AGI via Combining Logic and Deep Learning

AGI 2021 Conference

YKY

October 13, 2021

Table of contents

Part I

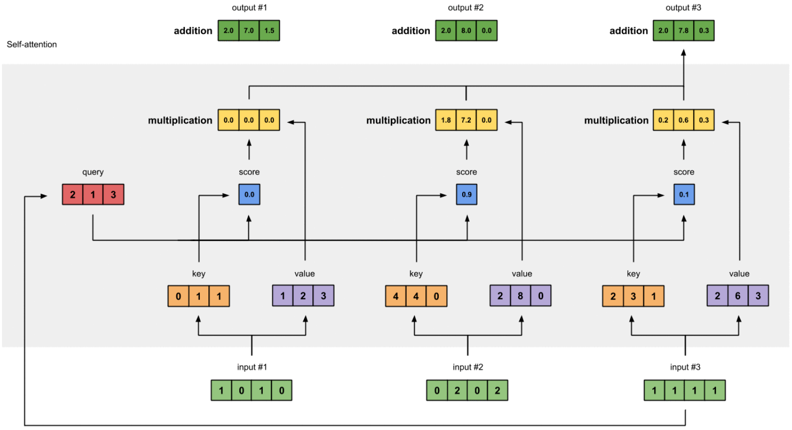
BERT / GPT

BERT is AGI

- BERT can generate texts, answer questions, and even write code
- BERT has learned human knowledge from text corpora
- Therefore, BERT is approximately AGI
- BERT also demonstrates that our computing power is in the **ballpark** of AGI
- I predict that human-level AGI will emerge within 5-6 years

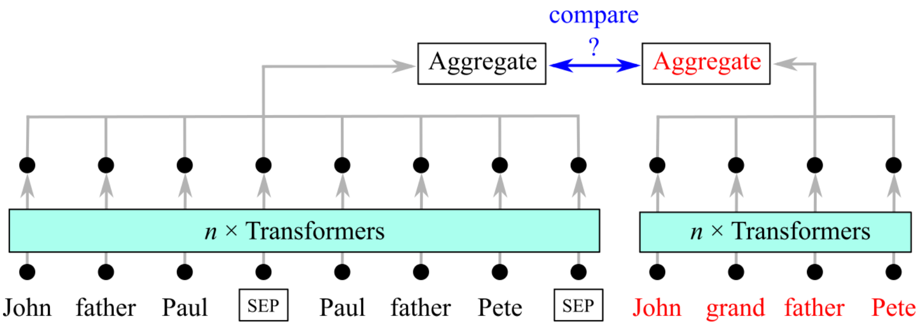
Equivariance of the Transformer

- Many people already know this:
  the Transformer module is **equivariant** in its inputs and outputs:



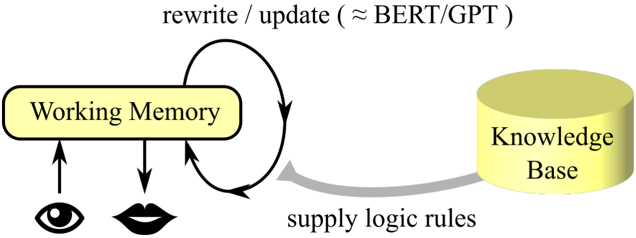
BERT's Hidden Representation

- A simplified view of how BERT performs **question-answering** or **logic deduction**:



BERT compared with AGI architecture

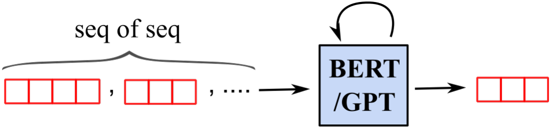
- A minimal AGI architecture is like this:



(1)

AGI as Sequence-to-sequence Rewriter

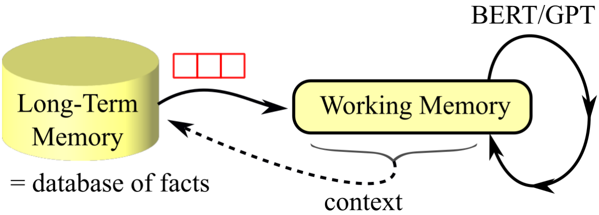
- The **state** of a logic system is a **set** of propositions
- So state = sequence of sequences,
  The 2nd level represents predicates and terms **inside** a proposition



(2)

