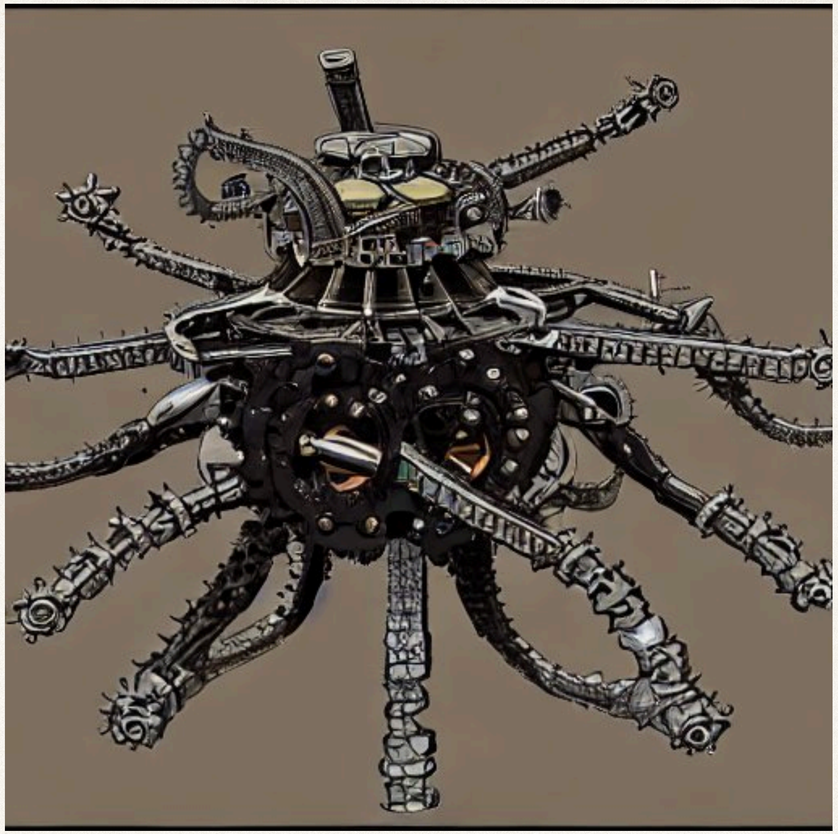


Defining agency

CHAIN Winter School, Lecture 1b - Manuel Baltieri

6th Feb 2023

ARAYA

Outline (second half)

What do people think about agency in:

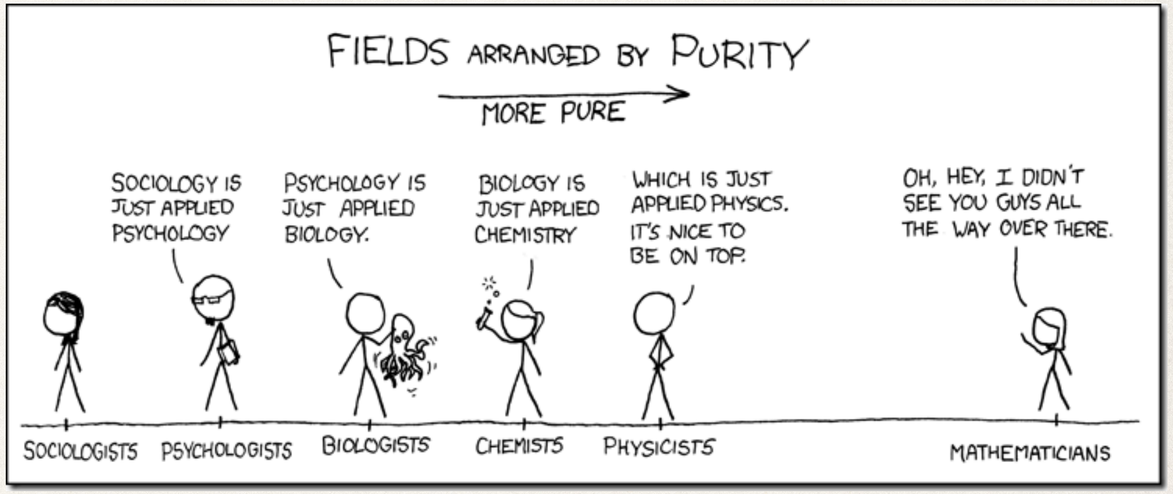
- physics
- computer science
- reinforcement learning
- control theory
- robotics

In physics, we question whether agency (even) exists.

