Learning By Abstraction:

The Neural State Machine

Drew Hudson & Christopher Manning

Stanford University

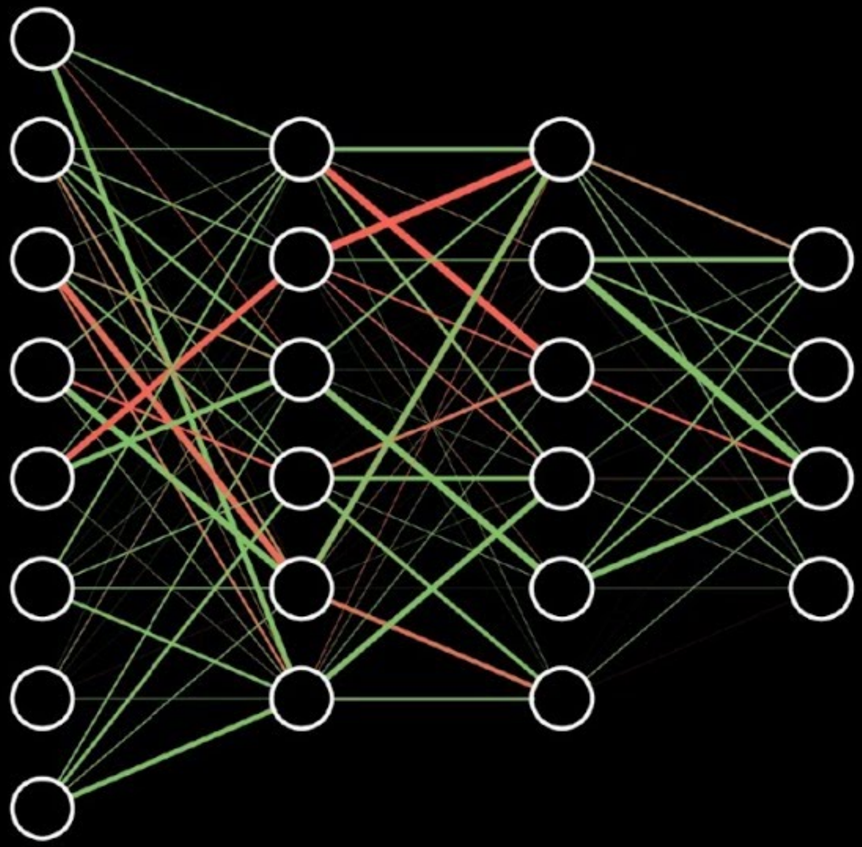
NeurIPS 2019

Stanford

The hope of deep neural models is to learn higher-level abstractions

Abstractions disentangle factors of variation, improving generalization

