①"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:

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

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:

$$(3)$$
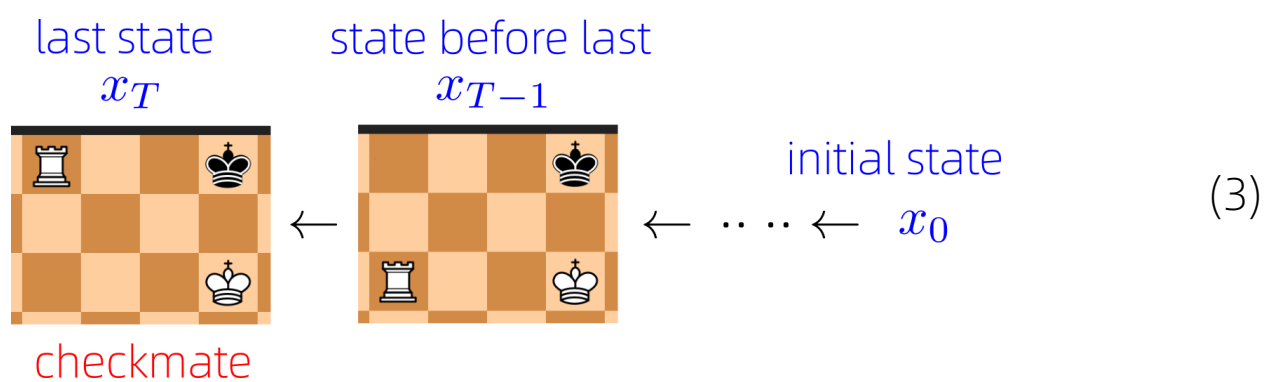
In other words, the reward of getting an apple back-propagates to the action of reaching out an arm for the apple. So far so good. But we can continue this process back to the **chain of thoughts** that decided to reach for an apple:

$$\text{hungry} \to \text{need to find food} \to \text{see an apple} \to \text{apple is food} \to .... \quad (4)$$

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


$$(5)$$
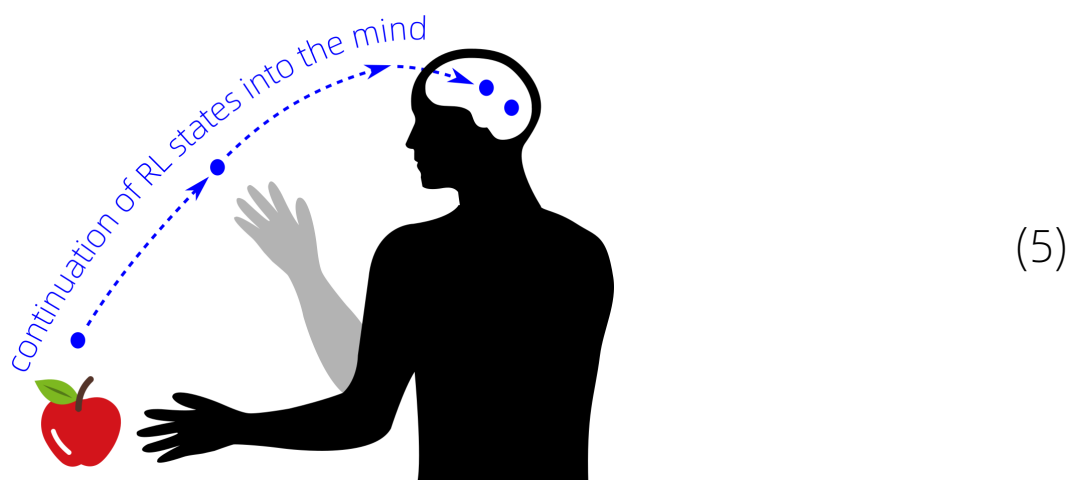
And this is <u>exactly analogous</u> to the propagation of rewards in a chess game. In other words, we can apply techniques of RL to learn the contents of mental space, thus providing a very rigorous foundation for AGI.

The approach of unifying internal and external states is philosophically entirely sound, because our "brain states" are actually physical states; they are electrical activations of neuronal populations, and their transitions are governed by synaptic weights between neurons, which weights are updated through **Hebbian learning**, at least according to our current best understanding.

