Abstraction and Analogy in Artificial Intelligence

Melanie Mitchell

Santa Fe Institute

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

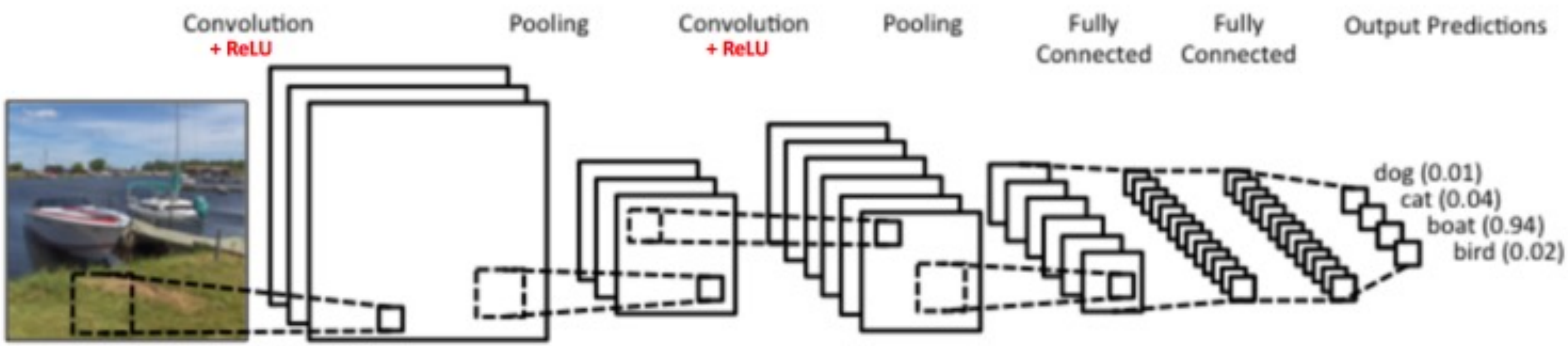
A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

Deep neural networks:



https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

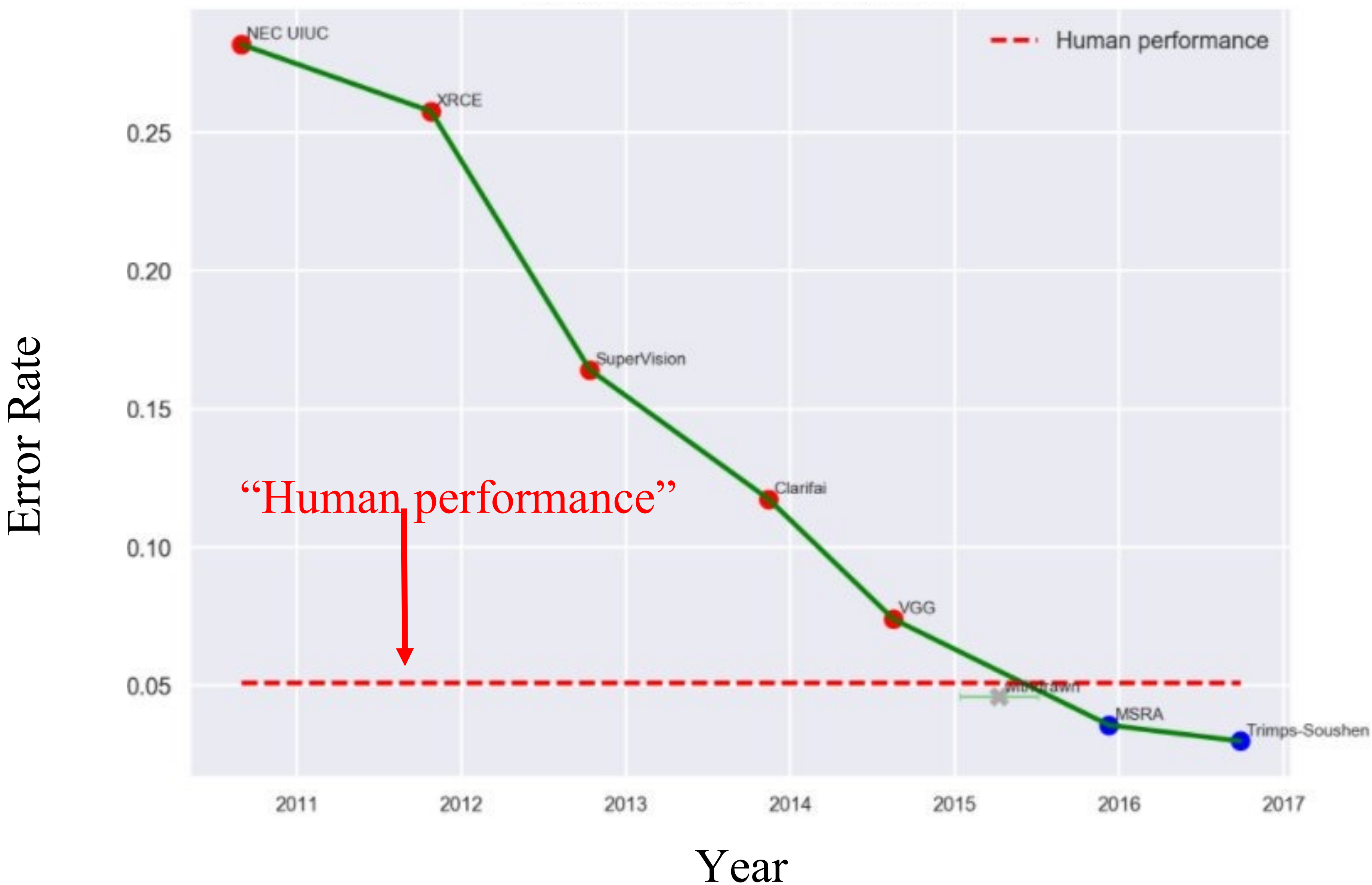
ImageNet Object-Recognition Competition

Train on 1.2 million human-labeled images

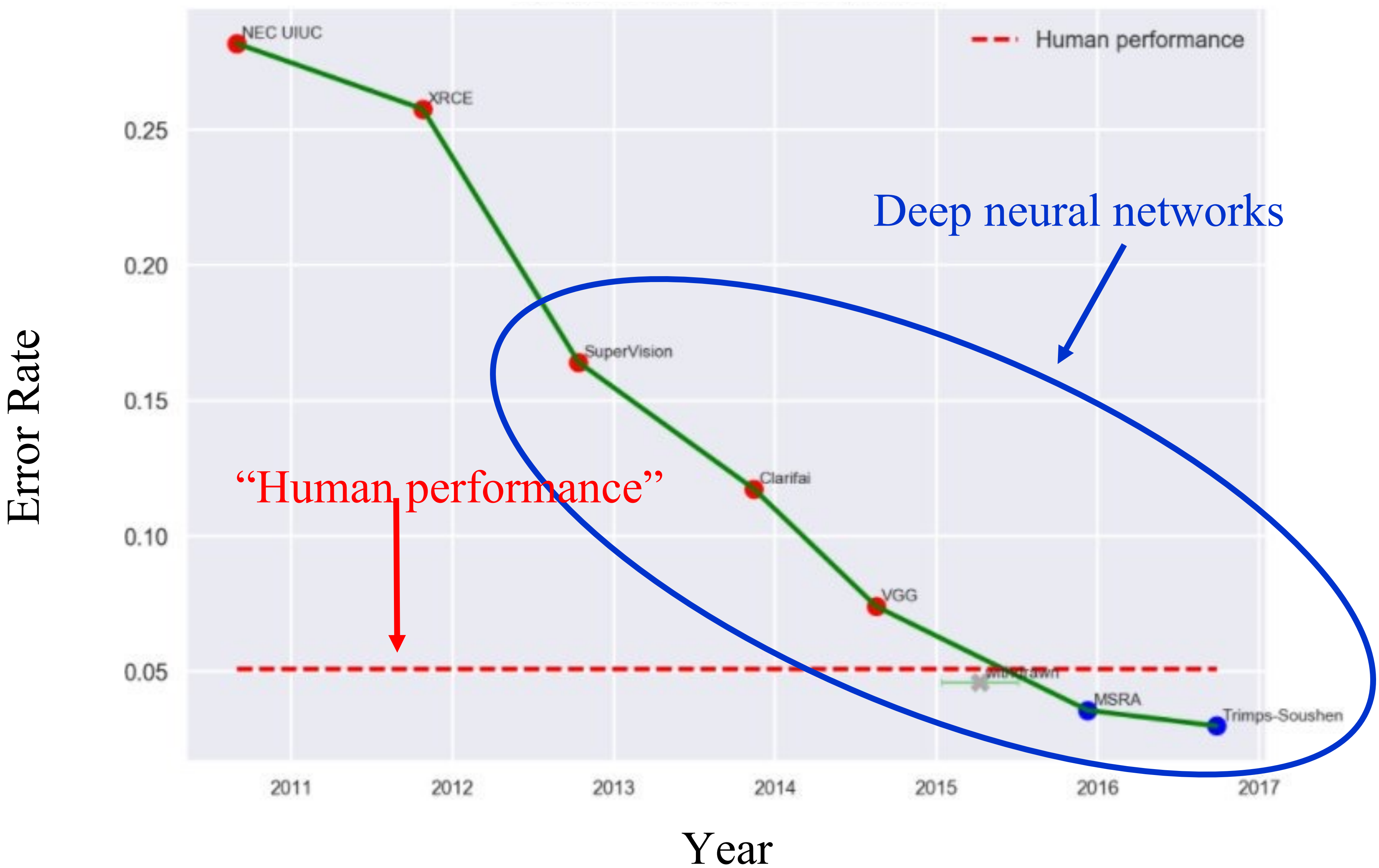
Test on 500K images

http://karpathy.github.io/assets/cnntsne.jpeg

ImageNet Object Recognition



ImageNet Object Recognition



What Are These Machines Learning?

What Are These Machines Learning?

“Animal”

“No Animal”

What Are These Machines Learning?

Alcorn, Michael A., et al. "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects." *arXiv preprint arXiv:1811.11553* (2018).

fire truck 0.99

What Are These Machines Learning?

Alcorn, Michael A., et al. "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects." *arXiv preprint arXiv:1811.11553* (2018).

fire truck 0.99 school bus 0.98

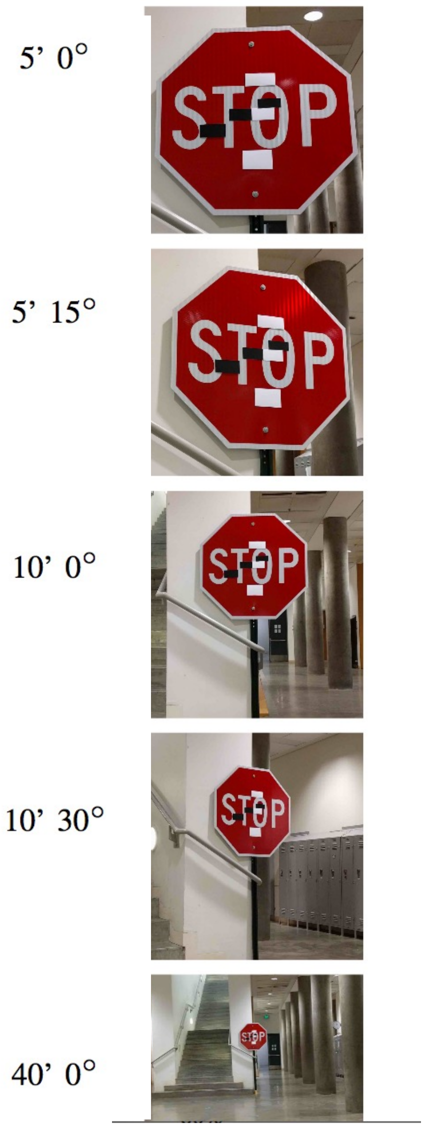
What Are These Machines Learning?

Alcorn, Michael A., et al. "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects." *arXiv preprint arXiv:1811.11553* (2018).

What Are These Machines Learning?

Alcorn, Michael A., et al. "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects." *arXiv preprint arXiv:1811.11553* (2018).

What Are These Machines Learning?

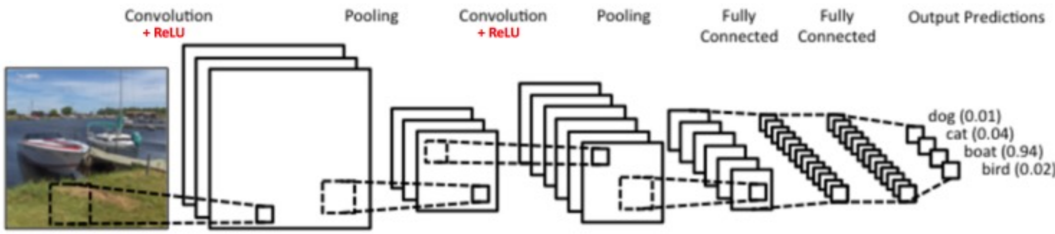


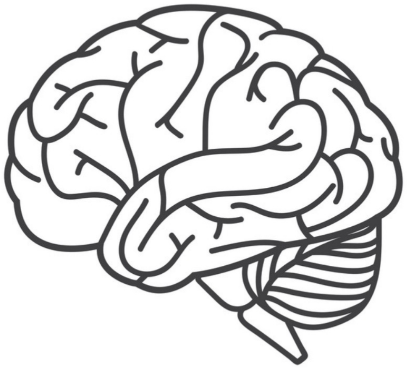
“Speed Limit 80”

Evtimov et al., “Robust Physical-World Attacks on Deep Learning Models”, 2017

“Perceptual Categories” versus **Concepts**

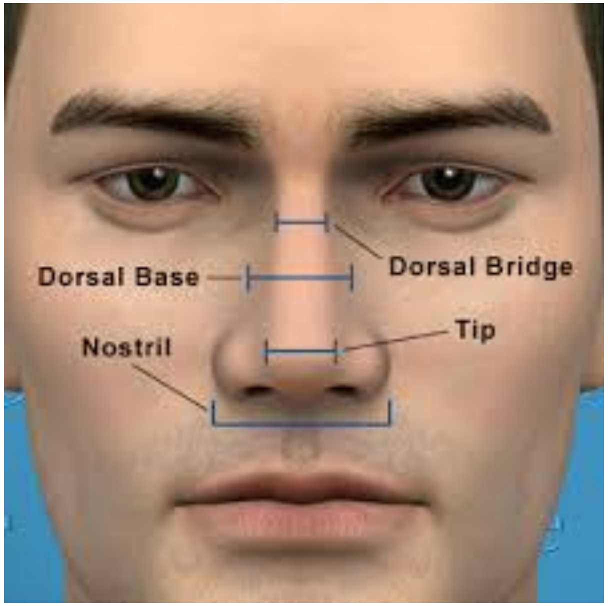
“Perceptual Categories” versus **Concepts**