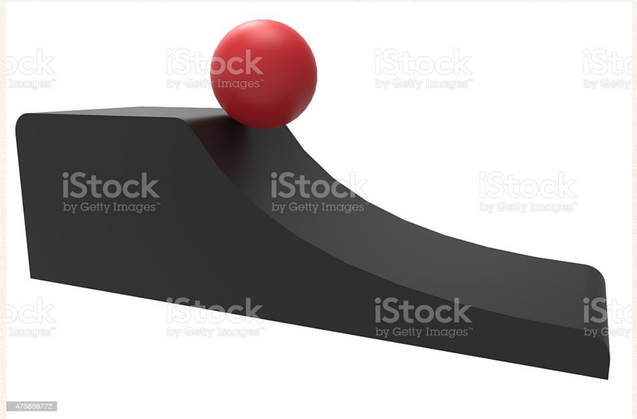
A reductionist approach to actions

Reductionism: everything can be decomposed into smaller components and their interactions

Emergence: “more than the sum of the parts”



Actions?



Happenings: outcomes of *purely mechanical* causes

Actions: outcomes *done* by an *agent*

McGregor, S. (2017). The Bayesian stance: Equations for 'as-if' sensorimotor agency. *Adaptive Behavior*, *25*(2), 72-82.

Photo by LOGAN WEAVER | @LGNWVR on Unsplash

Where are these actions?

