Reward is enough

David Silver, Satinder Singh, Doina Precup, Rich Sutton

Introduction

What is Intelligence ?

- One of the fundamental question of science
  - Natural Intelligence : how to understand existing intelligence
  - Artficial Intelligence : How to implement & make new Intelligence
- Expressions of intelligence in animal and human behaviour are so bountiful and so varied
  - social intelligence, language, perception, motor control ..
- What could drive agents (natural or artificial) to behave intelligently in such a diverse variety of ways?
  - how can it be that same animal both has knowledge, social intelligence, perception .. ?
  - how can we understand these rich diversity of Intelligence ?

What is Intelligence ?

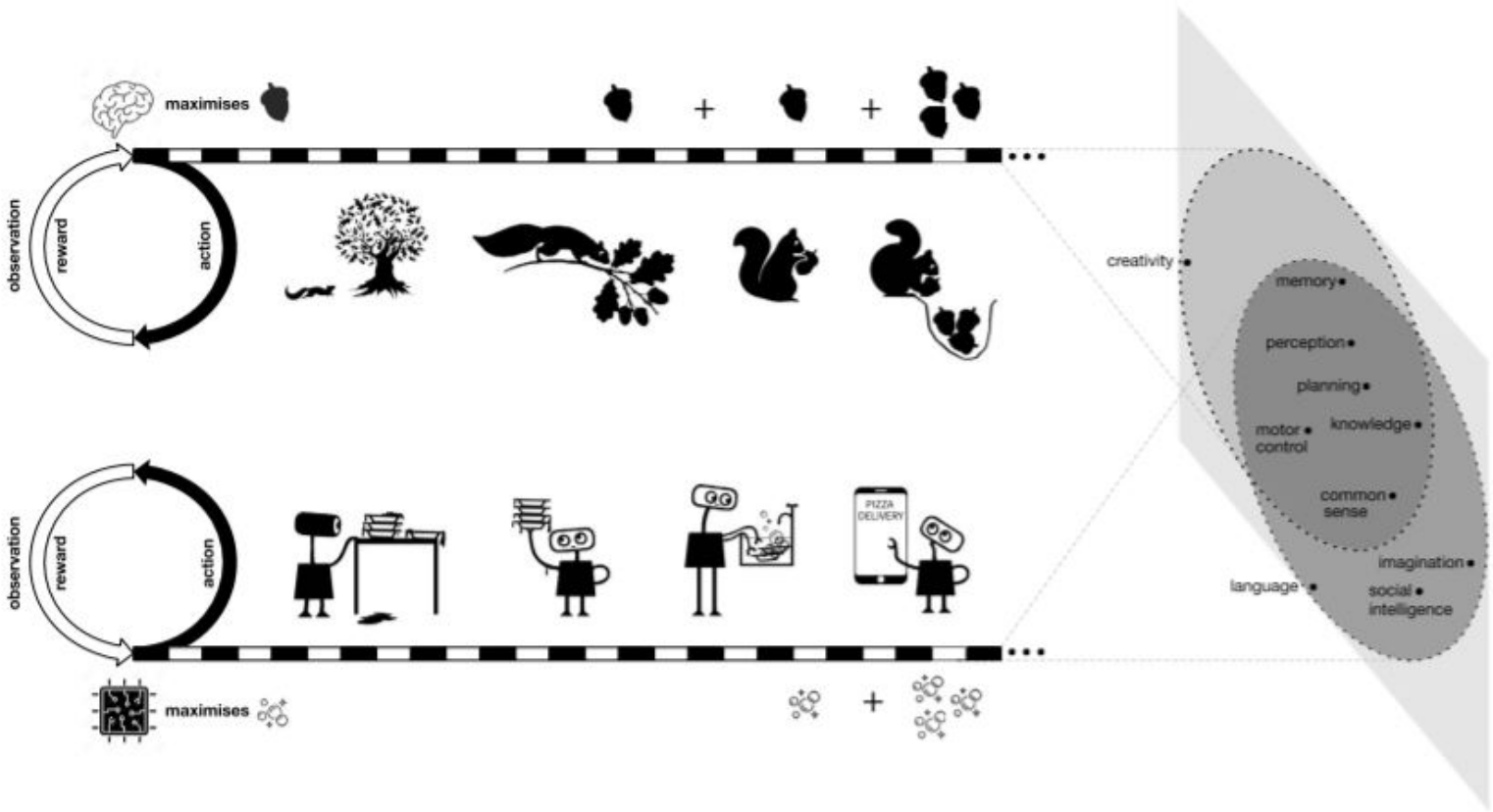
- One possible Answer : each ability arises from the pursuit of specialized goal (specialized problem formuation)

