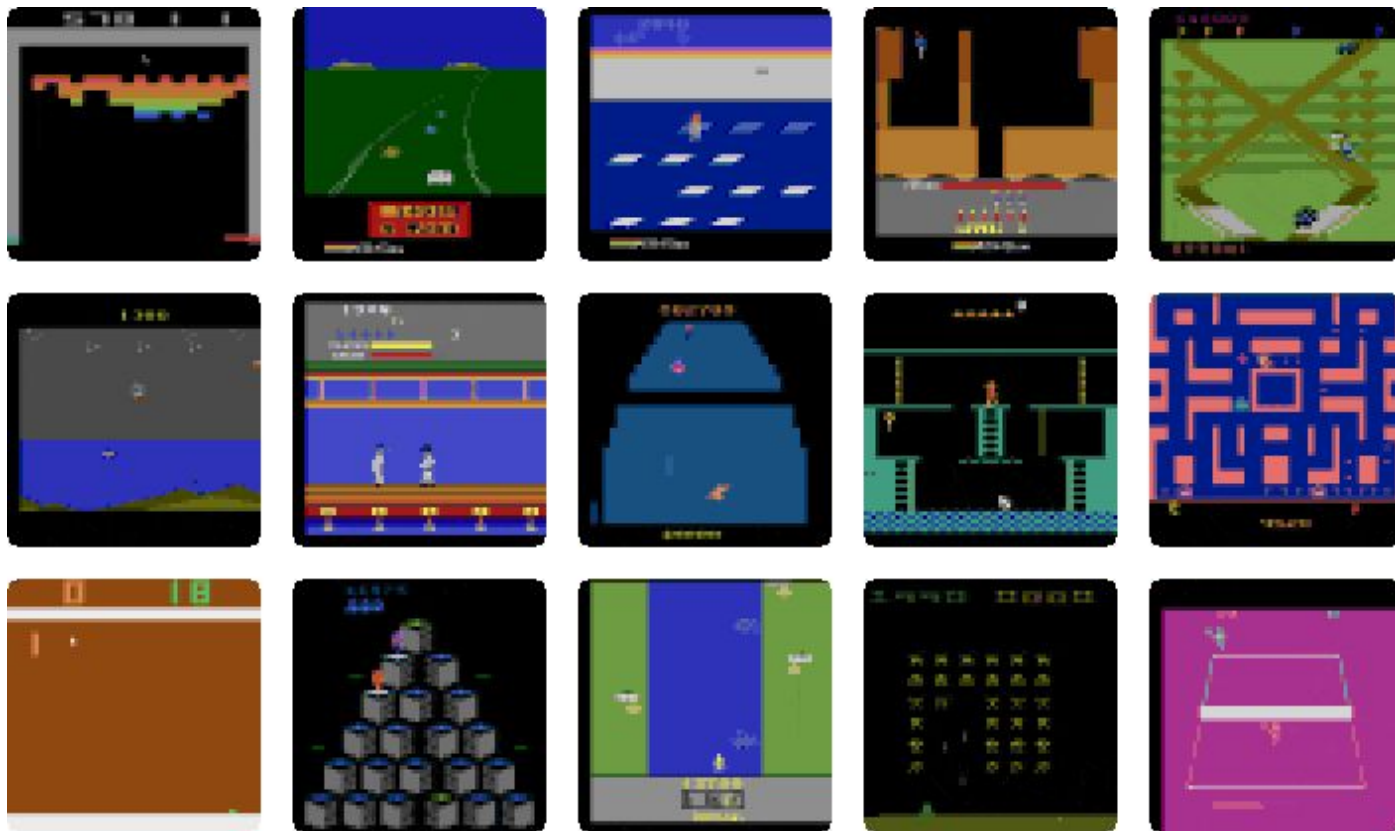
# Dreamer



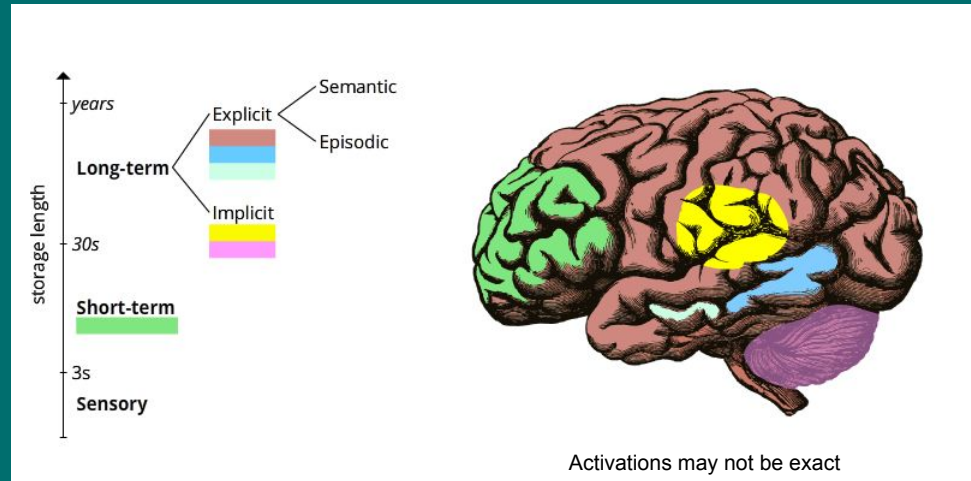
# Dreamer



# 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]

Lets Play a Game

Not very long-lasting

You need attention

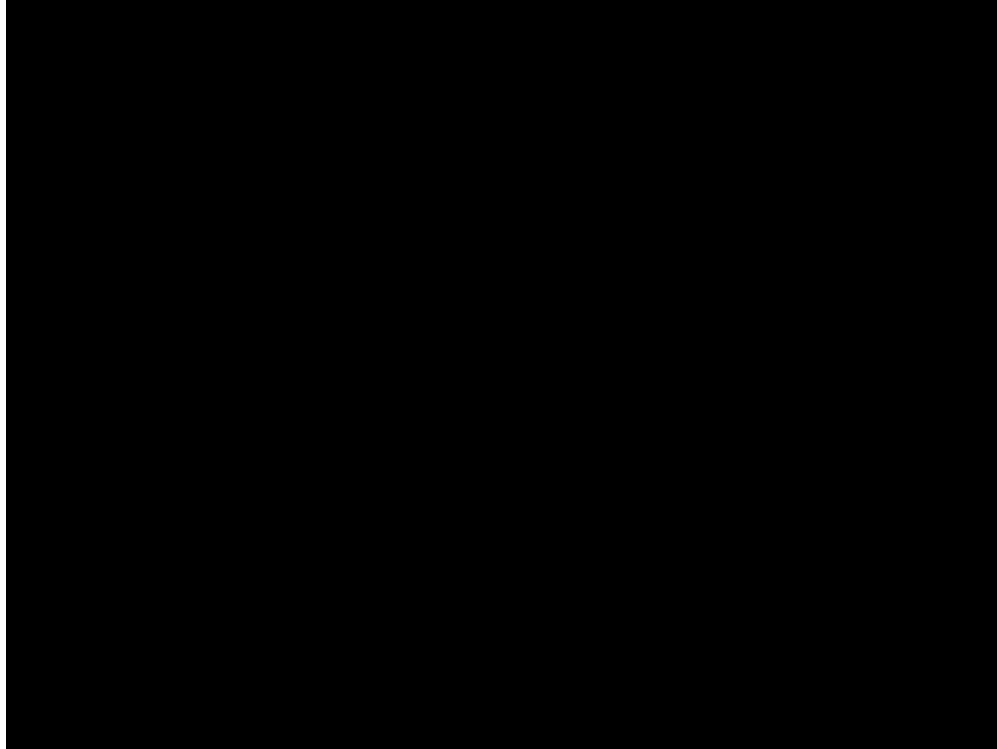
Numbers are much better and Relevancy matters [Jacobs]

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

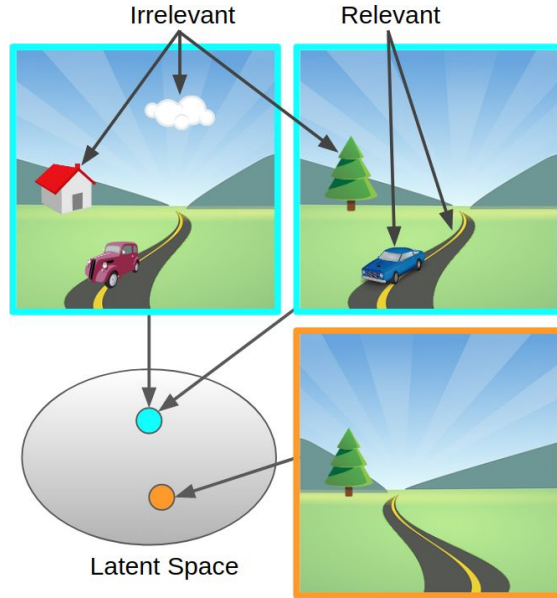
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97. <https://doi.org/10.1037/h0043158>