1st line: electricity, magnetism, strong and weak

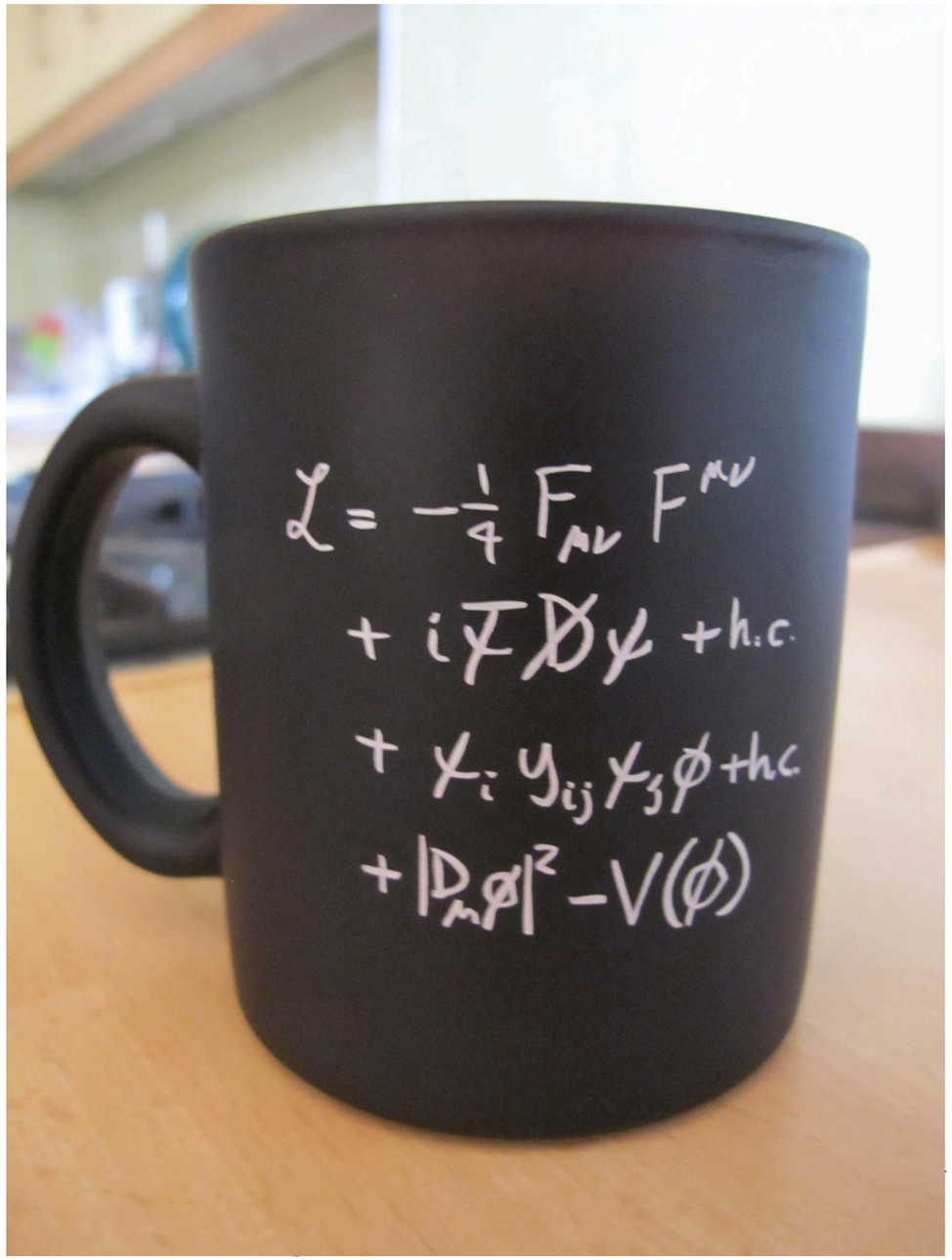
2nd line: how these forces act on quarks and leptons

3rd-4th lines: Higgs boson and how it gives mass to fundamental particles

Actions…

…as gravity?

No, really, no…

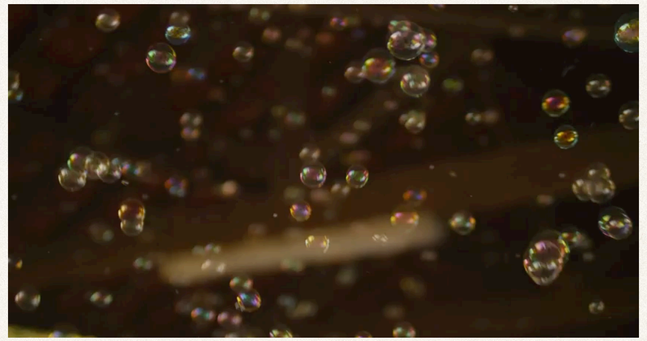


https://www.sciencealert.com/this-is-what-the-standard-model-of-physics-actually-looks-like

What's doing the acting anyway?

Entities: systems whose every part makes other parts more probable

Agents: a special kind of entities



https://www.youtube.com/watch?v=7gmEhb8qbTk&ab_channel=M%C3%A1rioJ.R.Matos

https://www.youtube.com/watch?v=V4f_1_r80RY&ab_channel=NationalGeographic

Biehl, M. A. (2017). *Formal approaches to a definition of agents* (Doctoral dissertation, University of Hertfordshire).