BERT with Long-Term Memory

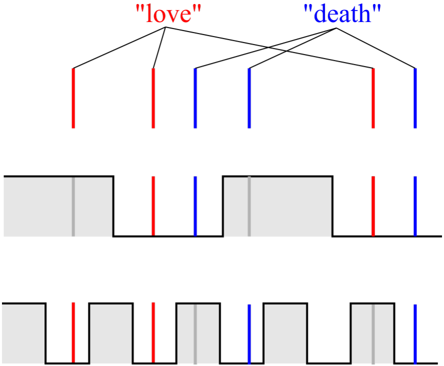
- With logic, it is easier to design cognitive architectures, eg: long-term memory module



(3)

Additivity of Word Embeddings

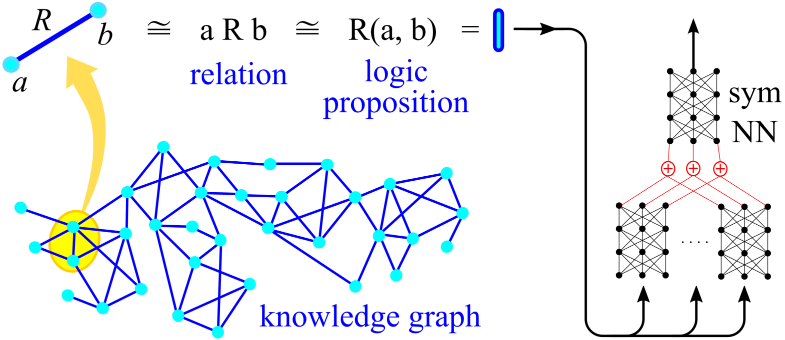
- The sine waves are **Positional Encodings**
- We can see that **Word Embeddings** can co-exist if embedding dimension is large enough



- This suggests that word embeddings can be added together to give **composite** meanings

Memory Recall

- Integrate seamlessly with **knowledge graphs**
- Graphs are made up of edges,
  edges = relations between nodes = propositions:



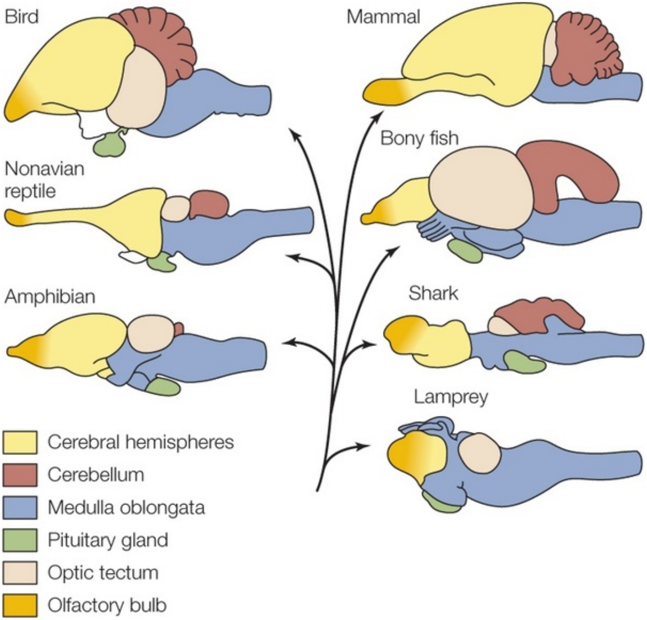
(4)