- Traditional approach to AI : break up AI into each field of Intelligence
  - Perception is driven by object recognition
  - Language is driven by tagging, parsing, sentiment analysis, next word prediction(GPT3)
  - Social intelligence is driven by the pursuit of a Nahs equilibrium
  - ..

Goal

- Intelligence may pursue a variety of diffrent goal
  - An animal may act to minimise hunger
  - A go playing agent may act to maximise win
- Reward provide a flexible representation of goals
- whatever the goal it is, Reward itself is enough to understand intelligence
  - why do we need sperate goal and framework to understand them?

Example



- the pursuit of one goal may generate complex behaviour that exhibits multiple abilities associated with intelligence.

Many Attributes, Common Goal

- A complex environment demands many attributes of intelligence
  - ex) perception, languages, social intelligence may be necessary to survie
- Reward maximization may then be viewd as a common goal
  - Perception, language, social intelligence emerge in the pursuit of reward maximization
- Reward maximization may be sufficient to understand intelligence
  - Explaining the diverse expressions intelligence found in natural intelligence

Reward is Enough Hypothesis

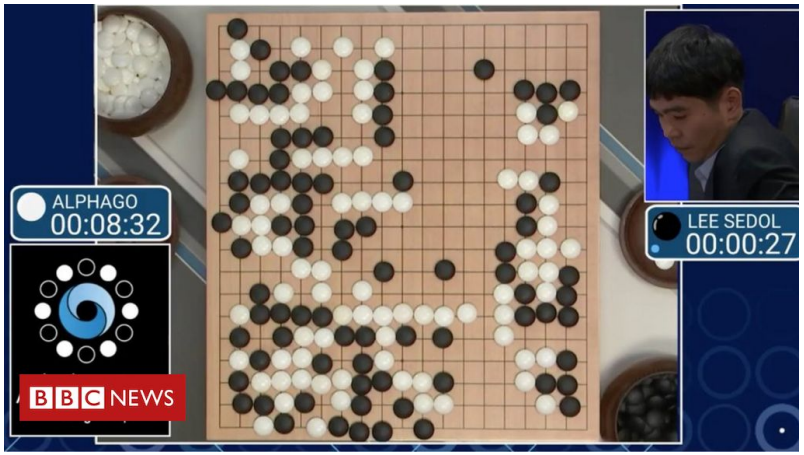
Reward is Enoguh Hypothesis

- Note : the reward hypothesis speculates that all goals may be represented by rewards
  - This should not be confused with our reward-is-enough hypothesis
- Sophisticated abilities may arise from the maximisation of simple rewards in complex environments

Hypothesis (Reward-is-Enough):

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

Example : Compute go



- Prior work focoused on seperate goals
  - Shape (pattern recognition)
  - Tactics (local search)
  - Endgames (combinatorial game theory)
- Alphago focoused on a common goal of maxmising game won
  - Led to deeper understaning of shape, tactics, endgames ..
  - Produced a broder set of Intelligence
    - Territory and Influence, thickness, attack-defence
- Integrated into a unified Intelligence

Why Reinforcement learning? (1)