Neural Networks

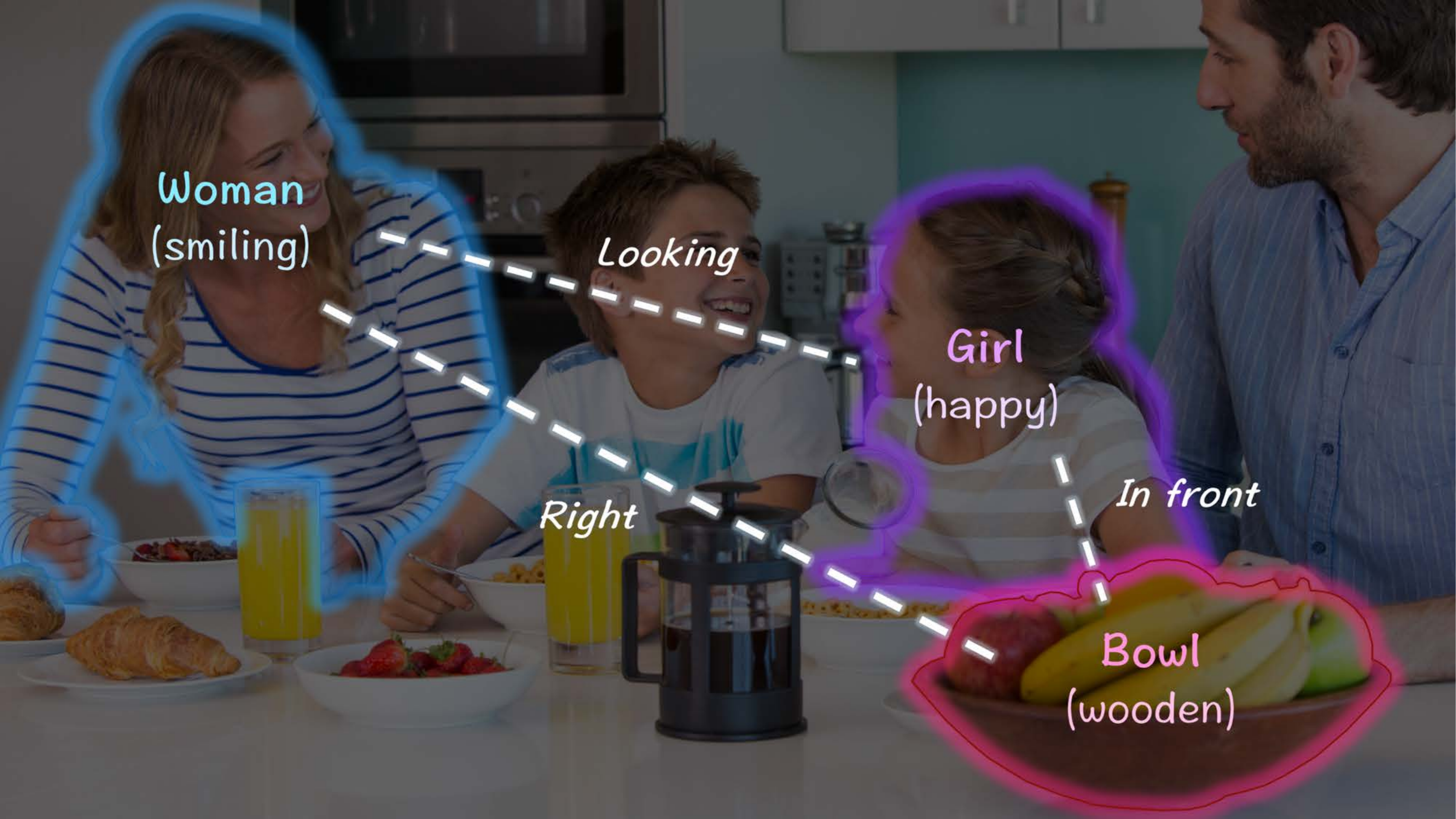


prediction

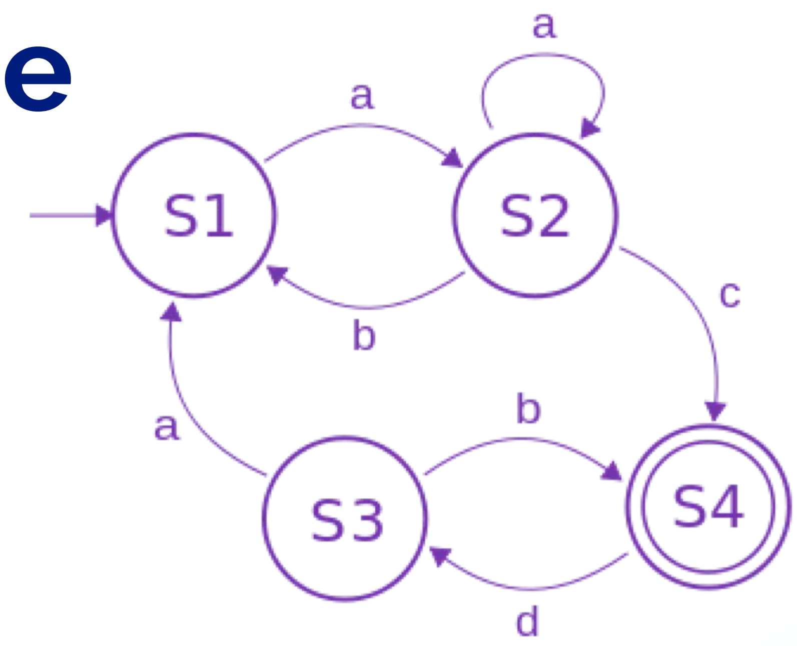
2

How many people are in the image?

Language of Thought

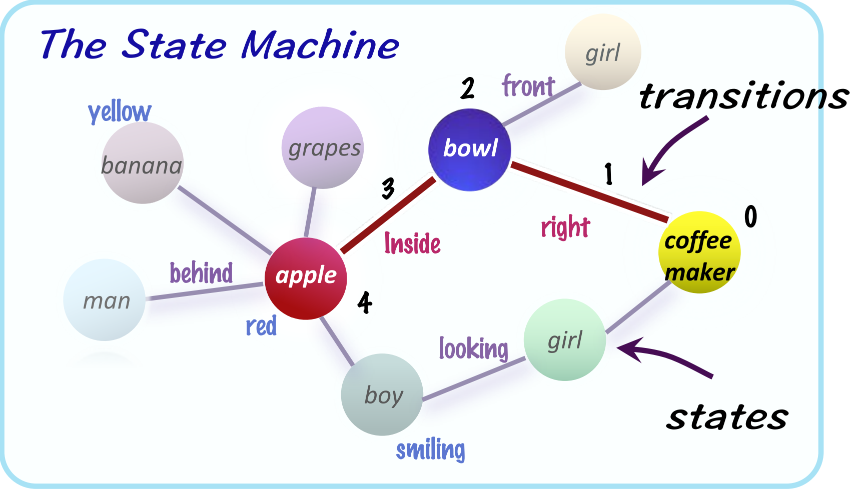

Woman
(smiling)
Looking
Girl
(happy)
Right
In front
Bowl
(wooden)

