DeepMind

Generalization and analogy in neural network models and agents

Felix Hill, DeepMind

Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks

Brenden Lake [1,2] **Marco Baroni** [2]

Abstract

Humans can understand and produce new utterances effortlessly, thanks to their compositional skills. Once a person learns the meaning of a new verb "dax," he or she can immediately un-

then dax again." This type of compositionality is central to the human ability to make strong generalizations from very limited data (Lake et al., 2017). In a set of influential and controversial papers, Jerry Fodor and other researchers have argued that neural networks are not plausible models of the mind because they are associative devices that cannot cap-

On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition

Steven Pinker

Massachusetts Institute of Technology

Alan Prince

Brandeis University

Acknowledgement

The authors contributed equally to this paper and listed their names in alphabetical order. We are grateful to Jane Grimshaw and Brian MacWhinney for providing transcripts of children's speech from the Brandeis Longitudinal Study and the Child Language Data Exchange System, respectively. We also thank Tom Bever, Jane Grimshaw, Stephen Kosslyn, Dan Slobin, an anonymous reviewer from Cognition, and the Boston Philosophy and Psychology Discussion Group for their comments on earlier drafts, and Richard Goldberg for his assistance. Preparation of this paper was supported by NSF grant IST-8420073 to Jane Grimshaw and Ray Jackendoff of Brandeis University, by NIH grant HD 18381-04 to Steven Pinker, and by a grant from the Alfred P. Sloan Foundation to the MIT Center for Cognitive Science. Requests for reprints may be sent to Steven Pinker at the Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139 or Alan Prince at the Linguistics and Cognitive Science Program, Brown 125, Brandeis University, Waltham MA 02254.

TSXS	TXS	TSSS
TSSXXVV	TSSSXXVV	TXXV
TXXTTVV	TSXXTVV	TSSX
TSXXTVPS	TXXTVPS	TXXV
PVV	PTTVV	PTVP
PVPXVV	PTVPXTVV	PVPX
PTVPS	PVPXTVPS	PTTT


BEST THING
January 1, 2000
Full-scale colonization
begins


MAKE GIFS AT GIFSOUP.COM

'Compositionality' in static vs temporally correlated training data

Find a ___

"blue guitar"
"red ball",
"green ladder".....

Train instructions

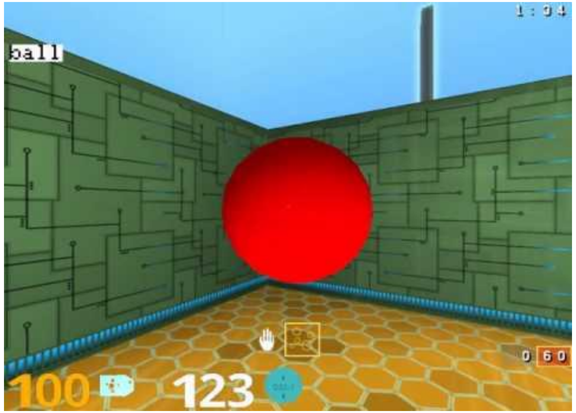
Find a ___

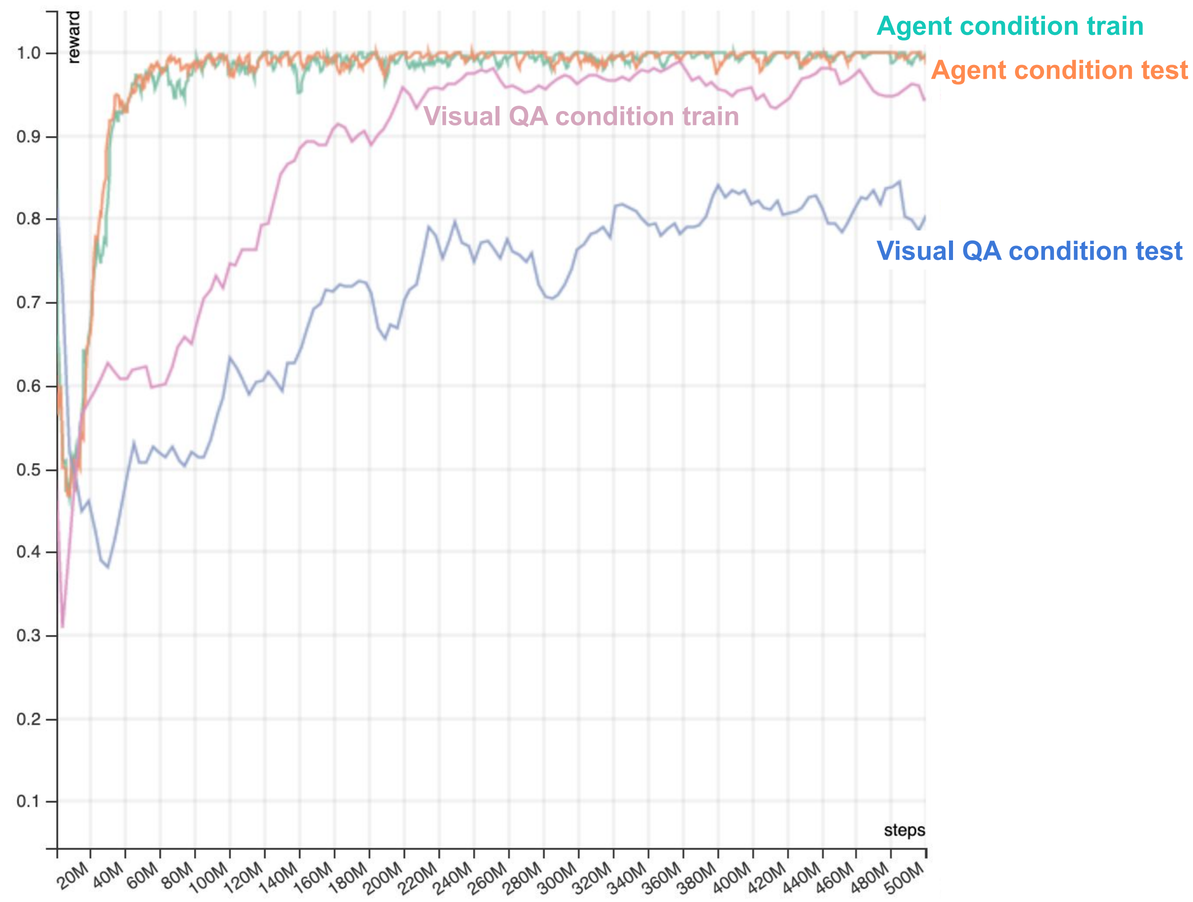
"red guitar"
"green ball",
"blue ladder"

Test instructions

Condition A

Condition B





reward
steps
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
20M
40M
60M
80M
100M
120M
140M
160M
180M
200M
220M
240M
260M
280M
300M
320M
340M
360M
380M
400M
420M
440M
460M
480M
500M
Agent condition train
Agent condition test
Visual QA condition train
Visual QA condition test

Training

Lift a ___

boat
bus car
helicopter
keyboard
plane robot
rocket train
racket candle
mug
hairdryer
picture frame
plate
potted plant
roof block
rubber duck

Put a ___ on a bed

boat
bus car
helicopter
keyboard
plane robot
rocket train
racket candle

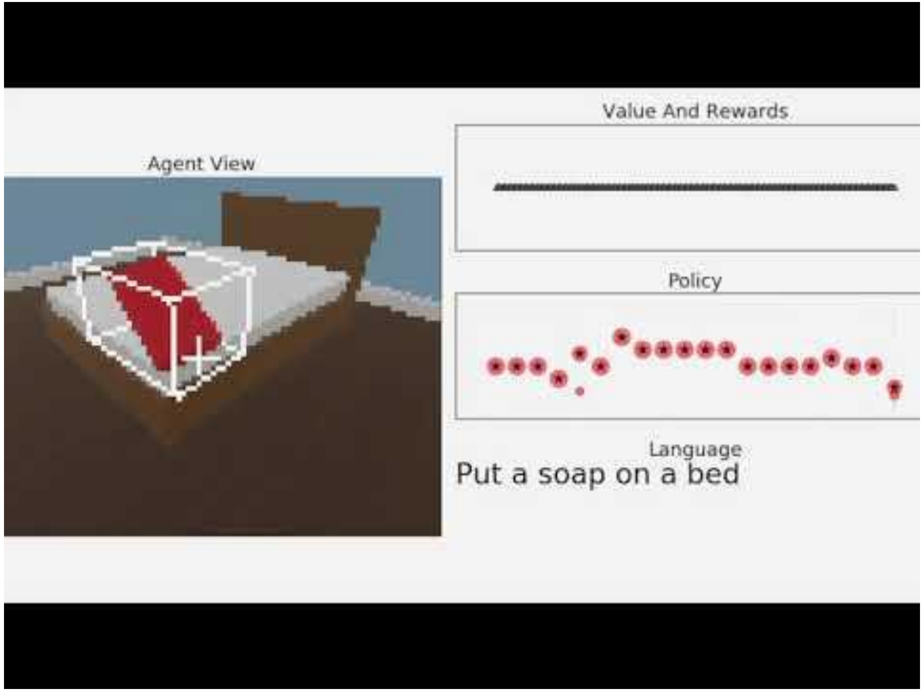
Testing

Put a ___ on a bed

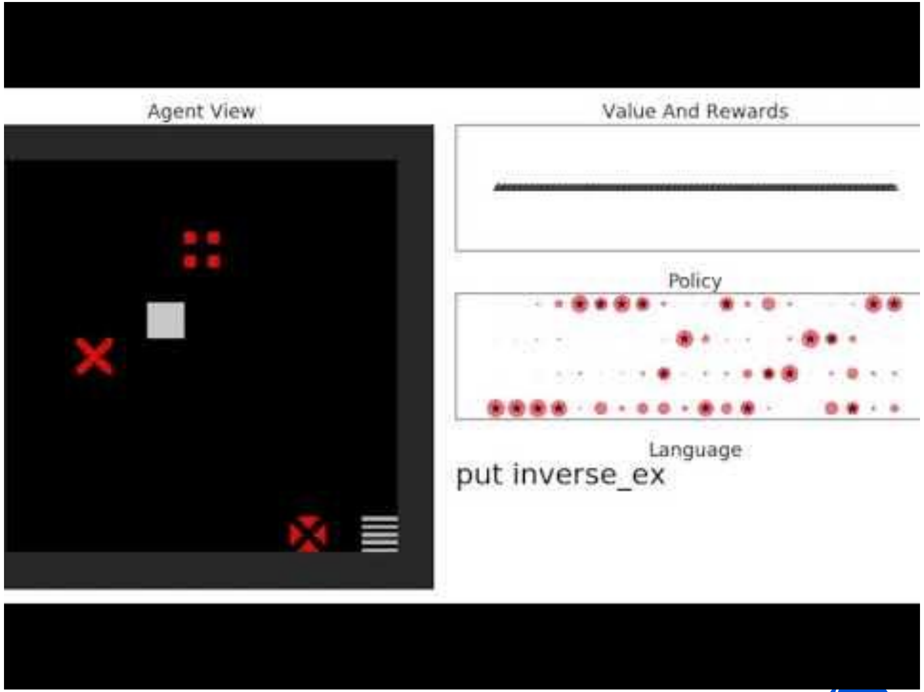
mug
hairdryer
picture frame
plate
potted plant
roof block
rubber duck