So how does RL learn logical content? An **action** here is a transition from one logic state to another; or an **inference** from some propositions to a new proposition, ie, a logic **rule**. We need to choose from among many actions (logic rules) the best ones. In other words, choose the best logic rule(s) from all the logic rules that are applicable to the current state. For example, "I'm hungry → need to find food" is a correct rule; "hungry → eat feces" would be a rather bad one.

Advantage of doing this: Theoretically, RL can 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:
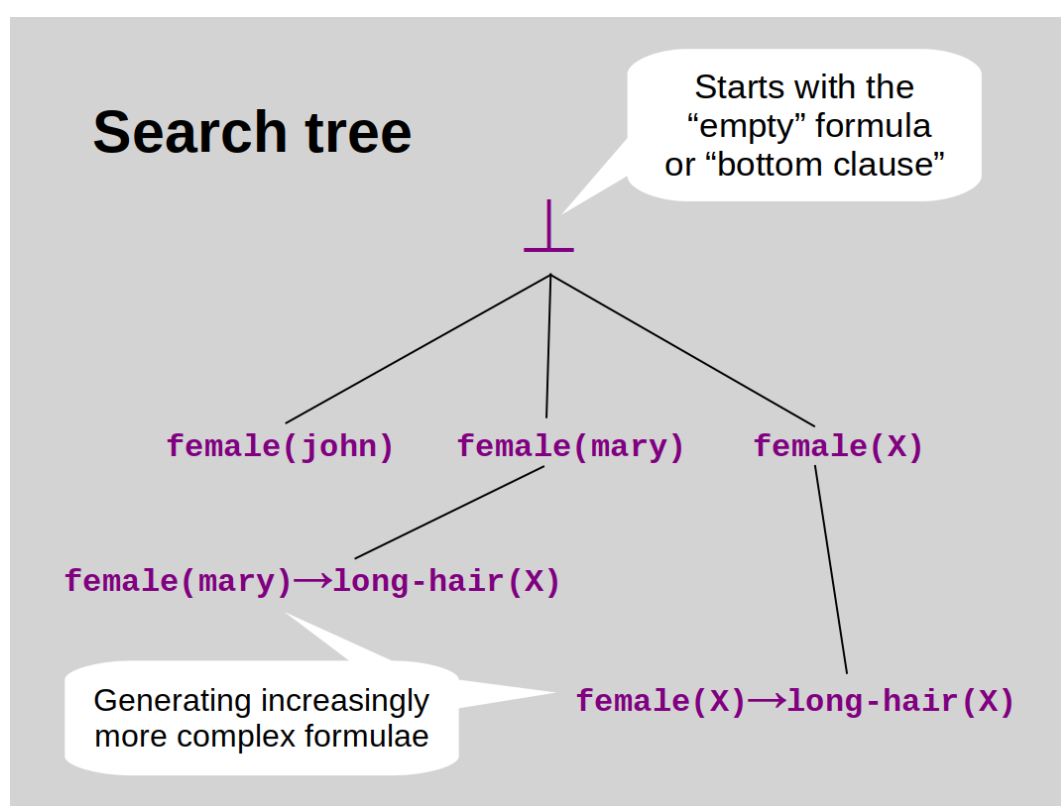
(6)

A special characteristic of **mental space** is that <u>any thought can potentially be deduced from any other thoughts</u>. 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:
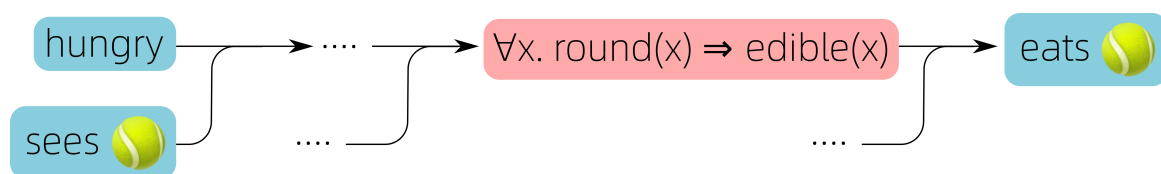

(7)