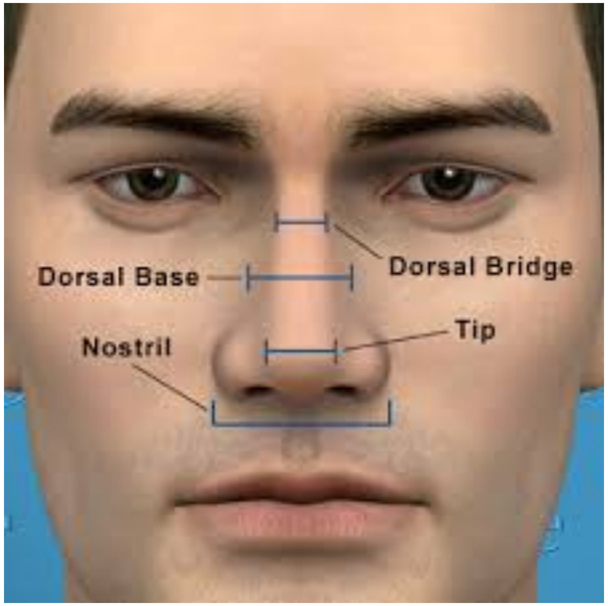


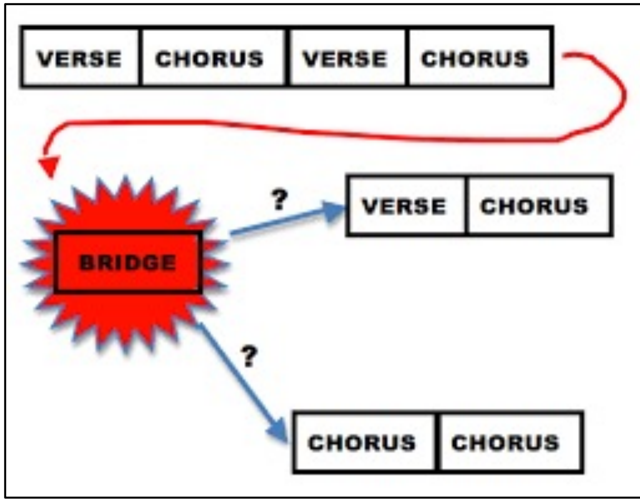


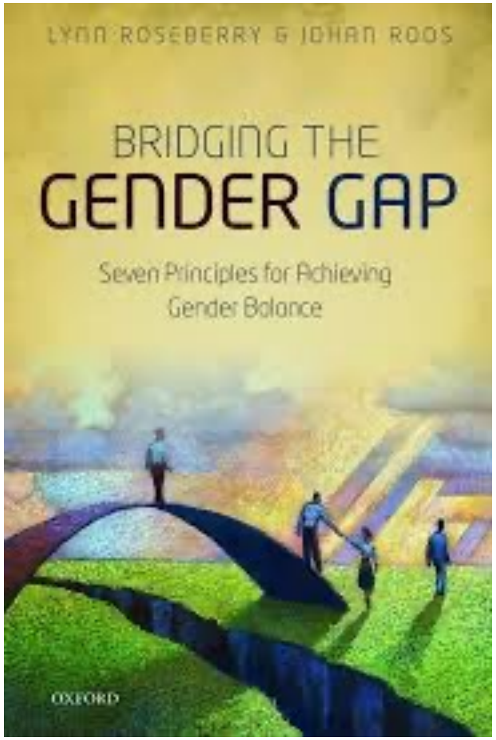
“Bridge”

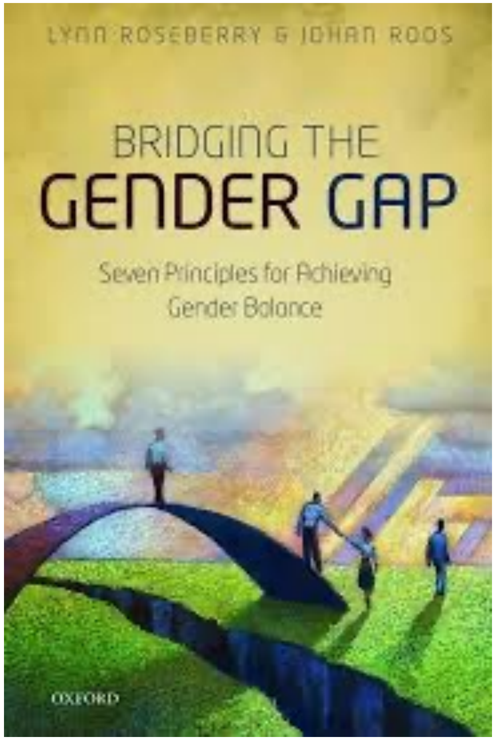



Dorsal Base
Dorsal Bridge
Nostril
Tip

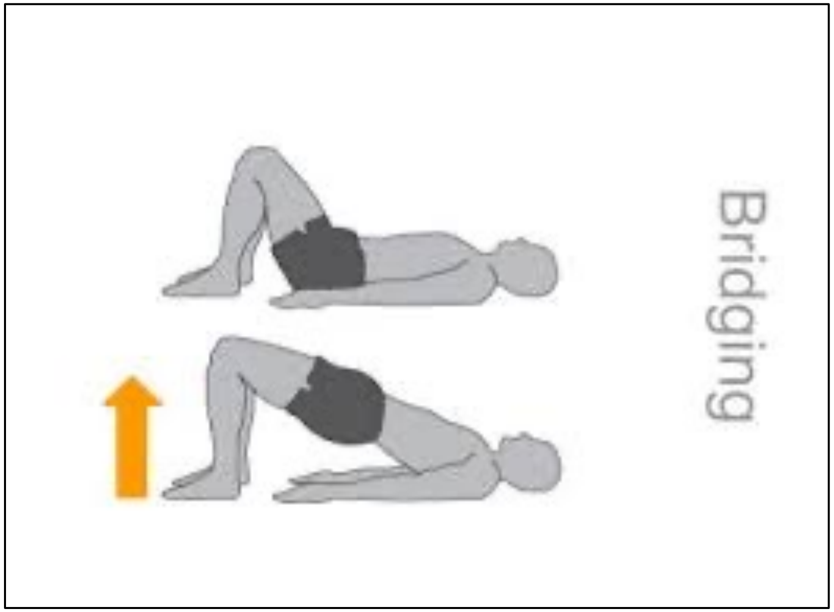

Dorsal Base
Dorsal Bridge
Tip
Nostril


VERSE
CHORUS
VERSE
CHORUS
BRIDGE
?
VERSE
CHORUS
?
CHORUS
CHORUS

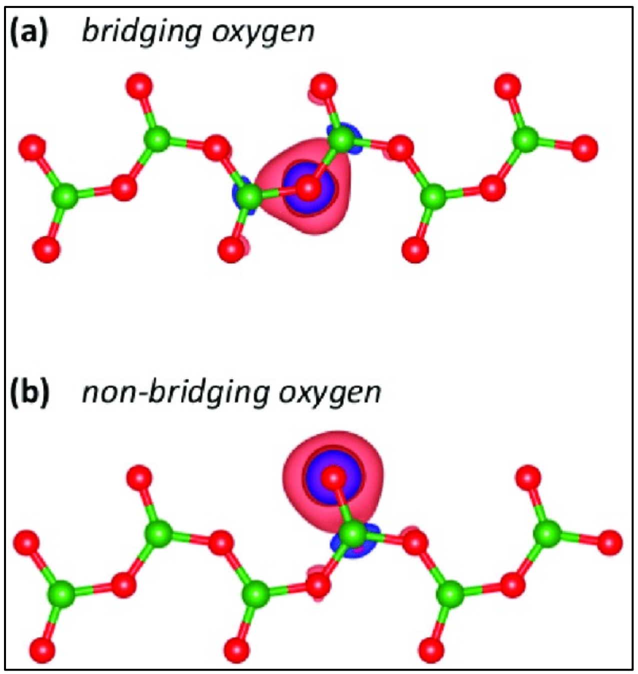

LYNN ROSEBERRY & JOHAN ROOS
BRIDGING THE
GENDER GAP
Seven Principles for Achieving
Gender Balance
OXFORD