"Same" test of generalization in 3D first-person and 2D top-down

Condition A

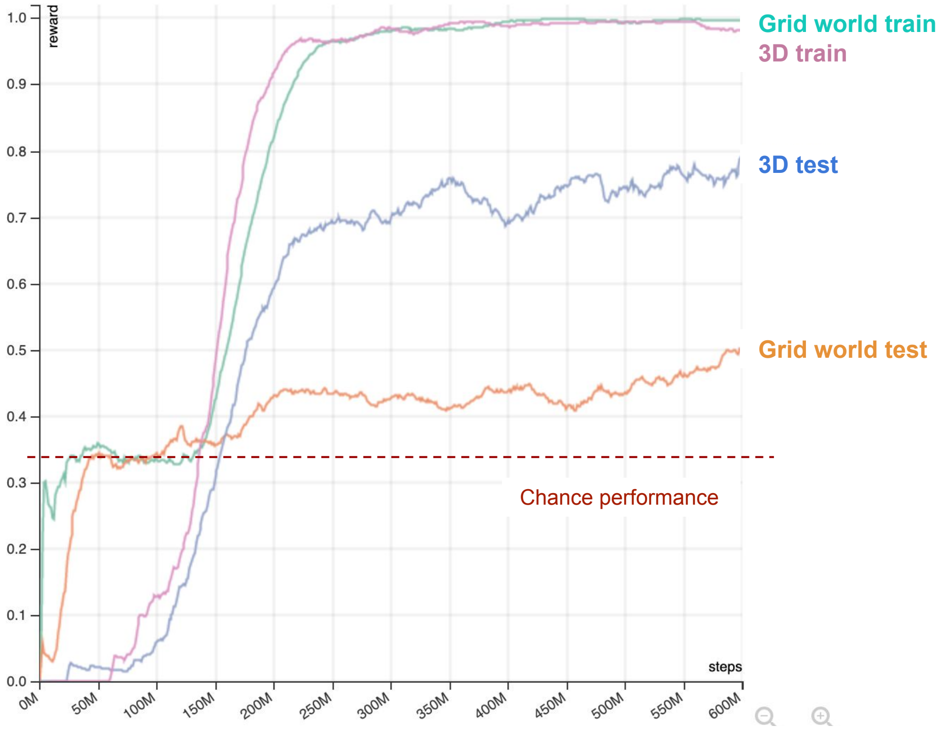


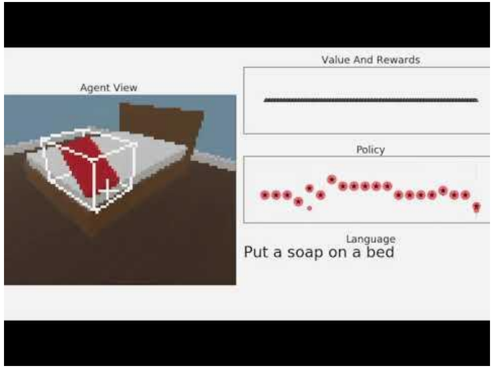
Condition B

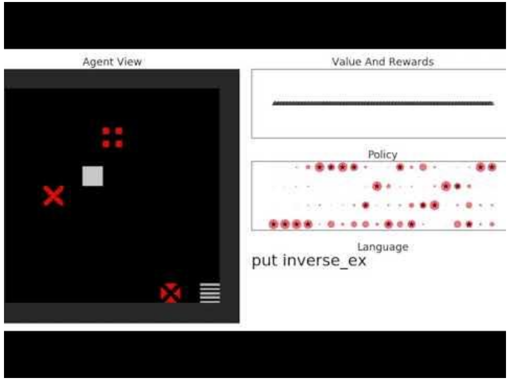


Environmental Drivers of Systematicity and Generalization in a Situated Agent. Hill et al. ICLR 2020

The agent's training experience affects 'compositionality'

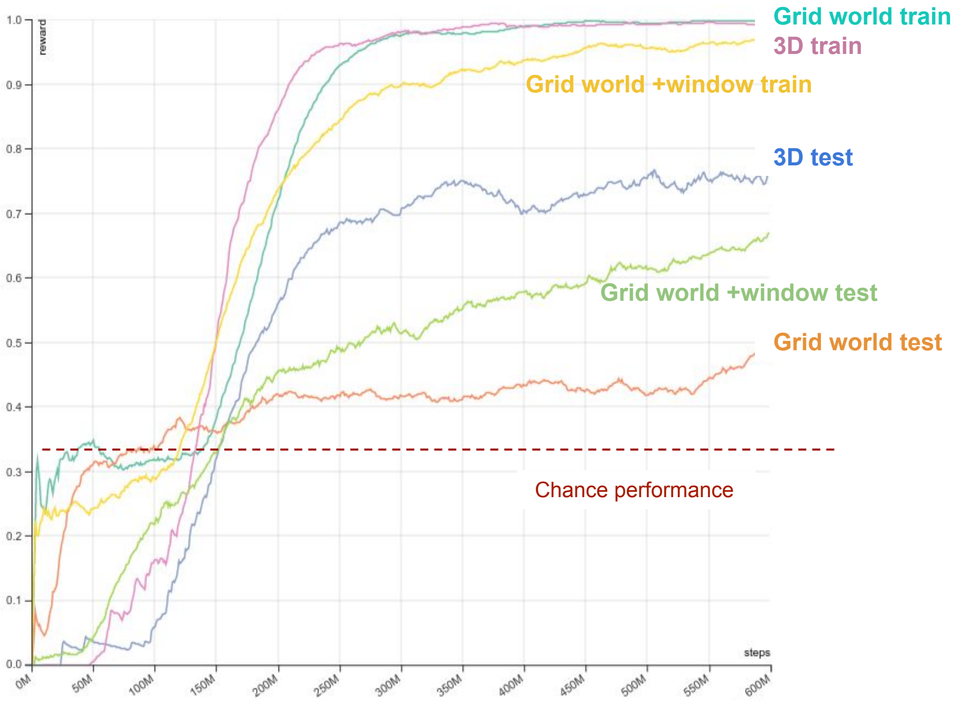


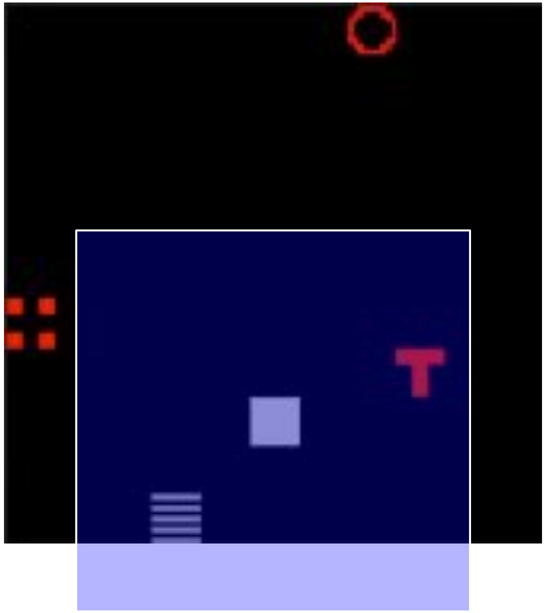




Environmental Drivers of Systematicity and Generalization in a Situated Agent. Hill et al. ICLR 2020

The agent's perspective affects 'compositionality'





DeepMind

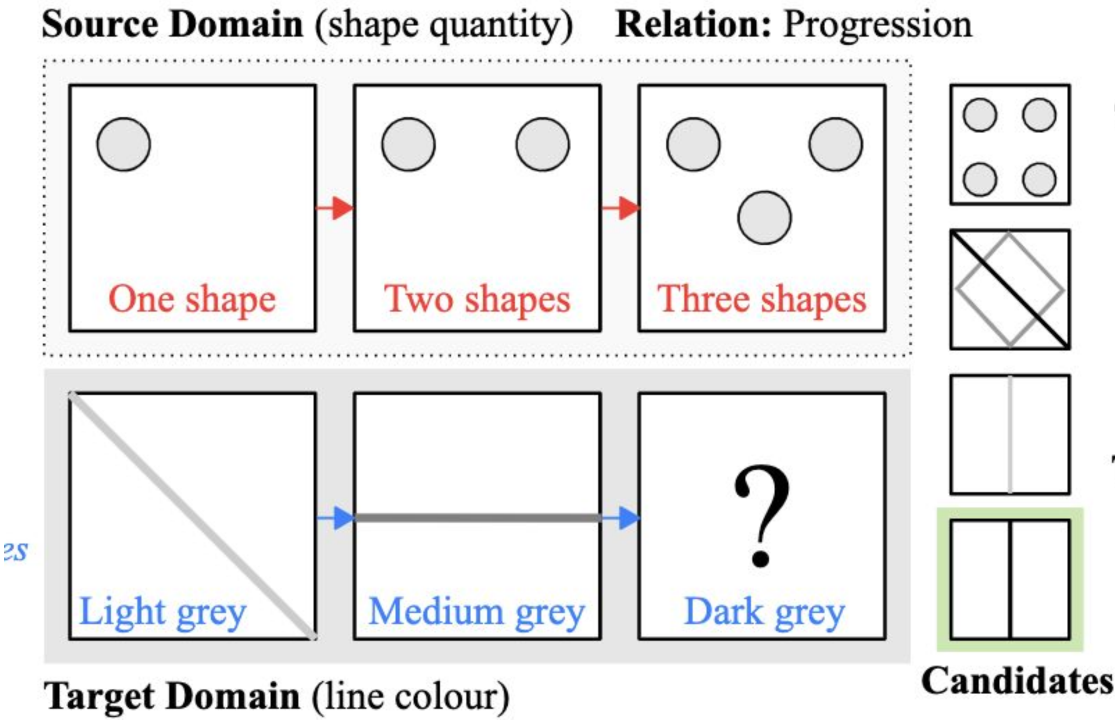
Some things to think about

- Why do some think that neural networks won't generalise in ways that humans do?
-
- What are some alternative approaches?

