Causal Reinforcement Learning

A Road to Artificial General Intelligence

Chaochao Lu

Department of Engineering University of Cambridge Cambridge, UK	Department of Empirical Inference Max Planck Institute for Intelligent Systems Tübingen, Germany

NOKIA Bell Labs | Social Dynamics Seminar | 28 Nov 2019

Google Books Ngram Viewer

Google Books Ngram Viewer

Ngrams not found: Causal Reinforcement Learning

The Bible Times

*The Bible, for example, tells us that just a few hours after tasting from the tree of knowledge, Adam is already an expert in **causal arguments**.*

When God asks: "Did you eat from that tree?"

This is what Adam replies: "The woman whom you gave to be with me, She handed me the fruit from the tree; and I ate."

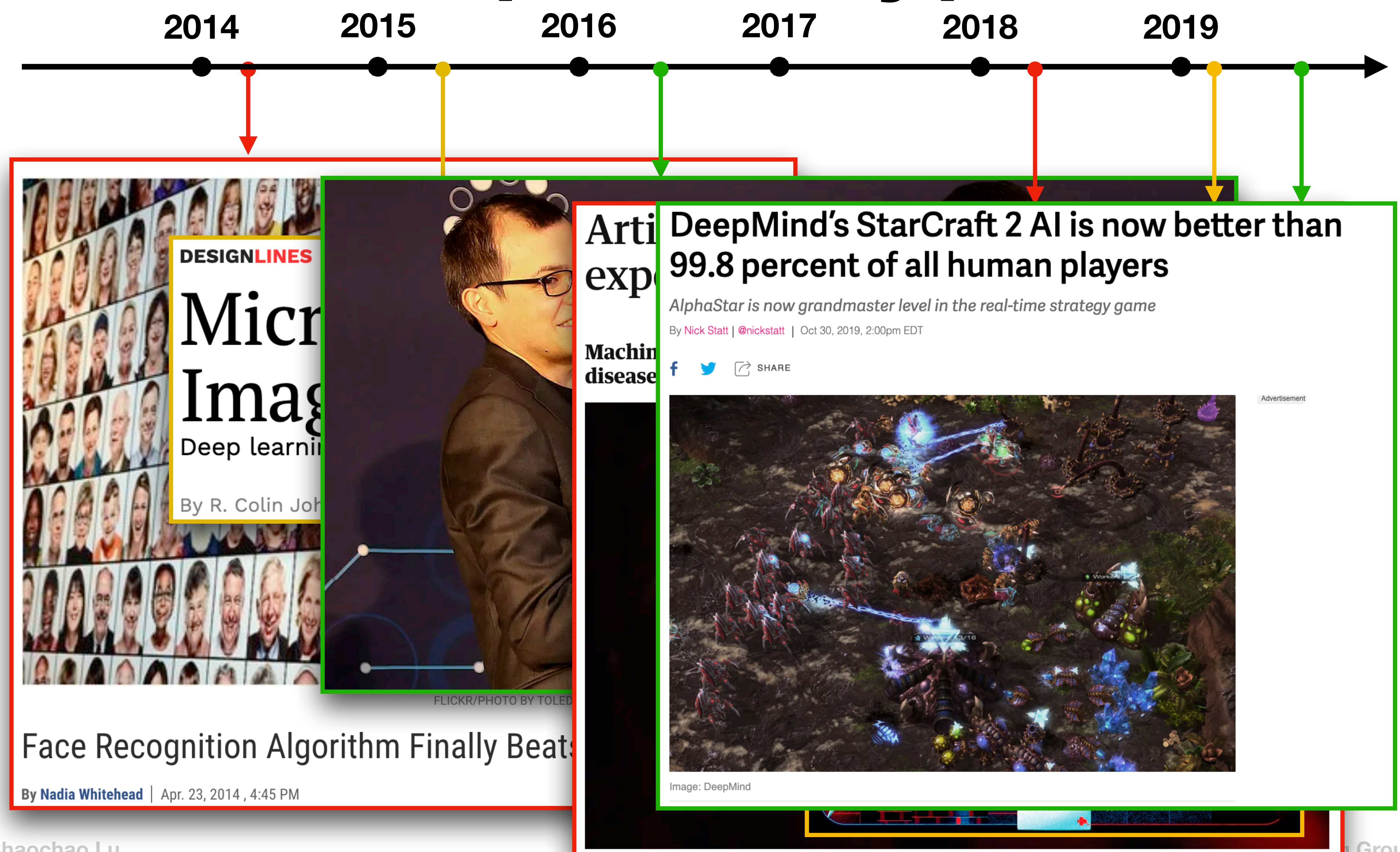
Eve is just as skilful: "The serpent deceived me, and I ate."

*The thing to notice about this story is that God did not ask for **explanation**, only for the **facts** – it was Adam who felt the need to explain. The message is clear: **causal explanation** is a man-made concept.*

Judea Pearl. Causality, 2009.

From AI to AGI

Hope or Hype


2014
2015
2016
2017
2018
2019
DESIGNLINES
Micr
Imag
Deep learni
By R. Colin Joh
FLICKR/PHOTO BY TOLED
Face Recognition Algorithm Finally Beats
By Nadia Whitehead | Apr. 23, 2014 , 4:45 PM
Arti
exp
Machin
disease
DeepMind's StarCraft 2 AI is now better than 99.8 percent of all human players
AlphaStar is now grandmaster level in the real-time strategy game
By Nick Statt | @nickstatt | Oct 30, 2019, 2:00pm EDT
SHARE
Advertisement
Image: DeepMind

Hope or Hype

DeepMind's StarCraft 2 AI is now better than 99.8 percent of all human players

AlphaStar is now grandmaster level in the real-time strategy game

By Nick Statt | @nickstatt | Oct 30, 2019, 2:00pm EDT

SHARE

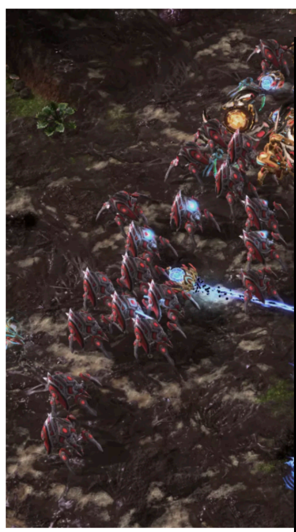


Image: DeepMind

DeepMind's latest Starcraft result with AlphaStar is a very impressive tour de force. But is it an important step towards general intelligence? Here are some questions

Starcraft is significantly harder than Atari games, and the new system is a significant advance beyond the previous system, interacting with complex coalitions of real-world actors. How general is the result? Here are some open questions.

- **Can AlphaStar beat other games, without modification**? Although AlphaStar is a descendant of the AlphaZero system that became a Go Champion, the highly structured models contains machinery (eg scatter connections and exploiter agents), representations (e.g, # of workers, cargo status) and training regimes that were specifically developed for and tuned to StarCraft.
- **Can training specific to StarCraft reduce the amount time of required to learn a closely-related game, such as Warcraft**? (cf a human transferring experience on real time strategy game to another).
- **Can they transfer between different maps nor between different "races" within the game**? a human would generalize at at least some of its experience between races and between maps.
- **Could a future iteration of the system succeed using only the amount of data that a human champion would get?** The massive number of "replays" required may not realistically obtainable in many real world situations.
- **How important is human expertise**? In 2017 DeepMind made a big deal in out of AlphaZero purportedly mastering Go "without human knowledge"; the StarCraft victory emerged in part from human insights into Starcraft and the dynamics of exploitation within that game. Also unlike with AlphaZero's self-play, human demonstrations have been critical. Perhaps, it is time to accept the value of innate and human-derived knowledge.
- **Would the AlphaStar system that worked for the closed-world domain of Starcraft would work in more open-ended domains,** such as natural language understanding, where there in an essentially infinite range of sentences**?**

Hope or Hype

NEWS · 07 NOVEMBER 2019

AI Copernicus ‘discovers’ that Earth orbits the Sun

A neural network th
quantum-mechanic

s the
m 100
r

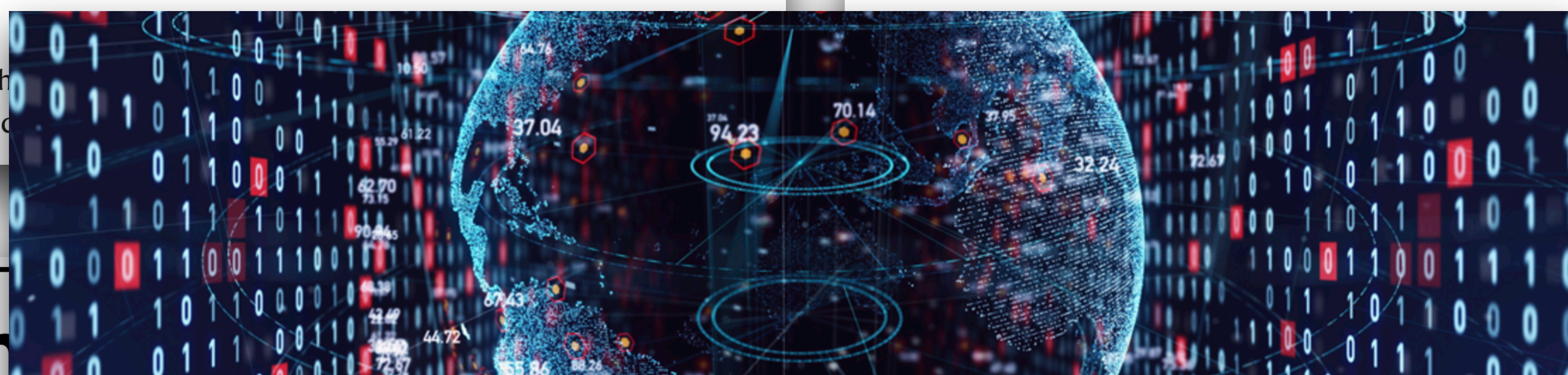
to tackle one of

Oct 26, 2019

**Who n
you ha
learni**

It took humanity
Now a neural ne
the same data, i

by **Emerging Technology from the arXiv** Aug 3, 2018



NUMBERS | MATH

Are Neural Networks About to Reinvent Physics?

The revolution of machine learning has been greatly exaggerated.

BY GARY MARCUS & ERNEST DAVIS
NOVEMBER 21, 2019

The Debate on AGI

Does AI Need More Innate Machinery?

