Moving Beyond Probabilities: Memory as World Modelling

John Tan Chong Min

The Long Tail

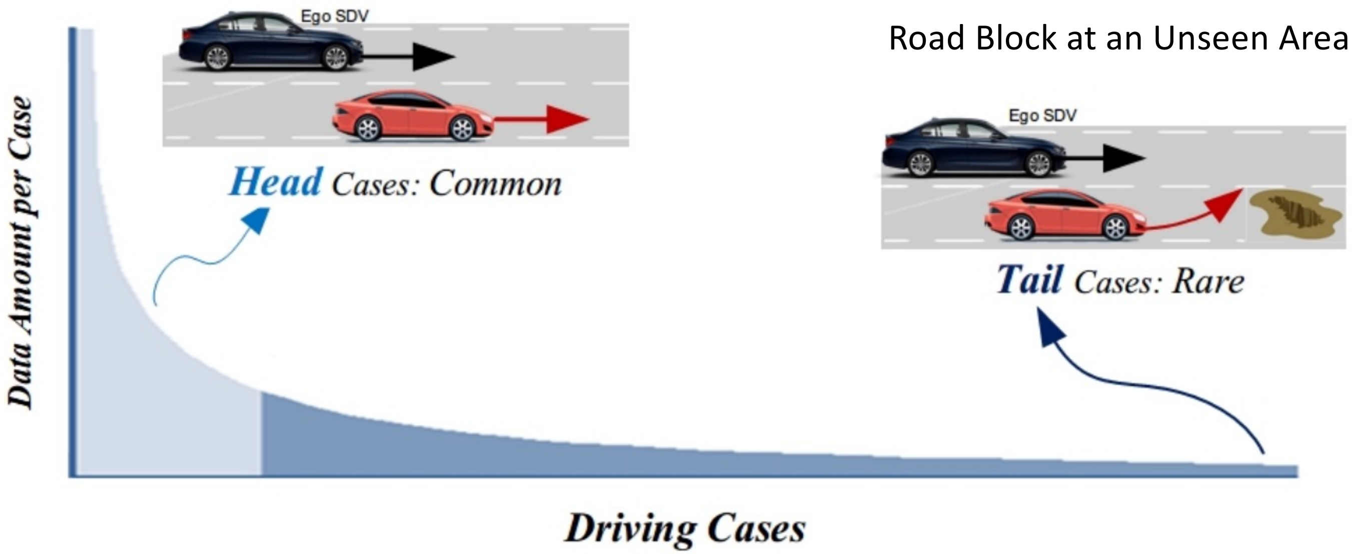
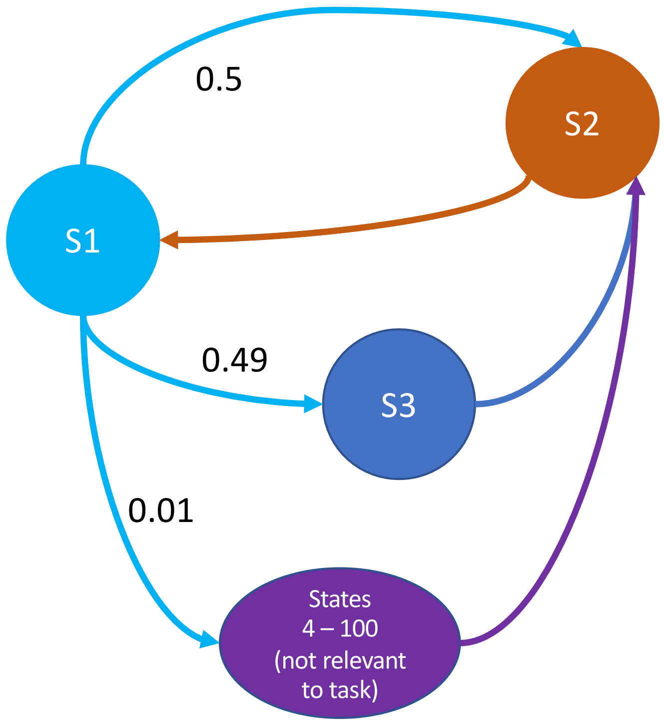


Fig. 1. The distribution of real-world driving data. There exists “long-tail” of driving cases where the amount of data is small.

Long-Tail Prediction Uncertainty Aware Trajectory Planning for Self-driving Vehicles. Zhou et al. 2022.

What's wrong with probabilities

- Long-tailed probabilities mean that rare events are not represented well
- May be **impossible to represent** infinite numbers of outcomes mathematically
- May be **time consuming to learn exactly** (requires experiencing all trajectories multiple times in order to learn the exact probabilities)
- Not all states are important and hence **no need to learn the entire world model**



Do Humans Compute Probability?

Daniel Kahneman and some psychological experiments

Which is more probable (Part 1)?

- In four pages of a novel (about 2000 words), how many words would you expect to find that have the form

- - - - i n g

Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. Tversky and Kahneman. 1972

Which is more probable (Part 1)?

- In four pages of a novel (about 2000 words), how many words would you expect to find that have the form

- - - - - n -

Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. Tversky and Kahneman. 1972

Which is more probable (Part 1)?

- In four pages of a novel (about 2000 words), how many words would you expect to find that have the form

- - - - i n g (Median: 8.8)

- - - - - n - (Median: 4.4)

People view results ending in –ing as more probable than –n–

Although statistically it is the other way round

Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. Tversky and Kahneman. 1972

Which is more probable (Part 2)?

- Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

- Rank from highest to lowest, what is the probability of each statement being true.
  - Linda works in a bookstore and takes Yoga classes.
  - Linda is active in the feminist movement.
  - Linda is a bank teller.
  - Linda is an insurance salesperson.
  - Linda is a bank teller and is active in the feminist movement.

Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. Tversky and Kahneman. 1972

Which is more probable (Part 2)?

- Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

- Rank from highest to lowest, what is the probability of each statement being true.
    - Linda works in a bookstore and takes Yoga classes.
    - Linda is active in the feminist movement. (F)
    - Linda is a bank teller. (T)
    - Linda is an insurance salesperson.
    - Linda is a bank teller and is active in the feminist movement. (T & F)

  88% of people rank F > T & F > T for Linda
  But by probability, T & F combined should be lower chance than T or F

  Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. Tversky and Kahneman. 1972

What is more probable (Part 3)?

On each round of a game, 20 marbles are distributed at random among five children: Alan, Ben, Carl, Dan, and Ed. Consider the following distributions:

	I		II
Alan	4	*Alan*	4
Ben	4	*Ben*	4
Carl	5	*Carl*	4
Dan	4	*Dan*	4
Ed	3	*Ed*	4

In many rounds of the game, will there be more results of type I or of type II?

Subjective Probability: A Judgement of Representativeness. Kahneman and Tversky. 1972

What is more probable (Part 3)?

On each round of a game, 20 marbles are distributed at random among five children: Alan, Ben, Carl, Dan, and Ed. Consider the following distributions:

	I		II
Alan	4	*Alan*	4
Ben	4	*Ben*	4
Carl	5	*Carl*	4
Dan	4	*Dan*	4
Ed	3	*Ed*	4

In many rounds of the game, will there be more results of type I or of type II?