Some exceptions to the standard view

Agency in Physics

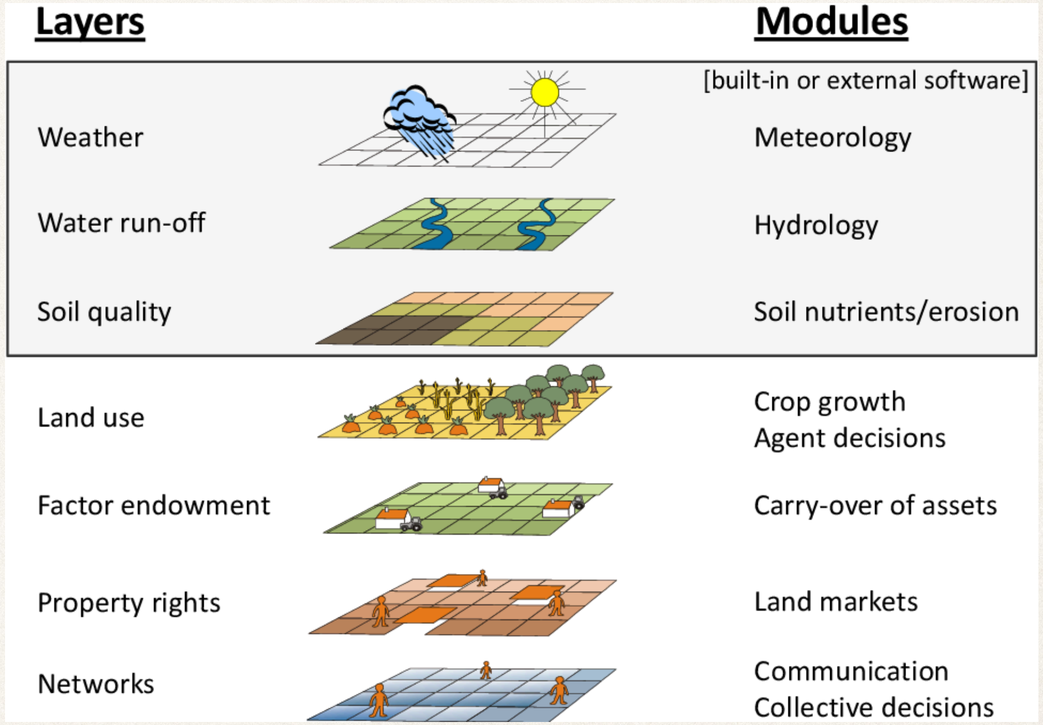
Carlo Rovelli
Aix Marseille University, Université de Toulon, CNRS, CPT, 13288 Marseille, France.
Perimeter Institute, 31 Caroline Street North, Waterloo, Ontario, Canada, N2L 2Y5.
The Rotman Institute of Philosophy, 1151 Richmond St. N London, Ontario, Canada, N6A 5B7.
(Dated: July 14, 2020)

I discuss three aspects of the notion of agency from the standpoint of physics: (i) what makes a physical system an agent; (ii) the reason for agency's time orientation; (iii) the source of the information generated in choosing an action. I observe that agency is the breaking of an approximation under which dynamics appears closed. I distinguish different notions of agency, and observe that the answer to the questions above differ in different cases. I notice a structural similarity between agency and memory, that allows us to model agency, trace its time asymmetry to thermodynamical irreversibility, and identify the source of the information generated by agency in the growth of entropy. Agency is therefore a physical mechanism that transforms low entropy into information. This may be the general mechanism at the source of the whole information on which biology builds.

In computer science, we try to model agency.

ABMs

Agent-based modelling: computer simulations used to study the interactions between agents over time