Why are people like me sceptical about these alternatives?

What about 'higher' cognition?

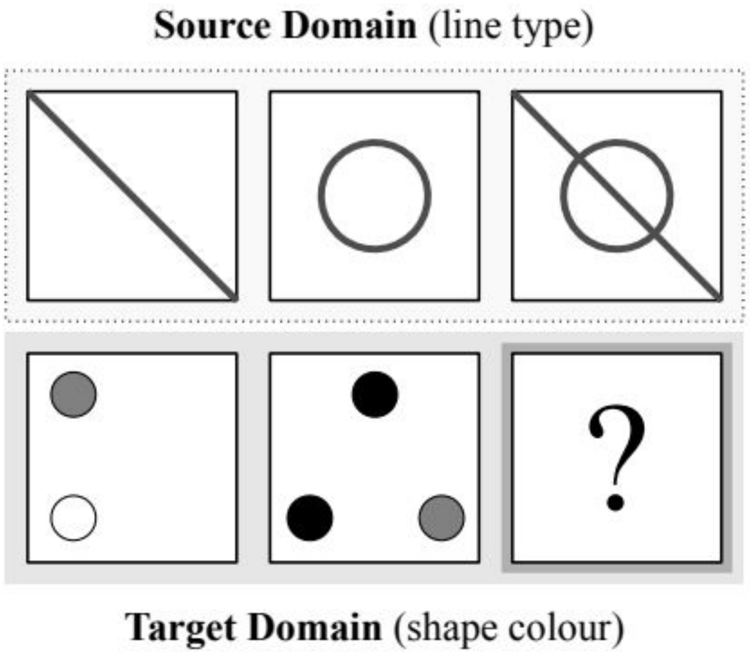
How does a model's training *experience* affect analogy learning?



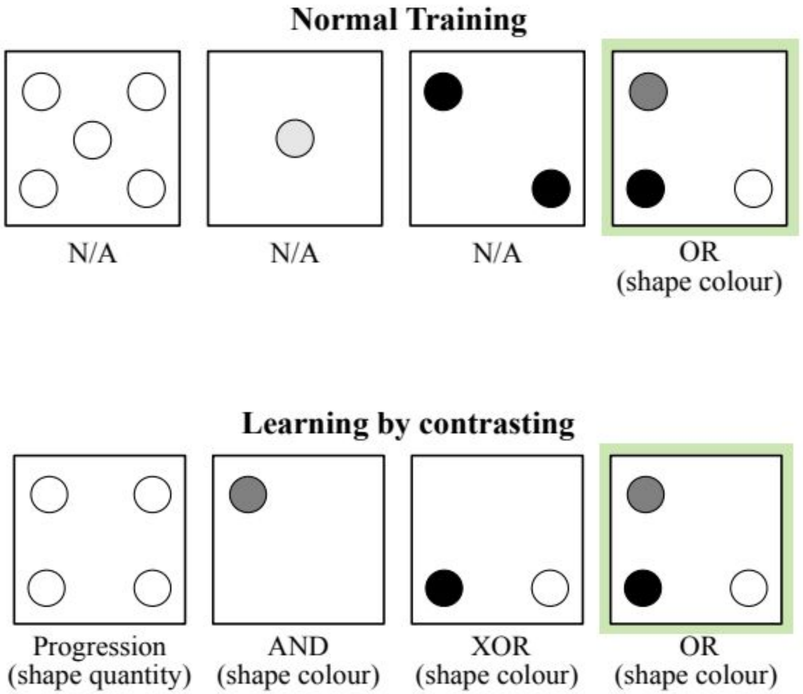
Types of Domains
shape quantity
shape colour
shape type
shape size
shape position
line type
line colour

Types of Relations
Progression
XOR
OR
AND

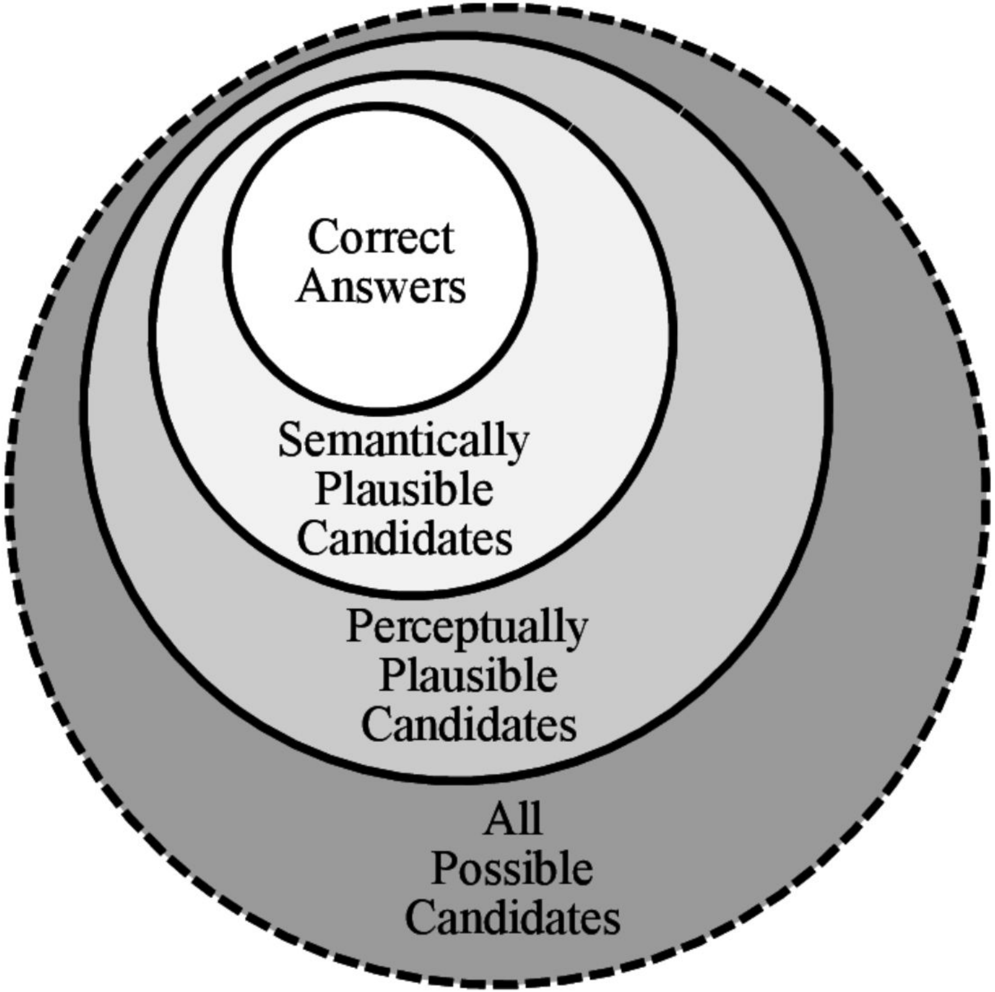
Learning to make analogies by contrasting abstract relational structure. Hill et al. ICLR 2019



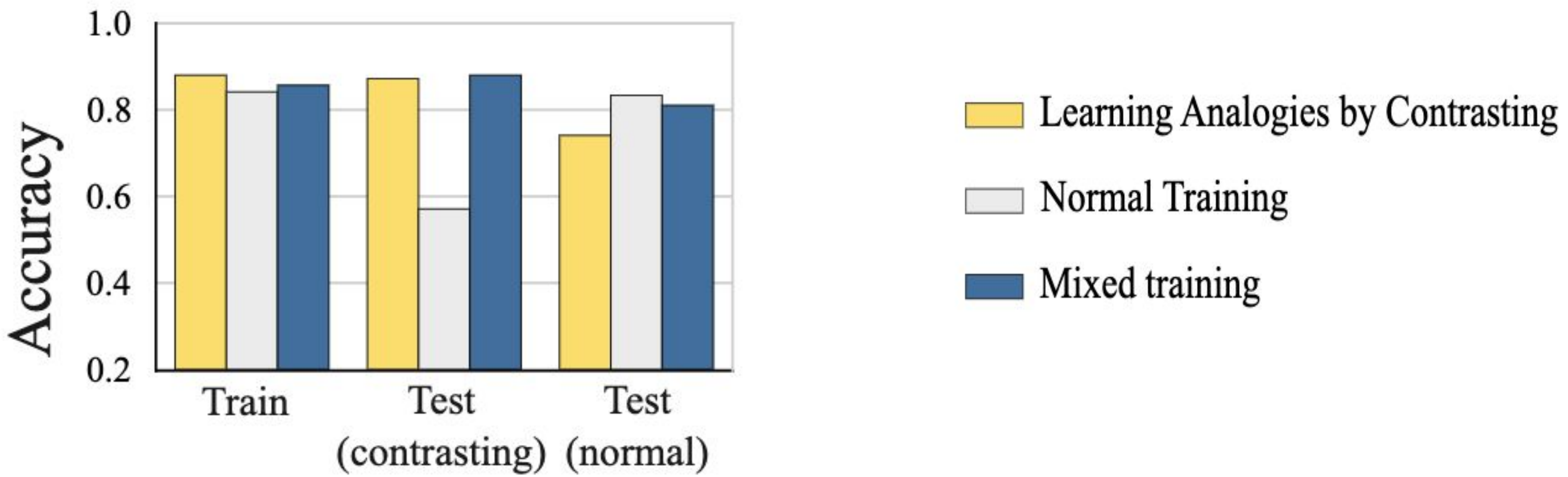
Source Domain (line type)
?
Target Domain (shape colour)