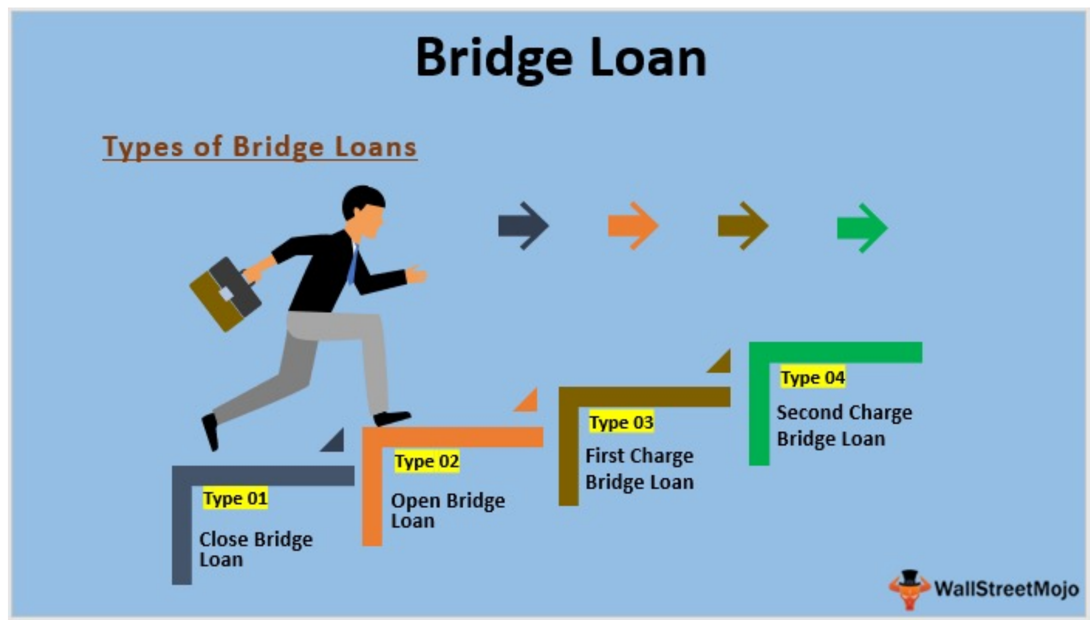


Biden says he's a 'bridge' to new 'generation of leaders' while campaigning with Harris, Booker, Whitmer



"Don't burn your bridges"





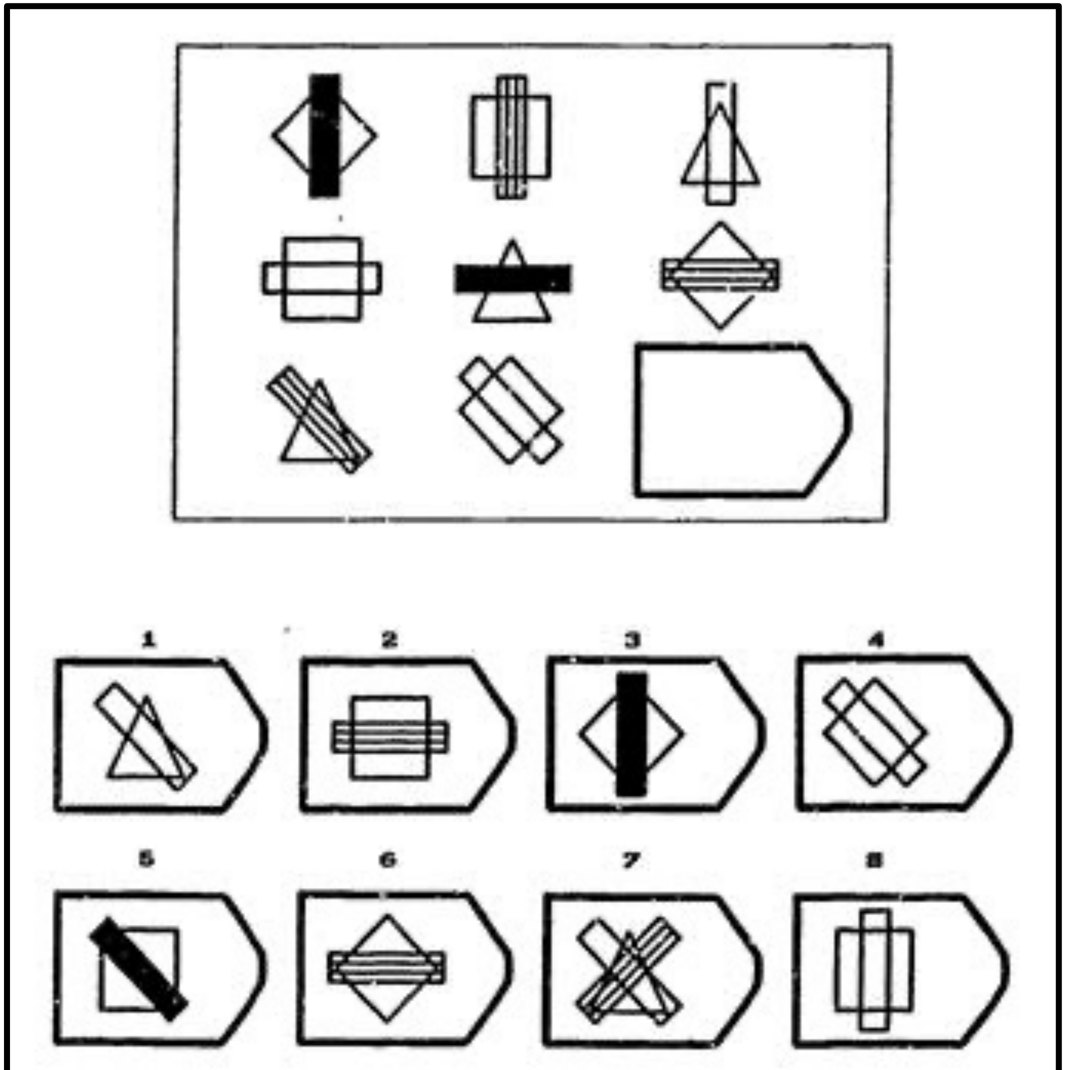
“A concept is a package of analogies.”

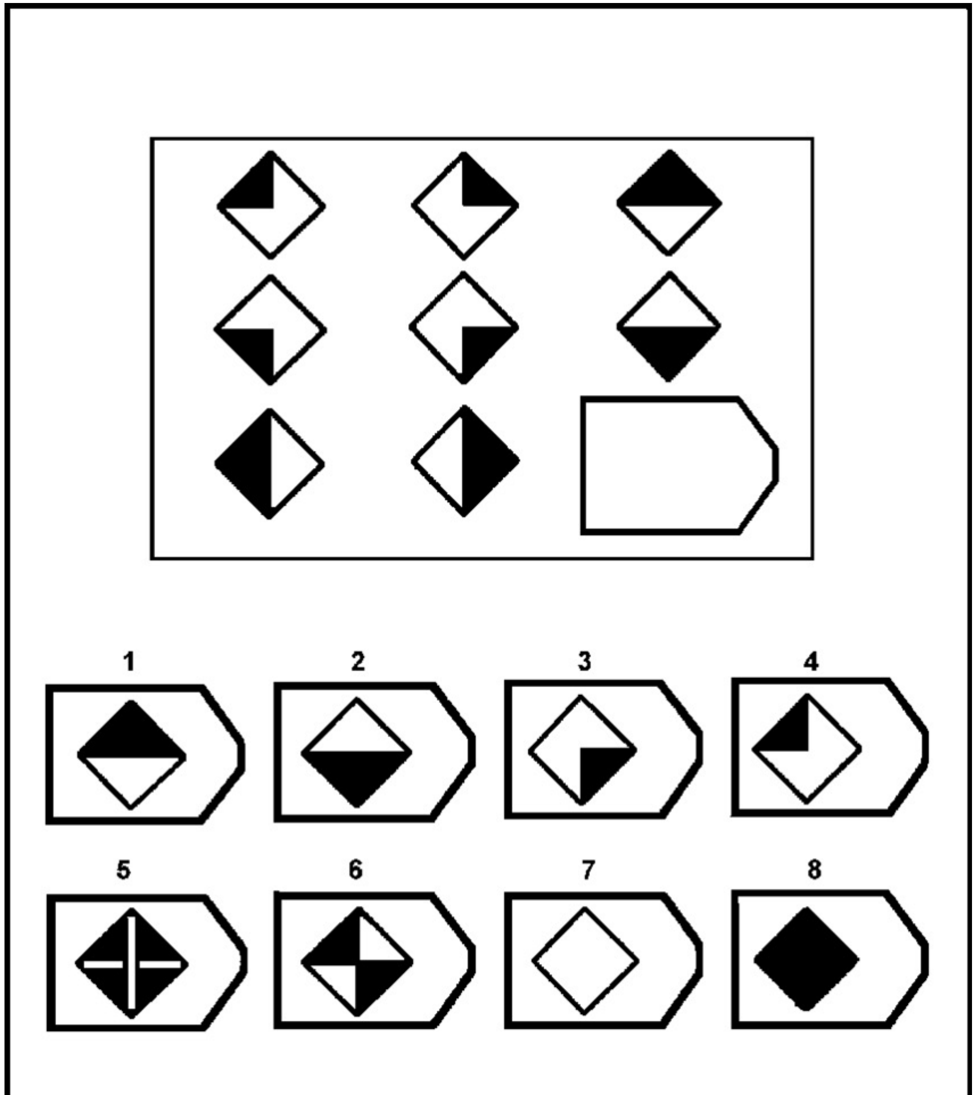
—D. Hofstadter, *Analogy as the Core of Cognition*