Part II

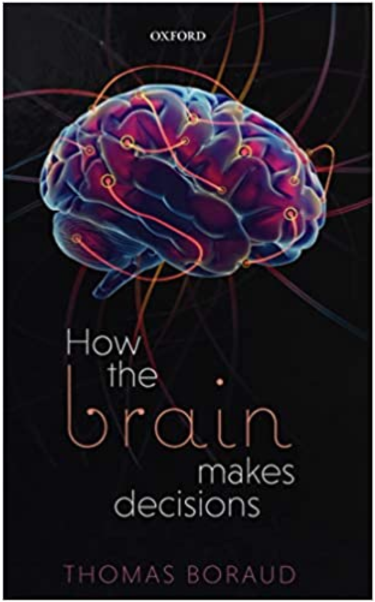
Reinforcement Learning

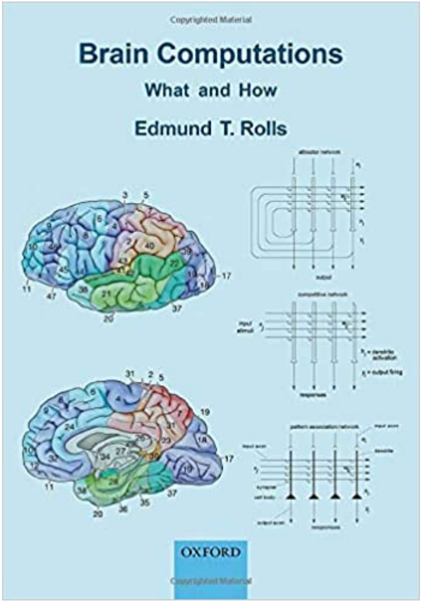
This is a lamprey

Evolution of the Neocortex

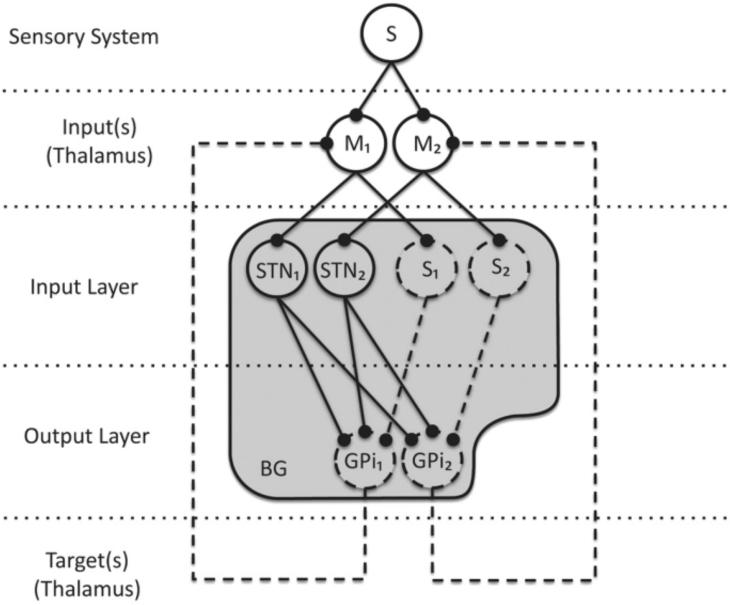


Books

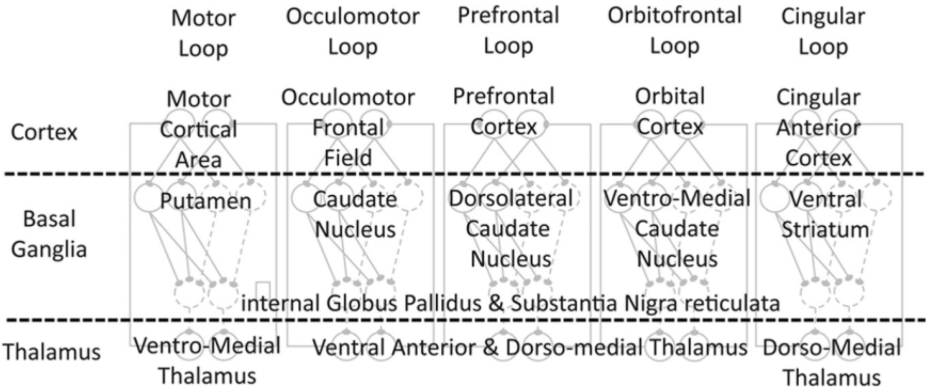

OXFORD
How the brain makes decisions
THOMAS BORAUD


Copyrighted Material
Brain Computations
What and How
Edmund T. Rolls
OXFORD

How the Basal Ganglia makes Decisions

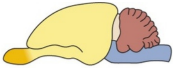


Sub-networks of the Neocortex



Neocortex vs Reinforcement Learning

Neocortical Learning



$+1$ $+100$ $+10$ $\longrightarrow$

food sex money

associations

action_1 action_2 action_3

Reinforcement Learning

reward R_x

$$Q \mathrel{+}= \eta\left[R_x + \gamma \max_{a'} Q' - Q\right]$$

(Q-learning)

action a

Part III

Model-based vs Syntax-based

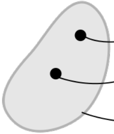
What is a Model?