Marcus and LeCun in Complete Agreement on Seven Points

October 2017

- AI is still in its infancy
- Machine learning is fundamentally necessary for reaching strong AI
- Deep learning is a powerful technique for machine learning
- Deep learning is not sufficient on its own for cognition
- [model-free] Reinforcement learning is not the answer, either
- AI systems still need better internal forward models
- Commonsense reasoning remains fundamentally unsolved

Some basics that evolution might have endowed humans with

- Representations of objects
- Structured, algebraic representations
- Operations over variables
- A type-token distinction
- A capacity to represent sets, locations, paths tracjectories, obstacles and enduring individuals
- A way of representing the affordances of objects
- Spatiotemporal contiguity / conservation of mass
- Causality
- Translational invariance
- Capacity for cost-benefit analysis

Questions

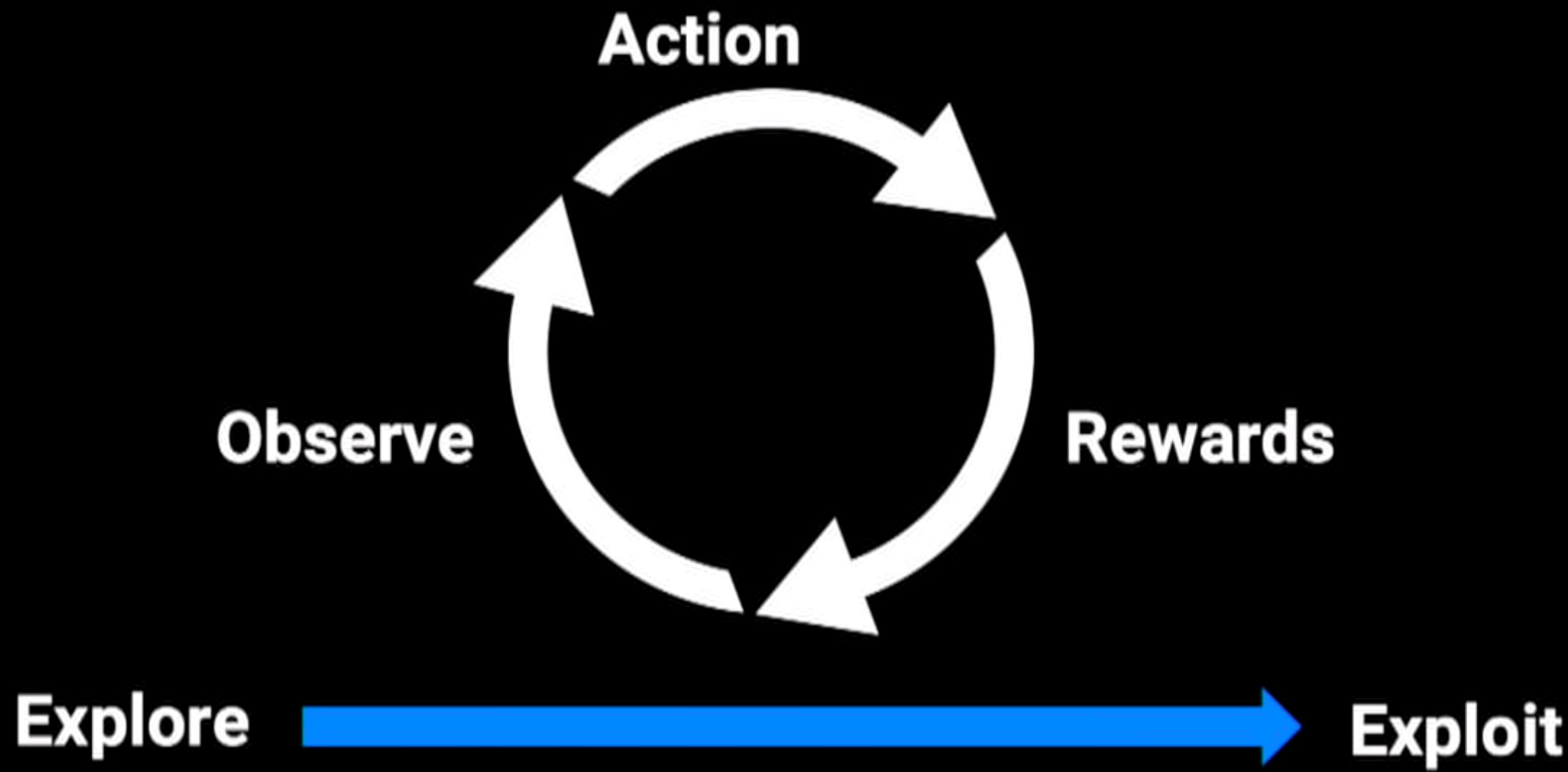
- **"All of these AI systems we see, none of them is 'real' AI**

  – Josh Tennenbaum at CCN 2017

  - I agree (Josh and I start our talks the same way).
- **The brain learns with an efficiency that none of our machine learning methods can match.**
  - Our supervised learning systems require large numbers of example
  - Our reinforcement learning systems require millions of trials
  - that's why we don't have robots that as agile as a cat or a rat
  - that's why we don't have dialog systems that have common sense
- **What is missing?**

  Learning paradigms that build (predictive) models of the world through observation and action.

Nature’s Learning Method: Reinforcement

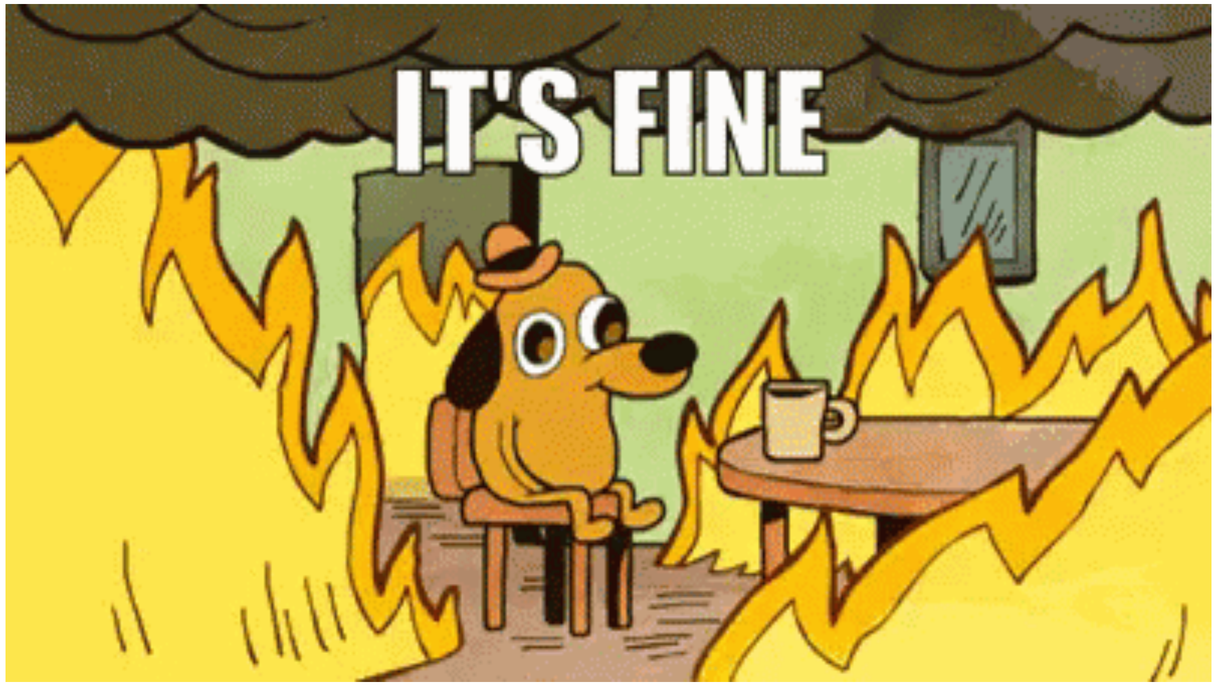


unity

GOTO 2018 • On the Road to Artificial General Intelligence • Danny Lange

What’s Wrong with RL?

Reinforcement Learning never worked, and 'deep' only helped a bit.



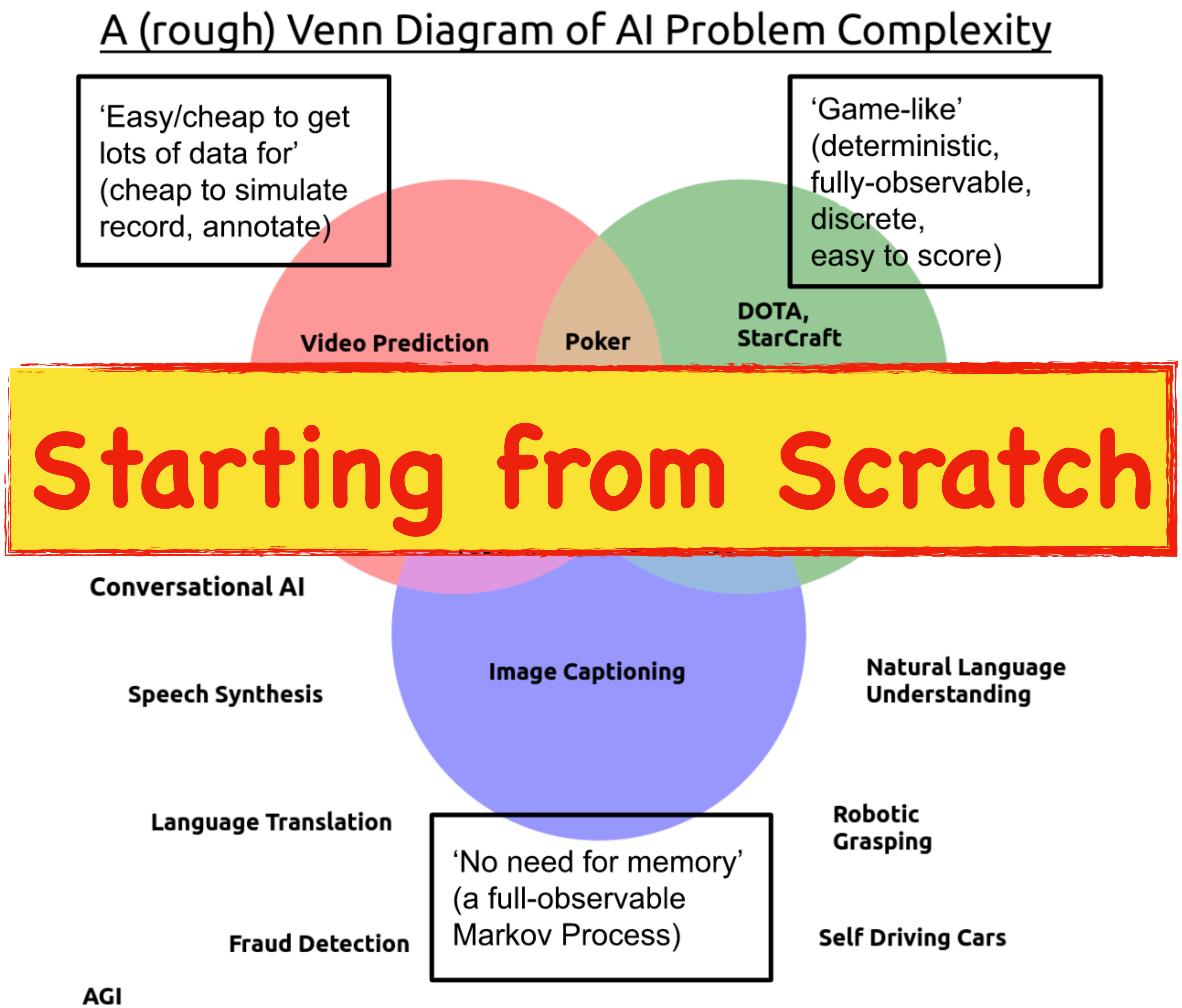
RL researchers all the time

Legit RL research request

Exploration and Long Term Credit Assignment

[Himanshu Sahni's Blog]

RL’s Fundamental Flaw



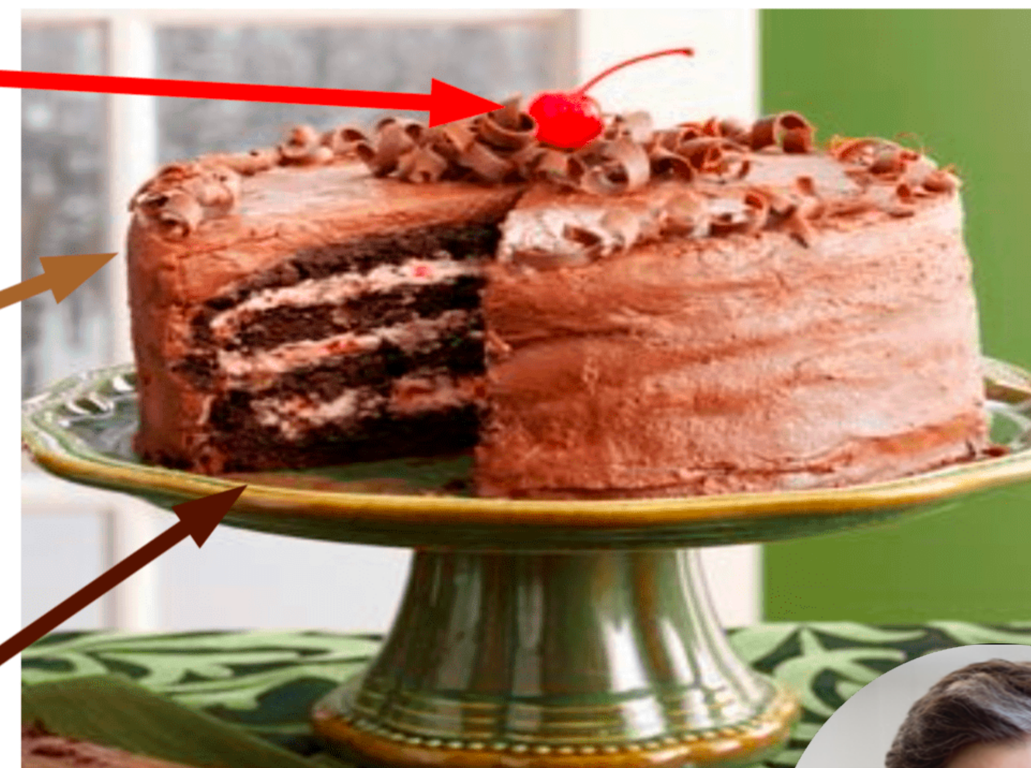
[Andrey Kurenkov's Blog]

RL is a Cherry

Y. LeCun

How Much Information is the Machine Given during Learning?