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

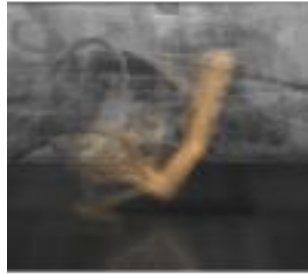


## Deep Bisimulation for Control

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).



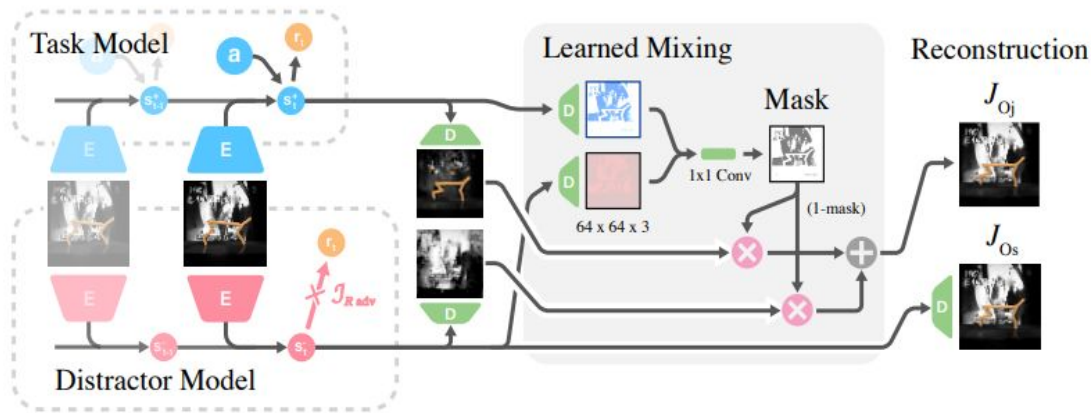


## Some Assumptions

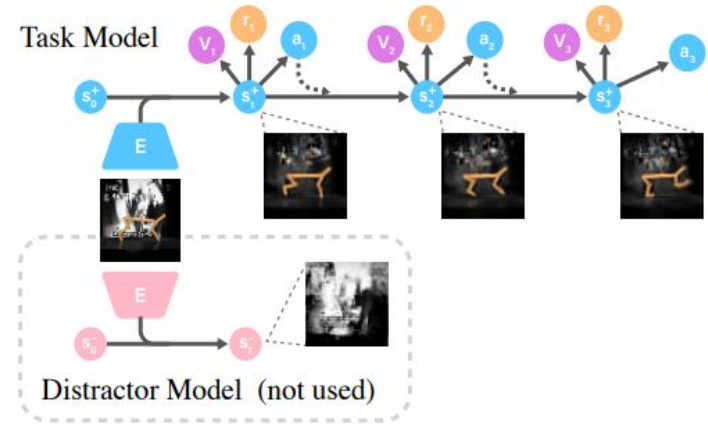
- Used same video for test and train
- No RGB background
- Used Ground Plane

# Research in relevancy prediction

## Learning Task Informed Abstractions



(a) Learning Task Informed World Models



(b) Policy Learning only unrolls in  $S^+$ .