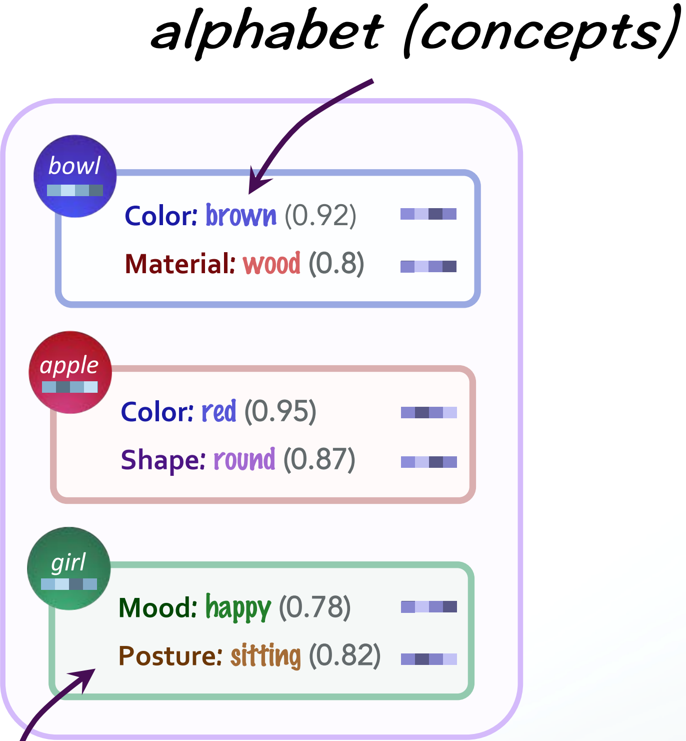
The Neural State Machine



- A **differentiable graph-based** model that simulates the operation of a **state machine**
- Uses **concepts** to represent visual information
- Reasons over semantic **world models** relating these concepts to move from facts to conclusions
- Combines the **neural** and **symbolic** approaches

The Neural State Machine



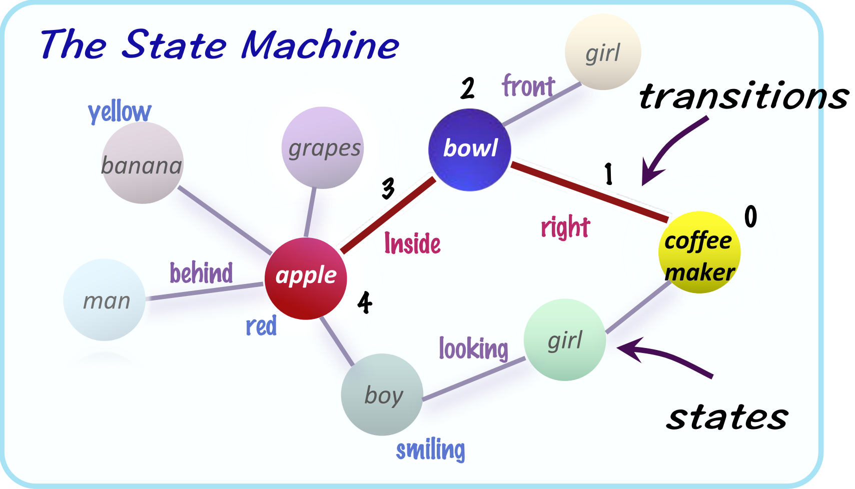


properties

Given an image, we construct a **scene graph** and treat it as a **state machine**:

- **States** correspond to **objects**
- **Transitions** correspond to **relations**
- States have **soft properties** – **attention** over **attributes**