36 out of 52 people view Distribution 1 as more probable than Distribution 2.
However, Distribution 2 is theoretically more probable as it is the uniform distribution.

Subjective Probability: A Judgement of Representativeness. Kahneman and Tversky. 1972

Which is more probable (Part 4)?

- All families of six children in a city were surveyed.
- In 72 families the exact order of births of boys and girls was G B G B B G.
- What is your estimate of the number of families surveyed in which the exact order of births was B G B B B B?

Subjective Probability: A Judgement of Representativeness. Kahneman and Tversky. 1972

Which is more probable (Part 4)?

- All families of six children in a city were surveyed.
- In 72 families the exact order of births of boys and girls was G B G B B G.
- What is your estimate of the number of families surveyed in which the exact order of births was B G B B B B?

- Theoretically both sequences have the same chance of occurring
- People's median guess is 30 for B G B B B B
- Representativeness shapes people's perceptions of likelihood

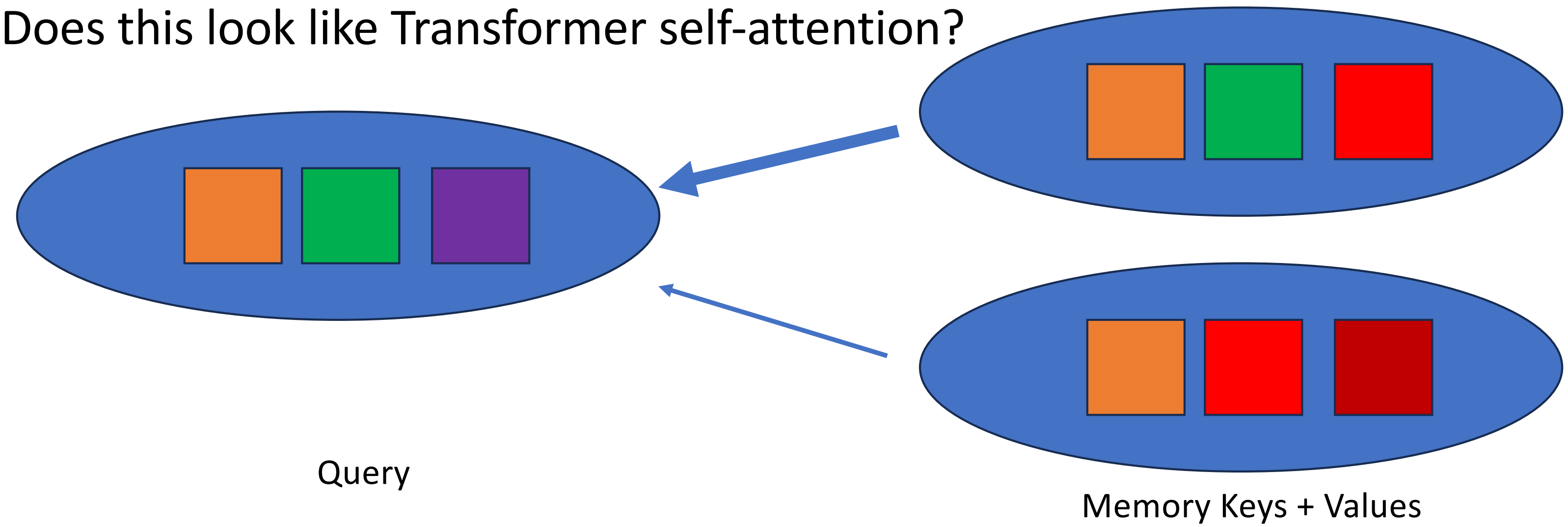
Subjective Probability: A Judgement of Representativeness. Kahneman and Tversky. 1972

From Psychology to Machine Learning

How to interpret Daniel Kahneman’s Results in a Modern Lens

Memory and Representativeness (Hypothesis)

- We query our memory and find out what matches
- Memories that match more of the query will be retrieved first
- Does this look like Transformer self-attention?



Transformers: Context as memory

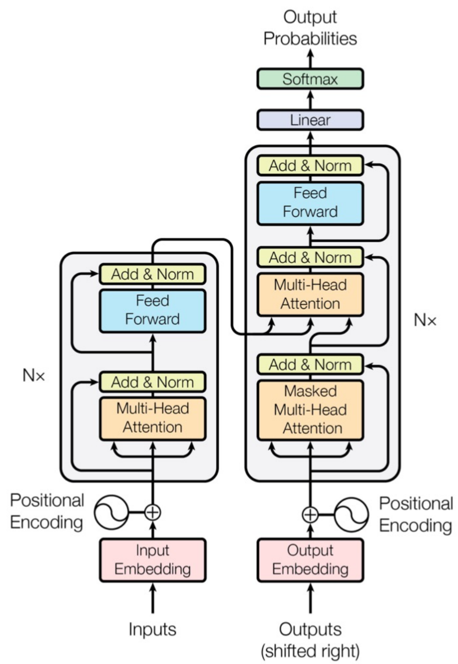


Figure 1: The Transformer - model architecture.

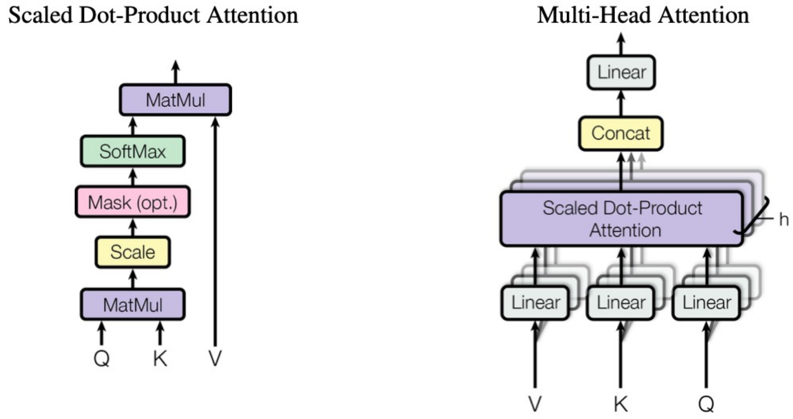


Figure 2: (left) Scaled Dot-Product Attention. (right) Multi-Head Attention consists of several attention layers running in parallel.