Syntactic Representation
ie. a set of logic formulas:

```
human (Plato)
human (Aristotle)
∀x. human(x) ⇒ mortal(x)
```

Model theory relates
syntax with semantics

Semantic Representation

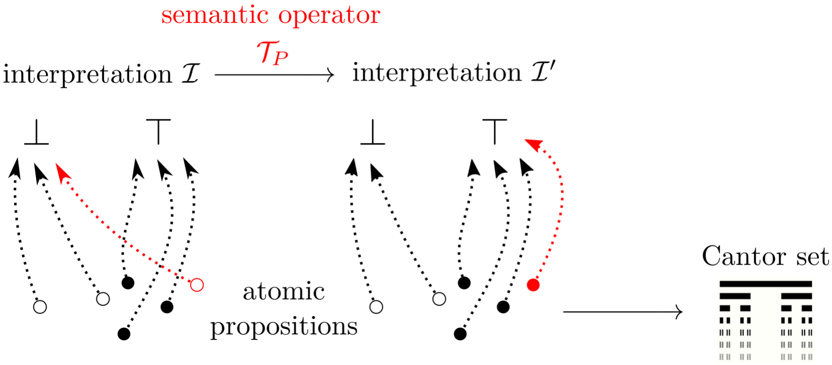


Language **describes** the real world

A **model** is a **representation** of the real world

Hitzler's Core Method

- The **semantic operator** $\mathcal{T}_P$ updates an interpretation to another interpretation:



Hitzler's Core Method

- The **level mapping** ι takes an interpretation to a real number:

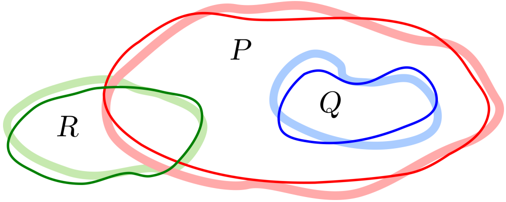
$$\begin{tikzcd} \mathcal{I} \arrow[r, "\mathcal{T}_P"] \arrow[d, "\iota"] & \mathcal{I} \arrow[d, "\iota"] \\ \mathcal{K} \arrow[r, "f_P"] & \mathcal{K} \end{tikzcd}$$

$$\iota(\mathcal{I}) = \sum_{\text{atom} \in \mathcal{I}} b^{-\ell(\text{atom})} \in \mathbb{R}$$

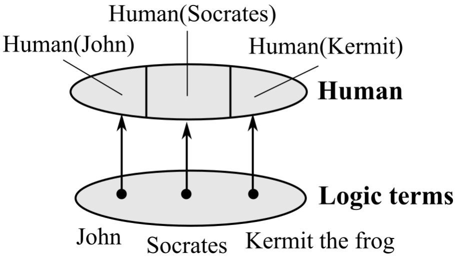
approximated by
neural network

Fibrations

- "Floating" regions and points as models



- "Fibration" structure of predicates

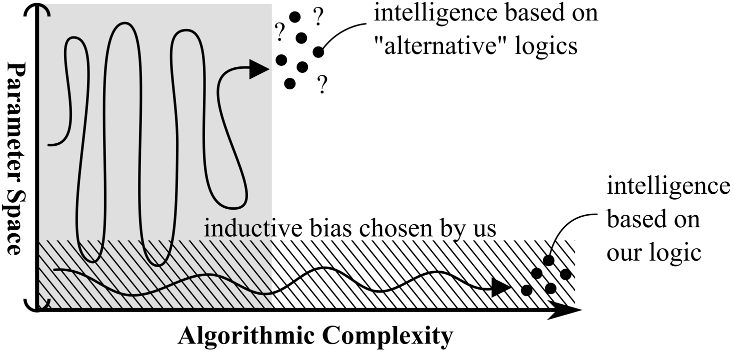


Part IV

No Free Lunch and Inductive Bias

Symmetry and inductive bias

- In mathematics, symmetry often simplifies computation, which is why mathematicians love to study symmetries
- In machine learning, one introduces **inductive bias** to narrow down the **search space**:



(5)

- Oftentimes, if inductive bias is chosen correctly, solution is found quickly, otherwise problem becomes **intractable**

Richard Sutton's view

- In contrast, Sutton expressed the view that AI can be solved merely by **increasing computing power**, under the reinforcement learning framework

(6)

References

Thanks for watching ☺️