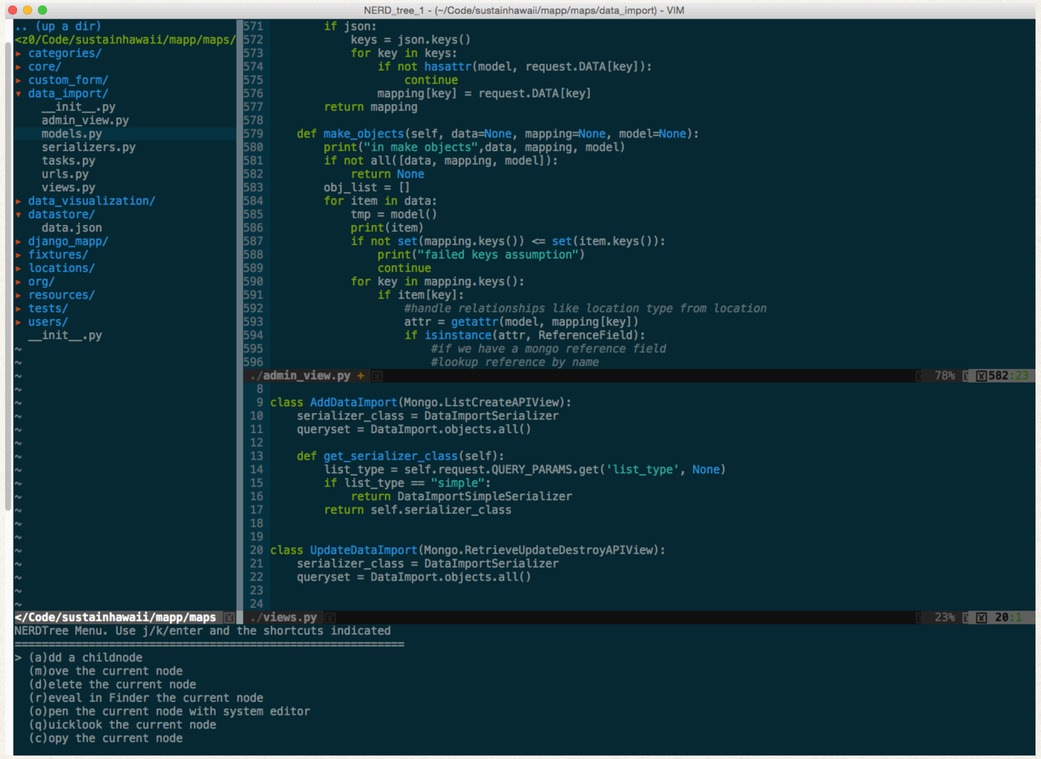


Berger, T., & Troost, C. (2012). Agent-based modelling in the agricultural economics tradition of recursive farm modelling and adaptive micro-systems.

Softwares as agents?

Software agent: “[something that] receives keystrokes, file contents, and network packets as sensory inputs and acts on the environment by displaying on the screen, writing files, and sending network packets.”

Russell, Stuart J, & Norvig, P.. Artificial intelligence a modern approach. Pearson Education, Inc., 2010.



https://realpython.com/vim-and-python-a-match-made-in-heaven/

What about chatbots?