- **“Pure” Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Self-Supervised Learning (cake génoise)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**



© 2019 IEEE International Solid-State Circuits Conference 1.1: Deep Learning Hardware: Past, Present, & Future

Why is Causal RL?

Why from RL

Is RL an exercise in causal inference? Of course! Albeit a restricted one. By deploying interventions in training, RL allows us to infer consequences of those interventions, but ONLY those interventions. A causal model is needed to go BEYOND, i.e., to actions not used in training.

The relation between RL and causal inference has been a topic of some debate. It can be resolved, I believe, by understanding the limits of each.

Question 1: why is RL on the original high-dimensional Atari games harder than on downsampled versions?
Question 2: why is RL easier if we permute the replayed data?

RL is closer to causality research than the machine learning mainstream in that it sometimes effectively directly estimates do-probabilities (on-policy learning). However, as soon as off-policy learning is considered, in particular in the batch (or observational) setting, issues of causality become subtle.

Why from Natural Science

AAAS: Machine learning 'causing science crisis'

By Pallab Ghosh
Science correspondent, BBC News, Washington

16 February 2019 | Science & Environment

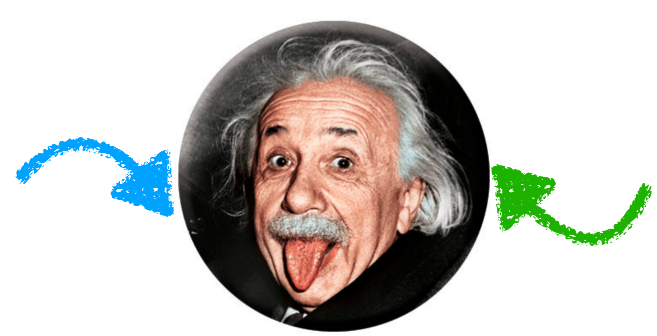
Reproducibility Crisis

Flawed Patterns

Experiments

Theory

Symmetry



Why from Cognition

Humans ***summarise rules or experience*** *from their* ***interaction*** *with nature and then* ***exploit this to improve their adaptation*** *in the next exploration.*

What ***Causal RL*** *does is exactly to* ***mimic human behaviours,*** *i.e.,* ***learning causal relations*** *from an agent that* ***communicates*** *with the environment and then* ***optimising its policy based on the learned causal structures****.*

"Our grasp of the world — the way we mirror its causal structure — is at the mercy of the inferential tools we have in the brain."

— JAKOB HOHWY

"Play is the answer to how anything new comes about."

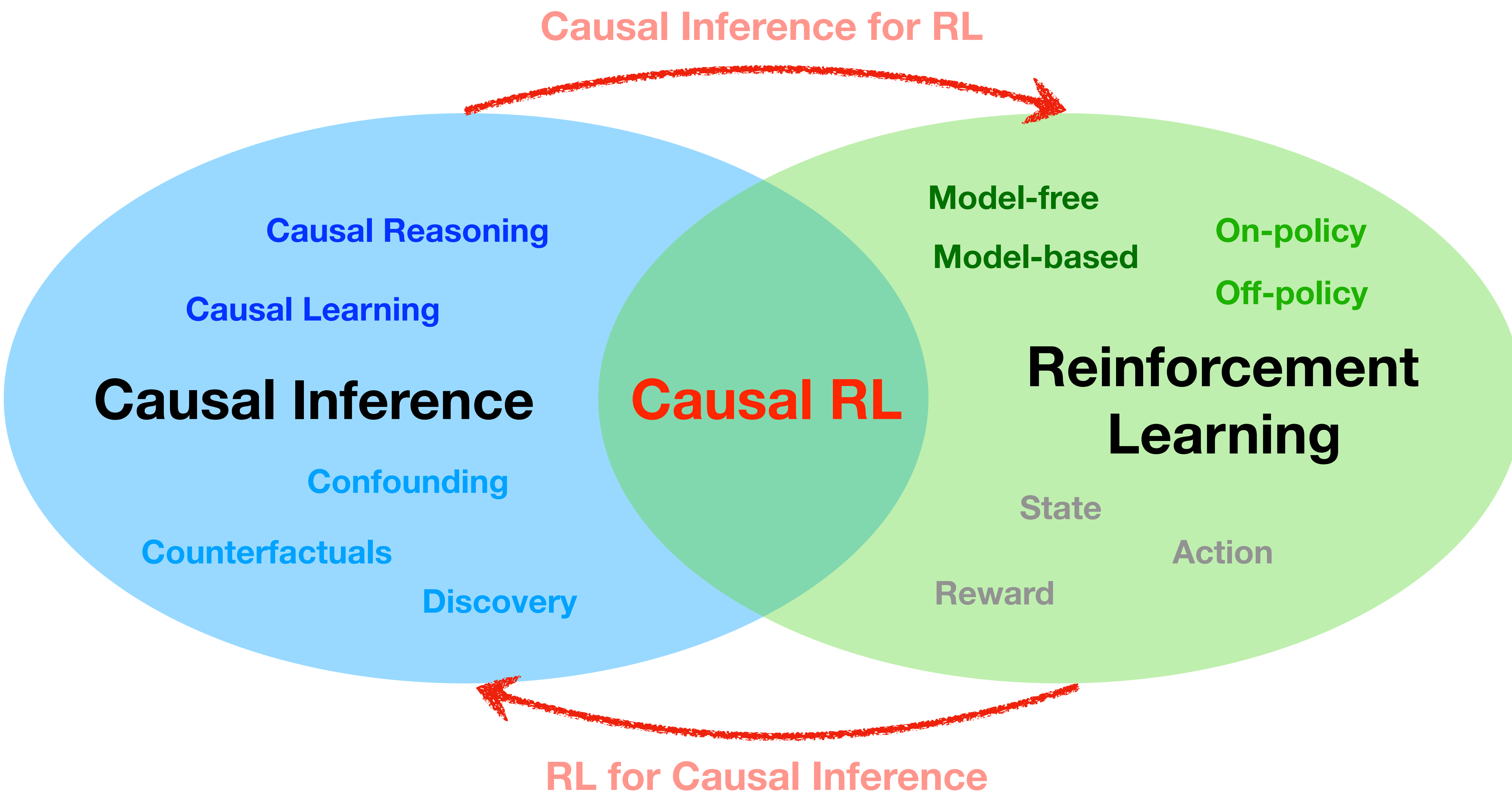
— JEAN PIAGET

"All reasonings concerning matter of fact seem to be founded on the relation of cause and effect. By means of that relation alone we can go beyond the evidence of our memory and senses."

— DAVID HUME

What is Causal RL?

Causal RL


Causal Inference for RL
Causal Reasoning
Causal Learning
Causal Inference
Confounding
Counterfactuals
Discovery
Causal RL
Model-free
Model-based
On-policy
Off-policy
Reinforcement Learning
State
Action
Reward
RL for Causal Inference

Reinforcement Learning (RL)

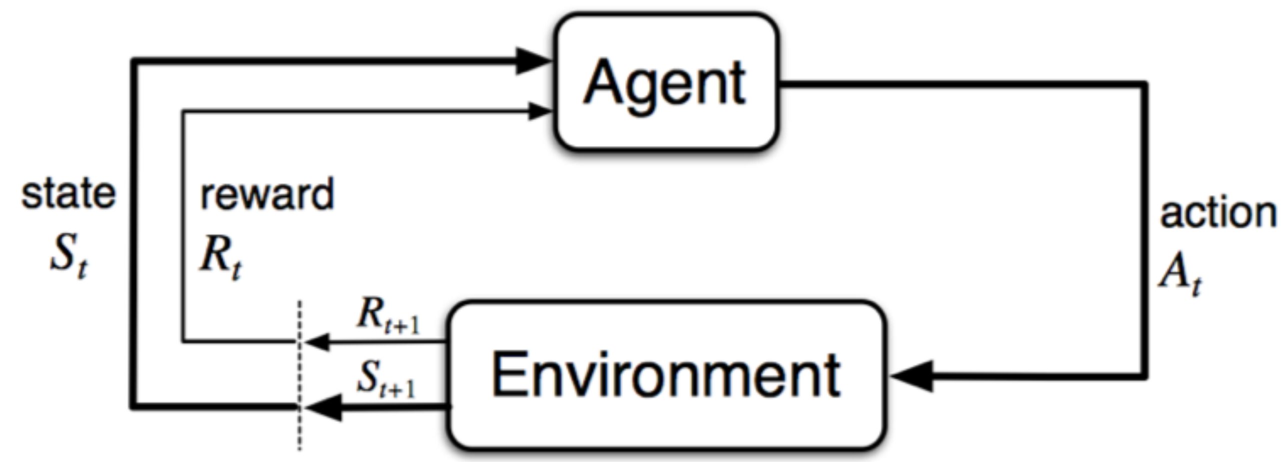


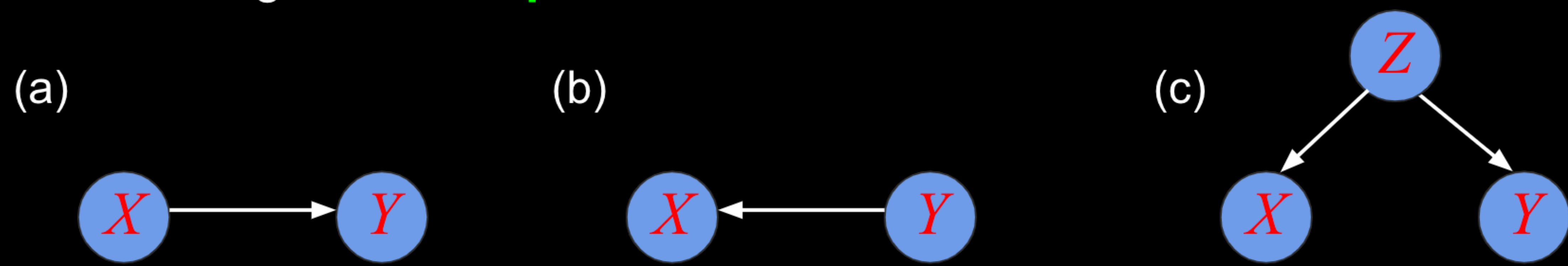
Figure 1: The agent-environment feedback loop [Sutton and Barto, 1998]