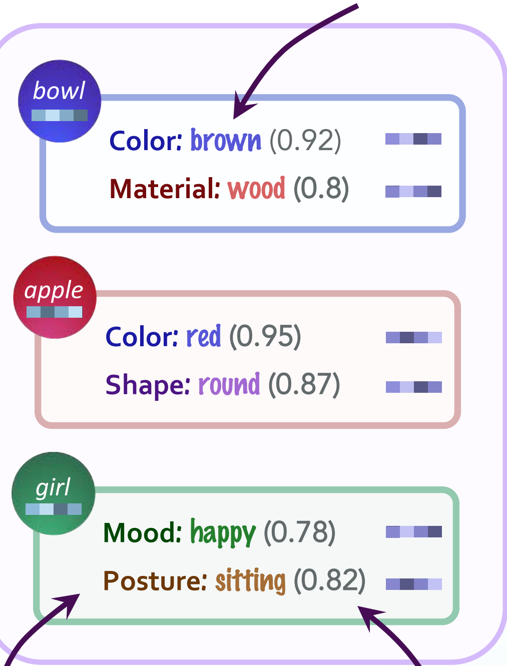
The Neural State Machine



transitions

states

alphabet (concepts)

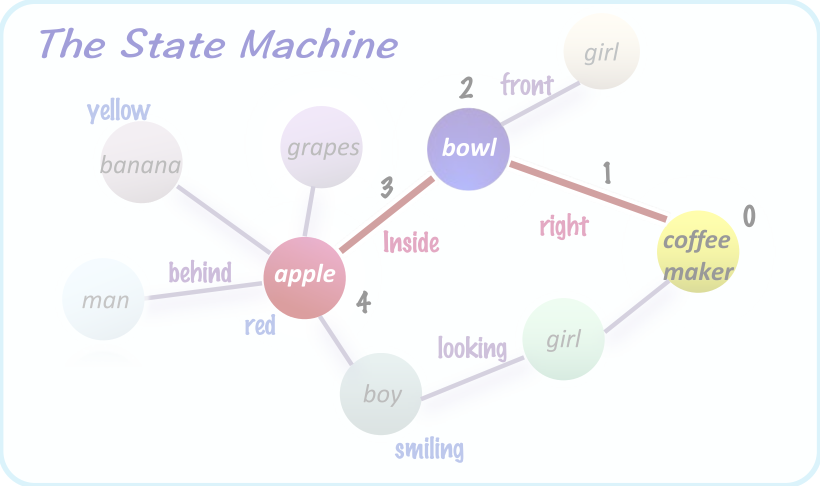


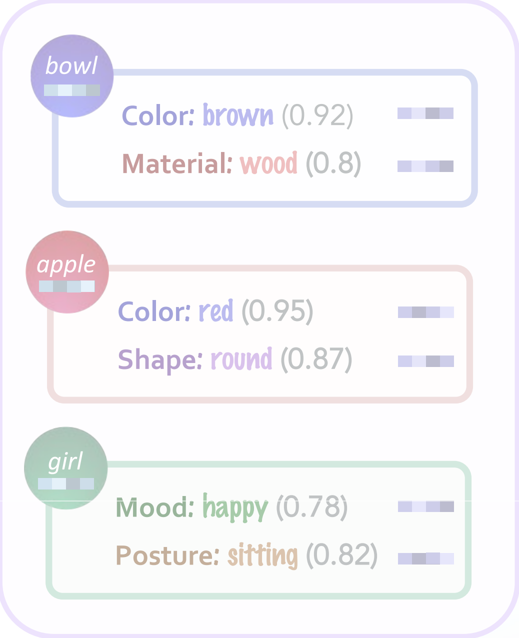
properties

disentangled representation

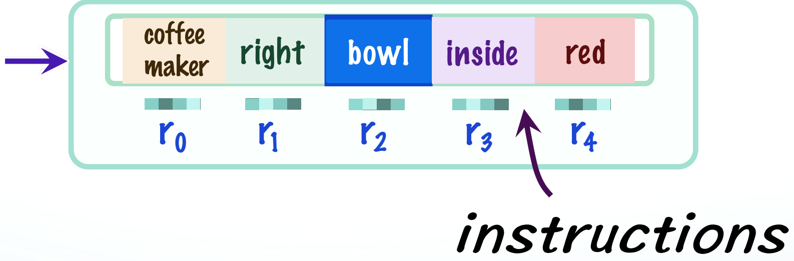
Objects are represented through a **factorized distribution** over **semantic properties** **(color, shape, material)**, defined over the **concept vocabulary**.