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Taken from: Attention is all you need. Vaswani et al. (2017)

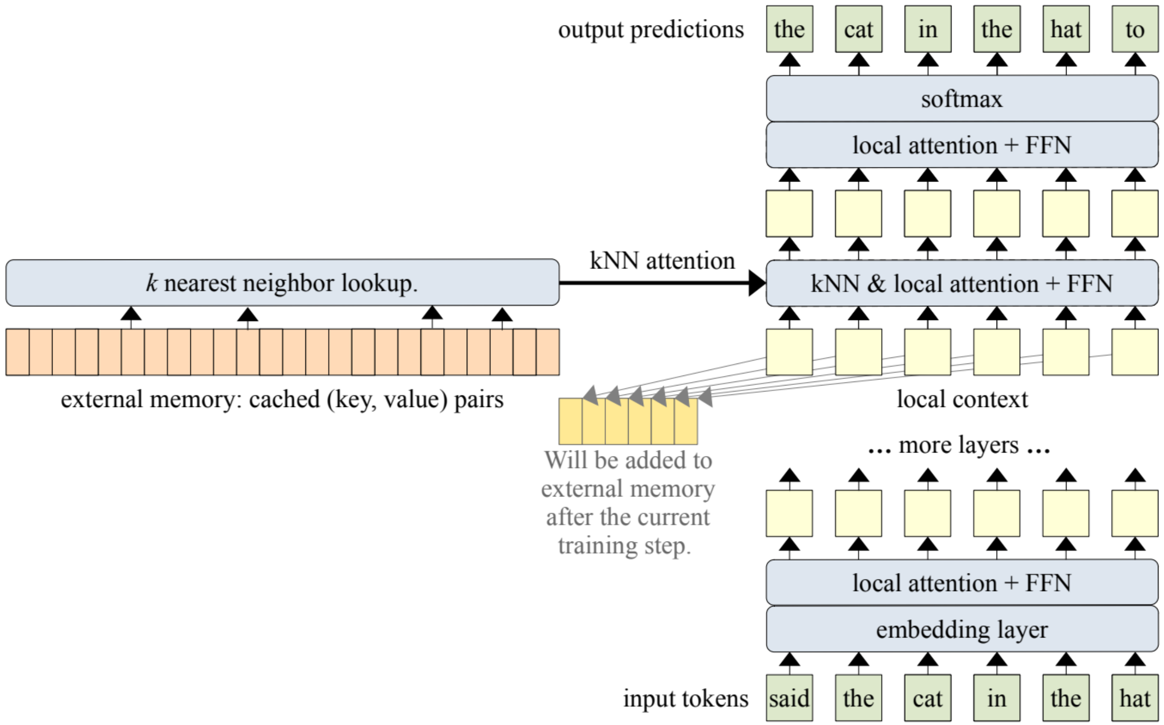
Retrieval Augmented Generation to boost context

- <Retrieved Memory 1>
- <Retrieved Memory 2>
- <Retrieved Memory k>

- <Query>

What about using external memory storage?

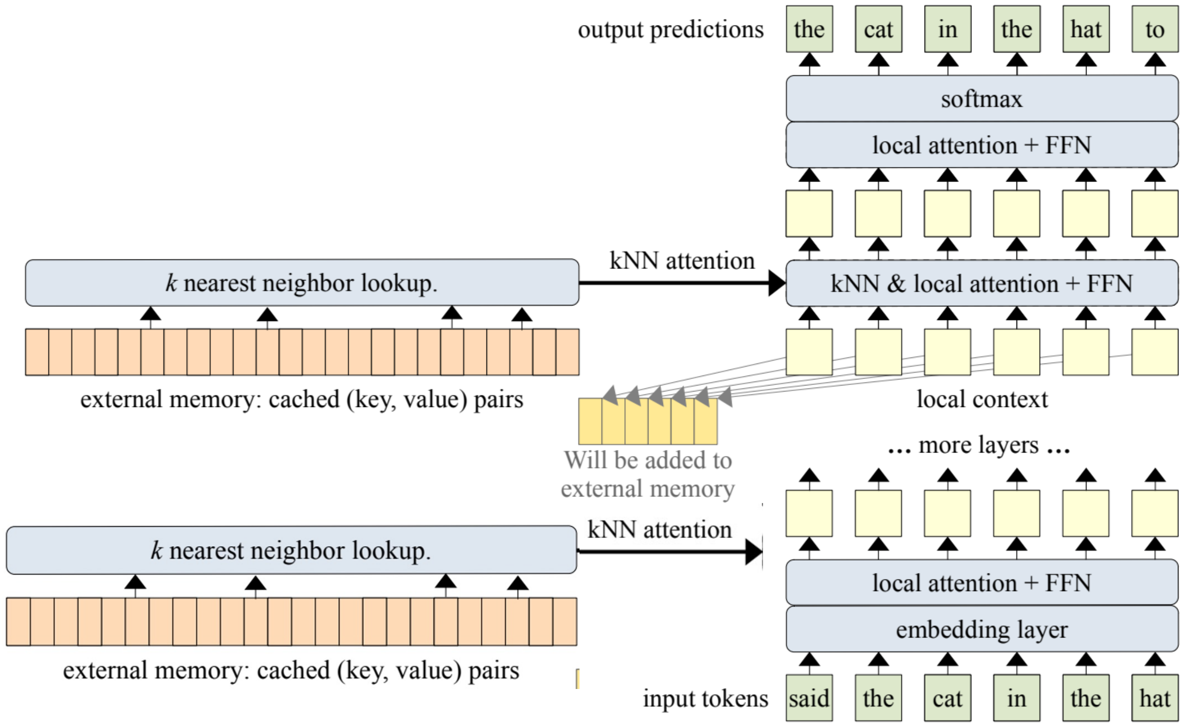
- Can theoretically attend to unlimited numbers of neighbours
- Memory is stored as embedding space of the layer before softmax
  - Only works for tokens of the same sequence, since attention is only for previous tokens



Memorizing Transformers. Wu et al. 2022.

Hierarchical Memory Referencing in Embeddings (my idea)

- Memory referencing should be done at multiple layers of the hierarchy
- Think about how we can go up and down scales of memory:
  - Tying a shoe (Base)
  - How do you tie a shoe? (One level lower)
  - Why do you tie a shoe? (One level higher)
- Accessing memory at various levels should lead to more accurate retrieval



Modified from: Memorizing Transformers. Wu et al. 2022.

