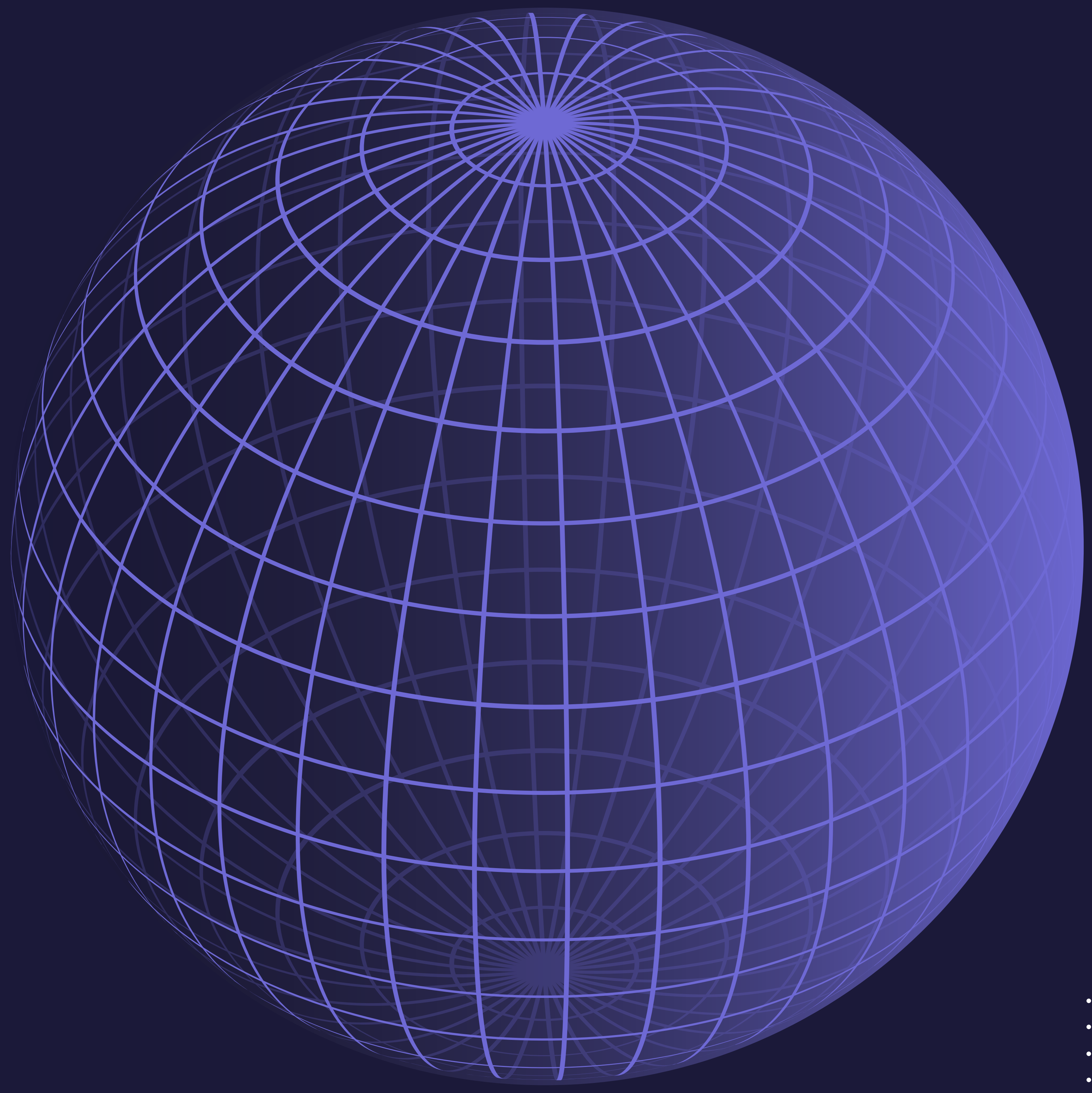
GROUNDED LANGUAGE LEARNING



VEDANT CHAVDA

WHAT IS GROUNDED LANGUAGE LEARNING?

The process of learning representations for words based on non-linguistic experience.

Where Robotics and NLP meet.

Treating language as a *Grounding Problem*.

If we can understand how humans understand language, we can hopefully figure out how to replicate it in computers.

Leading theory driving automatic grounded language learning:
– Language is a formal symbol system

HOW CAN MACHINES EXPERIENCE & LEARN LANGUAGE?

SENSORY INPUT MODALITIES!

Vision - Images and videos
Audio
Olfactory Perception
Haptic feedback
Multimodal - Using more than one input modality.