- Reinforcement learning defines the problem of reward maximisation
- RL address the practical problem faced by natural and artficial intelligence
  - Not a theoretical abstaction of the problem
- Agent is bounded by practical constraints
  - Agent has limited capacity determined by its machinery
    - Limited neurons in brain
    - Linited memory in a computer
  - Agent runs in real-time
    - synapses take time to fire
    - CPU takes time to process

$\therefore$ we need to solve reward maximisation problem subject to these practical constraints

Why Reinforcement learning? (2)

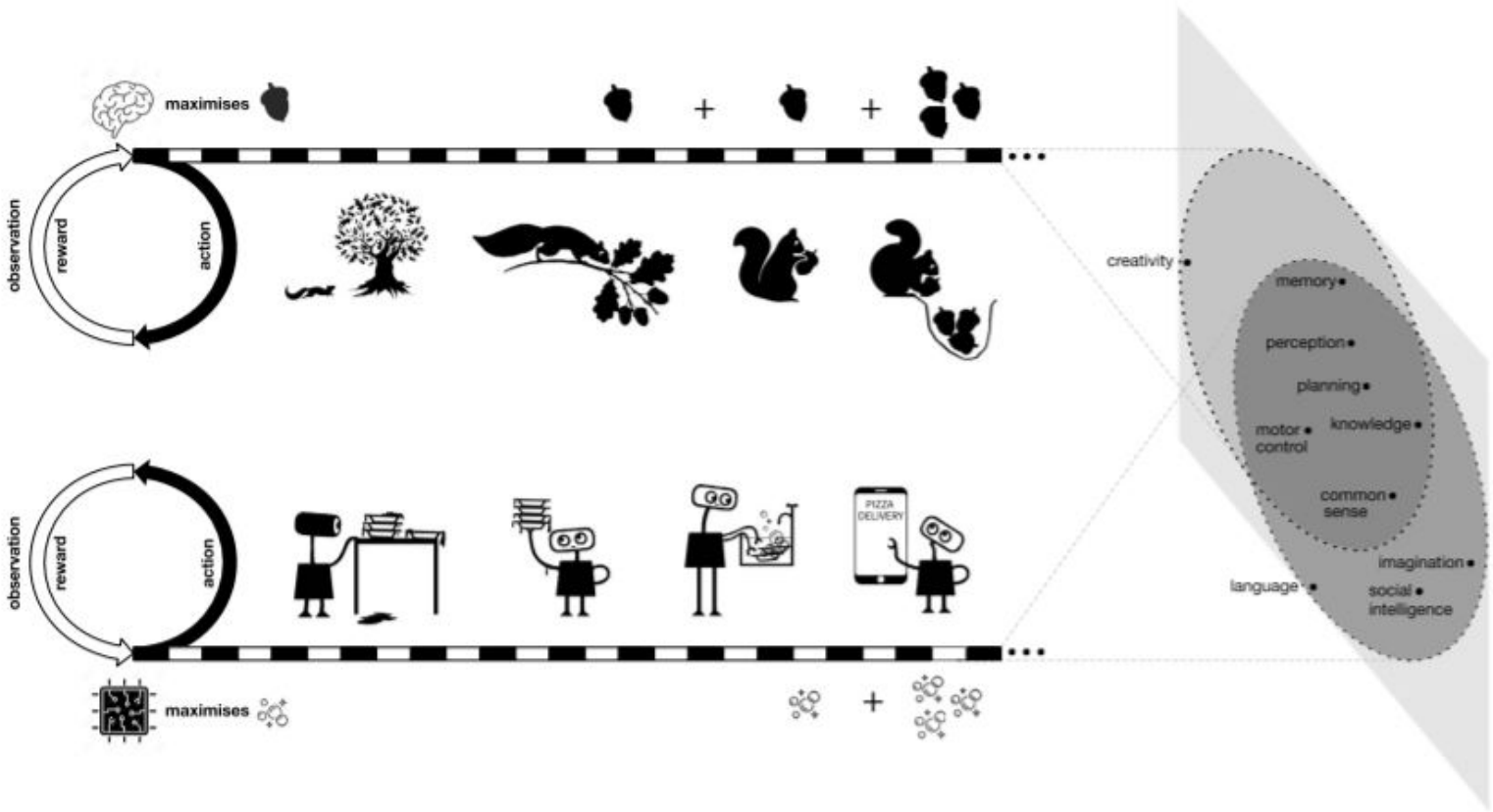
The definition of environment is broad and encompasses many problem dimensions.