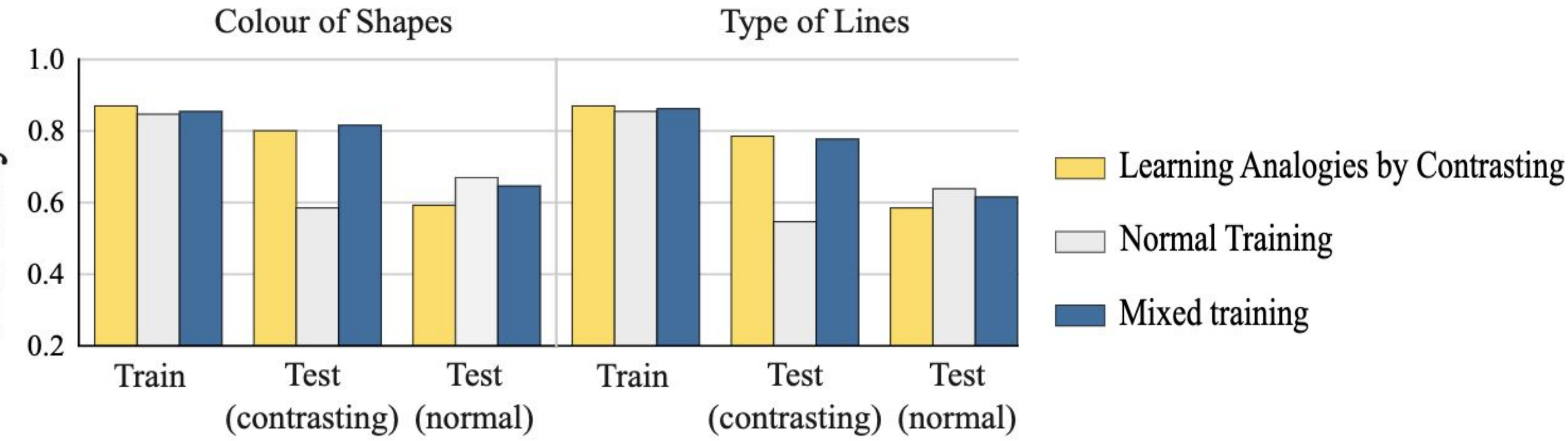
Normal Training
N/A
N/A
N/A
OR
(shape colour)
Learning by contrasting
Progression
(shape quantity)
AND
(shape colour)
XOR
(shape colour)
OR
(shape colour)


Correct Answers
Semantically Plausible Candidates
Perceptually Plausible Candidates
All Possible Candidates

Unseen domain transition



Unseen target domain



DeepMind

Some things to think about

What do these experiments suggest about ways to train (or 'educate') deep networks?

Could a symbolic model achieve the sort of generalisations observed here?

DeepMind

Back to a 3D world!

Learning to use a word in one shot

GPT-3

Prompt: To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

GPT3: One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduddles.

Language Models are Few Shot Learners. Brown et al. 2020. arXiv.

The 'dax' task to probe *fast-mapping*

Acquiring a Single New Word. Carey & Bartlett (1978).

SHARE

REPORT

Word Learning in a Domestic Dog: Evidence for "Fast Mapping"

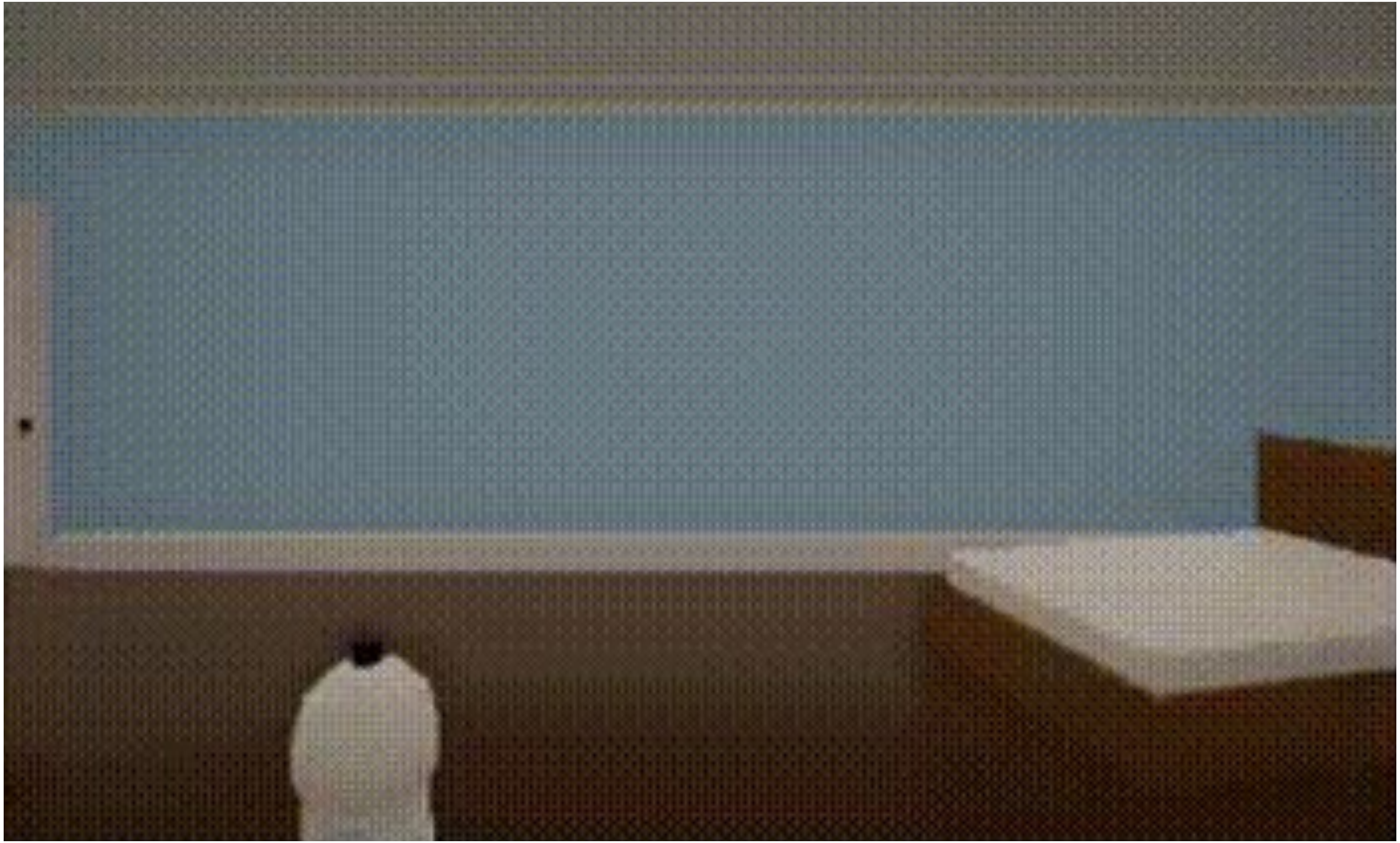
Juliane Kaminski, Josep Call, Julia Fischer*
+ See all authors and affiliations

Science 11 Jun 2004:
Vol. 304, Issue 5677, pp. 1682-1683
DOI: 10.1126/science.1097859

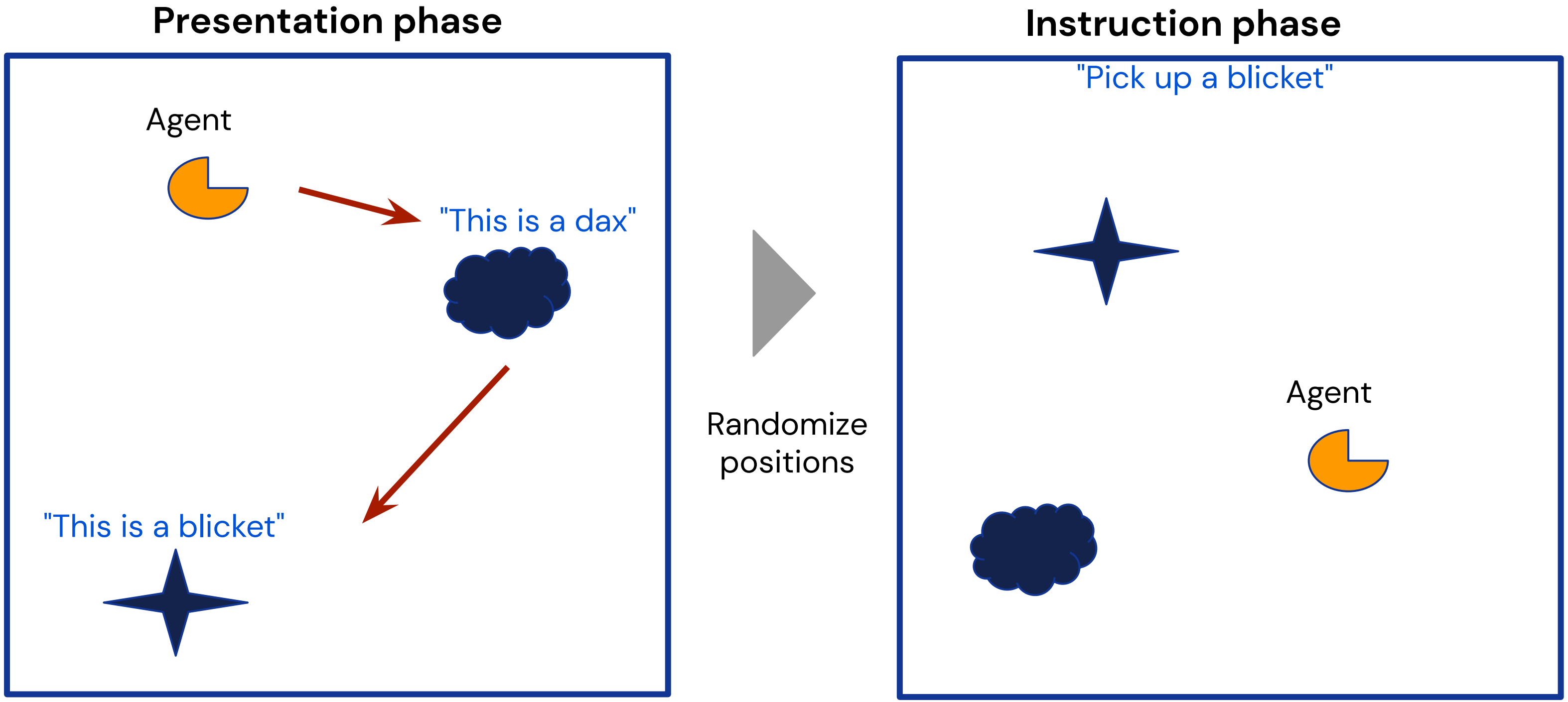
Abstract

During speech acquisition, children form quick and rough hypotheses about the meaning of a new word after only a single exposure—a process dubbed "fast mapping." Here we provide evidence that a border collie, Rico, is able to fast map. Rico knew the labels of over 200 different items. He inferred the names of novel items by exclusion learning and correctly retrieved those items right away as well as 4 weeks after the initial exposure. Fast mapping

Fast-mapping in a neural-network agent?