MODELS

In theory, any classification or regression algorithm could work.
Recently : Recently, neural networks:
– Autoencoders
– Boltzmann Machines
– LSTMs

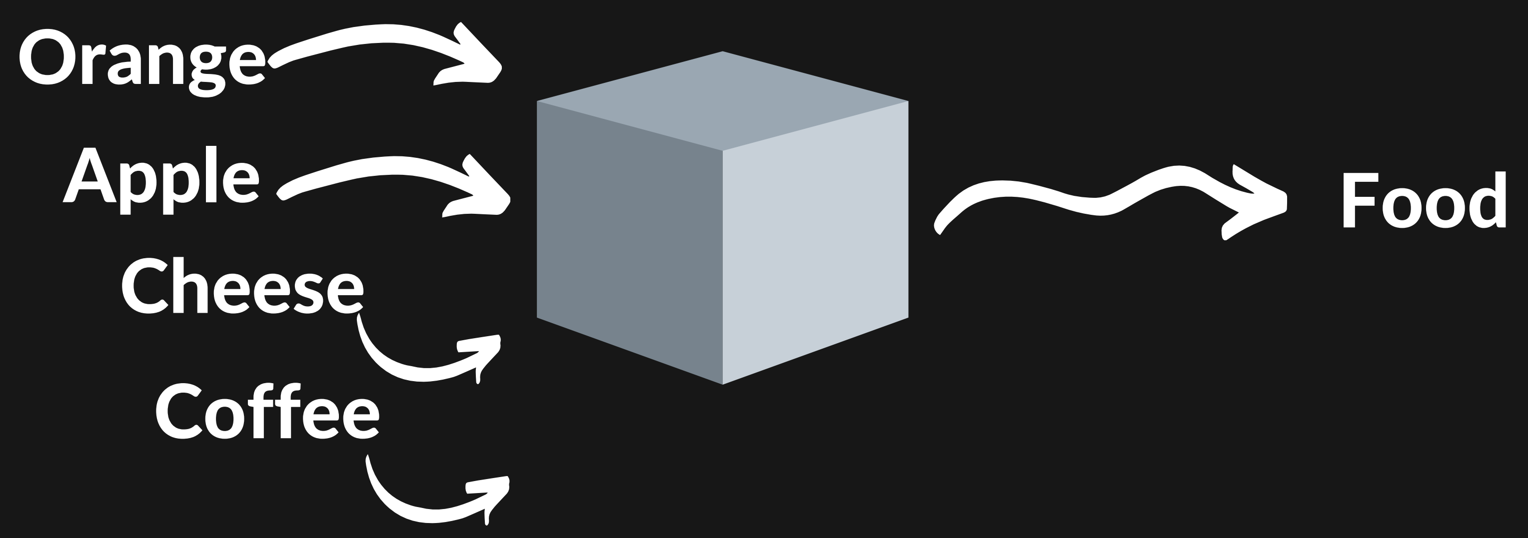
MODELS CAN LEARN MAPPINGS FROM ONE MODALITY TO THE OTHER, OR LEARN REPRESENTATIONS BASED ON MULTIPLE MODALITIES IN THE SAME SPACE.

LANGUAGE AS CLASSIFICATION

A statistical approach.

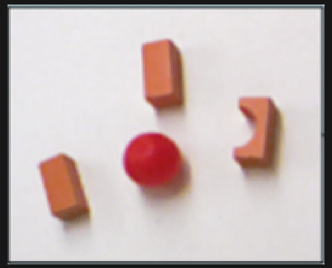
Understanding words and phrases → Classification problem

Language references physical objects and actions. Words and linguistic structures may be considered to denote a classifier.



TRAINING GROUNDED LANGUAGE CLASSIFIERS

A joint model of language and sensor data by combining language feature extraction from NL utterances with perception-based context that captures the physical setting.



Perceptually-derived world information (context) **C.**

$C_{color\text{-}orange}(x) = \{0,1\}$
$C_{color\text{-}red}(x) = \{0,1\}$
$C_{color\text{-}blue}(x) = \{0,1\}$
$S_{shape\text{-}round}(x) = \{0,1\}$
$S_{shape\text{-}cube}(x) = \{0,1\}$
...

Perceptual model that encodes the kind of knowledge that the system is expected to learn.

$C_{color\text{-}orange}(x) = \mathbf{0.87}$
$C_{color\text{-}red}(x) = 0.61$
$S_{shape\text{-}round}(x) = \mathbf{0.91}$
$S_{shape\text{-}triangle}(x) = 0.06$
...

An encoding of some particular context gives a formal, perceptual world representation ***w***.