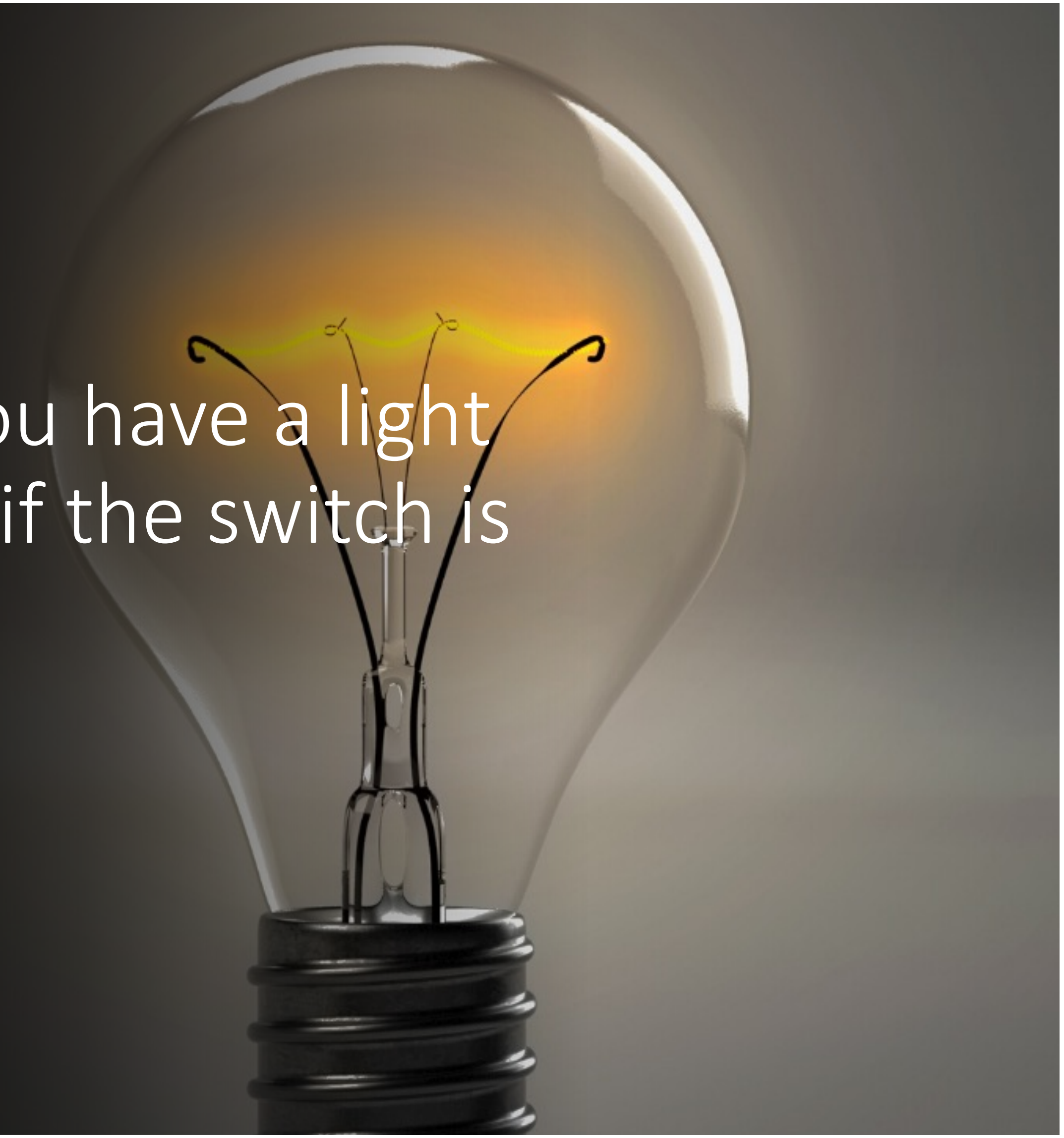
Thought Experiment

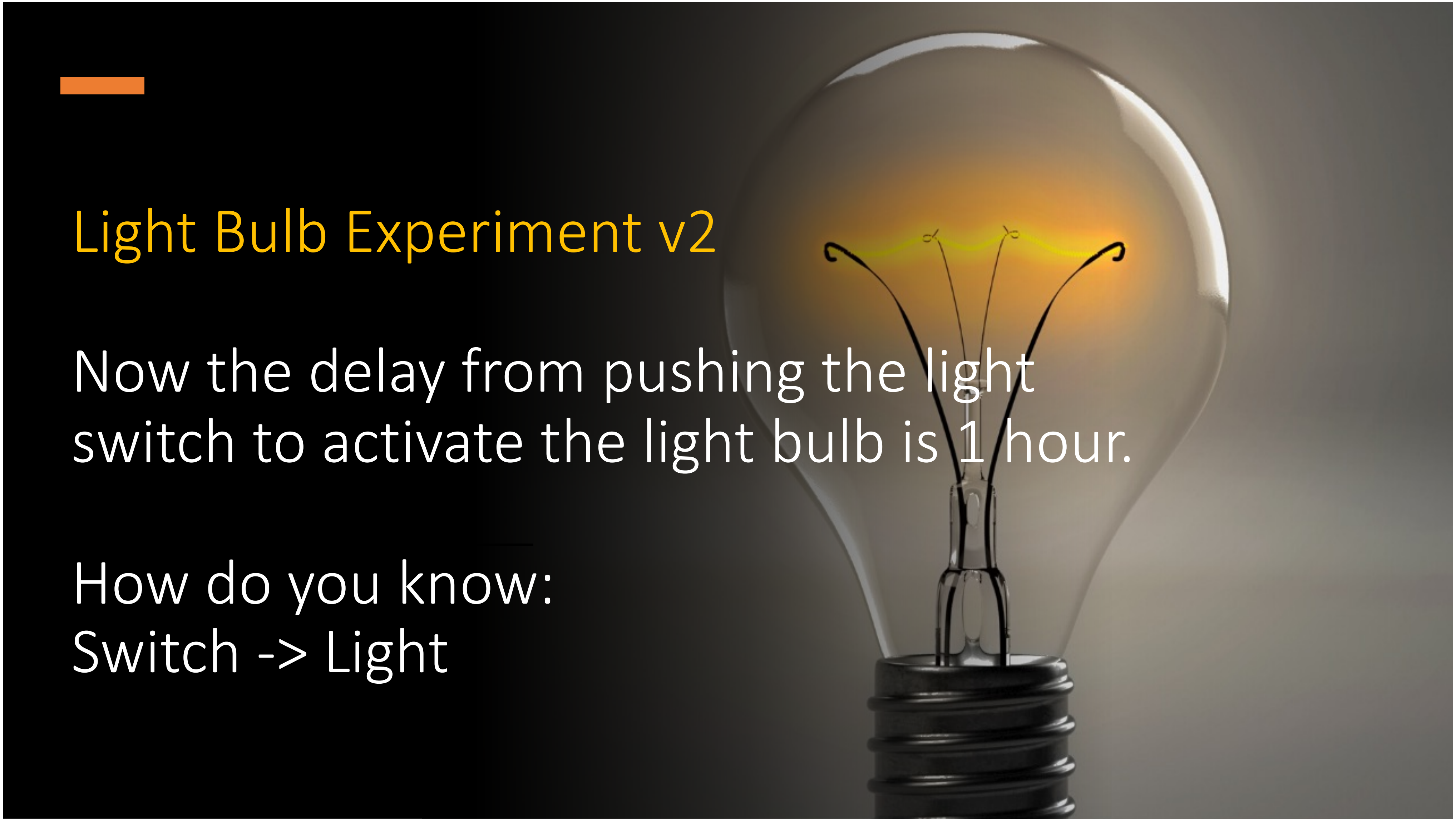
Light Bulb and Light Switch

Light Bulb Experiment

You have a light bulb, you have a light switch, you don't know if the switch is linked to the bulb.