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

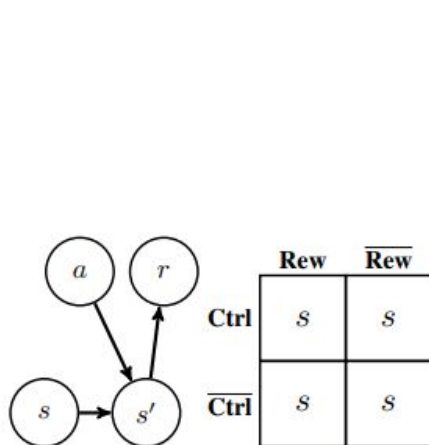
# Learning Task Informed Abstractions



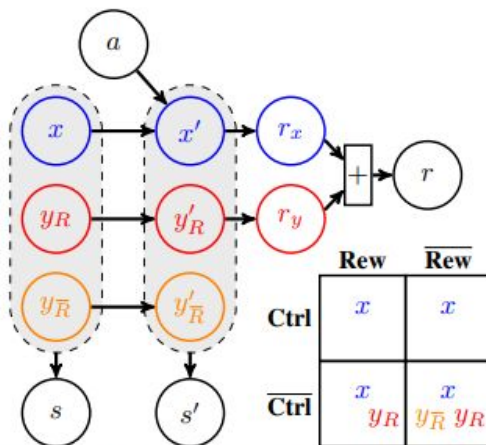
Fu, Xiang, et al. "Learning task informed abstractions." International Conference on Machine Learning. PMLR, 2021.

# Research in relevancy prediction

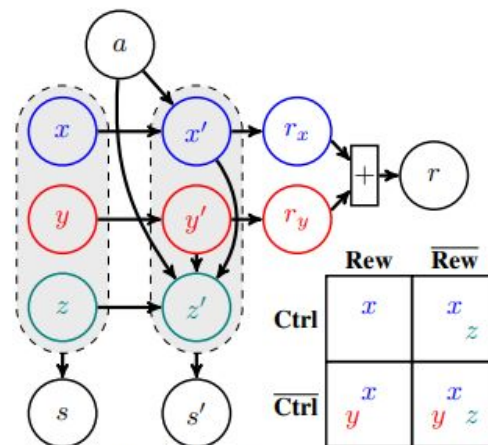
## Denoised MDPs



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



(b) Transition that factorizes out uncontrollable information in  $y_R$  and  $y_R$ .



(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 mdp: Learning world models better than the world itself. In International Conference on Machine Learning. PMLR, 2022.

# Denoised MDPs

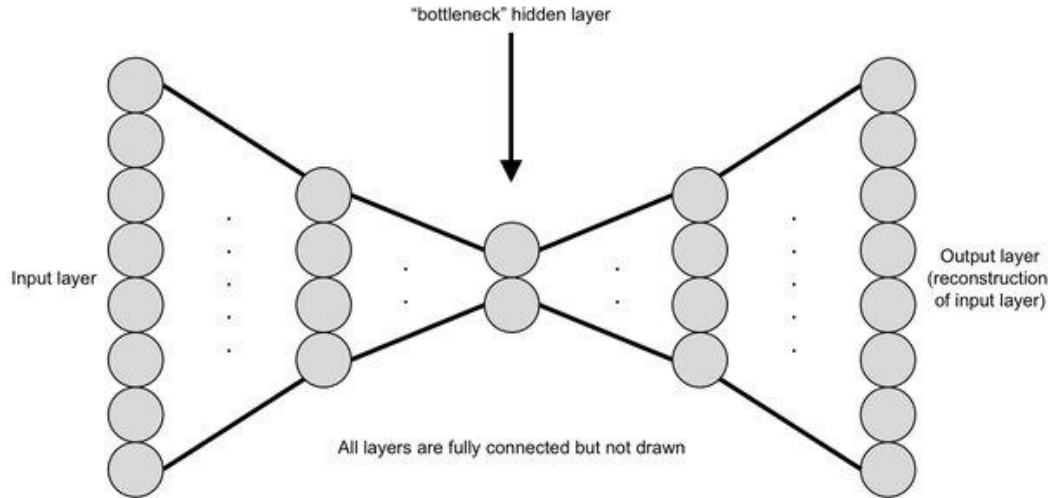


Tongzhou Wang, Simon S. Du, Antonio Torralba, Phillip Isola, Amy Zhang, and Yuandong Tian. Denoised mdps: Learning world models better than the world itself. In International Conference on Machine Learning. PMLR, 2022.