Hypothesis 1 (The Reward Hypothesis). *That all of what we mean by goals and purposes can be well thought of as the maximization of the expected value of the cumulative sum of a received scalar signal (called reward).*

Association vs. Causation

Principle of Common Cause [Reichenbach, 1991]

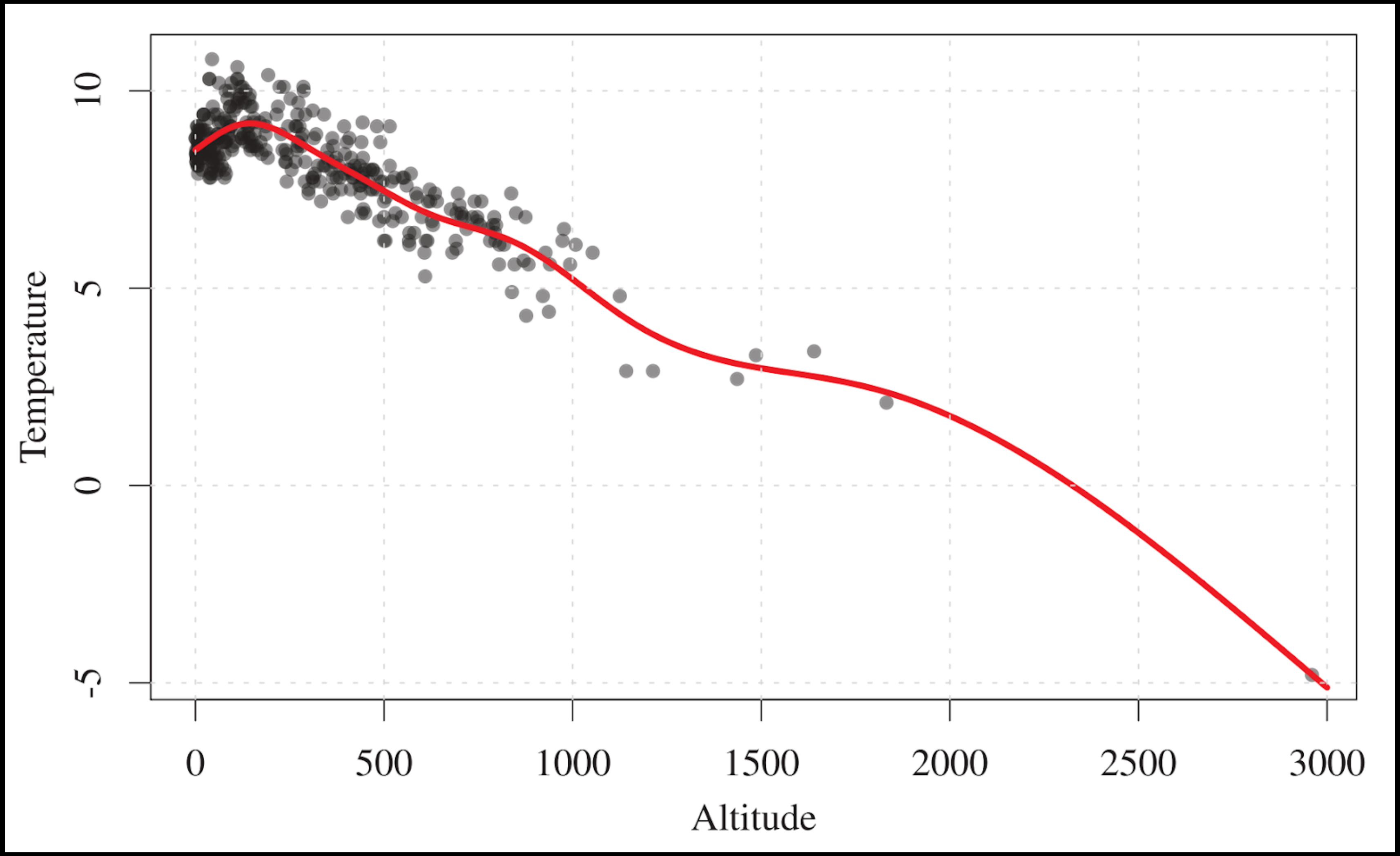
If two random variables X and Y are **statistically dependent**, then one of the following **causal explanations** must hold:



Causation has two obvious advantages:

1) Predict what would happen if some variables are *intervened*.
2) Predict the outcomes of cases that you *never observed before*.

Independent Causal Mechanism



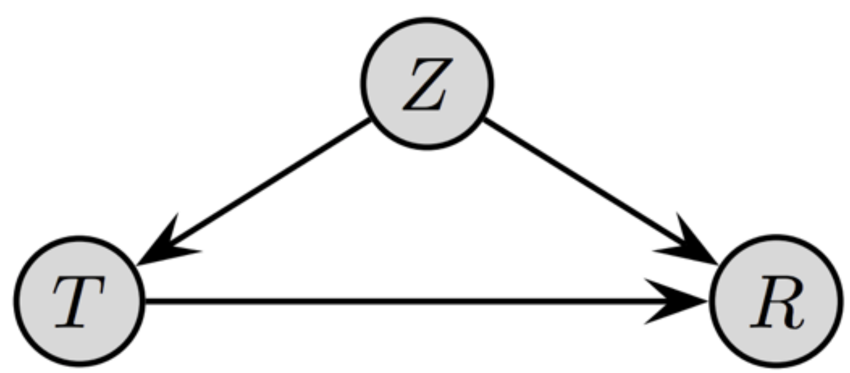
Credit: Elements of Causal Inference

$$\begin{aligned} p(a,t) &= p(a|t)p(t) \qquad T \to A \\ &= p(t|a)p(a) \qquad A \to T \end{aligned}$$

Confounder

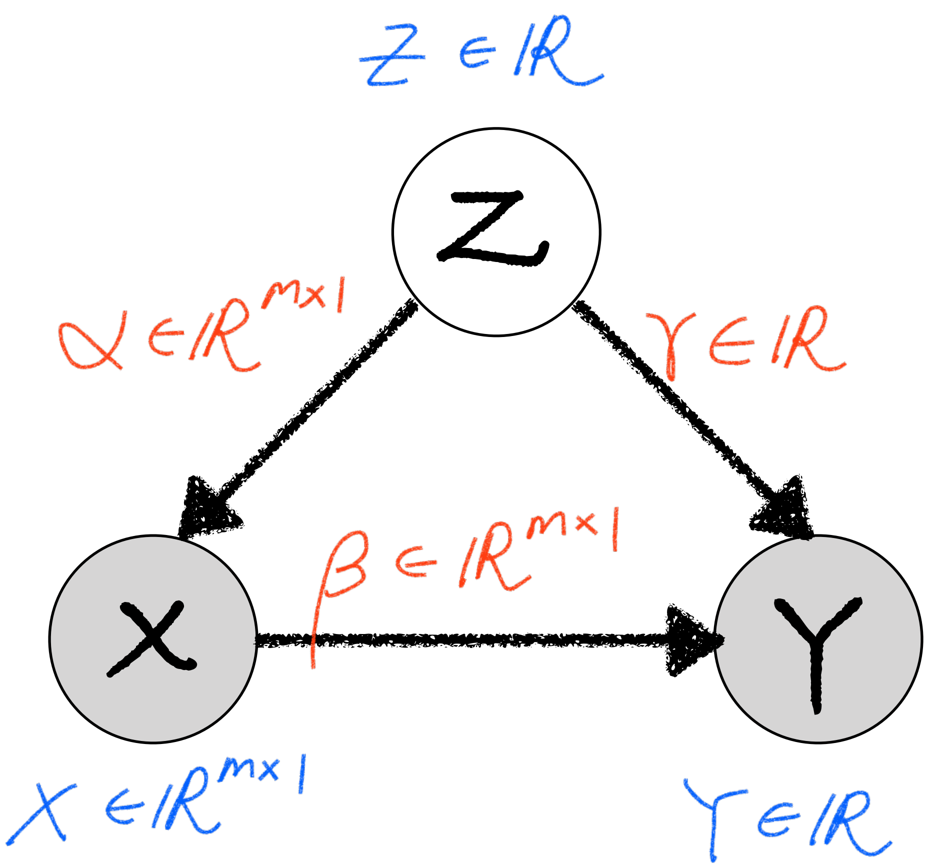
	Overall	Patients with small stones	Patients with large stones
Treatment *a*: Open surgery	78% (273/350)	**93%** (81/87)	**73%** (192/263)
Treatment *b*: Percutaneous nephrolithotomy	**83%** (289/350)	87% (234/270)	69% (55/80)

Credit: Elements of Causal Inference



$$P(R=1 \mid do(T=1)) = \sum_{z=\{0,1\}} P(R=1 \mid T=1, z)\, P(z)$$

Latent Confounders



$$\begin{aligned} Z &:= \epsilon_Z \\ X &:= \alpha' Z + \epsilon_X \\ Y &:= \beta' X + \gamma Z + \epsilon_Y \end{aligned}$$

$\epsilon_W \sim N(0, \sigma_W^2)$ for $W \in \{X, Y, Z\}$

$\epsilon_X \in \mathbb{R}^{m\times 1}$, $\epsilon_Z, \epsilon_Y \in \mathbb{R}$

$$\Sigma_{XYZ} = \begin{pmatrix} \Sigma_{ZZ} & \Sigma_{ZX} & \Sigma_{ZY} \\ \Sigma_{XZ} & \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YZ} & \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}$$

$\mathbb{R}^{m\times m}$

$\mathbb{R}^{m\times 1}$

$\mathbb{R}^{1\times 1}$

$$\begin{aligned} \Sigma_{XX} &= \alpha\alpha'\sigma_Z^2 + \mathrm{diag}(\sigma_X^2) \\ \Sigma_{XY} &= \Sigma_{XX}\beta + \gamma\sigma_Z^2\alpha \\ \Sigma_{YY} &= (\beta'\alpha + \gamma)^2\sigma_Z^2 + \beta'\,\mathrm{diag}(\sigma_X^2)\beta + \sigma_Y^2 \end{aligned}$$

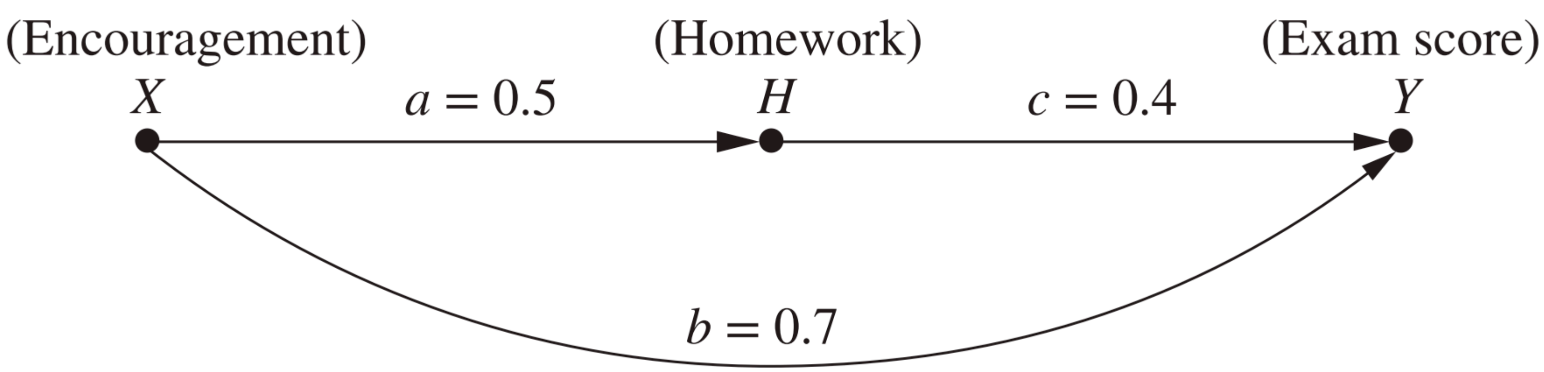
When $m > 3$

$$\boxed{(\alpha_1, \beta_1, \gamma_1, \sigma_{Z,1}^2, \sigma_{X,1}^2, \sigma_{Y,1}^2) \neq (\alpha, \beta, \gamma, \sigma_Z^2, \sigma_X^2, \sigma_Y^2)}$$