How do you know:
Switch -> Light

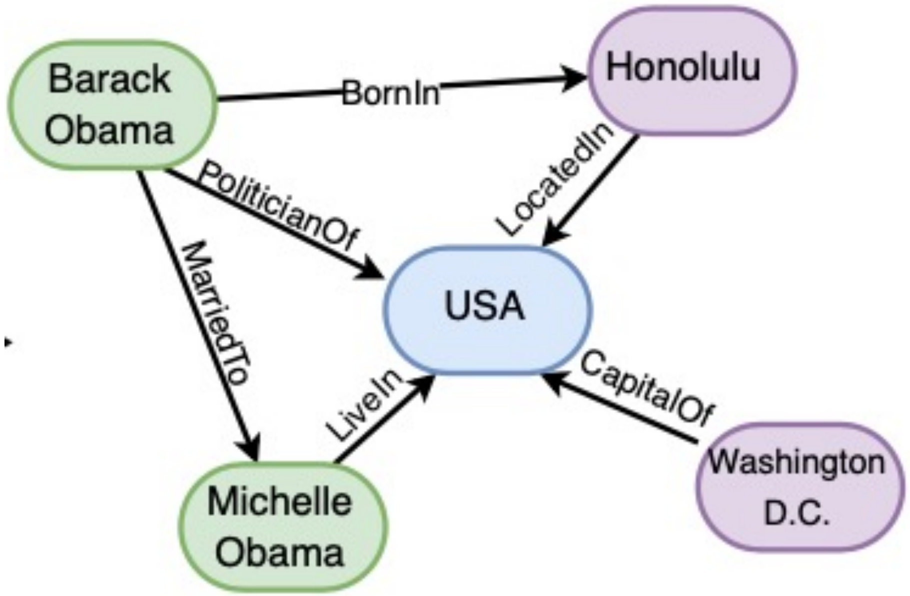



Light Bulb Experiment v2
Now the delay from pushing the light switch to activate the light bulb is 1 hour.
How do you know:
Switch -> Light

What is a Knowledge Graph?

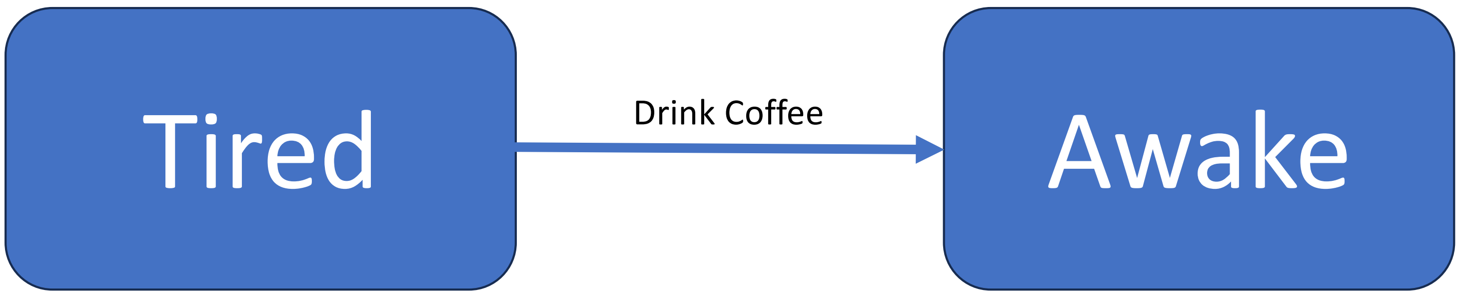
- Triplets:
  - {Source, Destination, Relation}

- Typically a Directed Graph



Knowledge Graphs and Causal Relations

- The relation between Source and Destination can be a causal link

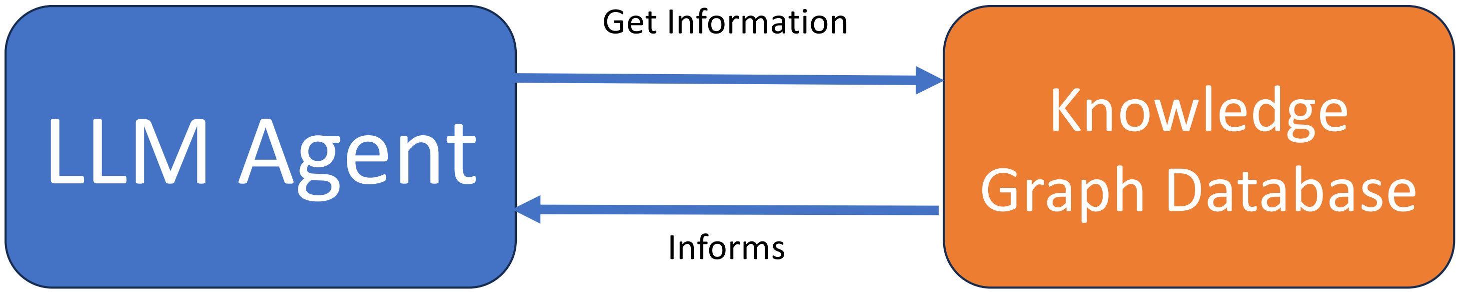


Creating a Causal Link

- If the time lag is instant, there should already be a causal link established with just the (State, Action, Next State) memory tuple
- A process of reflection might help to consolidate the causal link, especially when the time lag is huge
- Causal links help to process future occurrences of similar scenarios faster
  - But: What if a learned causal link is incorrect? Will it ever be unlearned?

Knowledge Graph + LLM Symbiosis

- Extracting relevant parts of a Knowledge Graph can serve as a way to retrieve context
- May embody causal relations easily
- Knowledge Graphs can grow dynamically, much like memory



Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

Meredith Ringel Morris
Google Research
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu

Generative Agents: Interactive Simulacra of Human Behavior. Park et al. 2023.

Memory Retrieval for Planning

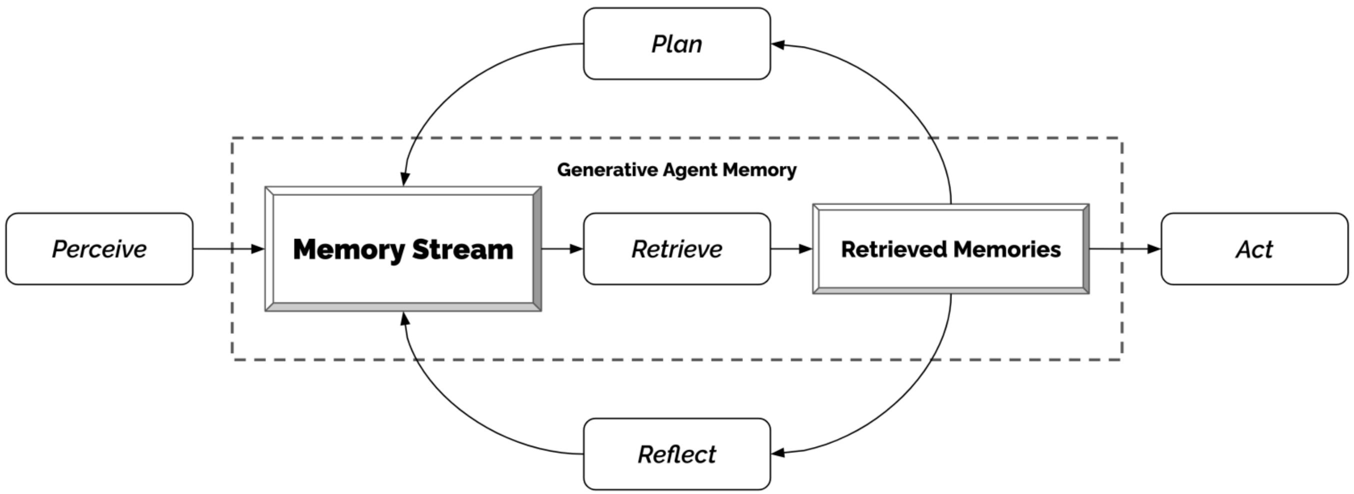


Figure 5: Our generative agent architecture. Agents perceive their environment, and all perceptions are saved in a comprehensive record of the agent's experiences called the memory stream. Based on their perceptions, the architecture retrieves relevant memories, then uses those retrieved actions to determine an action. These retrieved memories are also used to form longer-term plans, and to create higher-level reflections, which are both entered into the memory stream for future use.