# Our Research

## Denoised Predictive Imagination

- Information Bottleneck Principle

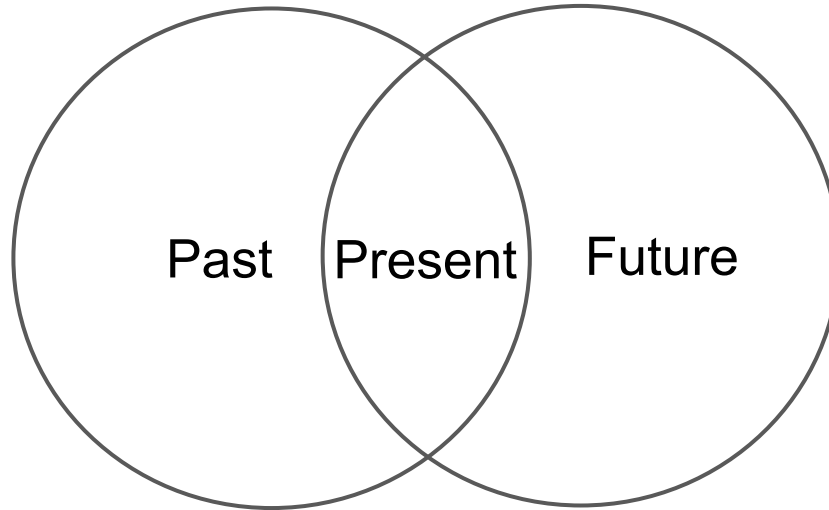


Tishby, Naftali, Fernando C. Pereira, and William Bialek. "The information bottleneck method." *arXiv preprint physics/0004057* (2000).