3m + 3 unknown parameters

[Alex D'Amour's Blog]

Counterfactuals



$$
\begin{aligned}
X &= U_X \\
H &= a \cdot X + U_H \\
Y &= b \cdot X + c \cdot H + U_Y
\end{aligned}
$$

Let us consider a student named Joe, for whom we measure X = 0.5, H = 1, and Y = 1.5. Suppose we wish to answer the following query: What would Joe's score have been had he doubled his study time?

$$
\begin{aligned}
U_X &= 0.5, \\
U_H &= 1 - 0.5 \cdot 0.5 = 0.75, \text{ and} \\
U_Y &= 1.5 - 0.7 \cdot 0.5 - 0.4 \cdot 1 = 0.75.
\end{aligned}
$$

$$
\begin{aligned}
Y_{H=2}(U_X = 0.5, U_H = 0.75, U_Y = 0.75) &= 0.5 \cdot 0.7 + 2.0 \cdot 0.4 + 0.75 \\
&= 1.90
\end{aligned}
$$

Judea Pearl. Causality, 2009.

Identification

- Identification in Causal Reasoning

Interventional prob. ⟶ Observational prob.

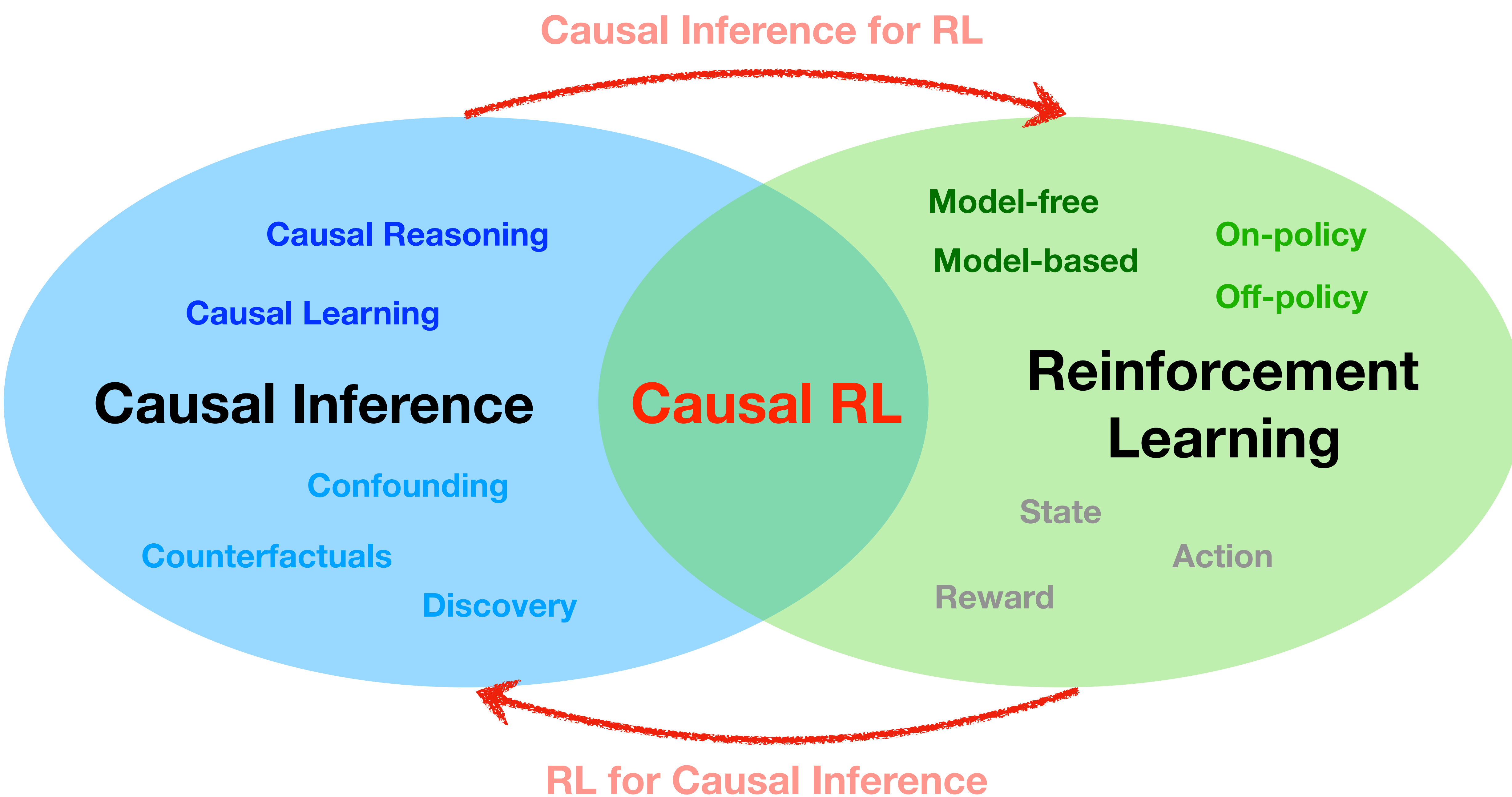
- Identification in Causal Learning

Uniqueness of Causal Orientation

- Identification in Latent Confounder Models

Uniqueness of Causal Strength

Causal RL

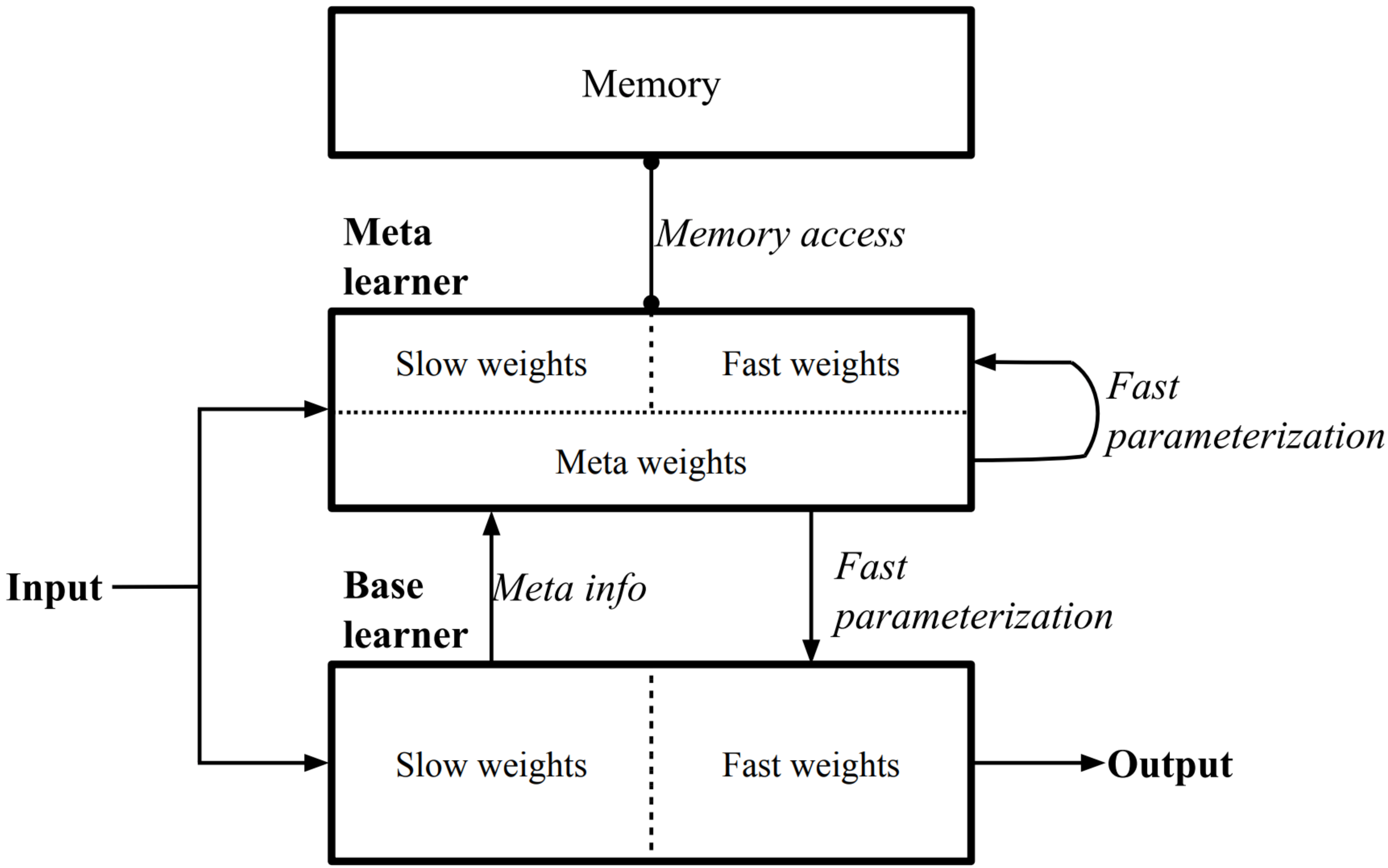

Causal Inference for RL
Causal Reasoning
Causal Learning
Causal Inference
Confounding
Counterfactuals
Discovery
Causal RL
Model-free
Model-based
On-policy
Off-policy
Reinforcement Learning
State
Action
Reward
RL for Causal Inference