https://openai.com/blog/chatgpt/

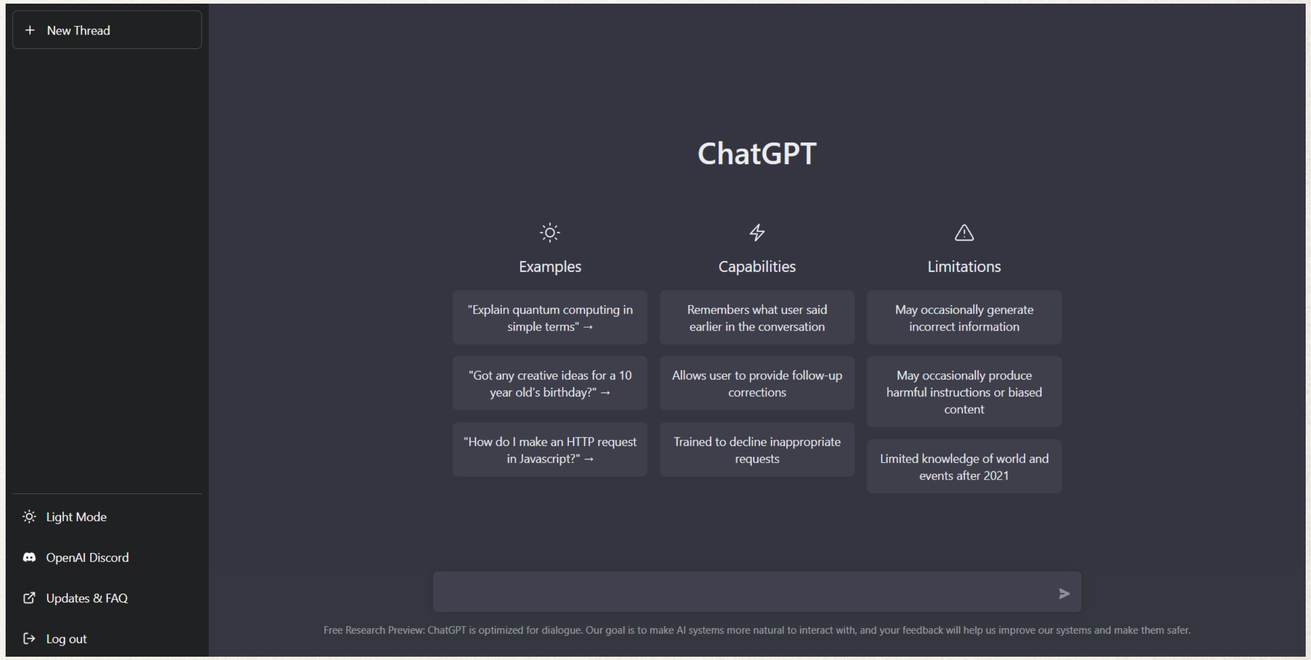


Photo: https://en.wikipedia.org/wiki/ChatGPT#/media/File:ChatGPT.png

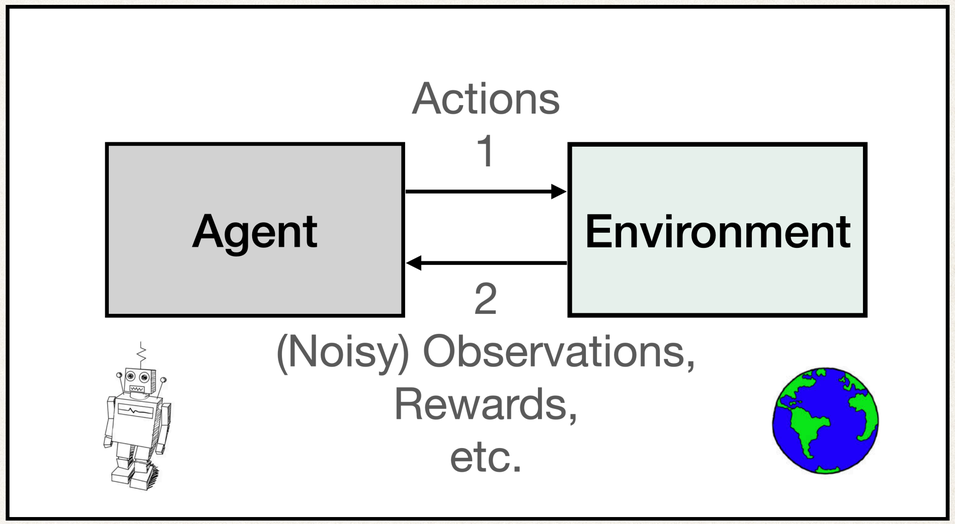
In reinforcement learning, we (don’t) explain agency by (not) defining other concepts.