Generative Agents: Interactive Simulacra of Human Behavior. Park et al. 2023.

Memory Stream - from observations to reflection

- Observations may be too generic to learn, need some higher level consolidation
- Reflection generation:
  - Generate reflections when the sum of the importance scores for the latest events perceived by the agents exceeds a certain threshold
  - Reflect on average two to three times a day
- Reflection process:
  - Retrieve 100 most recent observations/reflections, then prompt "Given only the information above, what are 3 most salient high-level questions we can answer about the subjects in the statements?”
  - For each question, retrieve most relevant X memories and generate reflections (also state the memories/reflections derived)

Generative Agents: Interactive Simulacra of Human Behavior. Park et al. 2023.

Trees of Reflection

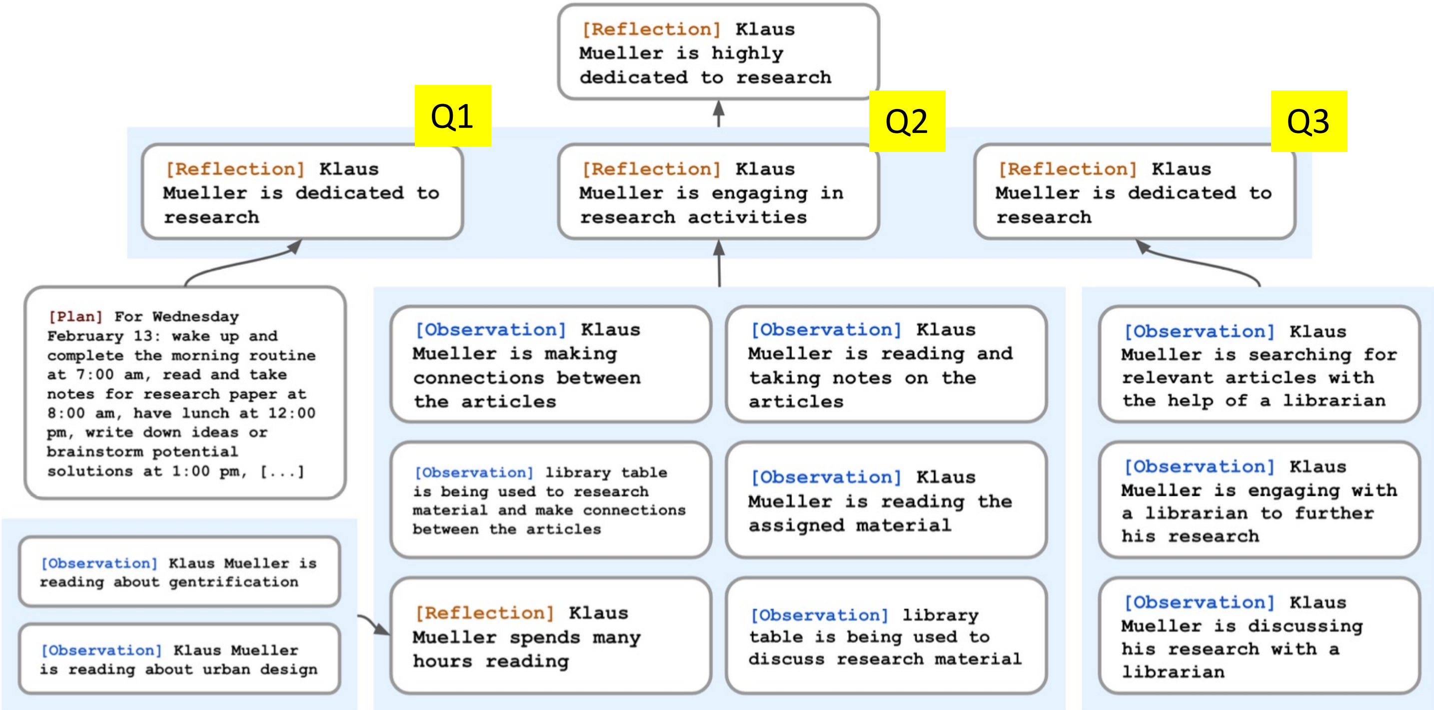


Figure 7: A reflection tree for Klaus Mueller. The agent's observations of the world, represented in the leaf nodes, are recursively synthesized to derive Klaus's self-notion that he is highly dedicated to his research.

Generative Agents: Interactive Simulacra of Human Behavior. Park et al. 2023.

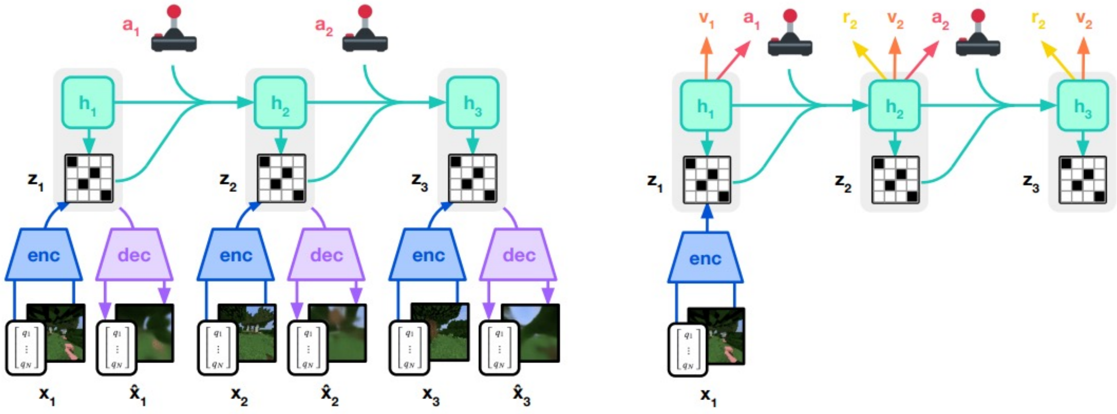
Memory as a way to approximate probabilities

Learning, Fast and Slow: A Goal-Directed Memory-Based Approach for Dynamic Environments

https://arxiv.org/abs/2301.13758

Dreamer v3

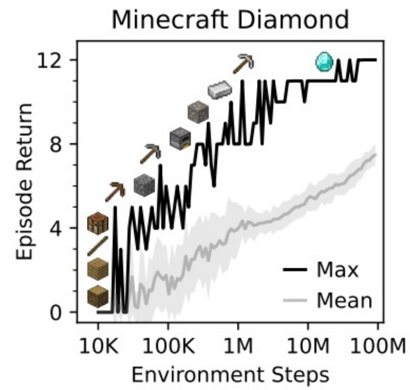
- Very difficult to generate world models



(a) World Model Learning **(b)** Actor Critic Learning

Figure 3: Training process of DreamerV3. The world model encodes sensory inputs into a discrete representation z_t that is predicted by a sequence model with recurrent state h_t given actions a_t. The inputs are reconstructed as learning signal to shape the representations. The actor and critic learn from trajectories of abstract representations predicted by the world model.