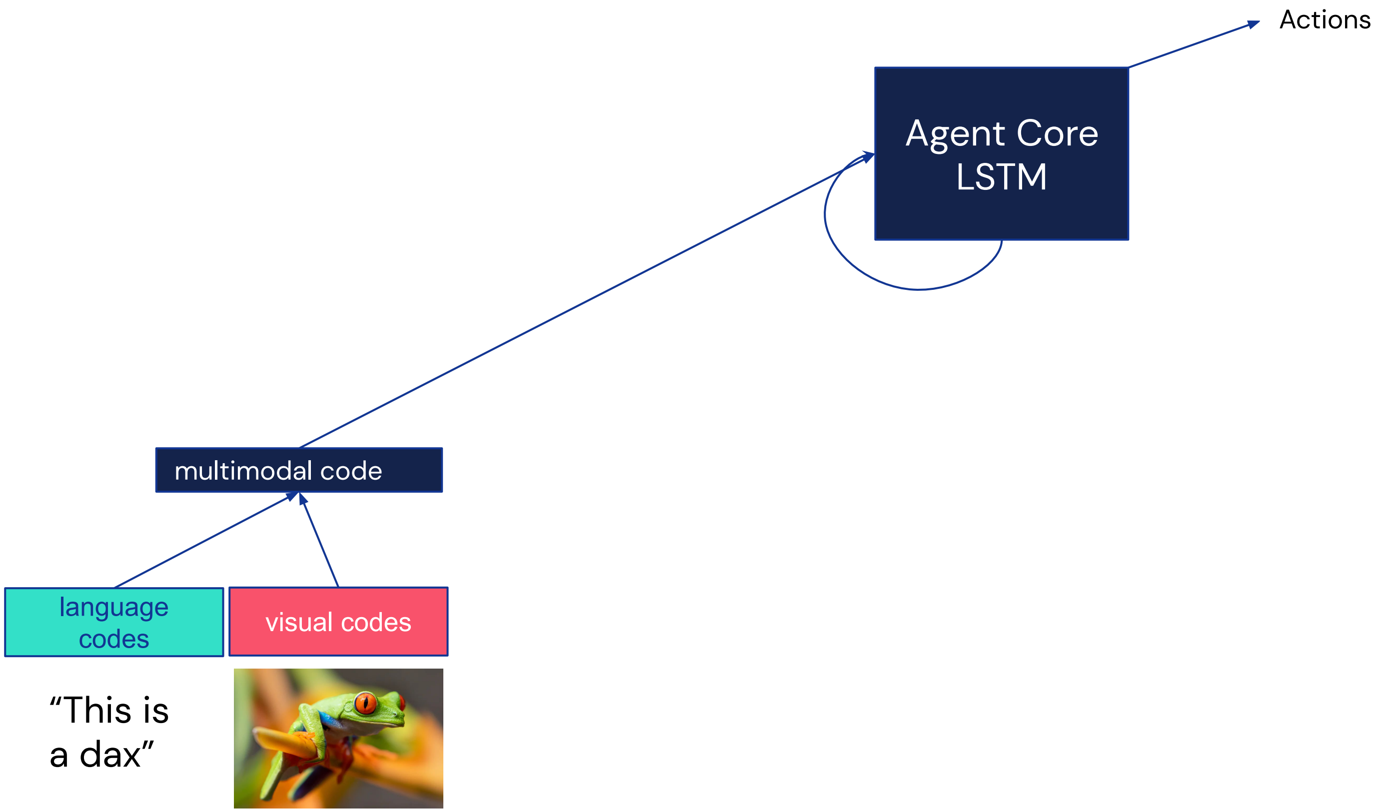


A simulated dax task



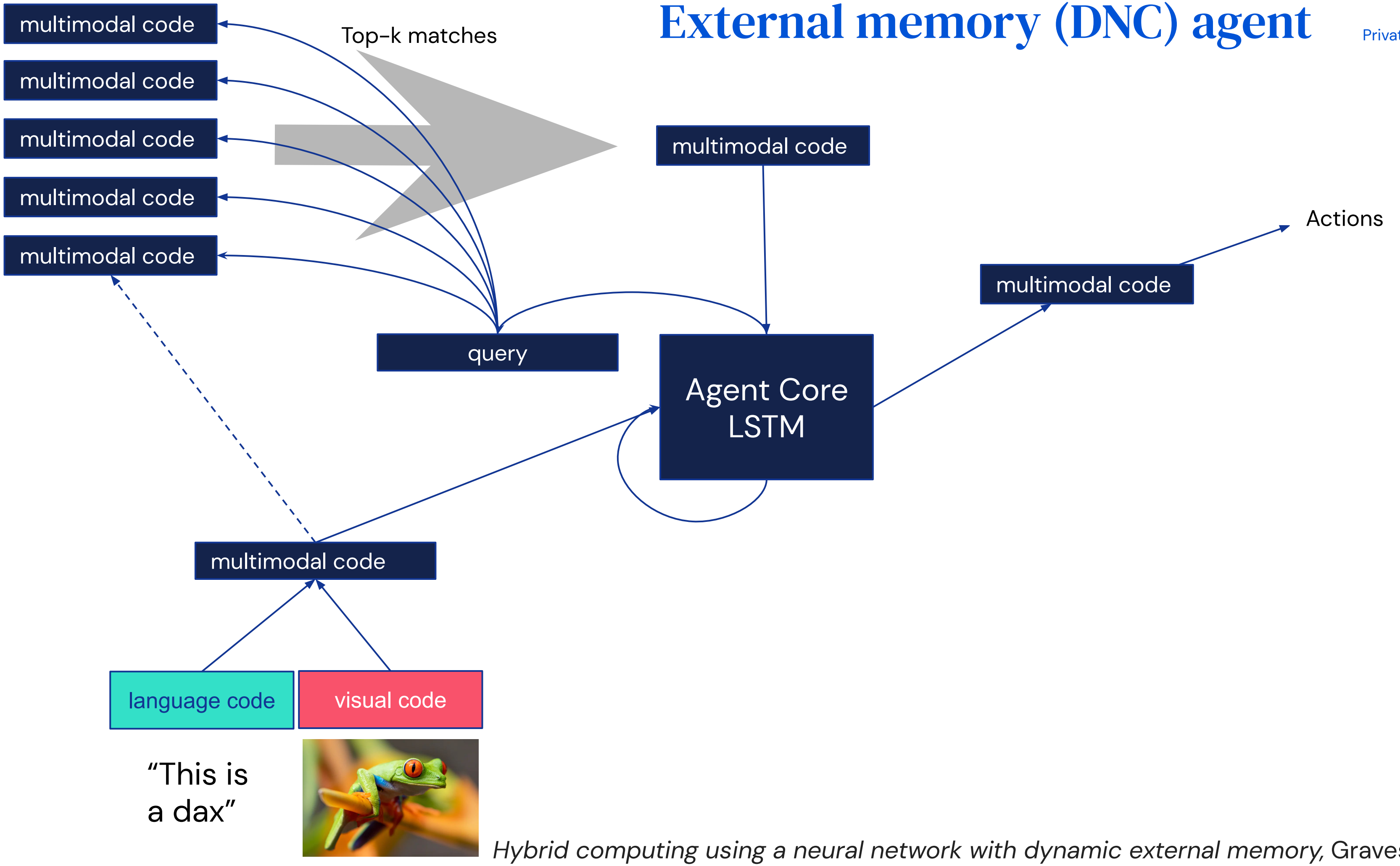
A simulated dax task

Default LSTM-based agent

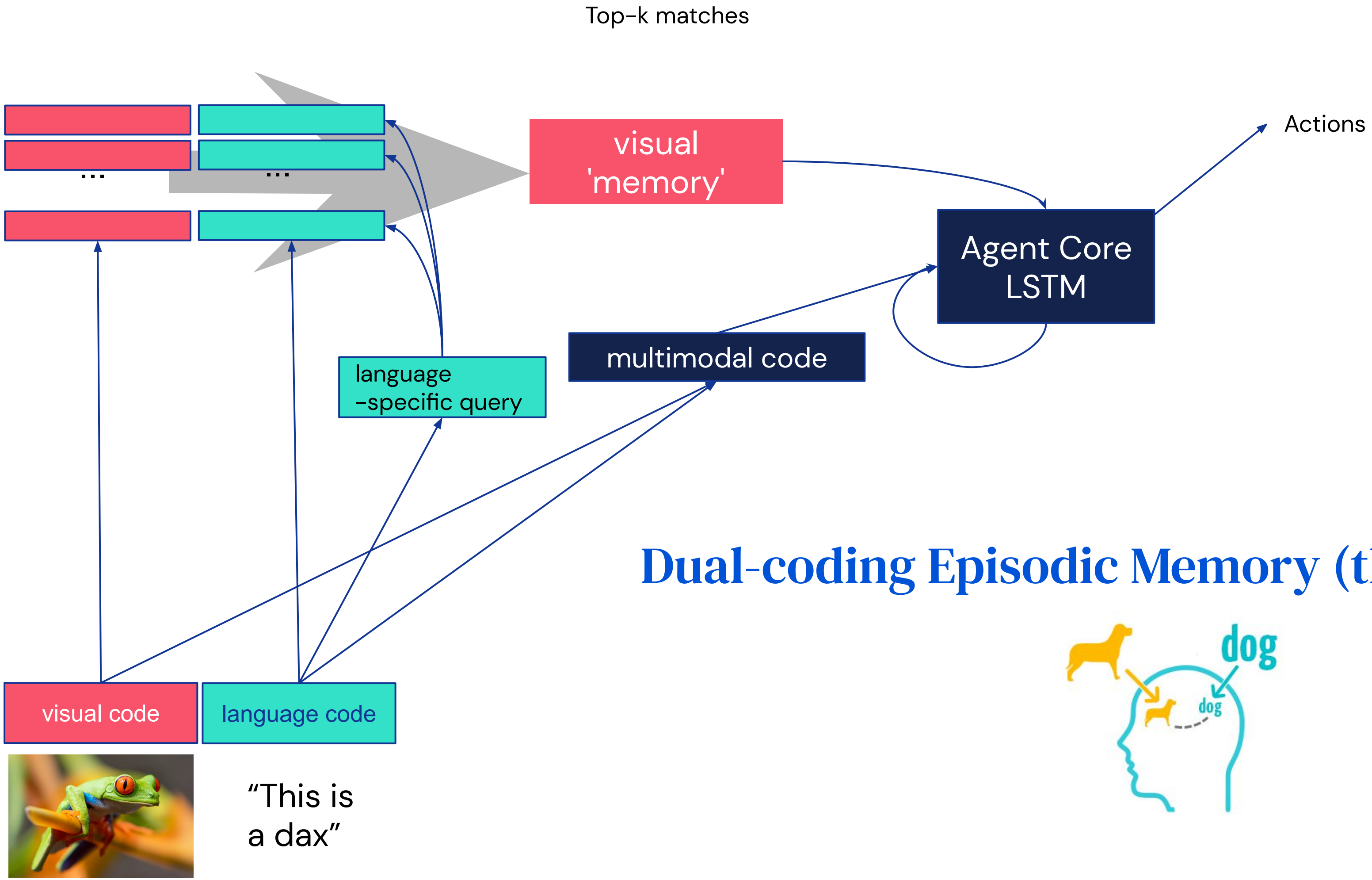


External memory (DNC) agent

Private & Confidential

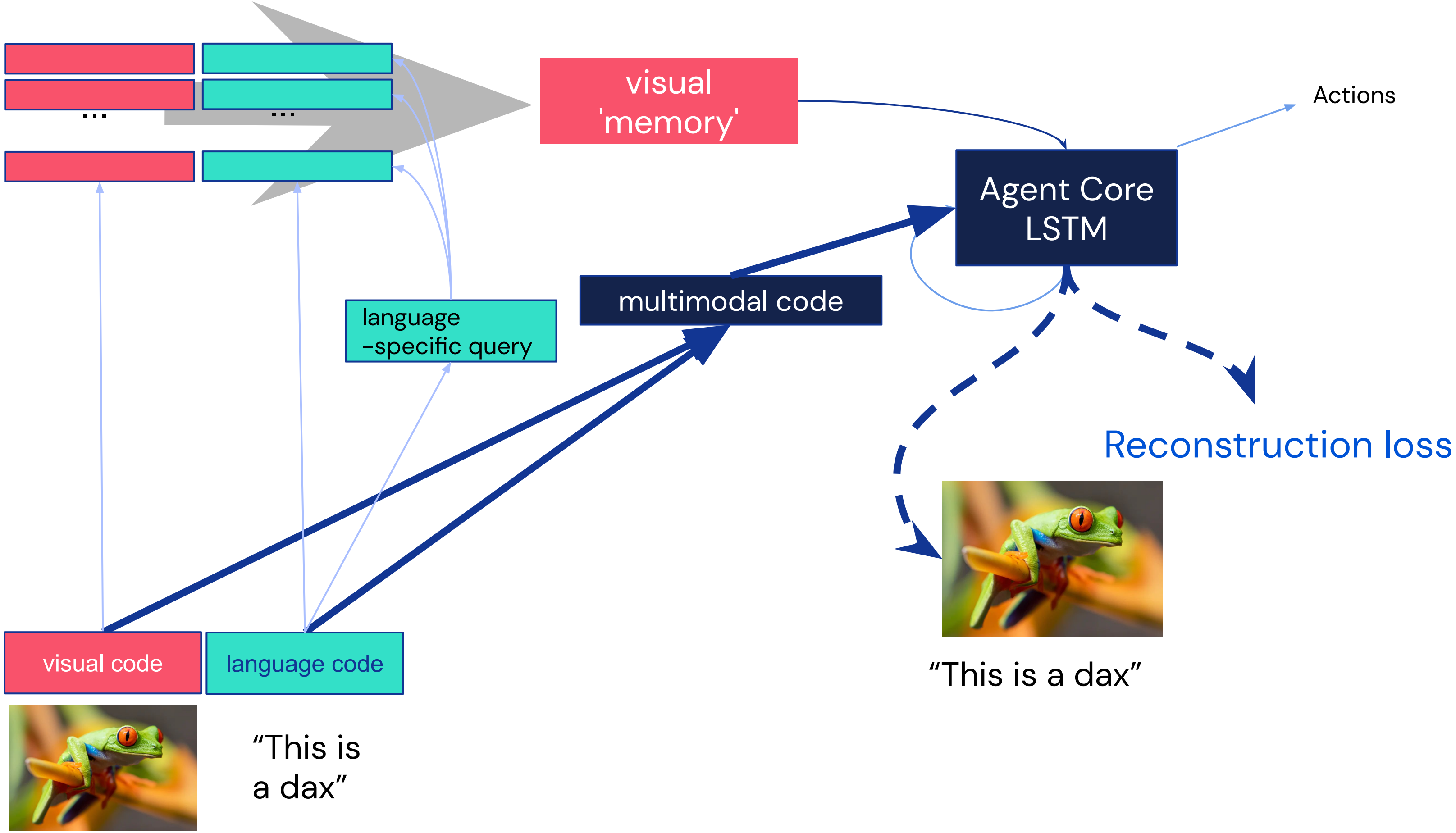


Hybrid computing using a neural network with dynamic external memory, Graves et al. 2016



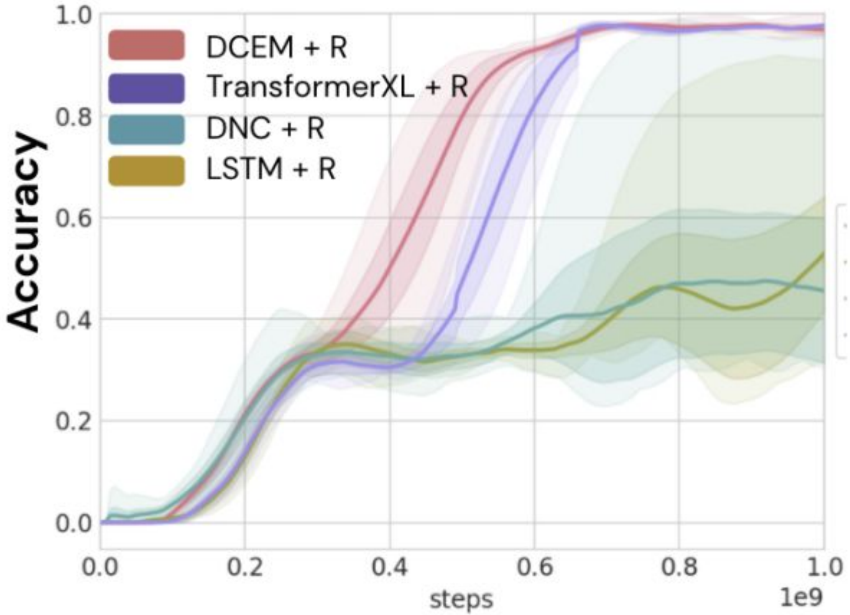