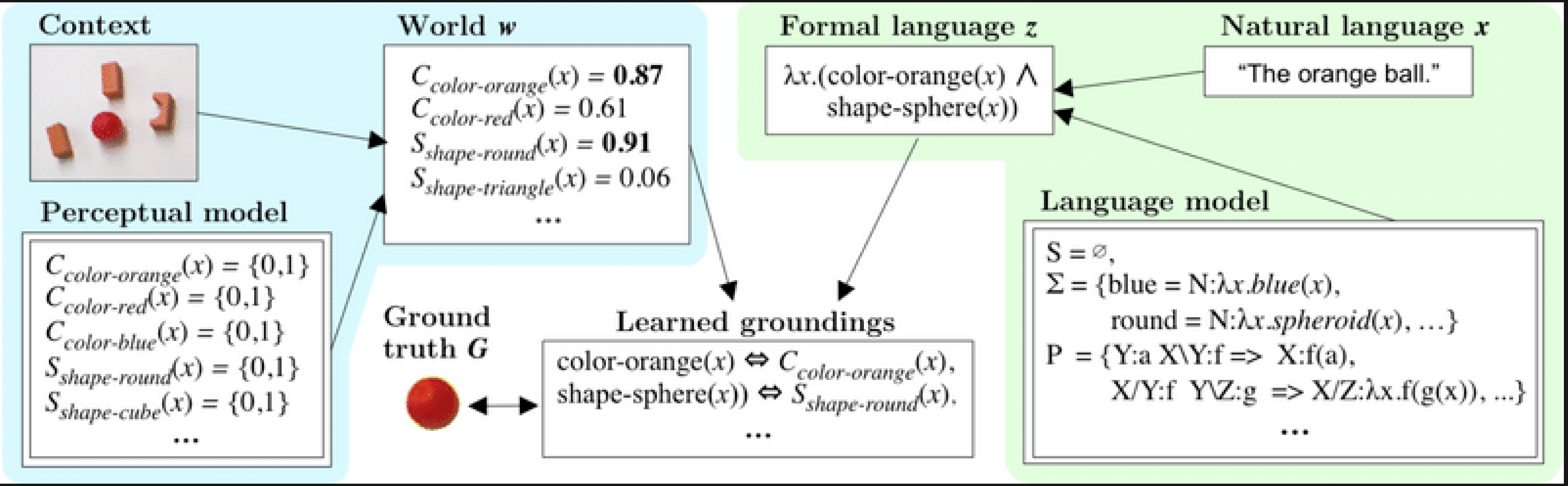
"The orange ball."

Similarly, a language model (here, a learned semantic parser) is applied to a natural language utterance ***x*.**

$\lambda x.(\text{color-orange}(x) \wedge \text{shape-sphere}(x))$

A formal semantic meaning representation, ***z***.

TRAINING GROUNDED LANGUAGE CLASSIFIERS



To learn the connection between language and perception, it is necessary to compute the probability of the correct grounding ***G*** conditioned on ***x*** and ***C*** by summing over the latent structures ***z*** and ***w***.

THE JOINT LEARNING PROBLEM

$P\,(z \mid x)$ → **Language Model**

$P\,(w \mid C)$ → **Perception Model**

Trained Independently

Joint learning problem :

$P\,(G \mid z, w)$ → Joint probability, specified by coupling above probabilities with a grounding ***G*.**

Holds the models in agreement.

Interpretation of language into world state is then given by maximizing

$$P\,(G, z, w \mid x, C) = P\,(z \mid x)\,P\,(w \mid C)\,P\,(G \mid z, w)$$

UNDERSTANDING EARLY WORD LEARNING IN SITUATED ARTIFICIAL AGENTS

FELIX HILL, STEPHEN CLARK, KARL MORITZ HERMANN & PHIL BLUNSOM

@ DEEPMIND.

INTRODUCTION

GLL - Something that can learn to locate ***referents*** of words in images.

Researchers currently lack a clear understanding of how models with no meaningful prior knowledge learn their first words.

Achieved - Neural network-based language learning agent, trained via policy-gradient methods, which can interpret single-word instructions in a simulated 3D world.

EXPERIMENTS

A simulated 3D world with a set of objects and properties, and symbolic linguistic stimuli.

In each episode - Agent presented with ***1*** word and ***2*** objects.

Agent must move by choosing between wight motor actions.

Recieves scalar ***+ve*** reward if it selects the correct object.

Early Word Learning -

- Agent success - fully learns a vocabulary of words from different semantic classes. Agent acquires new words increases rapidly after an initial slow period, like human infants.
- Investigate whether the agent exhibits a shape or colour bias.