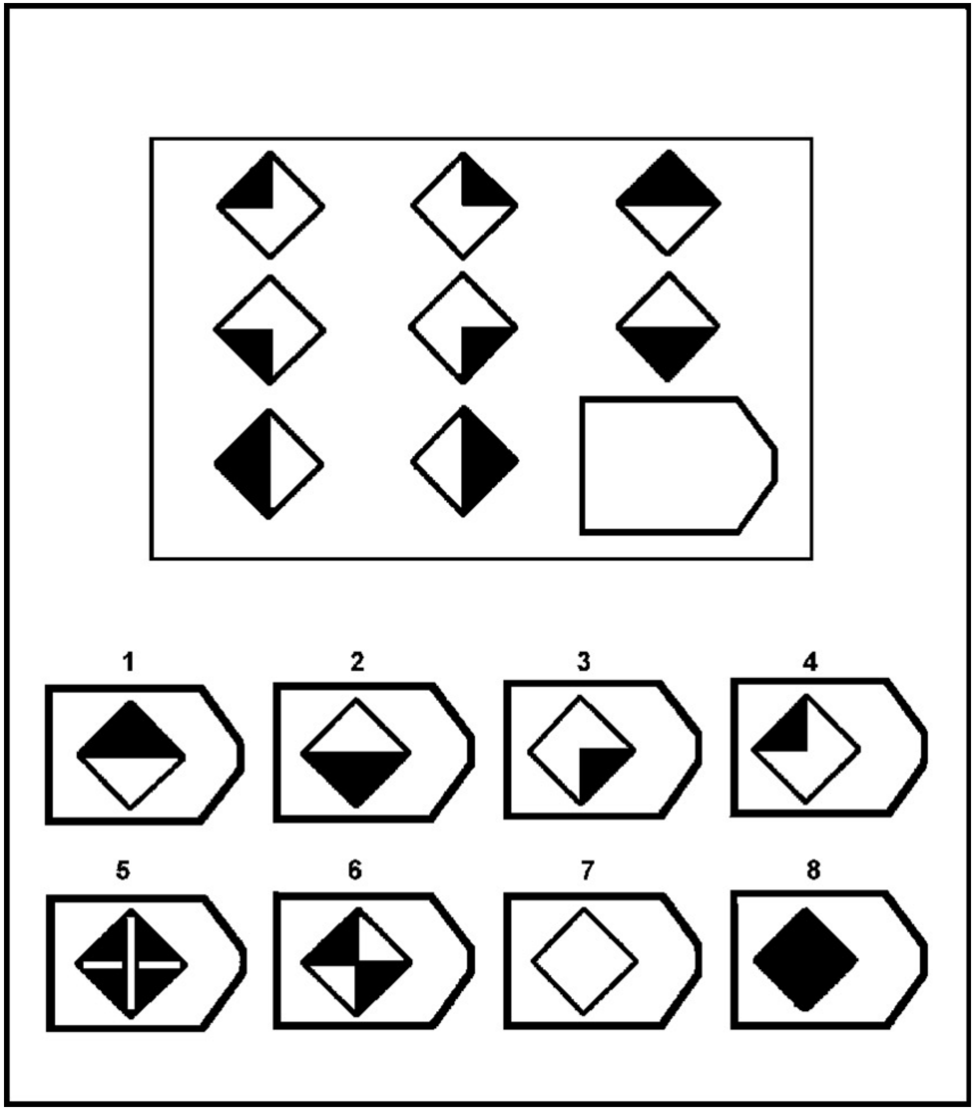
How can we get machines to learn ***concepts*** (rather than perceptual categories) and make analogies?

Deep Learning Approaches

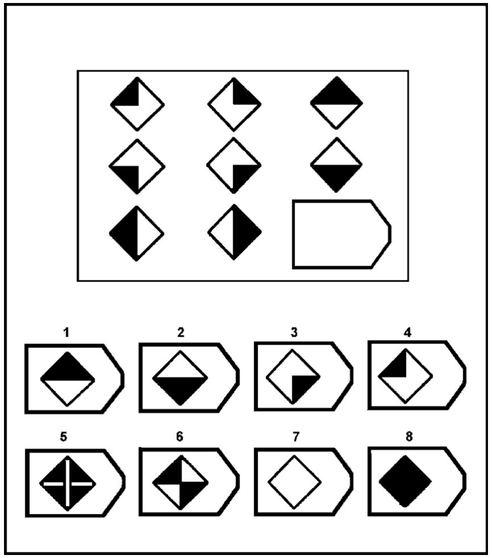
Raven’s Progressive Matrices




1
2
3
4
5
6
7
8



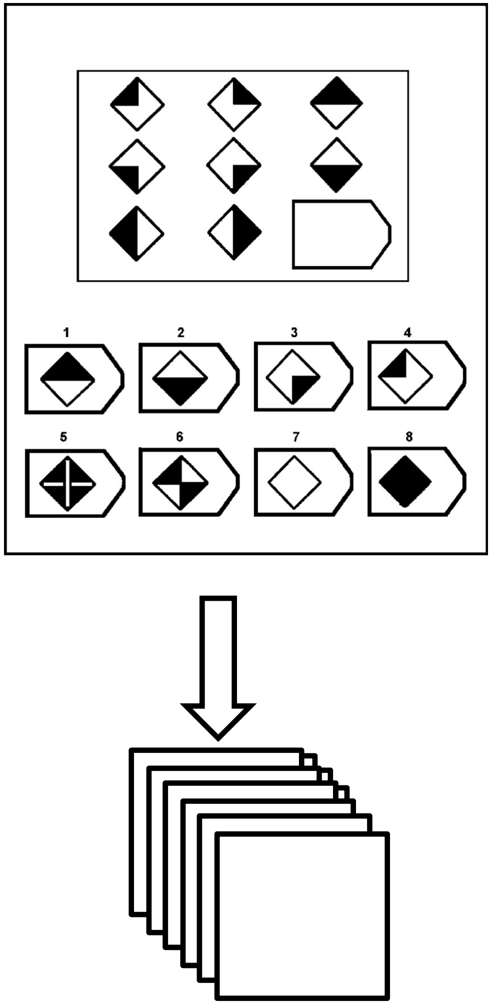
“highly correlated with human intelligence.”