Dual-coding Episodic Memory (this work)

Reconstruction loss (semi-supervised learning)

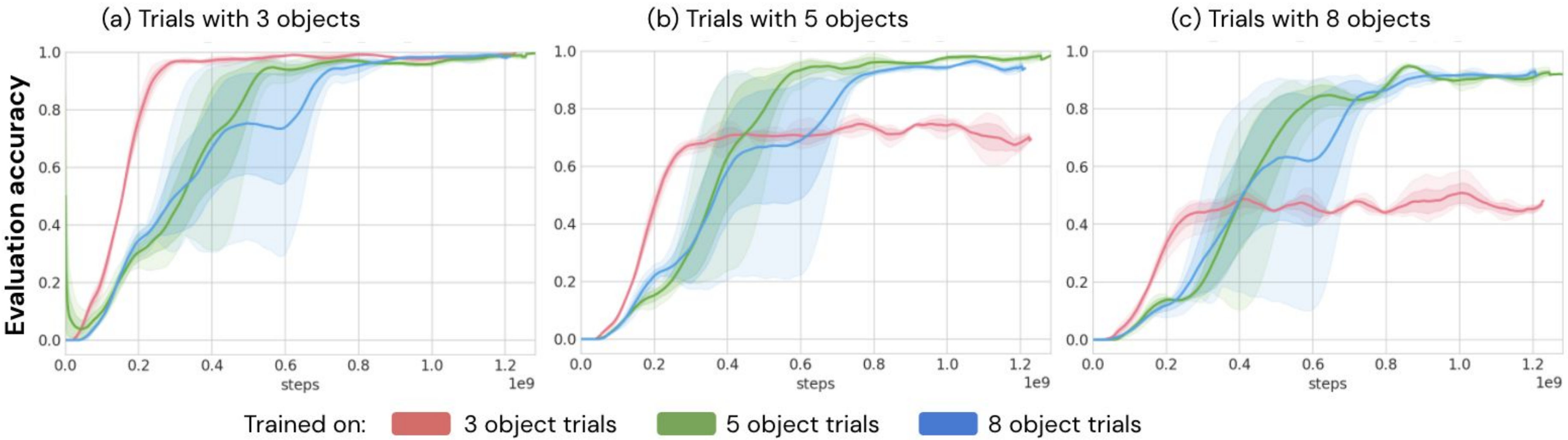


Performance on training tasks

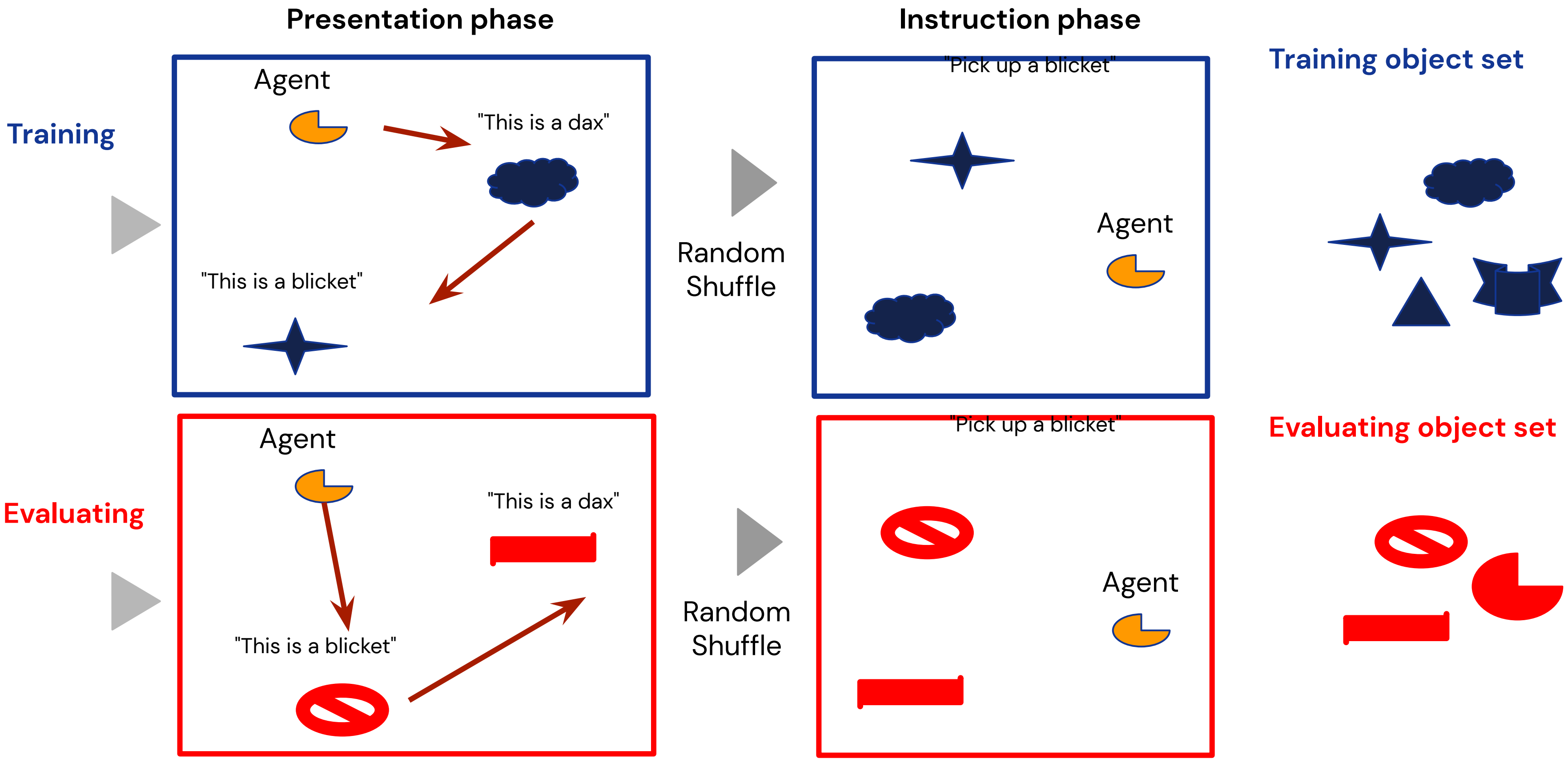
Architecture	Mean accuracy ($1e9$) steps
LSTM	0.33
LSTM + Recons	0.65
DNC + LSTM	0.33
DNC + LSTM + Recons	0.45
TransformerXL	0.34
TransformerXL + Recons	**0.98**
DCEM + LSTM	0.34
DCEM + LSTM + Recons	**0.98**
Random object selection	0.33