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.
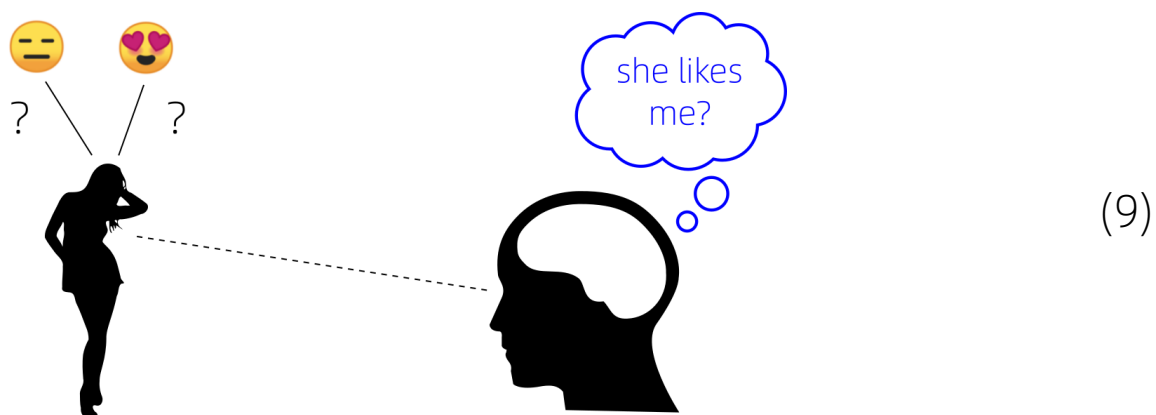
In traditional RL, there are **model-based** methods, and in our brain we build **mental models** of the world. These are actually one and the same concept, there is no conflict between the two.
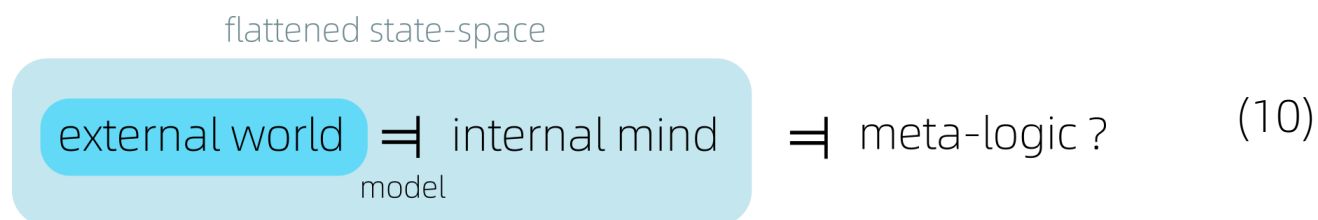
As is well-known in classical philosophy of logic, a logic proposition (in the mind) corresponds to certain **states** of the world:

$$(8)$$
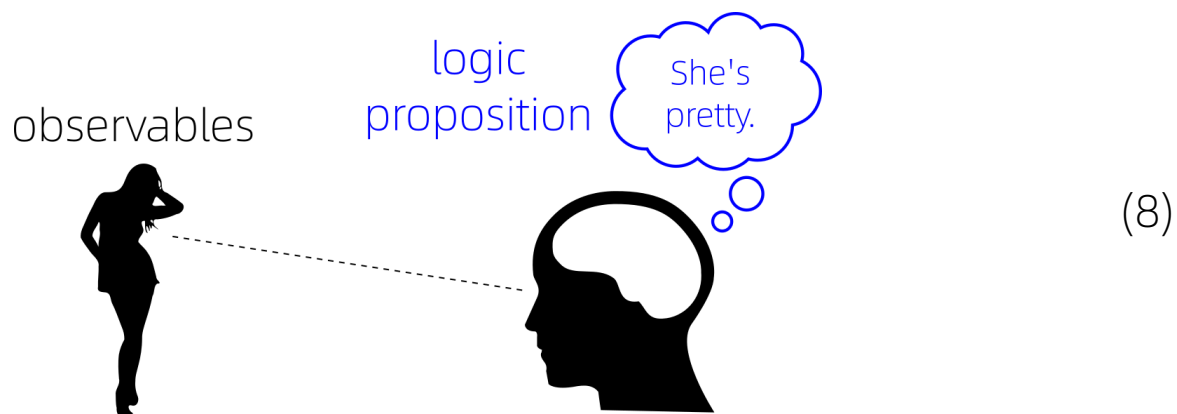
This can be extended to our understanding of "other minds":

$$(9)$$

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

$$(10)$$

Does the mixed state space has its own "theory"? That may be some kind of **meta-logic**.

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 $T \vDash M$ means: M is a **model** of T; T is a **theory** of M. This is rigorously defined in **model theory** in logic.