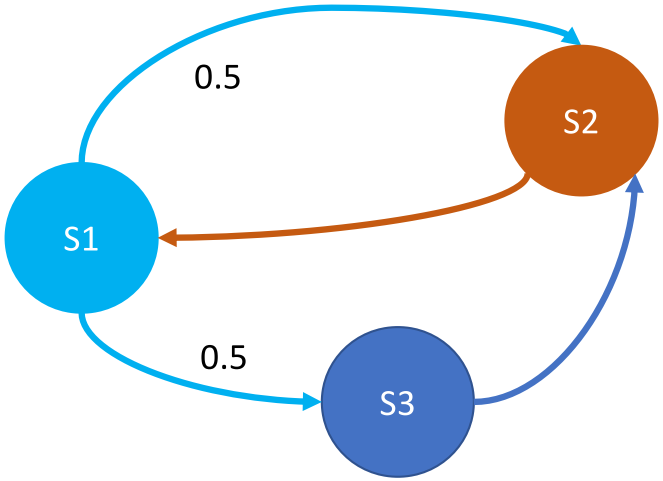
Sequence model:	$h_t = f_\phi(h_{t-1}, z_{t-1}, a_{t-1})$
Encoder:	$z_t \sim q_\phi(z_t \mid h_t, x_t)$
Dynamics predictor:	$\hat{z}_t \sim p_\phi(\hat{z}_t \mid h_t)$
Reward predictor:	$\hat{r}_t \sim p_\phi(\hat{r}_t \mid h_t, z_t)$
Continue predictor:	$\hat{c}_t \sim p_\phi(\hat{c}_t \mid h_t, z_t)$
Decoder:	$\hat{x}_t \sim p_\phi(\hat{x}_t \mid h_t, z_t)$



Mastering Diverse Domains through World Models. Hafner et al. 2023.

Should we model the Markov Decision Process of the world?

- Recent works have tried to utilize world models
  - MuZero
  - Dreamer v2/v3
- To get probability of transition, we would need many samples to learn it accurately
- Difficult to define state transition probability when number of possible states is unbounded



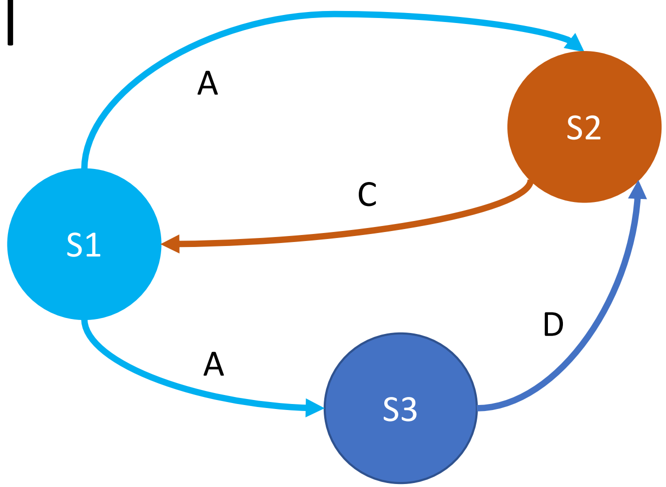
Learning, Fast and Slow: A Goal-Directed Memory-Based Approach for Dynamic Environments. John and Motani. 2023.

Using memory as world model

- **No need to model Markov Decision Process fully** – impractical to model environments with unbounded state and action spaces

- **No need to model probability of transition** – just need to see how often the next state is stored based on memory

- Just need to **remember experienced transitions** and then **retrieve it** the next time we encounter a similar state



Key (State)	Value 1 (Next State)	Value 2 (Action)
1	2	A
1	3	A
2	1	C
3	2	D

Learning, Fast and Slow: A Goal-Directed Memory-Based Approach for Dynamic Environments. John and Motani. 2023.

Insight: Next action prediction is fast to learn,
Next state prediction is slow to learn
(Epoch 50, x axis actions preferred to y axis actions)

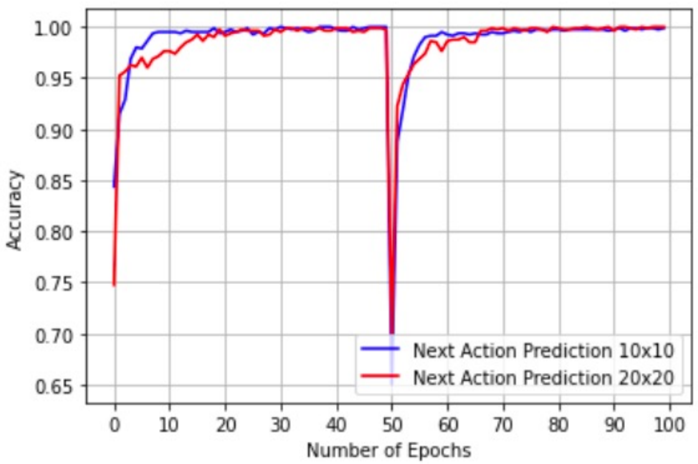


Figure 2: Accuracy of next action prediction for 10x10 grid (blue) and 20x20 grid (red) using 1000 samples

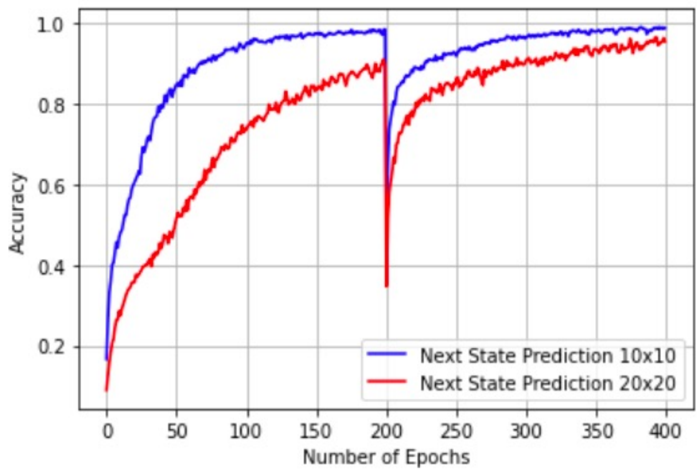
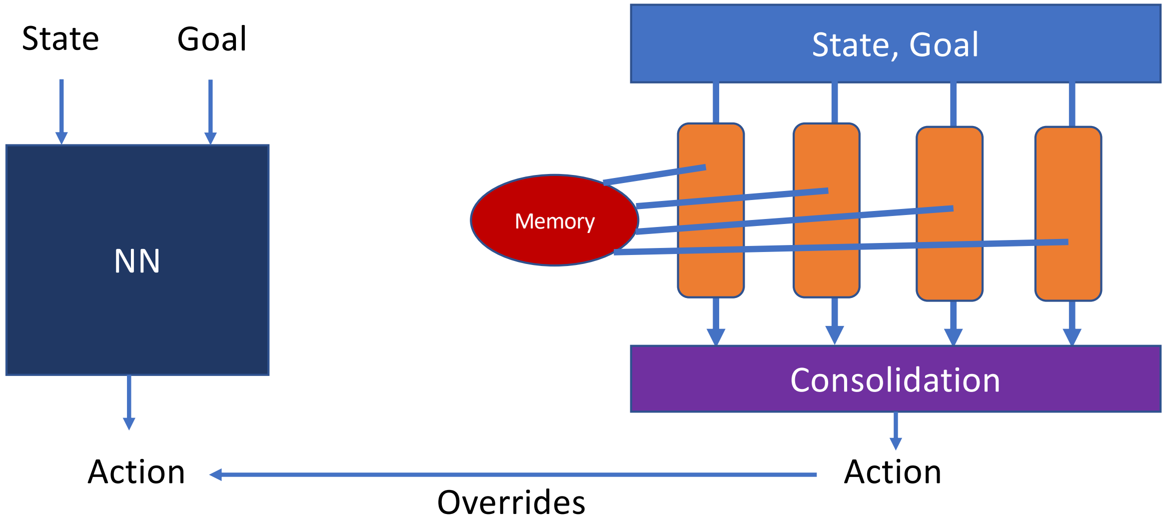