Connections to Machine Learning

Causal RL in Transfer Learning

https://youtu.be/hx_bgoTF7bs

Causal RL in Meta Learning



Munkhdalai et al. Meta Networks, 2017

Causal RL in Meta RL

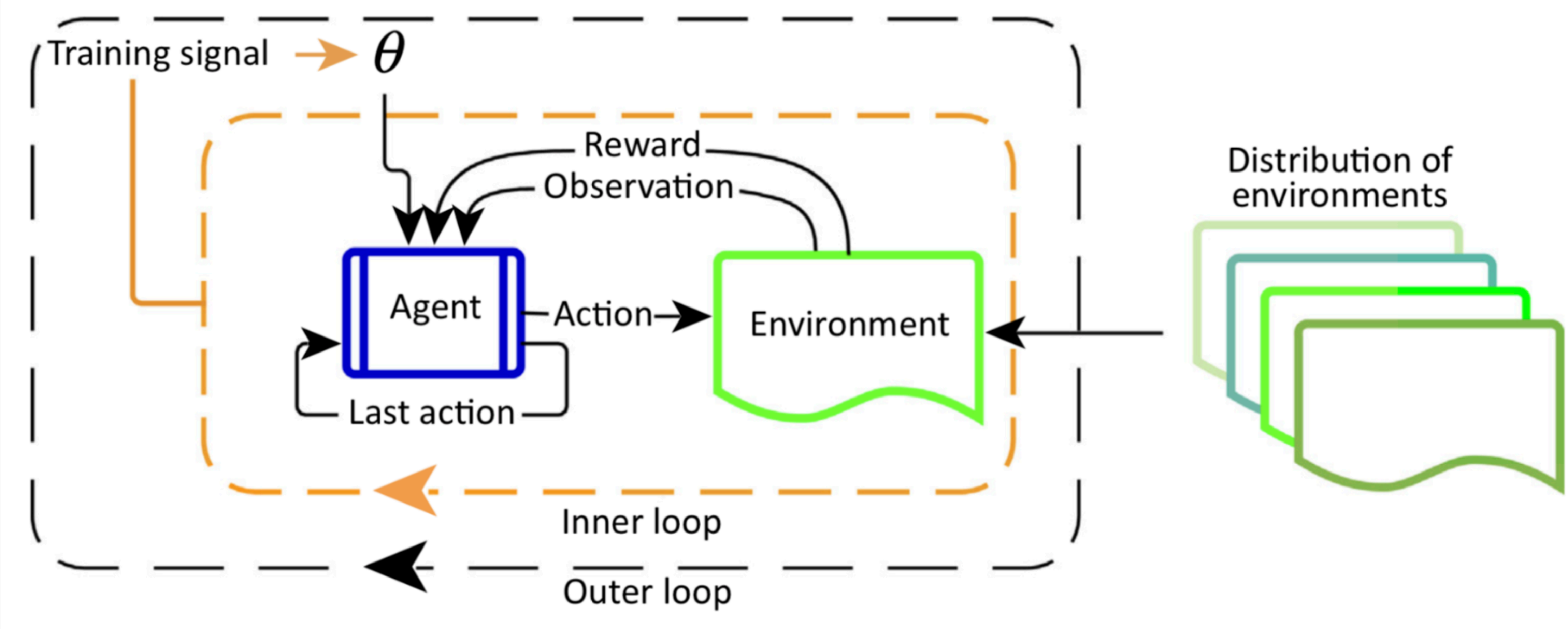


Fig. 2. Illustration of meta-RL, containing two optimization loops. The outer loop samples a new environment in every iteration and adjusts parameters that determine the agent's behavior. In the inner loop, the agent interacts with the environment and optimizes for the maximal reward. (Image source: Botvinick, et al. 2019

Causal RL in Multi-Agent RL

Challenge I: Joint Action Space

concerning result by [Lowe et al., 2017] shows that for a simple setting of binary actions, the probability of taking a gradient step in the correct direction decreases exponentially with the number of agents. Formally

$$Pr\left[\langle\hat{\nabla}J, \nabla J\rangle > 0\right] \propto 0.5^N \tag{26}$$

where the agent's policy is initialzed to an uninformed policy s.t. $\pi(a = 1|s) = 0.5$, N is the number of agents and $\hat{\nabla}J$ is the gradient estimate from a single sample.

Sanyam Kapoor. Multi-Agent Reinforcement Learning: A Report on Challenges and Approaches, 2018

Causal RL in Multi-Agent RL

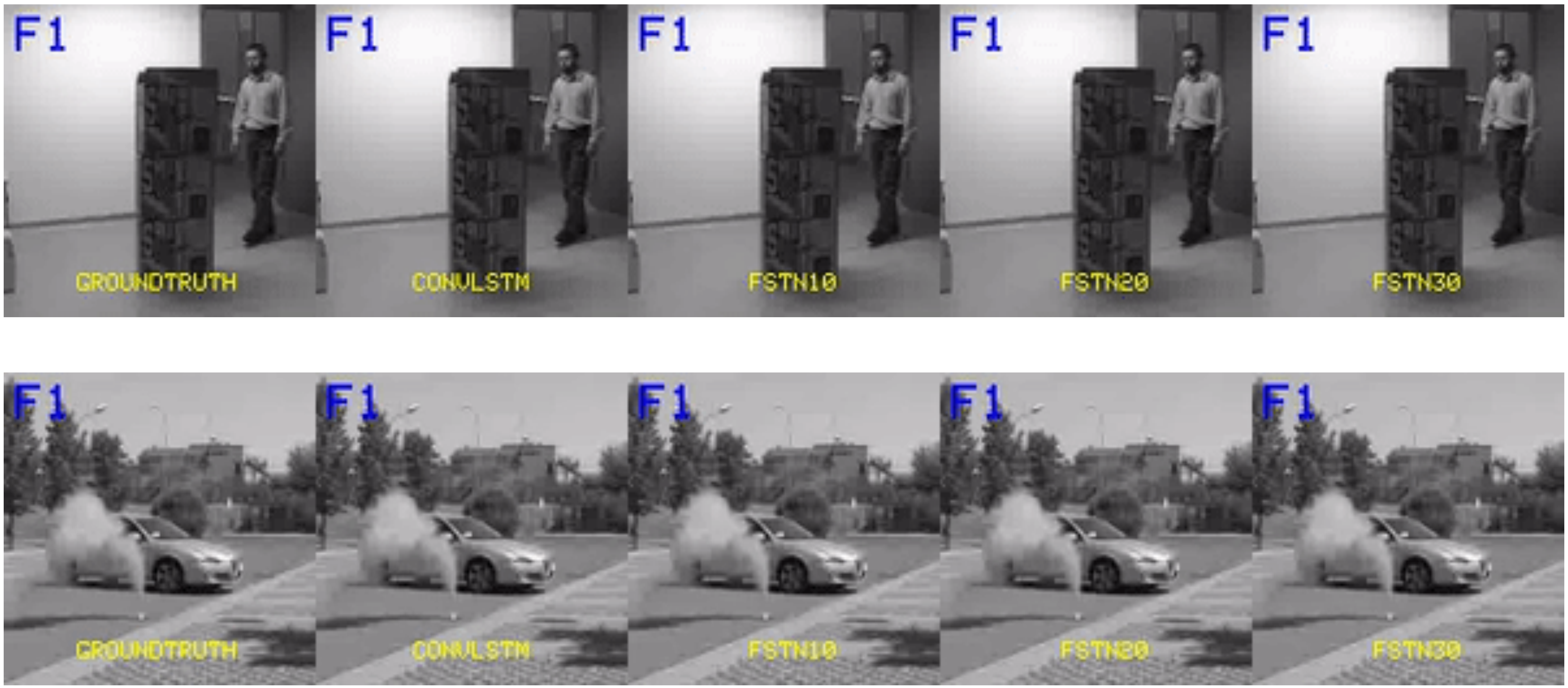
Challenge II: Common Knowledge of Rationality

Common knowledge of rationality is a more subtle requirement. Not only do we both have to be rational, but I have to know that you are rational. I also need a second level of knowledge: I have to know that you know that I am rational. I need a third level of knowledge as well: I have to know that you know that I know that you know I am rational. And so on to deeper and deeper levels. Common knowledge of rationality requires that we are able to continue this chain of knowledge indefinitely.

Pastine et al. Introducing Game Theory, 2017

Potential Applications in Various Areas

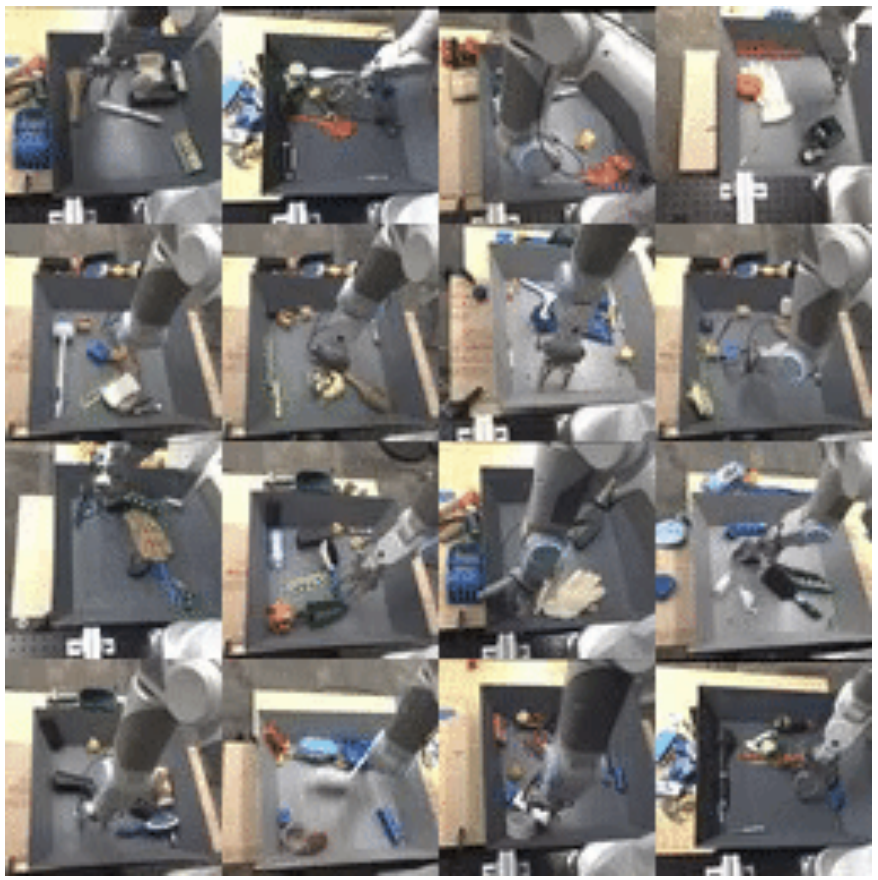
Causal RL for Vision



Flexible Spatio-Temporal Networks
(Lu et al. 2017)

Loving Vincent

Causal RL for Robotics



Video Pixel Networks (Kalchbrenner et al. 2016)

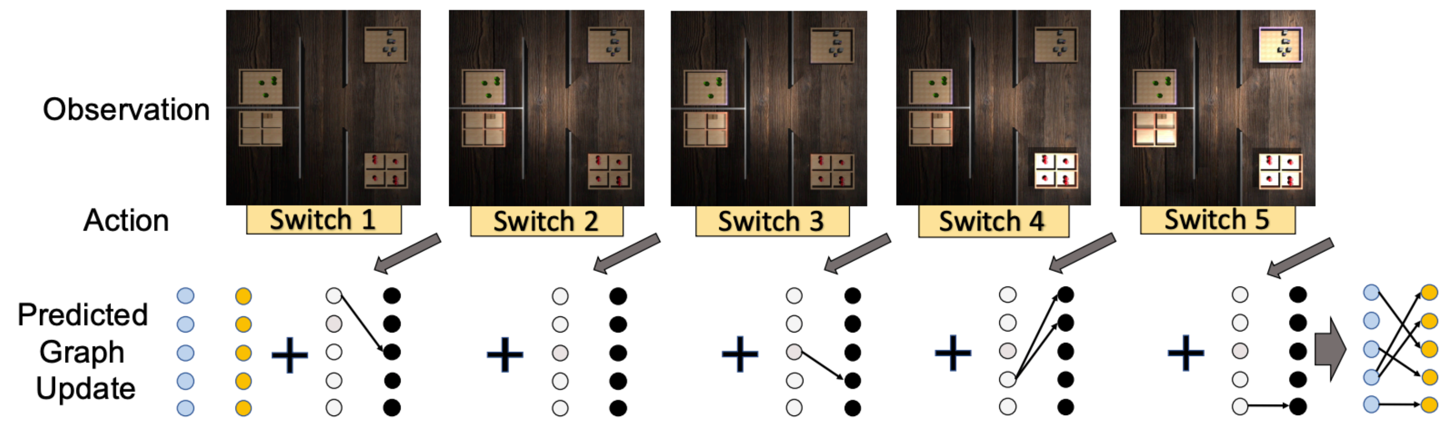


Figure 6: **Sample of Causal Induction**. Here we show an example of our Iterative Causal Induction Model for 5 switches, in the "One-to-Many" case. Given the trajectory of actions and images of the scene, the model needs to reason about which lights were turned on, and how what update this implies in the graph. In this example, the first observed action turns on one of the switches, and the model makes the corresponding update to the graph. The next switch does not change the lighting so the model outputs no update to the graph. The next action sees one light go on, and updates the corresponding switch. The next action turns on two lights, and the graph is updated to reflect this. Lastly, since one light remains unaccounted for, the model knows to add that edge to the graph. Note: The edges and updates are soft updates, but the model learns to predict close to exactly 1 for edges and exactly 0 for non-edges.