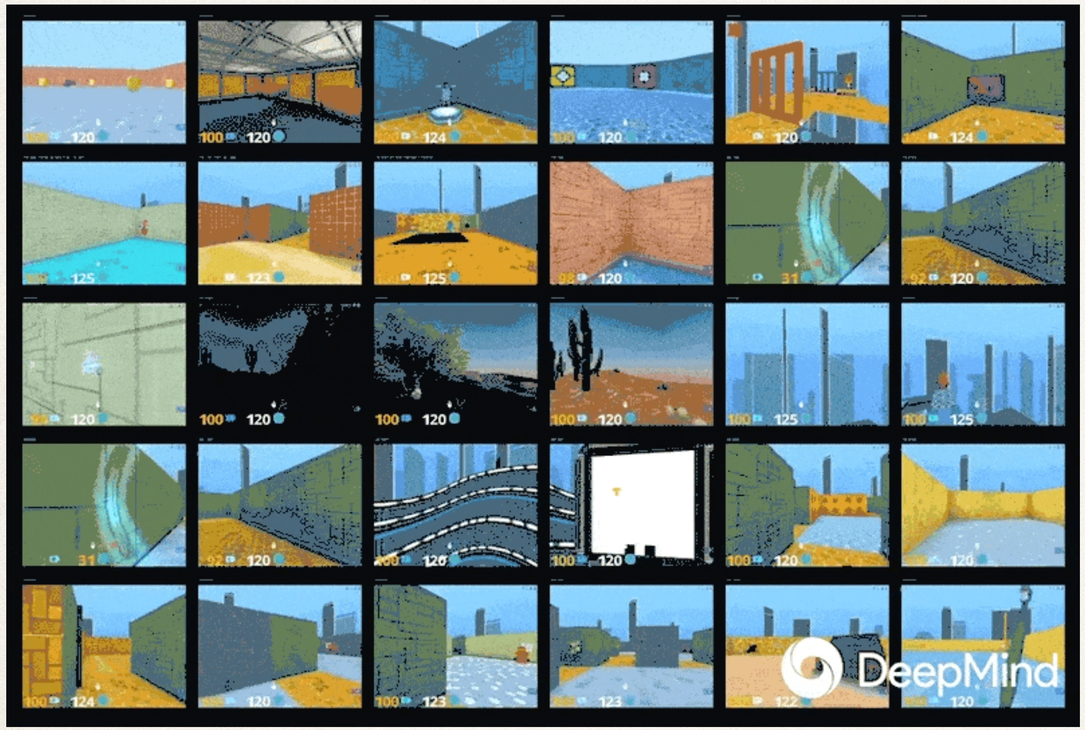
Agent-enrivonment interactions

Agents are distinct from their **environment** (cf. first part)

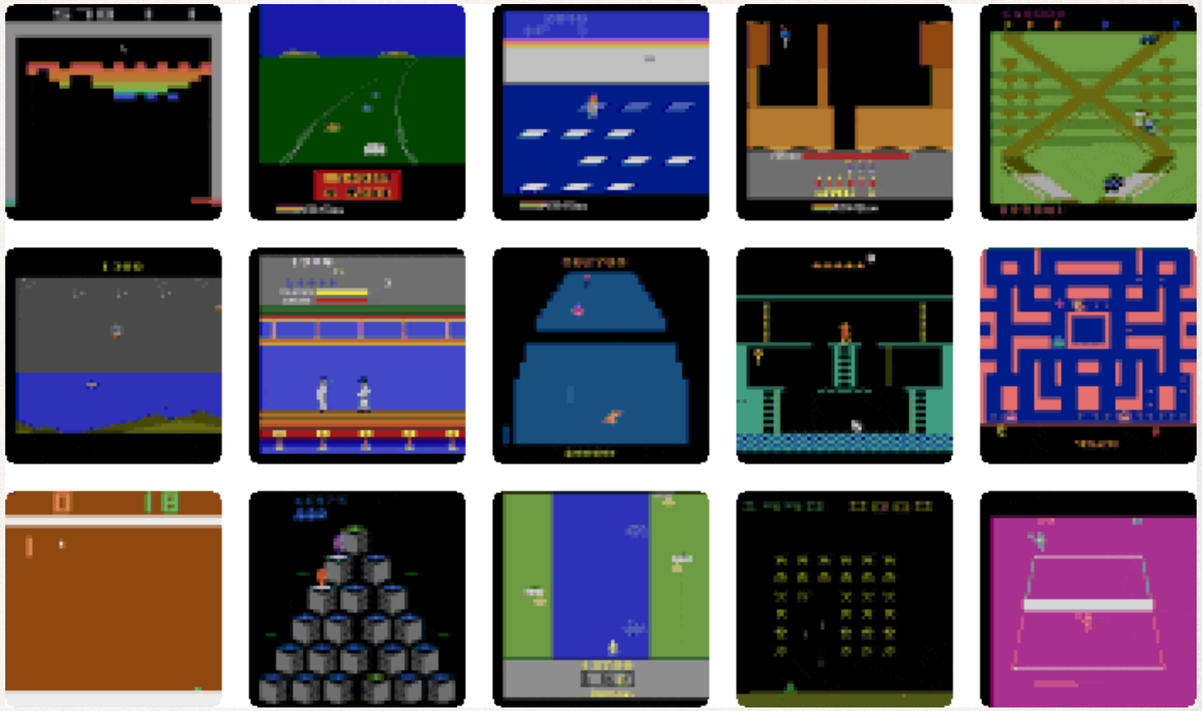
Agents as systems with **goals** (intrinsic or extrinsic)