The Neural State Machine



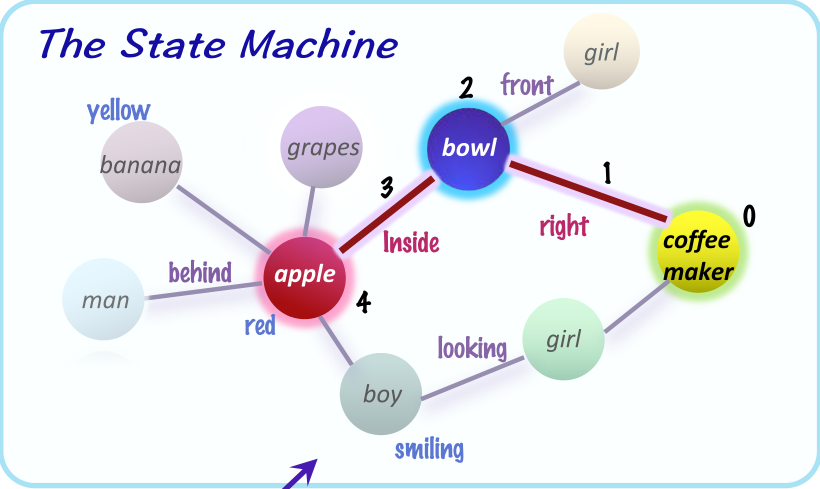


What is the red fruit inside of the bowl to the right of the coffee maker?

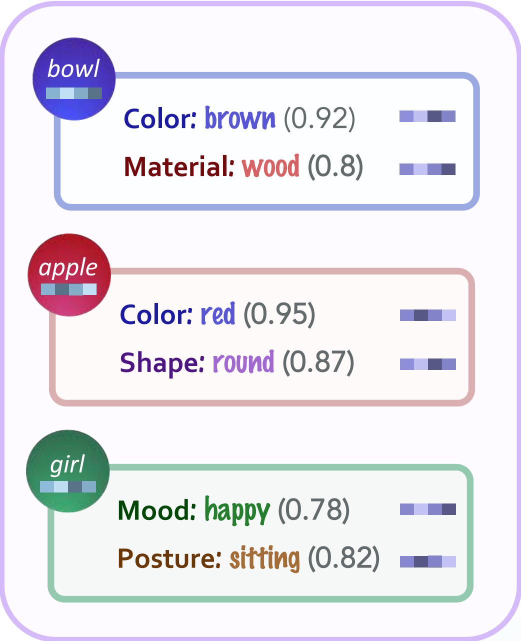


The **question** is translated into a **series of instructions** (with attention-based encoder-decoder), also defined over the **concepts**.

The Neural State Machine



The State Machine



Color: brown (0.92)

Material: wood (0.8)

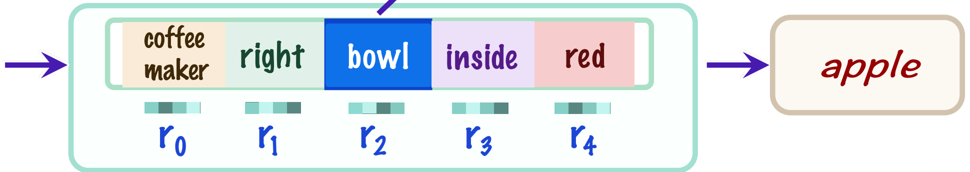
Color: red (0.95)

Shape: round (0.87)

Mood: happy (0.78)

Posture: sitting (0.82)

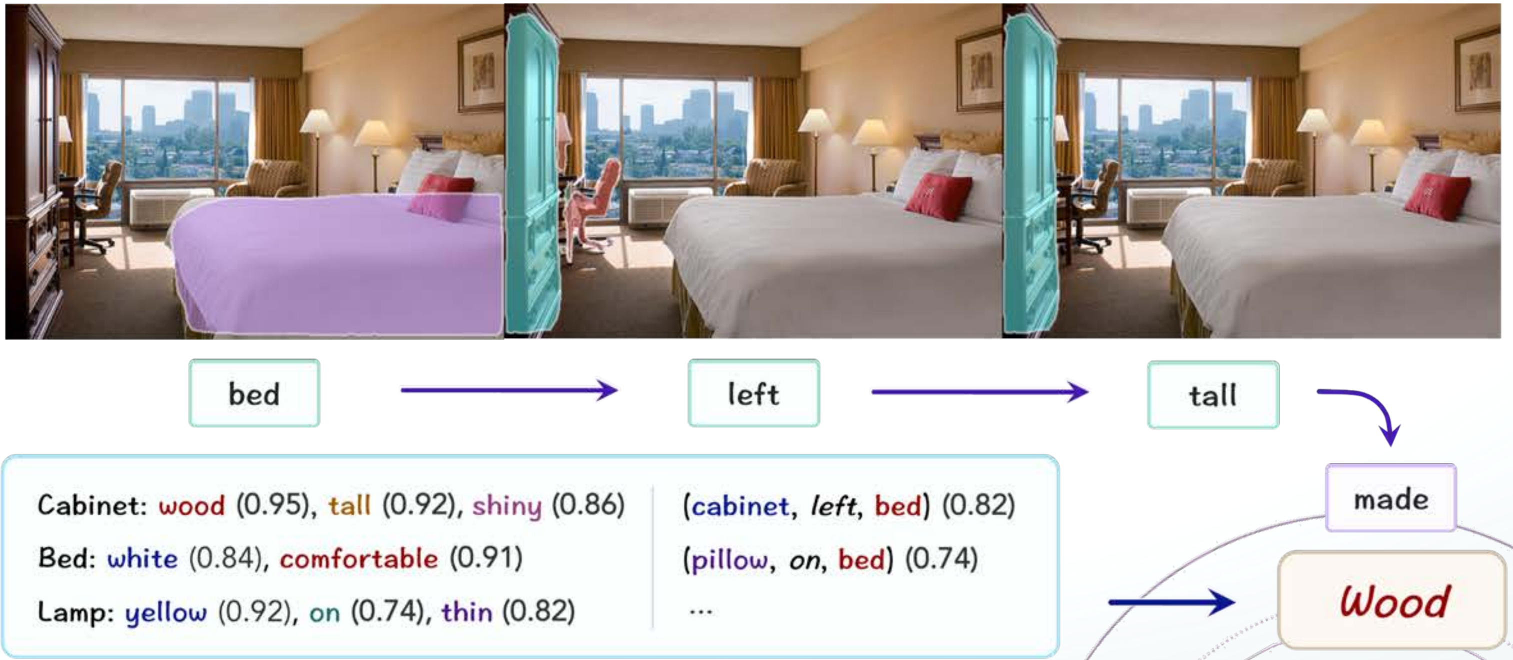
What is the red fruit inside of the bowl to the right of the coffee maker?