Causal Induction from Visual Observations for Goal Directed Tasks (Nair et al. 2019)

Causal RL for Self-driving

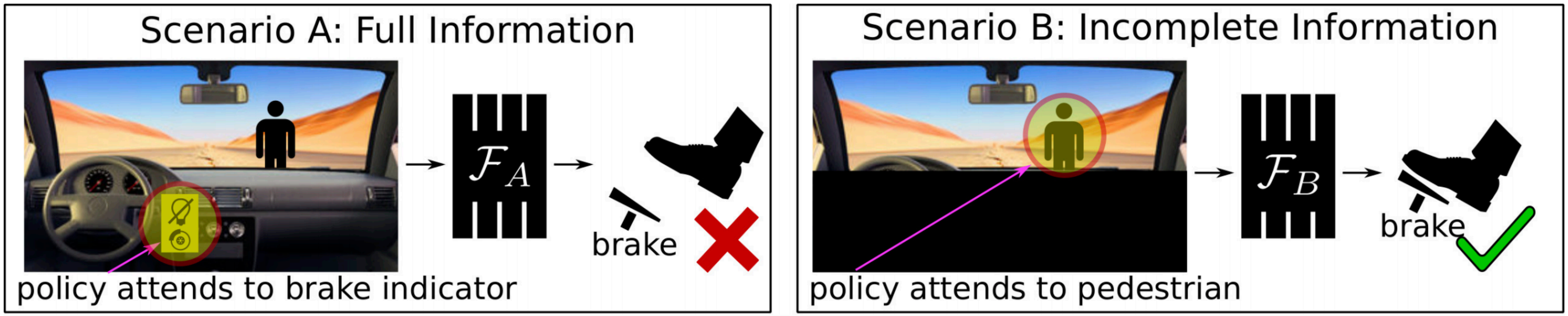


Figure 1: Causal misidentification: *more* information yields worse imitation learning performance. Model A relies on the braking indicator to decide whether to brake. Model B instead correctly attends to the pedestrian.

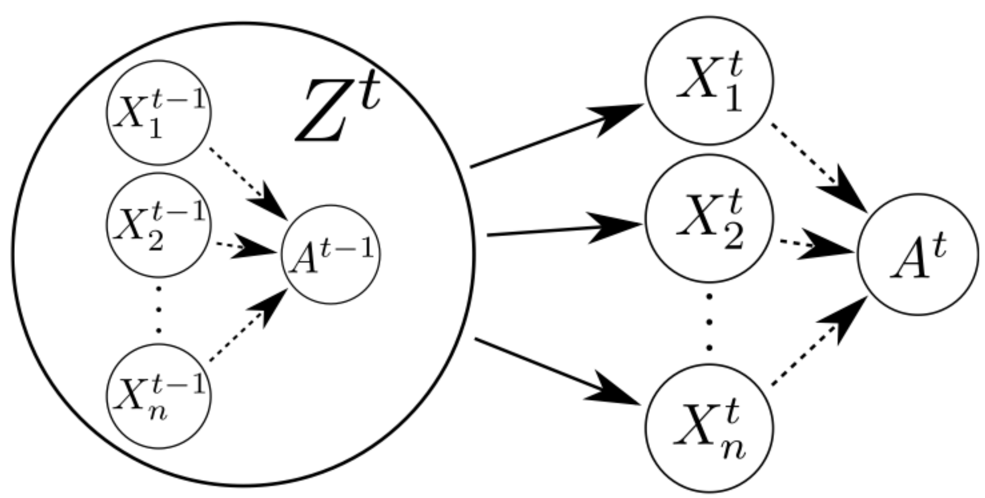
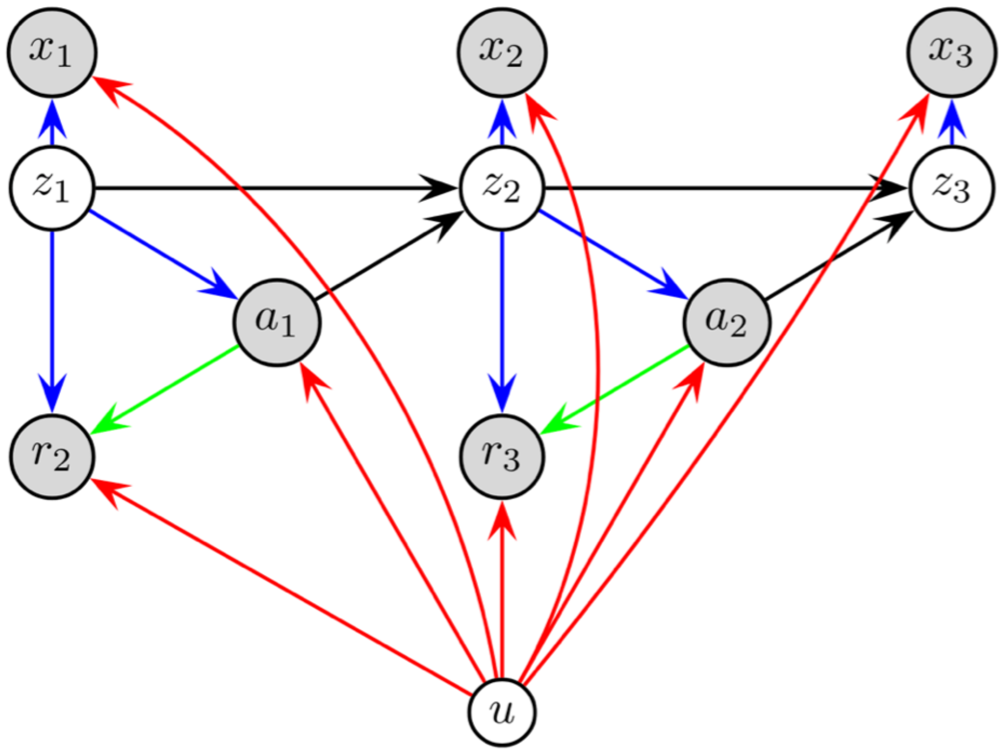


Figure 2: A graph of the underlying causal dynamics of imitation learning. Parents of a node represent its causes. State variables $\{X_i^t\}_{i=1}^n$ are fully observed.

de Hann et al. Causal Confusion in Imitation Learning, 2018

Causal RL for Healthcare/Medicine/Finance



$$p(r_{t+1} \mid z_t, a_t)$$

$$\downarrow$$

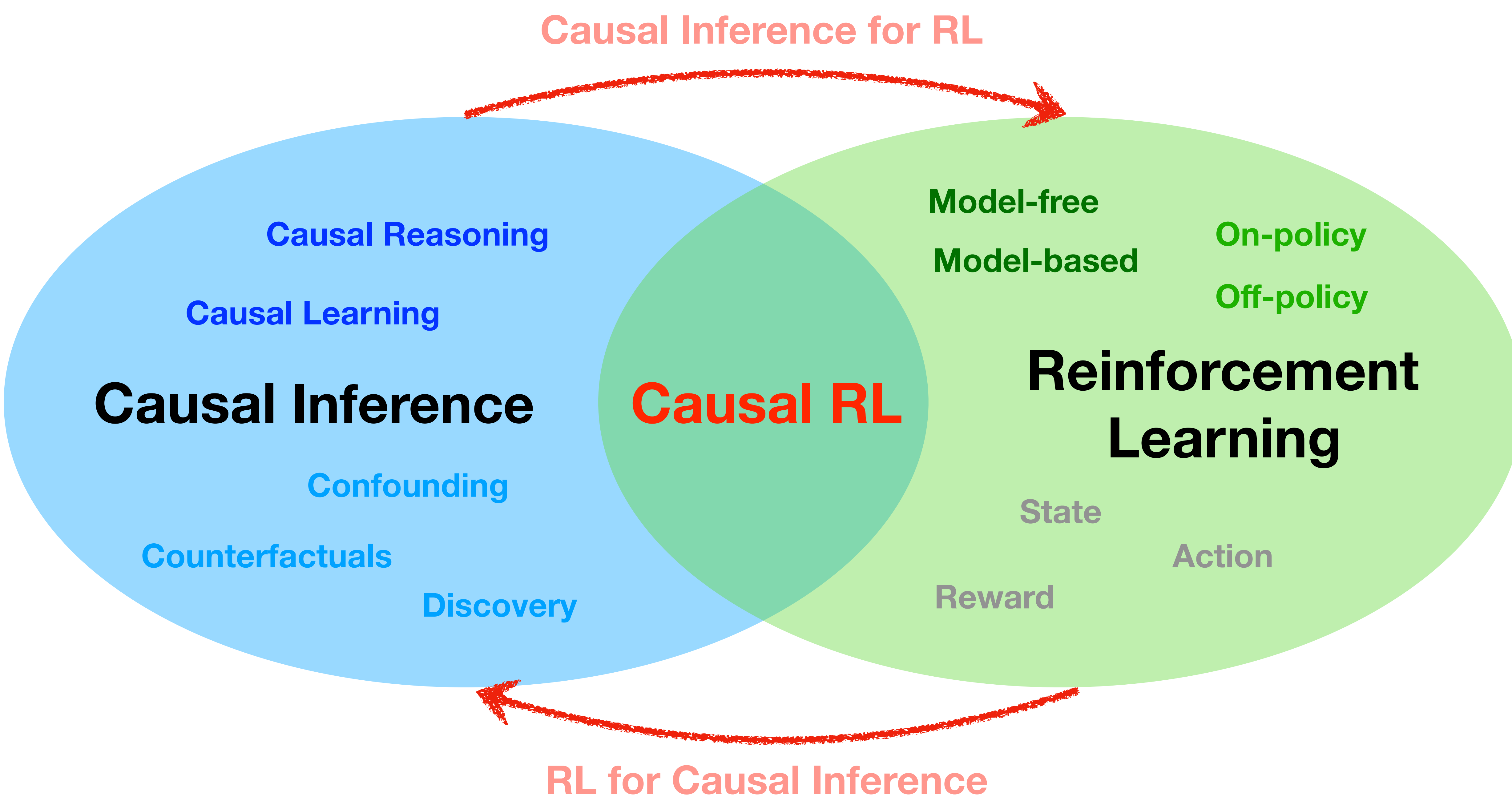
$$p(r_{t+1} \mid z_t, do(a_t))$$

Theorem 3. *Under Assumptions 1 & 2, let* $\pi_b^* = \arg\max_{\pi_b} v_{\pi_b}(s_t)$, $\pi_{t_i}^* = \arg\max_{\pi_{t_i}} v_{\pi_{t_i}}(b_t)$ *and* $\pi_{t_a}^* = \arg\max_{\pi_{t_a}} v_{\pi_{t_a}}(b_t^A)$ *where* $s_t = (z_t, u) \in \mathcal{S}$, b_t *and* b_t^A *are the belief state corresponding to* z_t *and the augmented belief state corresponding to* s_t, *respectively. For any* s_t, *the following statement holds:* $v_{\pi_{t_a}^*}(s_t) \leq v_{\pi_{t_i}^*}(s_t) = v_{\pi_b^*}(s_t)$.

Lu et al. Deconfounding RL, 2018 & Batch OPL, 2019

Conclusion

Causal RL


Causal Inference for RL
Causal Reasoning
Causal Learning
Causal Inference
Confounding
Counterfactuals
Discovery
Causal RL
Model-free
Model-based
On-policy
Off-policy
Reinforcement Learning
State
Action
Reward
RL for Causal Inference

The Take-home Message

Causal RL was born for AGI.

