Reward is Enough

Doina Precup

JBased on David Silver, Satinder Singh, Doina Precup and Rich Sutton

https://www.sciencedirect.com/science/article/pii/S0004370221000862

What is intelligence?

- A collection of abilities and attributes
  - perceive and predict
  - remember and use knowledge
  - plan
  - communicate and deal with other agents
  - ...
- *What drives agents to exhibit these attributes?*
  - Psychology / cognitive science: how do such attributes arise and manifest in natural agents
  - AI: how do we build agents that exhibit these attributes?

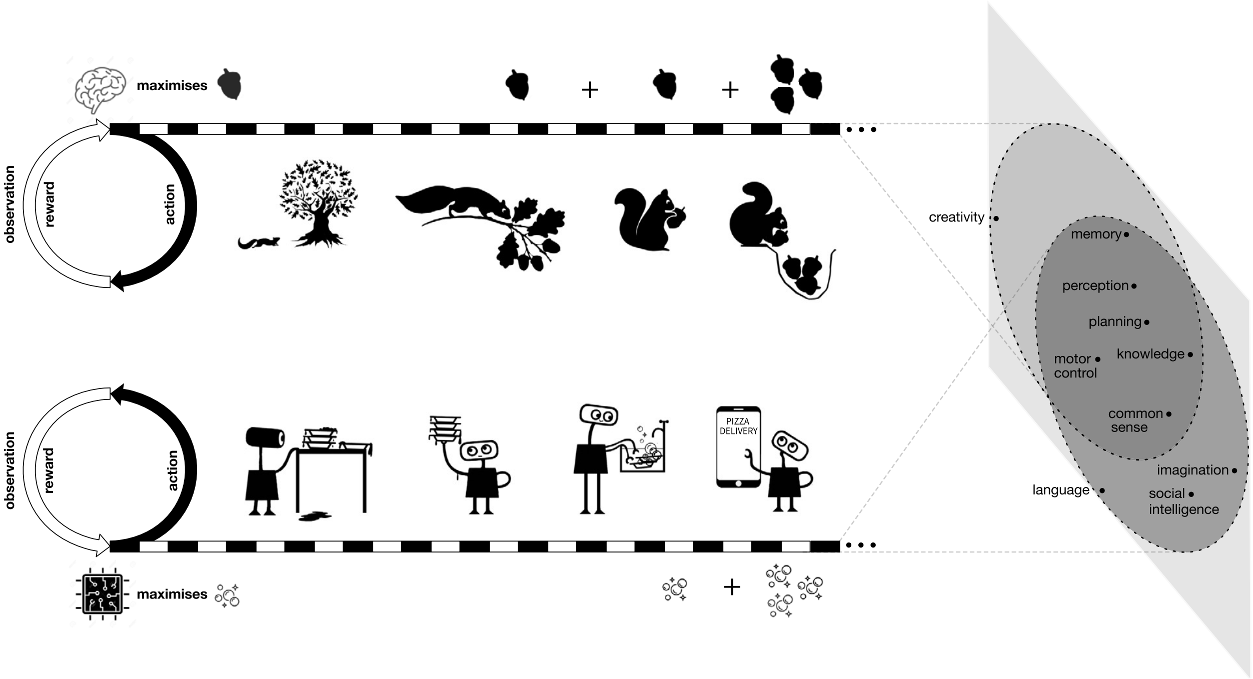
Traditional AI approach

- Each attribute of intelligence could have its own goal, leading to distinct problem formulations
- Examples:
  - Perception: driven by object recognition
  - Language: driven by next-word prediction, or parsing, or sentiment analysis
- This approach has lead to great progress in *specialized directions* of AI research, but great difficulty in putting pieces together
- This paper is a thought experiment: *could the pursuit of reward be viewed as a singular, overarching problem formulation?*

Why reward?

- Different intelligent agents may pursue different goals
  - An animal may want to minimize hunger
  - A Go playing agent may want to maximize wins
  - A kitchen robot way want to maximize cleanliness
- *Rewards provide a flexible representation of goals*

Reward Maximization as a Common Goal



Advantages of a Common Goal

- Deeper understanding: why is each attribute of intelligence important for a particular agent?
- Broader understanding: rich forms of the same attribute can understood in the same way

  Eg. seemingly rational vs irrational behavior
- *Integrated understanding and interpretation*

Example: AlphaGo

- Prior work focused on separate goals:
  - Shape (pattern recognition)
  - Tactics (local search)
  - Endgames (combinatorial game theory)

  AlphaGo focused on a *common goal: maximize number of wins*

  - Led to a deeper understanding of shape, tactics and endgames
  - Produced a broader set of attributes, eg. territory and influence, attack and defence
  - All attributes integrated seemingly into a *unified agent*

Reward-is-Enough Hypothesis

All attributes of intelligence can be understood as subserving the maximization of reward by an agent acting in its environment

Moreover, this may be true for many simple reward signals in many environments

Example: Reward is enough for perception

- Rich, real-world environments may demand various perceptual skills: image recognition, scene parsing, speech recognition...
- Cumulative reward maximization can lead to agents that:
  - Learn from sequences of action and observation (eg find the keys in the pocket)
  - Can optimize for the cost of perception (eg turning the head may take time and be costly)
  - Can specialize to context-dependent data distributions (eg city vs forest)

Example: Reward is enough for language

- Major recent advances have come from a common goal: predicting the next word in a large corpus of data
- However, language modelling may not produce broader linguistic attributes:
  - Language intertwined with other actions and observations
  - Purpose-driven and situated conversations
- Richer language may emerge from a common goal of reward maximization!

What else could be enough?

- Supervised learning: addresses *teacher's environment and goal*

  See also Dewey's work on the importance of experiential learning in human education

- Unsupervised learning: provides no goal for action selection

  Moreover, the world may be too complex to model in its entirety without the focus produced by goals

- Offline learning from large dataset: the real world is much larger than any dataset!!! And much richer

- Evolution: natural intelligence has emerged from maximizing reproductive fitness, but problems faced by AIs can include a much broader range of goals

Are there rich enough environments so that reward is enough?

- Yes! The natural world
- Maybe some similarly rich simulated worlds (though most environments we use in RL research are definitely NOT rich enough)

Does reward maximization make everything too hard?

- Can we even come up with the right reward functions? Is this a very hard problem?

  Under certain circumstances, this is surprisingly easy (see Abel et al, 2021, presented in a couple of weeks)

- Can we maximize reward efficiently?

  This is our challenge as RL researchers!