We **simulate a computation** of the **state machine**, feeding one **instruction** at a time and **traversing the states** until completion.

Qualitative Results

What is the tall object to the left of the bed made of?



Cabinet: wood (0.95), tall (0.92), shiny (0.86)

Bed: white (0.84), comfortable (0.91)

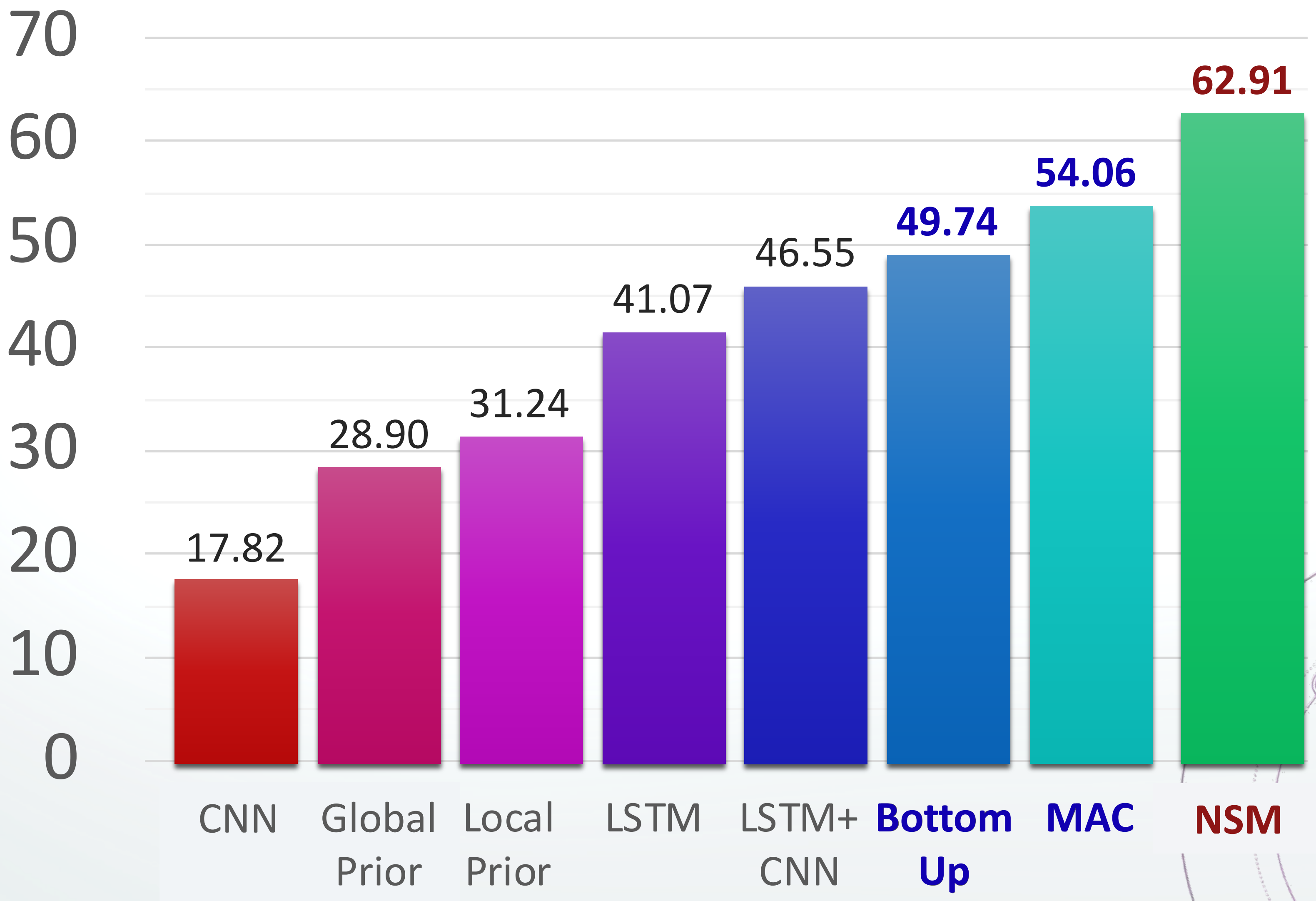
Lamp: yellow (0.92), on (0.74), thin (0.82)