Figure 3: Accuracy of next state prediction for 10x10 grid (blue) and 20x20 grid (red) using 1000 samples

Learning, Fast and Slow: A Goal-Directed Memory-Based Approach for Dynamic Environments. John and Motani. 2023.

Two Networks – Fast and Slow



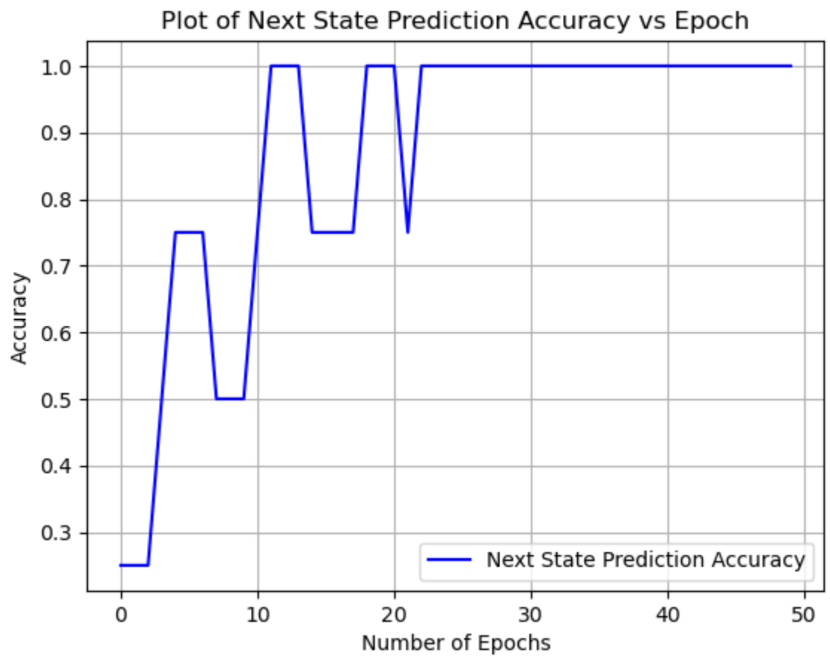
Neural Networks: Fast retrieval, slow learning

Memory: Slow retrieval, fast learning
(World Model planning as Memory Retrieval)

Learning, Fast and Slow: A Goal-Directed Memory-Based Approach for Dynamic Environments. John and Motani. 2023.

Comparison between Memory and Probabilities

- Experiment to see how fast each method takes to learn a world
- Memory learns instantly, next state prediction with neural network takes time
  - NN details: 1 input layer, 128 nodes in hidden layer, 4 output nodes, Adam learning rate 0.1, cross-entropy loss
- Deterministic: States 0, 1, 2, 3 -> 1, 2, 3, 0




```
Logits at epoch 1:
 [[0.43 0.16 0.23 0.19]
 [0.77 0.05 0.08 0.1 ]
 [0.95 0.01 0.01 0.03]
 [0.99 0.   0.   0.01]]
```

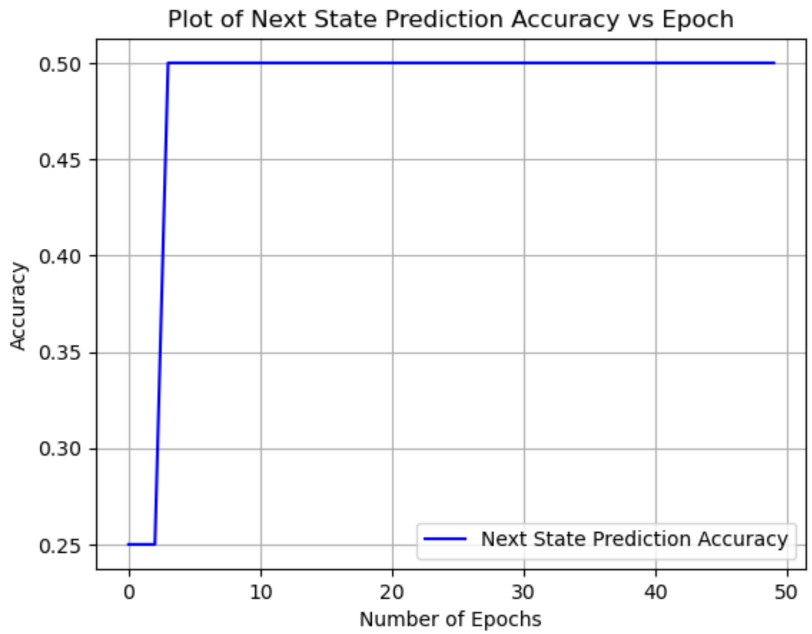
```
Logits at epoch 50:
 [[0.   0.96 0.03 0.01]
 [0.   0.   0.99 0.01]
 [0.02 0.   0.   0.97]
 [0.99 0.   0.   0.01]]
```

Input State: 0, 1, 2, 3

Output State: 0, 1, 2, 3

Comparison between Memory and Probabilities

- Experiment to see how fast each method takes to learn a world
- Memory learns instantly, next state prediction with neural network takes time
  - NN details: 1 input layer, 128 nodes in hidden layer, 4 output nodes, Adam learning rate 0.1, cross-entropy loss
- Stochastic: States 0, 0, 2, 2 -> 2, 3, 0, 1




```
Logits at epoch 1:
 [[0.34 0.35 0.16 0.15]
 [0.34 0.35 0.16 0.15]
 [0.48 0.52 0.   0.  ]
 [0.48 0.52 0.   0.  ]]
Logits at epoch 50:
 [[0.01 0.01 0.49 0.49]
 [0.01 0.01 0.49 0.49]
 [0.52 0.48 0.   0.  ]
 [0.52 0.48 0.   0.  ]]
```

Input State: 0, 1, 2, 3

Output State: 0, 1, 2, 3

Questions to Ponder

- When we add causal links to the Knowledge Graph, how do we know when to unlearn it?
- Does this mean that humans who explore only parts of the world are not able to be optimal?
- How does knowledge sharing occur? In humans, we can learn from each other's experiences? How can we code this inside AI?
- How often should we reflect and consolidate higher level concepts?
- How is hierarchy constructed in the knowledge graph for causal linkages?