# Our Research

## Denoised Predictive Imagination

- Predictive Information



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

# Our Research

## Denoised Predictive Imagination

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

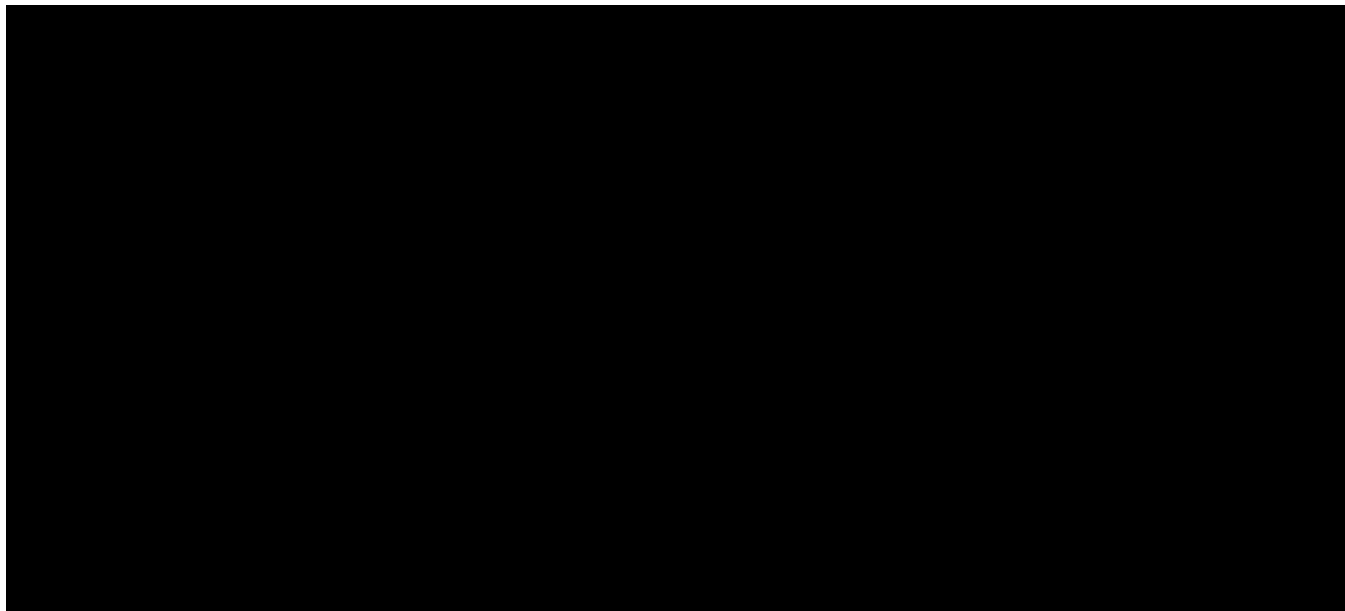


Submitted to International Conference of Learning Representations (ICLR) 2024



# Our Research

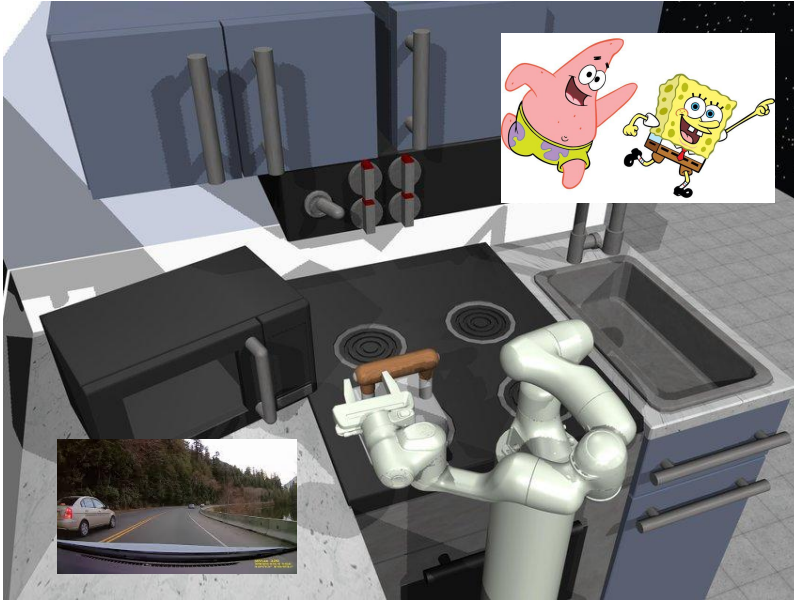
## **Denoised Predictive Imagination**



Submitted to International Conference of Learning Representations (ICLR) 2024

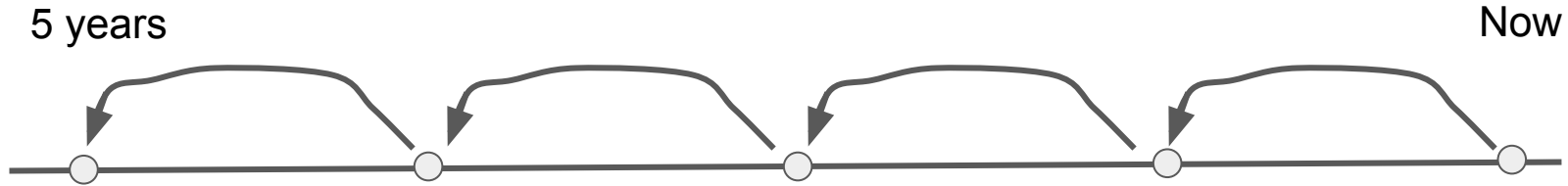
# New Ideas

## Why not Attention?



Done by ChatGPT-4

# Curiosity/Intrinsic Motivation

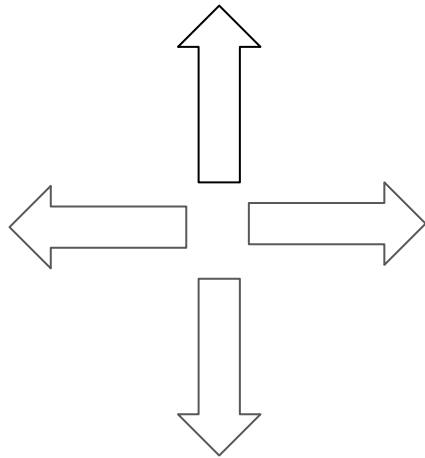


# Curiosity/Intrinsic Motivation

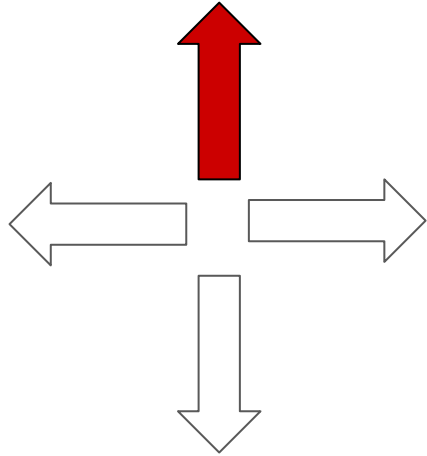


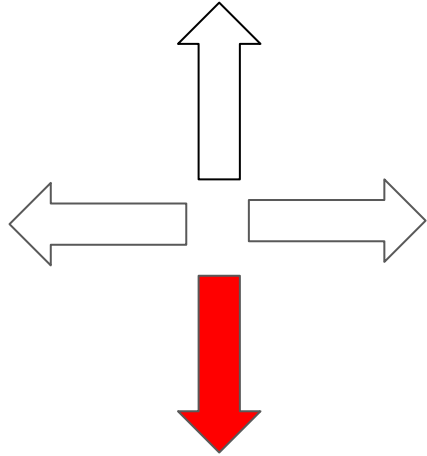
Silvia, P. J. (2012). Curiosity and motivation. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 157–166). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195399820.013.0010>