Deep learning approaches

Zhou et al, 2020, “Solving Raven's Progressive Matrices with Neural Networks”

Deep learning approaches

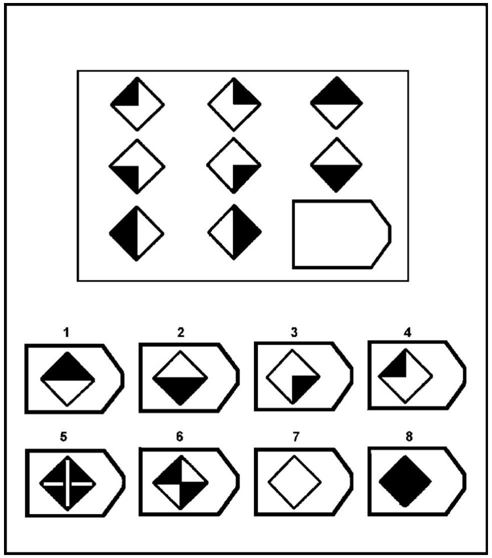


16 images (8 in problem matrix, 8 possible answers)

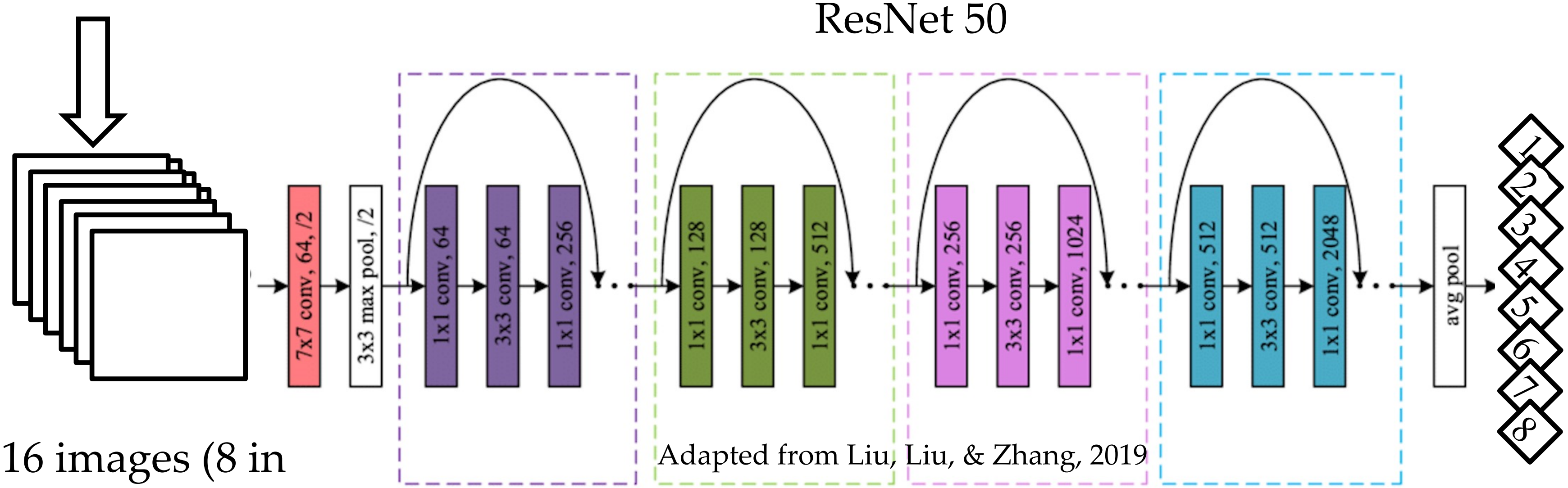
Zhou et al, 2020, "Solving Raven's Progressive Matrices with Neural Networks"

Deep learning approaches

Zhou et al, 2020, “Solving Raven's Progressive Matrices with Neural Networks”



ResNet 50



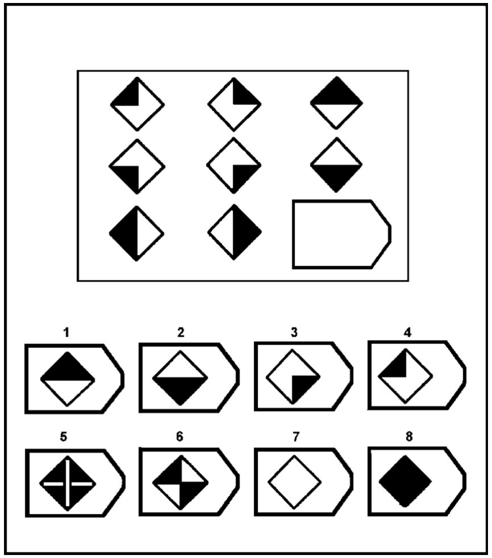
Adapted from Liu, Liu, & Zhang, 2019

16 images (8 in problem matrix, 8 possible answers)

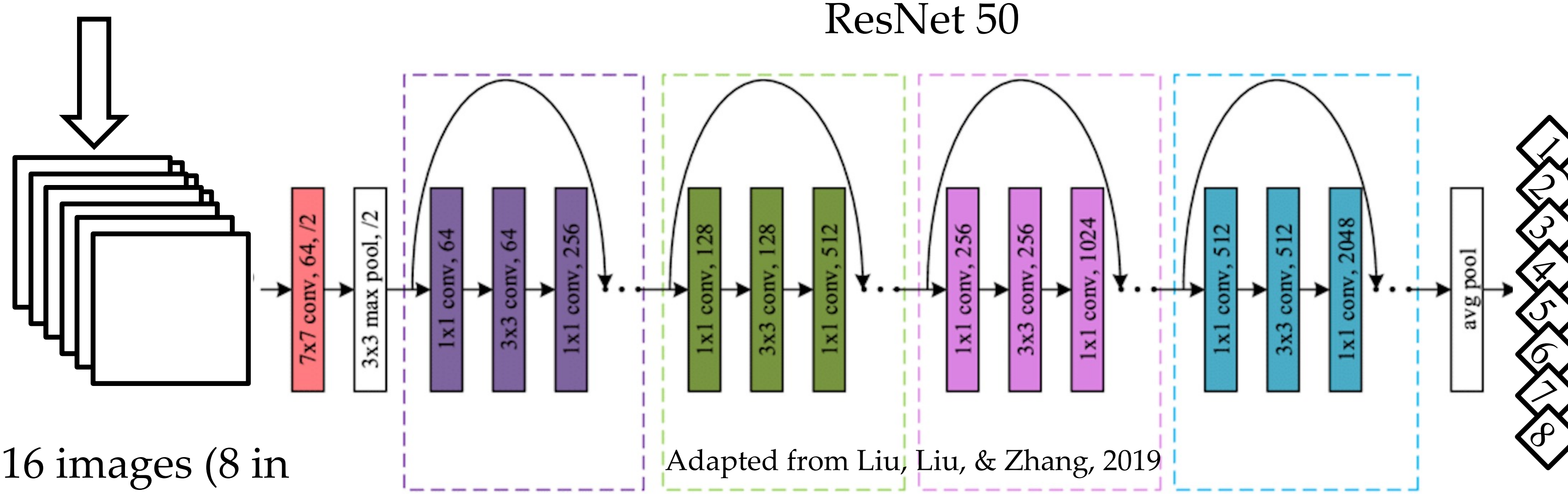
Probability distribution over the 8 possible answers

Deep learning approaches

Zhou et al, 2020, "Solving Raven's Progressive Matrices with Neural Networks"



ResNet 50



Adapted from Liu, Liu, & Zhang, 2019

16 images (8 in problem matrix, 8 possible answers)