A SITUATED WORD-LEARNING AGENT

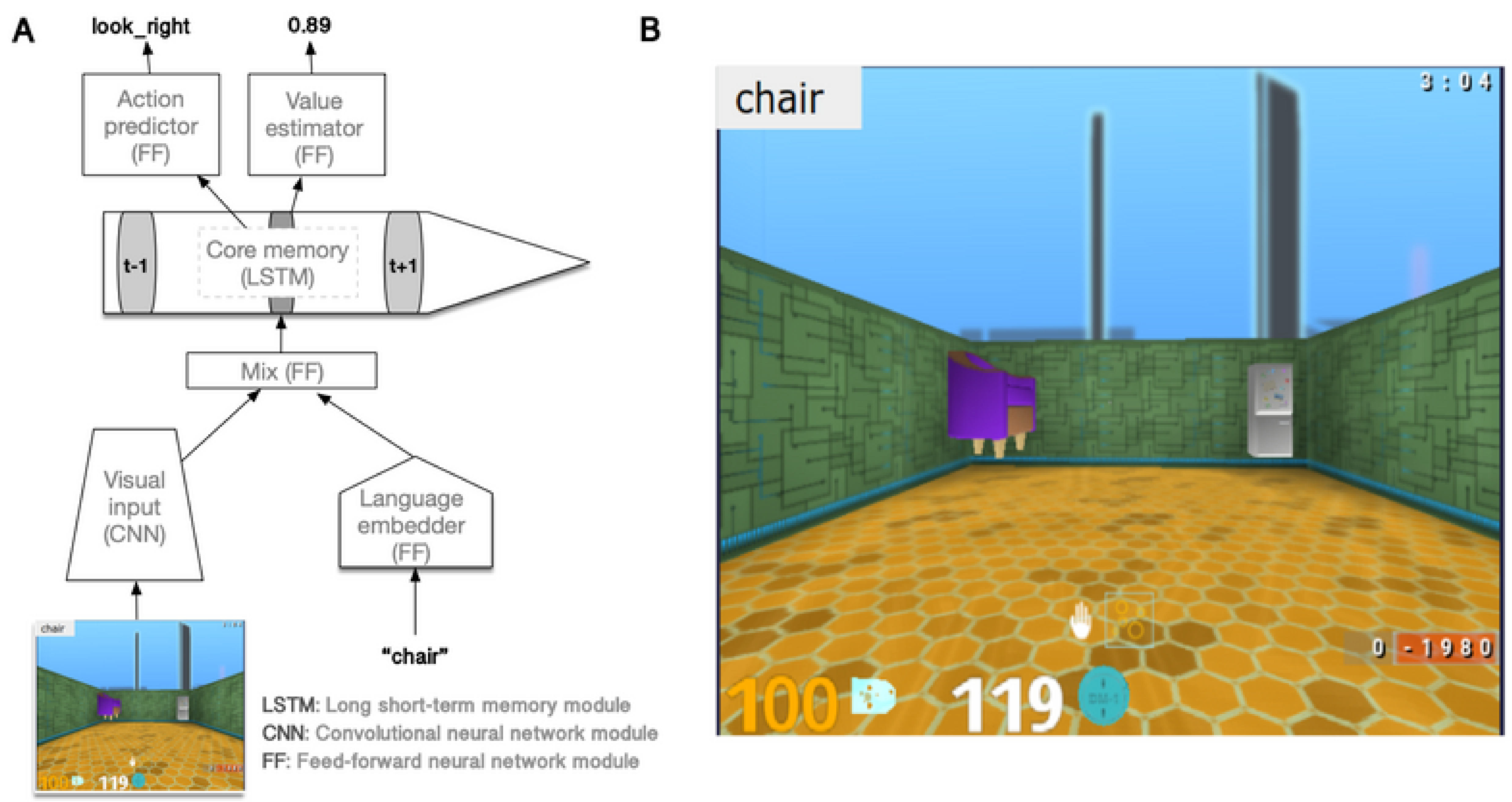


Figure 1: **A:** Schematic agent architecture. **B:** An example of the word learning environment.

FINDINGS

Initially slow learning; later word learning accelerated rapidly; an effect observed in young learners.

Why delay initially? - The need to acquire relatively language-agnostic capacities such as useful sequences of motor actions or the distinction between objects and walls.
Other possible reason - Accusing semantic knowledge. Early exposure to clear linguistic input helps child language acquisition.

Word learning biases. Regarding human learners, the simulations accord with accounts of the shape bias that emphasize the role of environmental factors in stimulating the development of a bias.

THANK YOU

Feel free to make this an open discussion for questions or clarifications.