Causality for Machine Learning

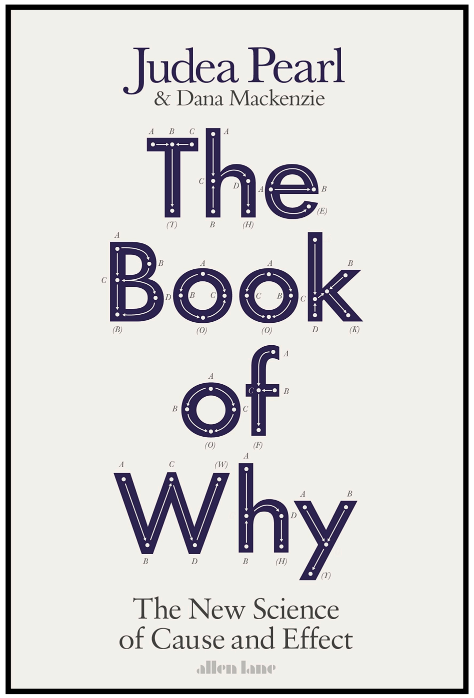
Bernhard Schölkopf
Max Planck Institute for Intelligent Systems, Max-Planck-Ring 4, 72076 Tübingen, Germany
bs@tuebingen.mpg.de

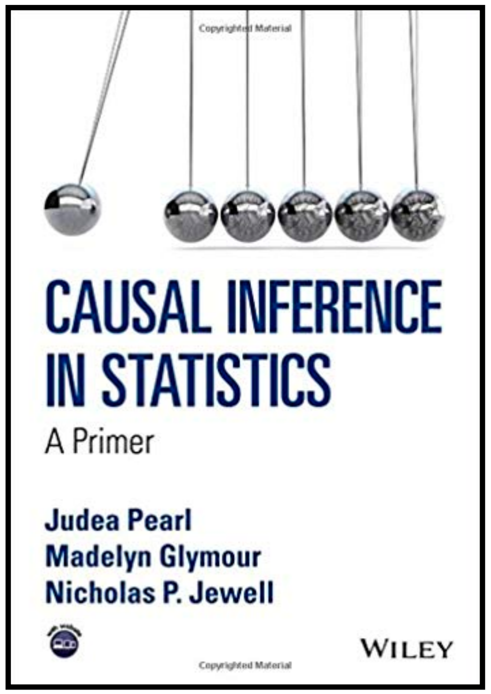
Abstract

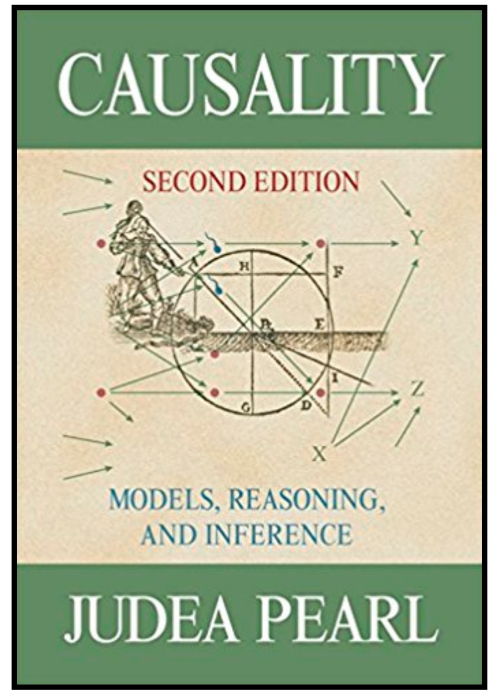
Graphical causal inference as pioneered by Judea Pearl arose from research on artificial intelligence (AI), and for a long time had little connection to the field of machine learning. This article discusses where links have been and should be established, introducing key concepts along the way. It argues that the hard open problems of machine learning and AI are intrinsically related to causality, and explains how the field is beginning to understand them.

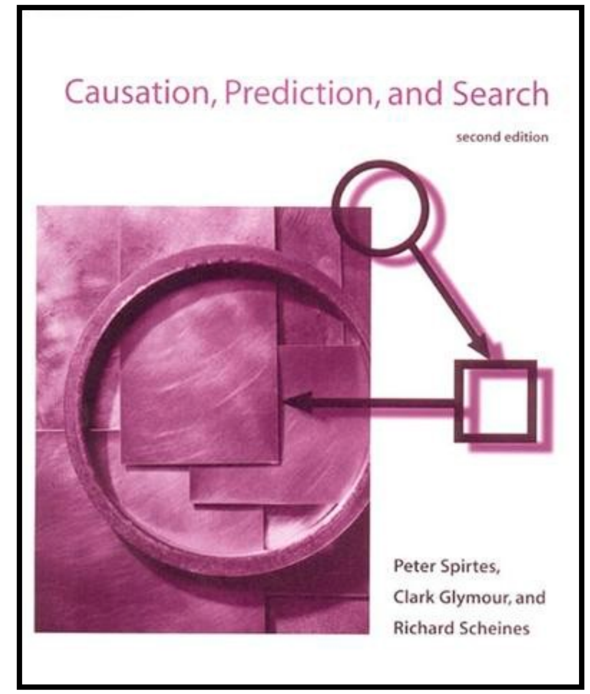
arXiv:1911.10500

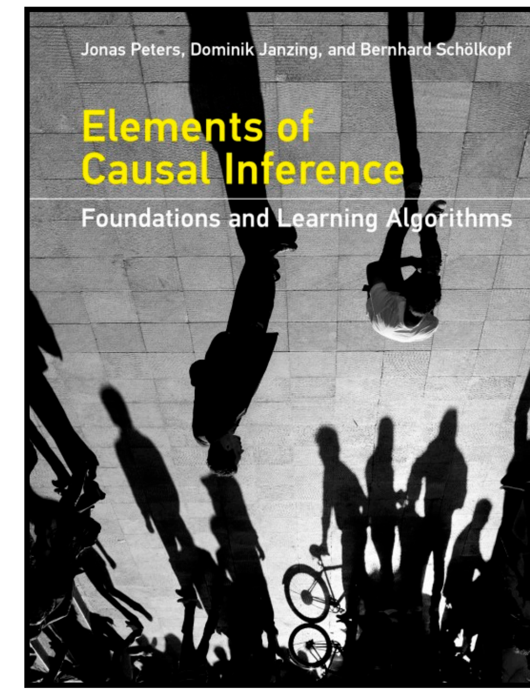
Recommendation

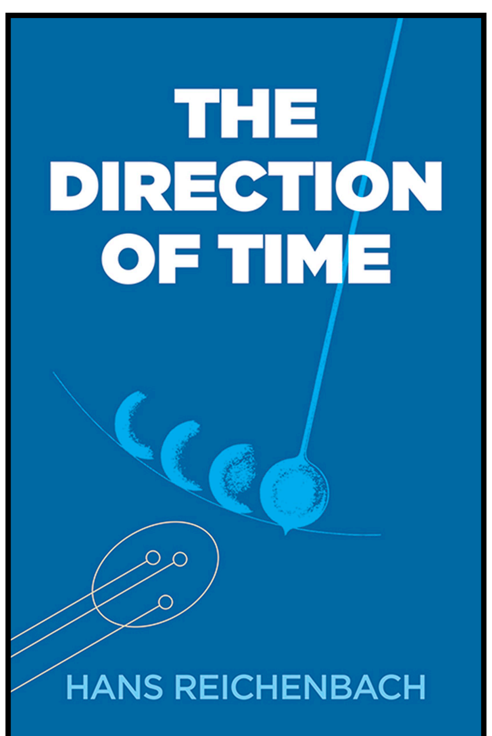


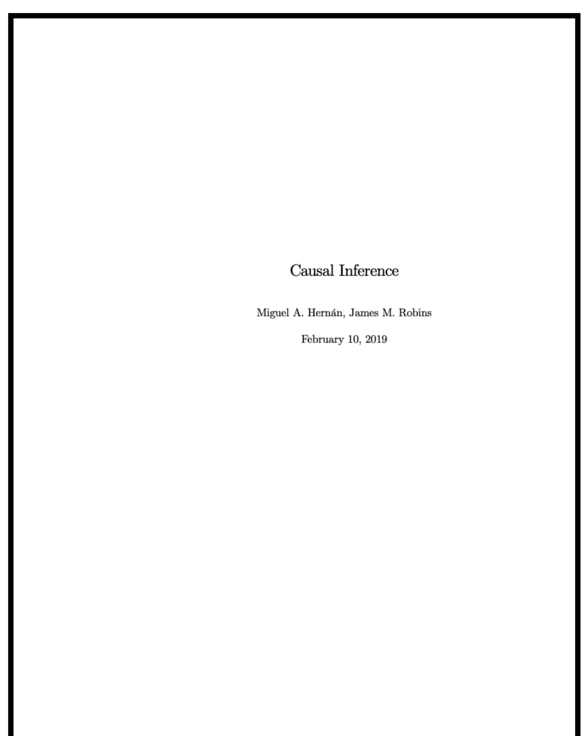


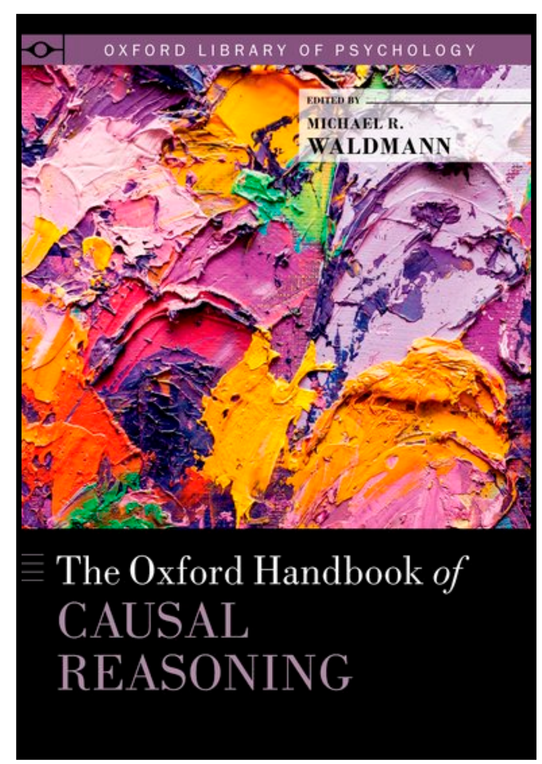


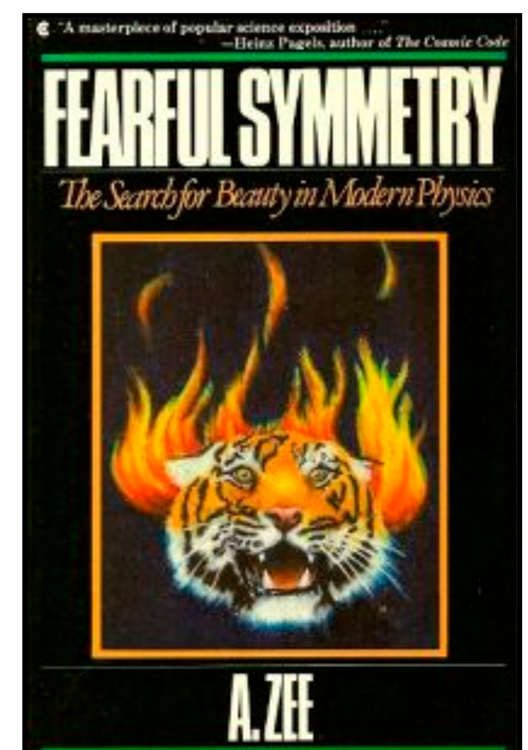












Thank You
&
Question