Agents take **actions** that maximise **value** (**rewards** over times)





https://www.deepmind.com/blog/scalable-agent-architecture-for-distributed-training



https://ai.googleblog.com/2021/02/mastering-atari-with-discrete-world.html

In control theory, we mix all the previous practices.

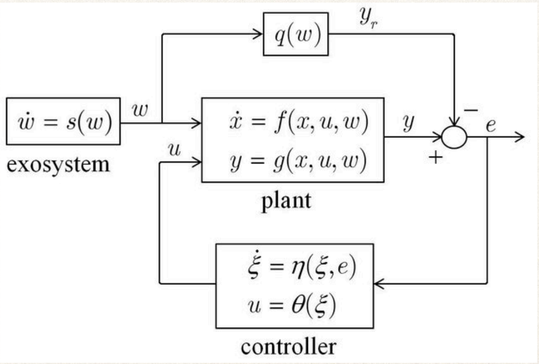
Control theory

Optimal control theory generalises classical mechanics by introducing goals and actions (but does **agency even exist**?)

Control theory can be used to **model agents** and their actions (decisions, plans, policies)

Same structure as reinforcement learning (**environment**/exosystem, agent/ plant+controller with **goals** and **actions**)

Photograph: Andrew Matthews/PA.



Natarajan, V., & Weiss, G. (2019). Minimal order controllers for output regulation of nonlinear systems. IFAC Journal of Systems and Control, 7, 100028.

In robotics, we build agents but don’t worry too much about definitions.

Adaptation and autonomy

Same definition as with R
to adaptation and autono

Atlas, by Boston Dynamics. https://www.youtube.com/watch?v=-e1_QhJ1EhQ&t=5s&ab_channel=BostonDynamics

Roomba.
t.co.uk/en_GB/irobot-roomba-j7/J715840.html

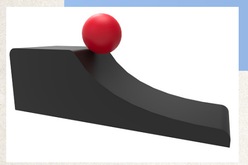
Agency on a spectrum

No agents:
physical laws
describe everything
in the universe

Agents, maybe?:
steering a system towards
a goal (physical laws
+ inputs/parameters)

Agents: systems
with adaptation
AND autonomy

Agents:
systems that
take actions for
their own sake



Physics

Reinforcement learning,
Control theory,
Computer science

Robotics

?

Take-home messages

- In physics, we question whether agency (even) exists
- In computer science, we try to model agency
- In reinforcement learning, we (don't) explain agency by (not) defining other concepts
- In control theory, we mix all the previous practices
- In robotics, we build agents but don't worry too much about definitions

Open questions (part 2)

What are agents?

Are they real or just products of our imagination? And do we need them?

How do we reconcile different understandings of agency / agents (biology, physics, robotics, etc.)?