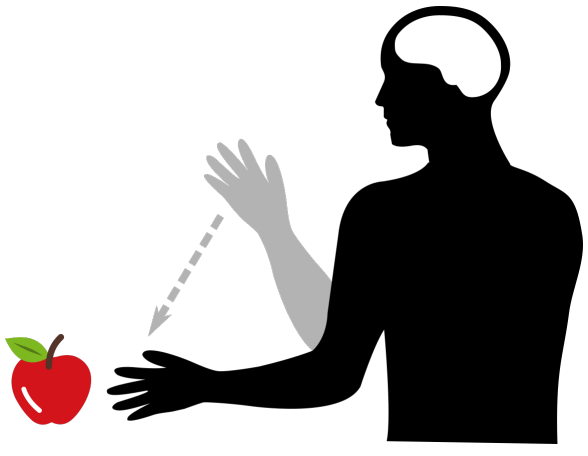
"Mental-Space Continuation" of Reinforcement Learning

In conventional RL, the **environment** is physically observable. I propose to extend it to the **internal** mental space.

From the traditional RL perspective: an agent reaches out for an apple, the apple is the **reward**, moving the arm is an **action**. These are all **observables**:

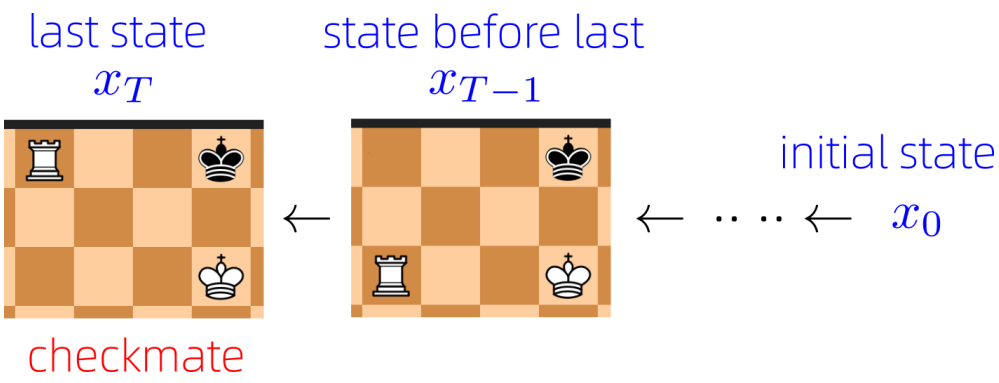


(1)

The foundation of RL is the **Bellman equation**. It can be viewed as a **recursive** formula:

$$\boxed{\text{Current state}} \quad V(x_0) = \max_a\{R + \gamma V(x_1)\} \quad \boxed{\text{Next state}} \tag{2}$$

It propagates the final state's reward **back** to the previous state, and the state before that... just like in a chess game... and so on until the very first move:

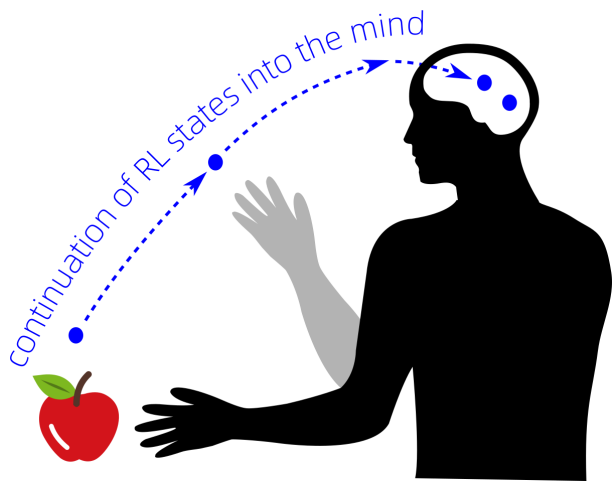


(3)

In other words, the reward of getting an apple *back-propagates*[1] to the action of reaching out an arm for the apple. So far so good. But we can continue this process back to the **chain of thoughts** that decided to reach for an apple:

$$\text{hungry} \to \text{need to find food} \to \text{see an apple} \to \text{apple is food} \to \ldots. \tag{4}$$

In other words, we turn our **internal** mental states "inside-out", viewing them on an equal footing as **external** states:



(5)

And this is exactly analogous to the propagation of rewards in a chess game. In other words, we can apply techniques of RL to learn the contents of mental space, with a very rigorous foundation.

[1] Note that this is not the same as "back-prop" in deep learning.

Learning of logic rules under RL

The approach of unifying internal and external states is philosophically entirely sound, as "brain states" are physical states too, and their transitions are learned via **Hebbian learning**.