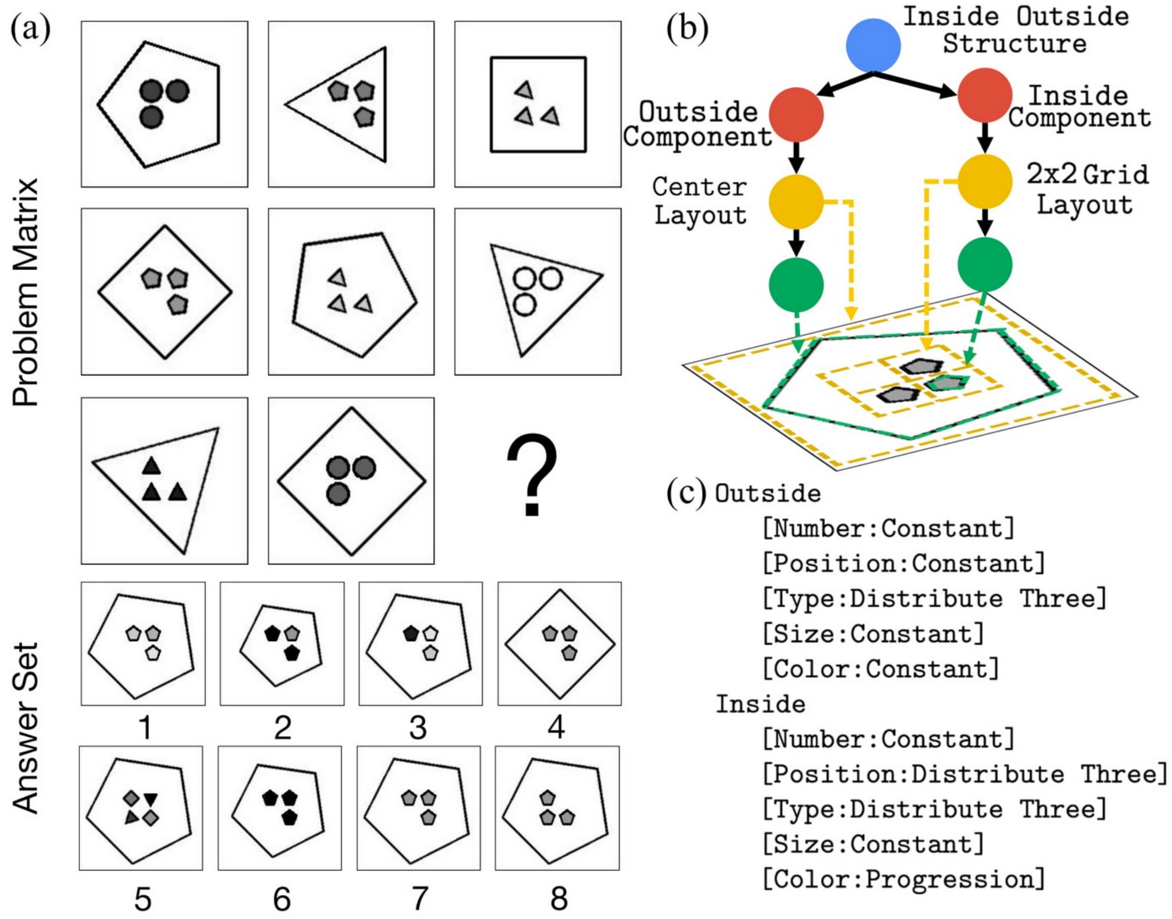
Probability distribution over the 8 possible answers

42,000 training examples (problems)

14,000 test examples

https://github.com/WellyZhang/RAVEN

RAVENS dataset is generated using a stochastic image grammar



Results:

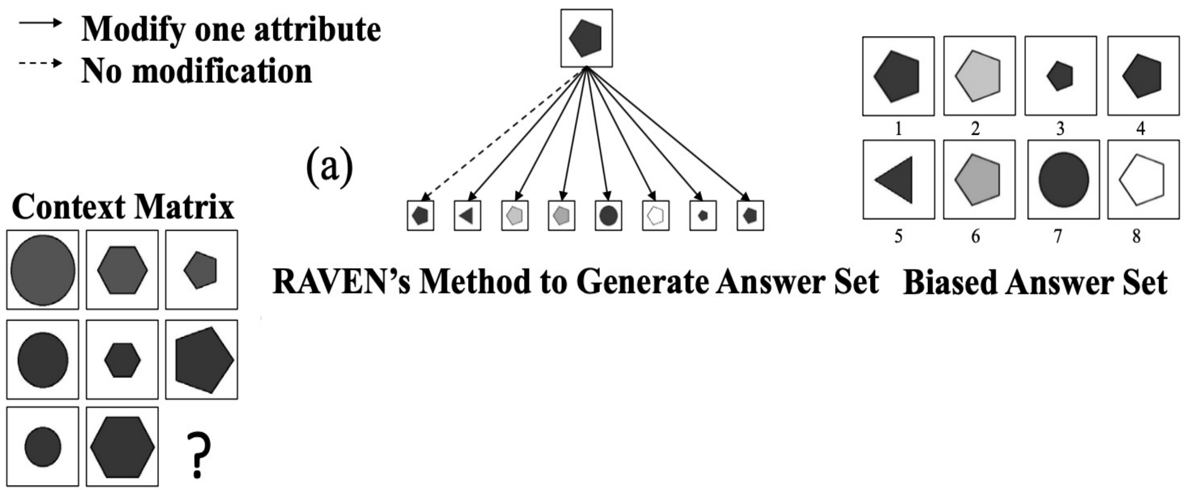
Zhou et al, 2020, "Solving Raven's Progressive Matrices with Neural Networks"

Table 2. Testing accuracy of different models in supervised manner. *Avg* denotes the average accuracy of each model.

Method	*Avg*	*Center*	*2*2Grid*	*3*3Grid*	*L-R*	*U-D*	*O-IC*	*O-IG*
LSTM	13.07	13.19	14.13	13.69	12.84	12.35	12.15	12.99
WReN	14.69	13.09	28.62	28.27	7.49	6.34	8.38	10.56
CNN	36.97	33.58	30.30	33.53	39.43	41.26	43.20	37.54
ResNet-18+MLP	53.43	52.82	41.86	44.29	58.77	60.16	63.19	53.12
LSTM+DRT	13.96	14.29	15.08	14.09	13.79	13.24	13.99	13.29
WReN+DRT	15.02	15.38	23.26	29.51	6.99	8.43	8.93	12.35
CNN+DRT	39.42	37.30	30.06	34.57	45.49	45.54	45.93	37.54
ResNet-18+MLP+DRT	59.56	58.08	46.53	50.40	65.82	67.11	69.09	60.11
RseNet-18 (ours, w/o pre-train)	77.18	72.75	57.00	62.65	91.00	89.60	88.40	78.85
RseNet-50 (ours, w pre-train)	86.26	89.45	66.60	67.95	97.85	98.15	96.60	87.20
CoPINet	**91.42**	**95.05**	**77.45**	**78.85**	**99.10**	**99.65**	**98.50**	**91.35**
Human	84.41	95.45	81.82	79.55	86.36	81.81	86.36	81.81