(cabinet, *left*, bed) (0.82)

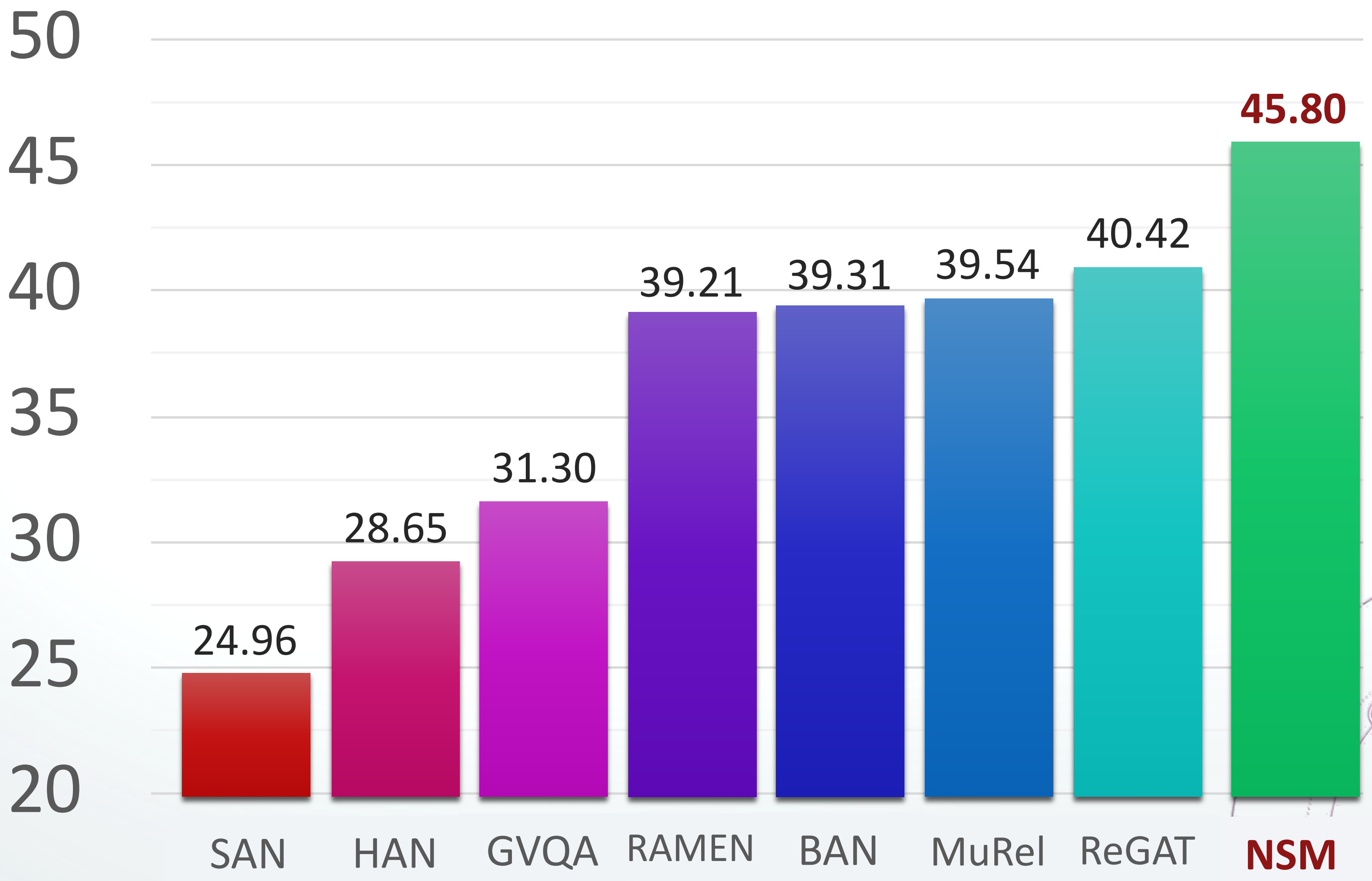
(pillow, *on*, bed) (0.74)

...