# How to formulate Curiosity?

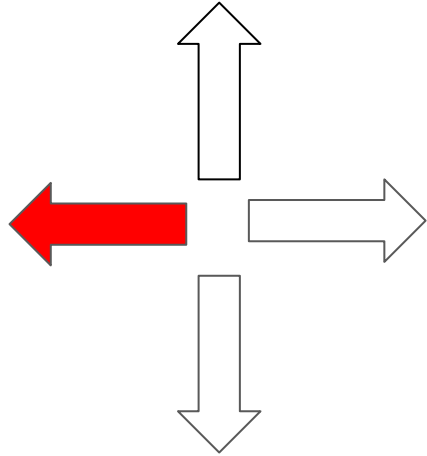


Source: <https://supermario-game.com/de>



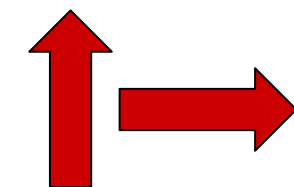






## Expected Behaviour

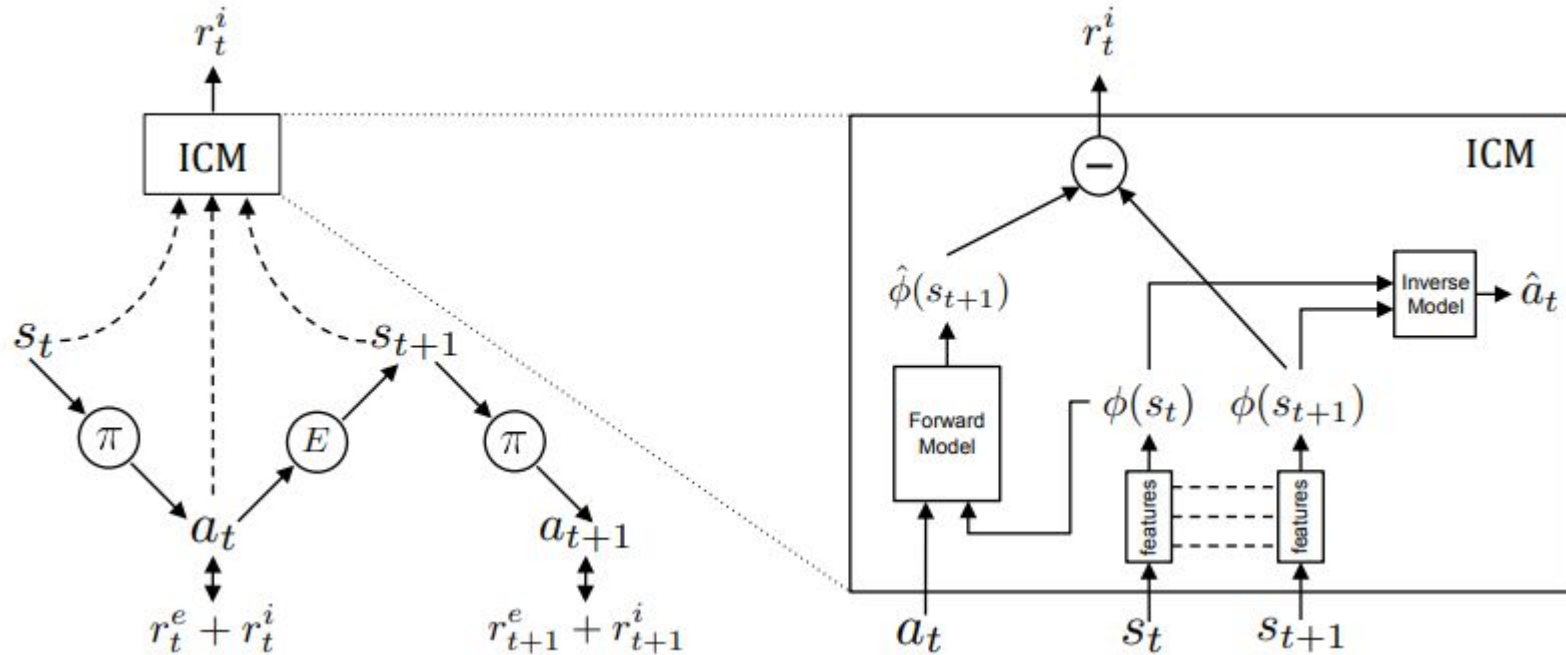
Curiosity = Prediction Error



## Actual Behaviour



# Curiosity/Intrinsic Motivation



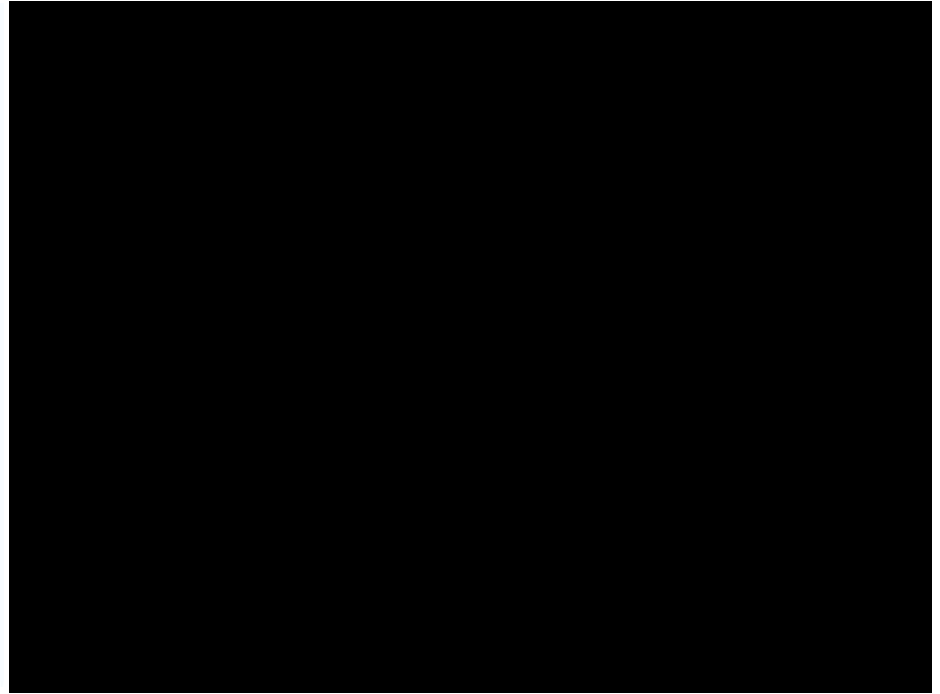
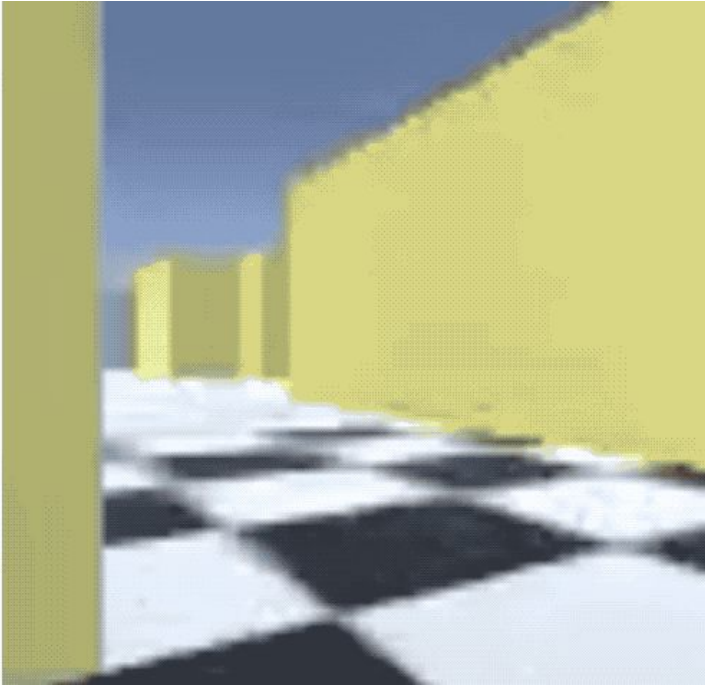
Pathak, Deepak, et al. "Curiosity-driven exploration by self-supervised prediction." *International conference on machine learning*. PMLR, 2017.

# Curiosity/Intrinsic Motivation



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

# Curiosity/Intrinsic Motivation



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 (<https://www.youtube.com/watch?v=Yr1mOzC93xs&t=1529s>)

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