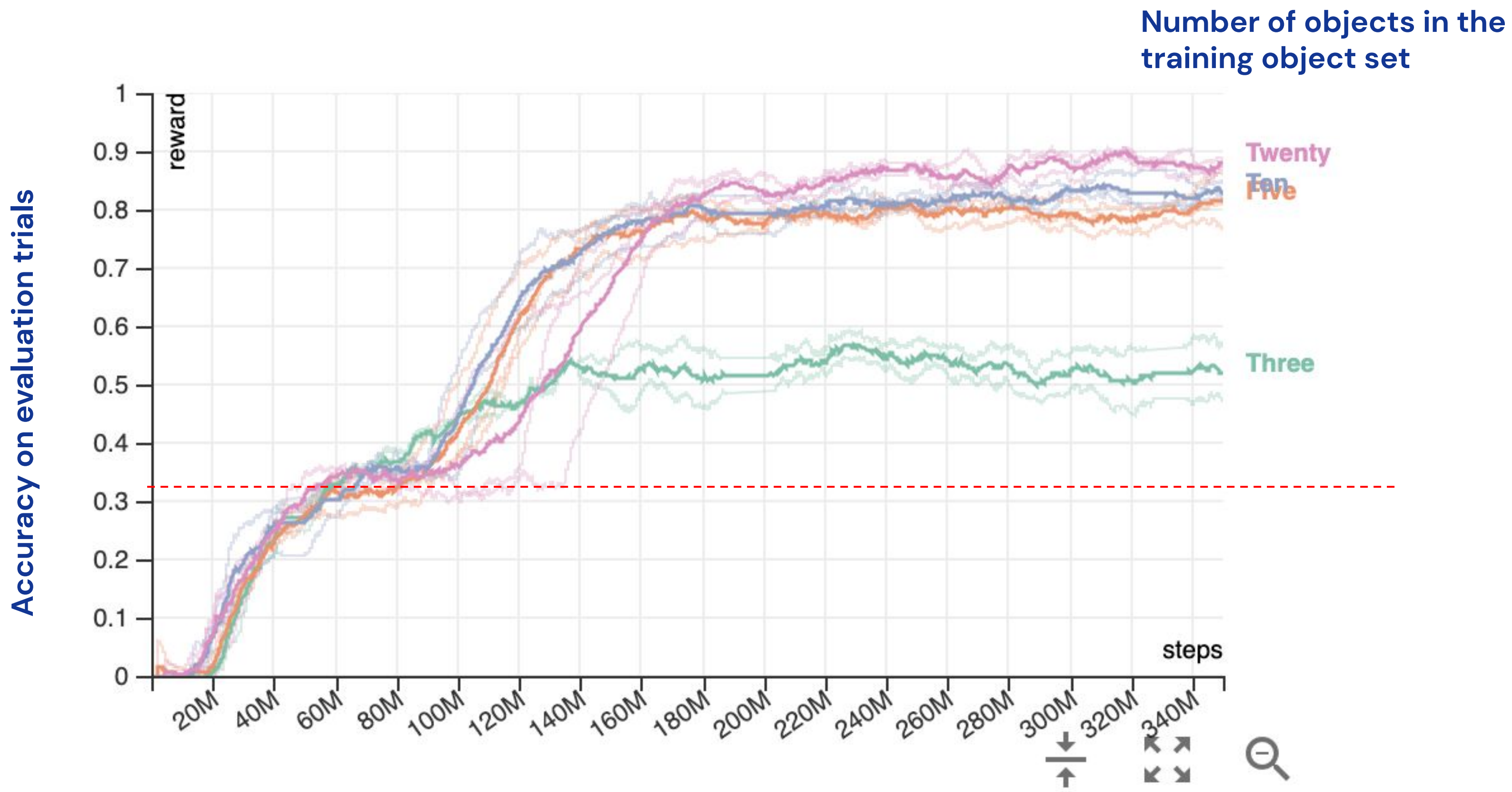
Generalization of object quantity



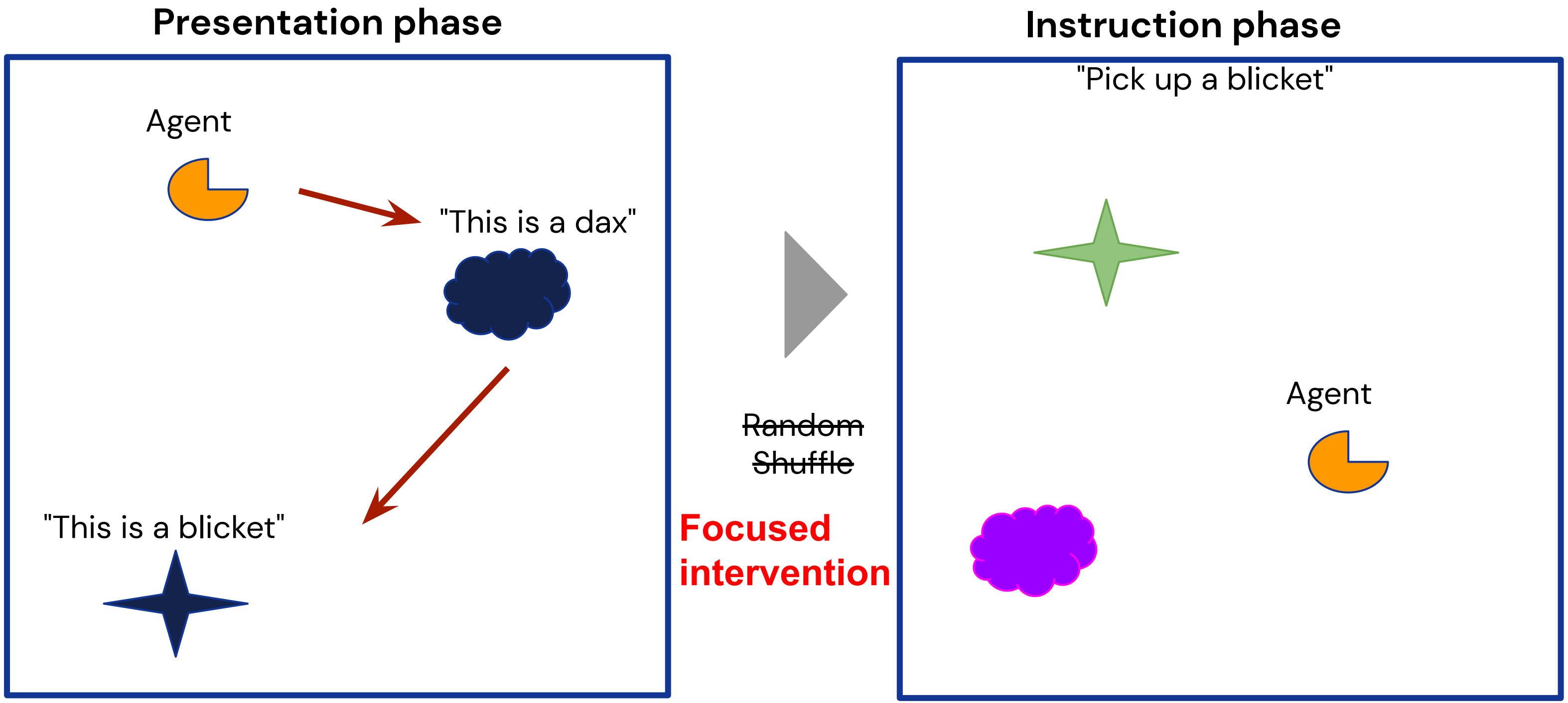
Generalization to unfamiliar objects



Generalization to unfamiliar objects



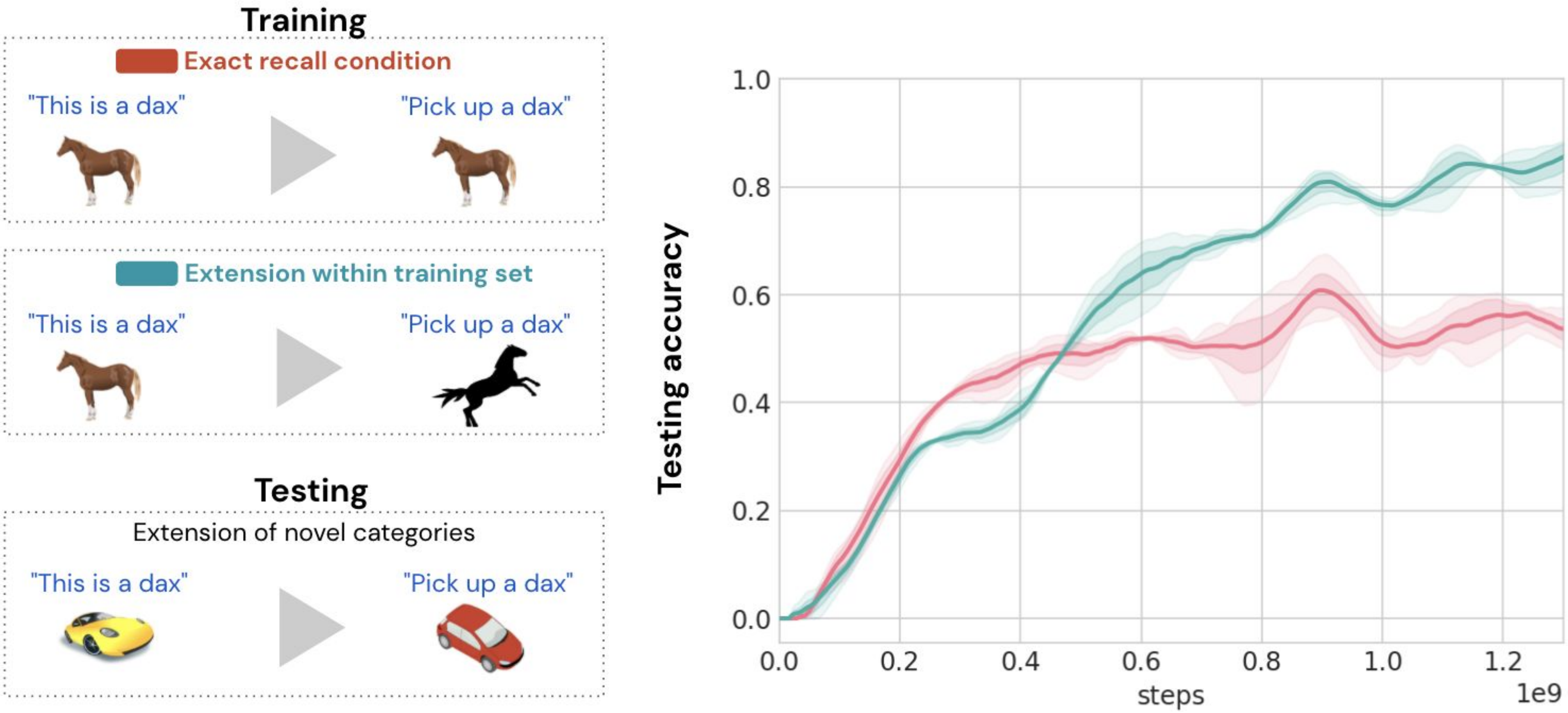
Fast 'category' learning



ShapeNet categories can contain diverse exemplars

ShapeNet: An Information-rich 3D Model Repository. Chang et al. (2015).

Zero-shot category extension



Integrating fast and slow knowledge

GPT-3

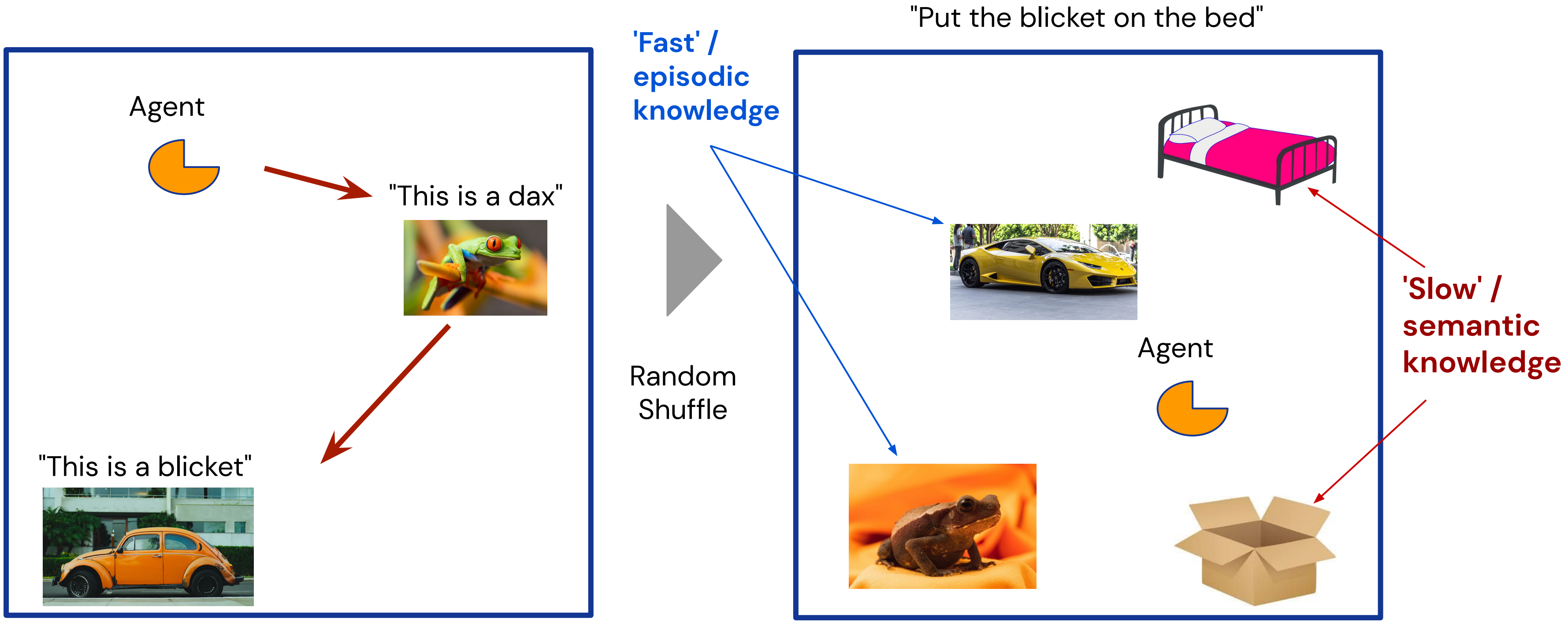
Prompt: To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

GPT3: One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduddles.

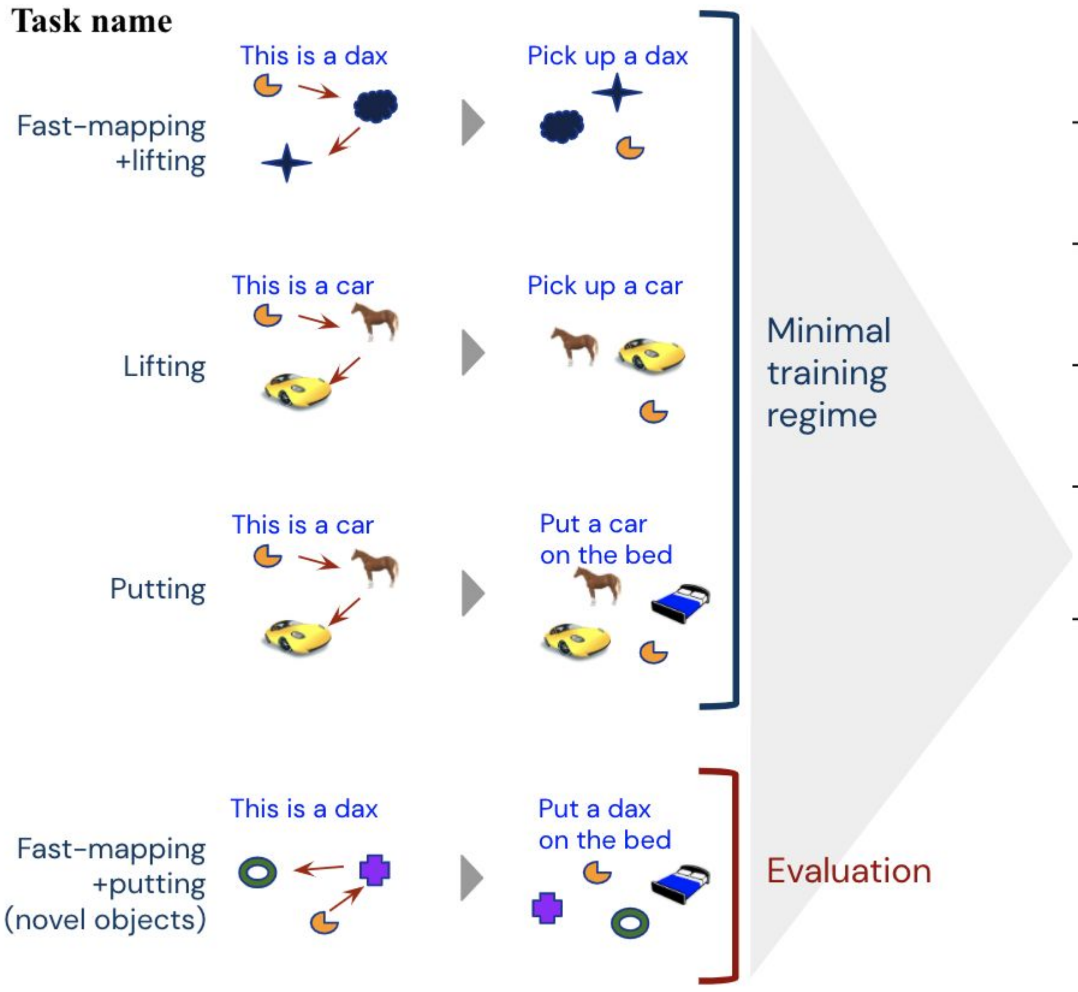
"Slow" lexical knowledge

"Fast" lexical knowledge

Integrating fast and slow knowledge




Put a rusteak on a bed



Train/test relationship	Evaluation accuracy *fast-mapping + putting*
Familiar objects, familiar task	**0.97** ± 1.8
Familiar objects, unfamiliar task	**0.96** ± 2.8
Unfamiliar objects, familiar task	**0.71** ± 3.8
Unfamiliar objects, unfamiliar task	**0.68** ± 6.1
Placing movable objects at random (chance)	**0.17**

DeepMind

Some things to think about

What is realistic/unrealistic about these simulations when considering human learners?

In what way could the agent's ability and memory be further improved?

What effect does being surrounded by language have on the way we learn, think and remember?

To conclude / discuss

- **We have shown various examples of strong / systematic / compositional / out-of-distribution generalization in neural nets**
  - An embodied agent that learns compositional generalization of nouns (objects) and verbs (motor-processes)
  - A model that can be taught to make visual analogies in a general way by exemplifying important contrasts in the task domain
  - An agent that can learn to fast-map new words in a highly general and flexible way

- **The training 'experience' seems to be a critical factor in the emergence of these capacities**
  - Greater ecological realism often implies better generalization
  - Thoughtful curricula or training methods clearly make a difference

- **How far we can get using this 'developmental approach' to AI is unknown. Few people attempt it**

DeepMind

Thank you