Compositional Reasoning



Generalization (VQA-CP)

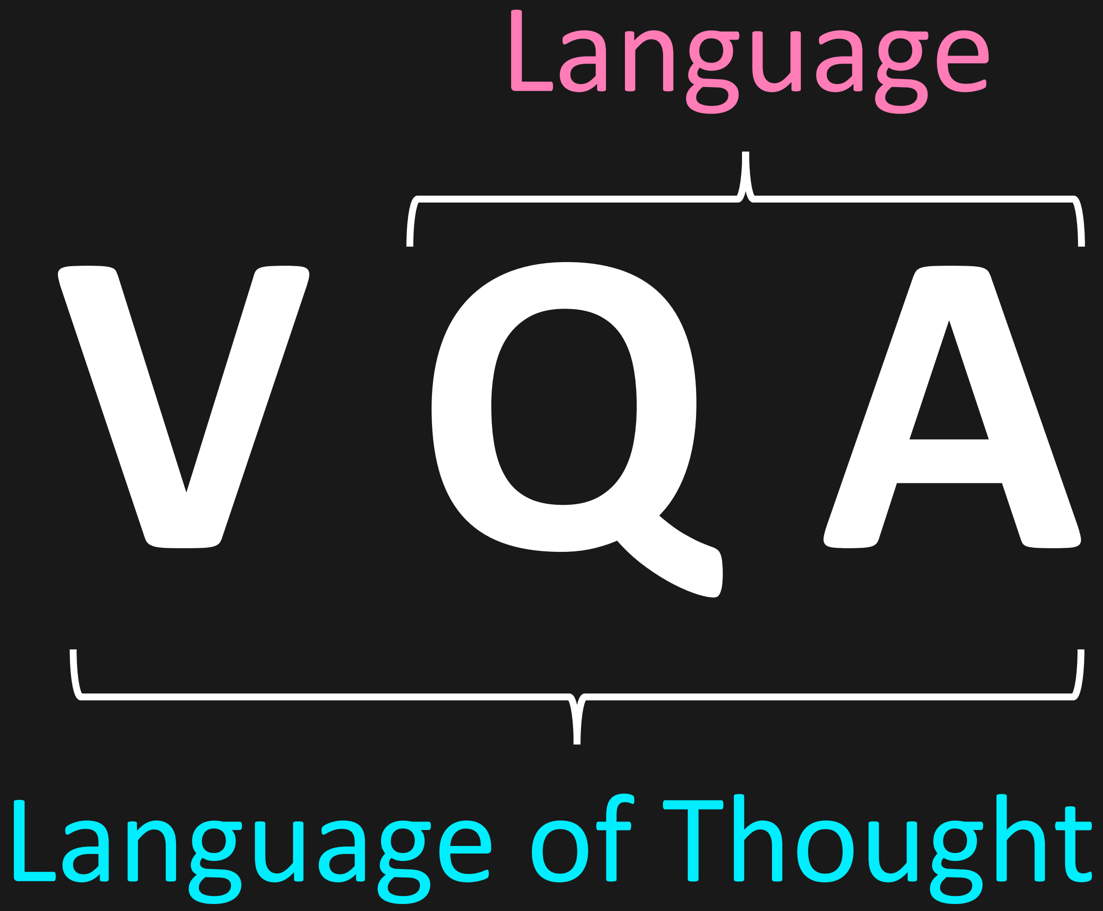


Generalization

	training	testing
structure	What is the <obj> **covered by**? **Is there a** <obj> in the **image**? What is the <obj> **made of**? **What's the name** of the <obj> **that is** <attr>?	What is **covering the** <obj>? **Do you see any** <obj>s in the **photo**? What **material makes up** the <obj>? **What is the** <attr> <obj> **called**?
content	Only questions that **do not** refer to any type of **food** or **animal** (do not have any word from these categories)	Only questions that refer to **foods** or **animals** (have a word from one of these categories)

Generalization

Model	Content	Structure
Global Prior	8.51	14.64
Lobal Prior	12.14	18.21
Vision	17.51	18.68
Language	21.14	32.88
Lang+Vision	24.95	36.51
BottomUp	29.72	41.83
MAC	31.12	47.27
NSM	**40.24**	**55.72**



Language

V Q A

Language of Thought

Let’s build networks that reason!

By iterative attention in an abstract space

over disentangled concepts

Thank you! ☺

130