Dimension	Alternative A	Alternative B	Notes
Observations	Discrete	Continuous	
Actions	Discrete	Continuous	
Time	Discrete	Continuous	Time-step may be infinitesimal
Dynamics	Deterministic	Stochastic	
Observability	Full	Partial	
Agency	Single agent	Multi-agent	Other agents are part of environment, from perspective of single agent (see Section 3.3)
Uncertainty	Certain	Uncertain	Uncertainty may be represented by stochastic initial states or transitions
Termination	Continuing	Episodic	Environment may terminate and reset to an initial state
Stationarity	Stationary	Non-stationary	Environment depends upon history and hence also upon time
Synchronicity	Asynchronous	Synchronous	Observation may remain unchanged until action is executed
Reality	Simulated	Real-world	May include humans that interact with agent

- Environment definition is incredibly general
- it allows vast range of different problem to be expressed

Reward is Enough for ..(somtehing)

Reward is Enough for ..(somtehing)



- Reward is enough for knowledge and learning
- Reward is enough for perception
- Reward is enough for language
- ..

Reward is Enough for Knowledge and Learning

- Some environment demand prior knowledge
  - A newborn gazell may have to run away from a lion
  - Prior knowledge limited by agent's capacity, and difficuity to construct
- Some environmnet demand learned knowledge
  - The future is uncertain (unkown/ stochasticity / complexity)
  - Total space of potential knowledge » capacity of agent
    - lions in savannah / polar bears
    - knowledge must adapt to circumstance
- in complex environment, Learned knowledge will become increasingly important
- ∴ Reward is Enough for Knowledge and Learning

Reward is Enough for Perception

- Rich real world environment may demand various perceptual skills
  - Image segmentation (avoding falling off a cliff)
  - Object recognition (to identify good and bad foods)
  - Face recognition (to differentiate friend from foe)
  - Speach recognition (to understand verbal warning)
- Alternatively suggest that perception may instead be understood as subserving the maximisation of reward.
- RL Problem include richer variety of perceptual task
  - There is no labelled data
  - Action and observation are intertwined (haptic perception identifying content of pocket)
  - Utility is policy dependent (misclassifying a crocodile when swimming vs walking)
  - Distribution of data is context dependent
    - classifying ice and polar bears vs savanah and lions

Reward is Enough for Language

- Advances have com from viewing languages as solution to a common goal
  - Predicting the next word in a large corpus of data (GPT3)
- Lagauge modelling may not produce broader linguistic Intelligence
  - Language may be intertwined with other modalities of action and observation
    - two agent discussing how to carry an awkward obejct
  - Language is situational and consequential and tailored to acheive a purpose
    - Salesperson maximize sales , politician maximise votes
- Richer language may emerge from a common goal of reward maximization
  - An agent must leran to comprehend ‘danger’ warning to survive ..
- $\therefore$ The ability of language in its full richness, including all of these broader abilities, arises from the pursuit of reward.

Discussion

What Else Could be Enough

- Unsupervised Learning
  - Provides no goals for acion-selection
  - it can‘t be enough itself for intelligent behavior
- Supervised Learning
  - it can be enough to achieve the goal that the teacher want to
  - but Can only address teacher‘s environment and goal
- Offline learning from Large Data-set
  - real problems in complex environment far larger than any data-set
  - Data-set of chimp playing Lego not enough o build a bridge
- ..

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

```
" "
```