So how does RL learn logical content? An **action** here is a transition from one logic state (set of propositions) to another, ie, a logic **rule**. The agent tries to learn the best actions (= logic rules) among all possible rules that are *applicable* to the current state. For example:

$$\text{I'm hungry} \rightarrow \text{need to find food} \tag{6}$$

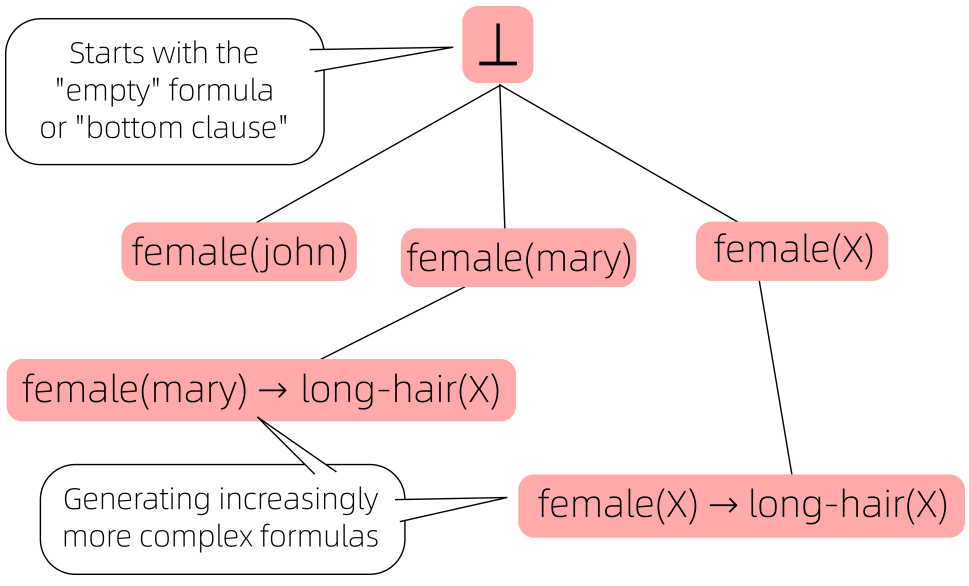
is a good rule;

$$\text{hungry} \rightarrow \text{eat feces} \tag{7}$$

is not so good.

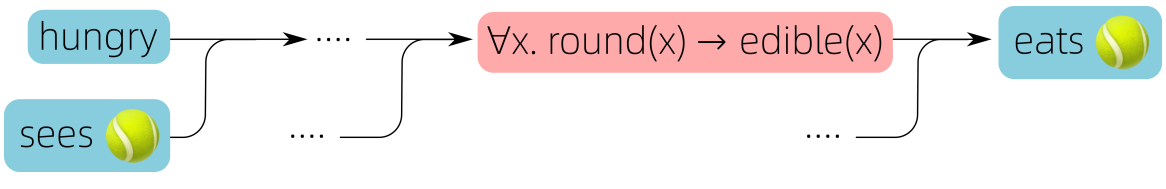
The advantage of this RL framework is that it may be able to learn highly **abstract** rules from the vast mental space, just like it learns to find solutions out of complicated **mazes** in Atari games.

In classical AI, the **combinatorial** search of logic rules has been studied, with **search trees** like this:



(8)

A special characteristic of **mental space** is that any thought can potentially be deduced from any other thoughts. In other words, any two points in thoughts space may be connected by a **path** (= logic rule = action). For example, if an agent is hungry and sees a tennis ball, the rule that "round implies edible" may satisfy the objective of finding food, only to be punished when one actually eats the object:



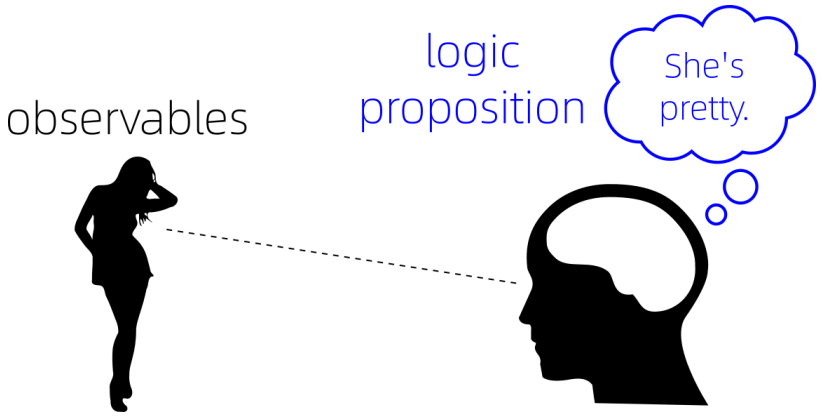
(9)

This negative reward would back-propagate through the **inference chain**, and after many iterations, the system would eventually find the culprit rule. In short, a logical system can get "as crazy as it wants" without other prior constraints. Indeed, some geniuses arrive at non-obvious conclusions because they are a bit "crazy". This property of mental space is common to all learning-reasoning systems, not just my architecture.

The mind as a "model" of the world

In traditional RL, there are **model-based** methods, and in our brain we build **mental models** of the world. These are actually one and the same concept.

As is well-known in classical philosophy of logic, a logic proposition (in the mind) corresponds to certain **states** of the world:



(10)

However, in the unified or "flat" view, both internal and external states belong to the same state space, within which, the internal states "model" the external states:[3]

flattened state-space

(observed states of) external world $\models$ internal mind $\models$ meta-logic ?

model

(11)

Does the unified cognitive space has its own "theory"? That may be some kind of **meta-logic**. This meta-theory is a form of **inductive bias** that may be important in accelerating learning on the first-order level.

Picture credits:
Human figure from www.onlinewebfonts.com licensed by CC BY 3.0
Thought bubble created by Catherine Please from the Noun Project

[3]The notation $T \models M$ means: M is a **model** of T; T is a **theory** of M. This is rigorously defined in model theory.