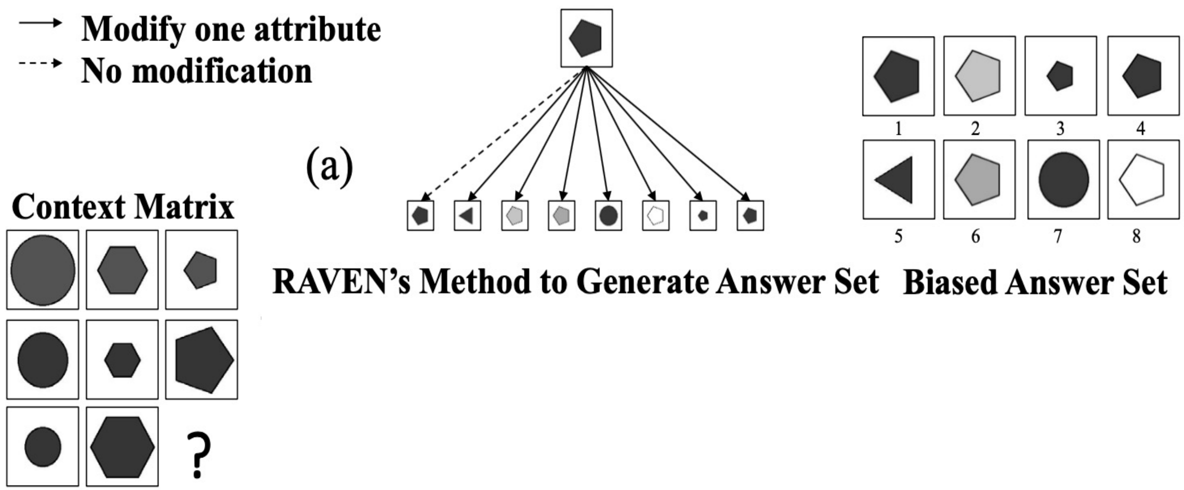
Bias in RAVENS dataset

(Hu et al, Hierarchical rule induction network for abstract visual reasoning)



Bias in RAVENS dataset

(Hu et al, Hierarchical rule induction network for abstract visual reasoning)



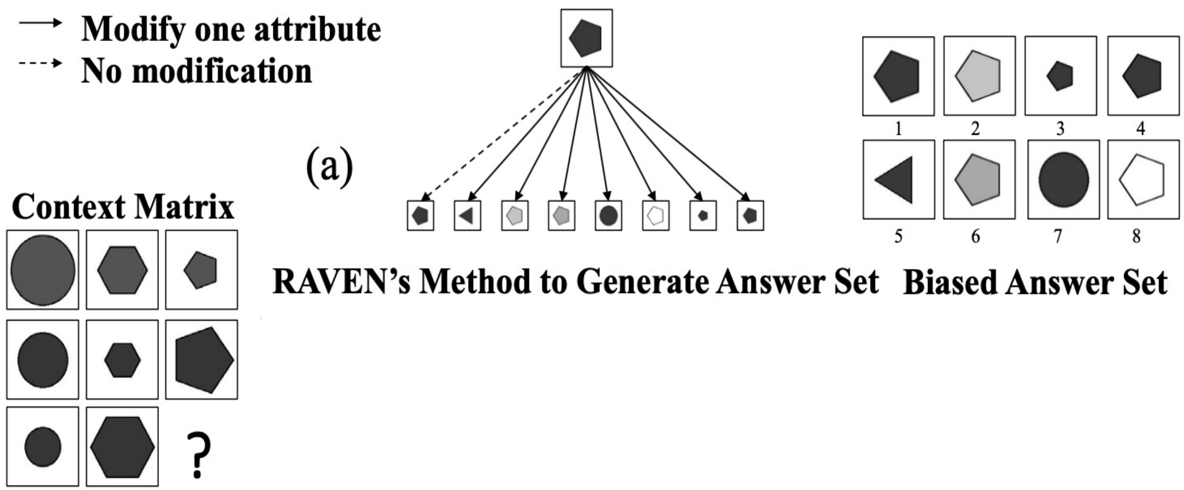
Model	RAVEN	Balanced-RAVEN
ResNet	89.2%	40.3%
Context-blind ResNet	90.1%	12.5%

Table 1. Test on RAVEN and Balanced-RAVEN.

Train on candidate answers only!

Bias in RAVENS dataset

(Hu et al, Hierarchical rule induction network for abstract visual reasoning)



Model	RAVEN	Balanced-RAVEN
ResNet	89.2%	40.3%
Context-blind ResNet	90.1%	12.5%

Table 1. Test on RAVEN and Balanced-RAVEN.

Train on candidate answers only!

Many types of deep learning approaches for Ravens-like problems

Wild Relation Network, Barrett et al. 2018

Disentangled Feature Representations, Steenbrugge et al. 2018

Attention Relation Network, Hahne et al. 2019

Contrastive Perceptual Inference Network, Zhang et al, 2019

Logic Embedding Network, Zheng et al., 2019

Multi-Layer Relation Network, Jahrens & Martinetz, 2020

Hierarchical Rule Induction Network, Hu et al., 2020

.

.

.

Deep Learning Approaches

Limitations

Deep Learning Approaches
Limitations

- Requires very large corpus of training examples. Need to generate automatically. Makes NNs susceptible to shortcuts.

Deep Learning Approaches
Limitations

- Requires very large corpus of training examples. Need to generate automatically. Makes NNs susceptible to shortcuts.

- Trained networks are not transparent. What did they learn?

Deep Learning Approaches
Limitations

- Requires very large corpus of training examples. Need to generate automatically. Makes NNs susceptible to shortcuts.

- Trained networks are not transparent. What did they learn?

- If goal is general humanlike abstraction abilities, it doesn't make sense to have to train on tens of thousands of examples.

Deep Learning Approaches
Limitations

- Requires very large corpus of training examples. Need to generate automatically. Makes NNs susceptible to shortcuts.

- Trained networks are not transparent. What did they learn?

- If goal is general humanlike abstraction abilities, it doesn't make sense to have to train on tens of thousands of examples.

The essence of abstraction and analogy is few-shot learning!

Copycat Architecture

Hofstadter & Mitchell, 1995

Letter-String Analogies

(Hofstadter and Mitchell, 1995)

abc $\longrightarrow$ **abd**

pqrs $\longrightarrow$ **?**

Letter-String Analogies

(Hofstadter and Mitchell, 1995)

abc $\longrightarrow$ **abd**

pqrs $\longrightarrow$ **?**

abc $\longrightarrow$ **abd**

ppqqrrss $\longrightarrow$ **?**

Letter-String Analogies

(Hofstadter and Mitchell, 1995)

abc $\longrightarrow$ **abd**

pqrs $\longrightarrow$ **?**

abc $\longrightarrow$ **abd**

ppqqrrss $\longrightarrow$ **?**

abcd $\longrightarrow$ **dcba**

srqp $\longrightarrow$ **?**

Letter-String Analogies

(Hofstadter and Mitchell, 1995)

abc $\longrightarrow$ **abd**

pqrs $\longrightarrow$ **?**

abc $\longrightarrow$ **abd**

ppqqrrss $\longrightarrow$ **?**

abcd $\longrightarrow$ **dcba**

srqp $\longrightarrow$ **?**

axxd $\longrightarrow$ **abcd**

xqxxx $\longrightarrow$ **?**

Letter-String Analogies

(Hofstadter and Mitchell, 1995)

abc → **abd**
pqrs → **?**

abc → **abd**
ppqqrrss → **?**

abcd → **dcba**
srqp → **?**

axxd → **abcd**
xqxxx → **?**

- Idealized "situations", with objects, relations, groups, actions, events

Letter-String Analogies

(Hofstadter and Mitchell, 1995)

abc $\longrightarrow$ **abd**
pqrs $\longrightarrow$ **?**

abc $\longrightarrow$ **abd**
ppqqrrss $\longrightarrow$ **?**

abcd $\longrightarrow$ **dcba**
srqp $\longrightarrow$ **?**

axxd $\longrightarrow$ **abcd**
xqxxx $\longrightarrow$ **?**

- Idealized "situations", with objects, relations, groups, actions, events
- Meant to be a tool for exploring general issues of abstraction and analogy-making

Copycat Architecture

(Mitchell & Hofstadter, 1995, "The Copycat project: A model of mental fluidity and analogy-making")

Copycat Architecture

(Mitchell & Hofstadter, 1995, “The Copycat project: A model of mental fluidity and analogy-making”)

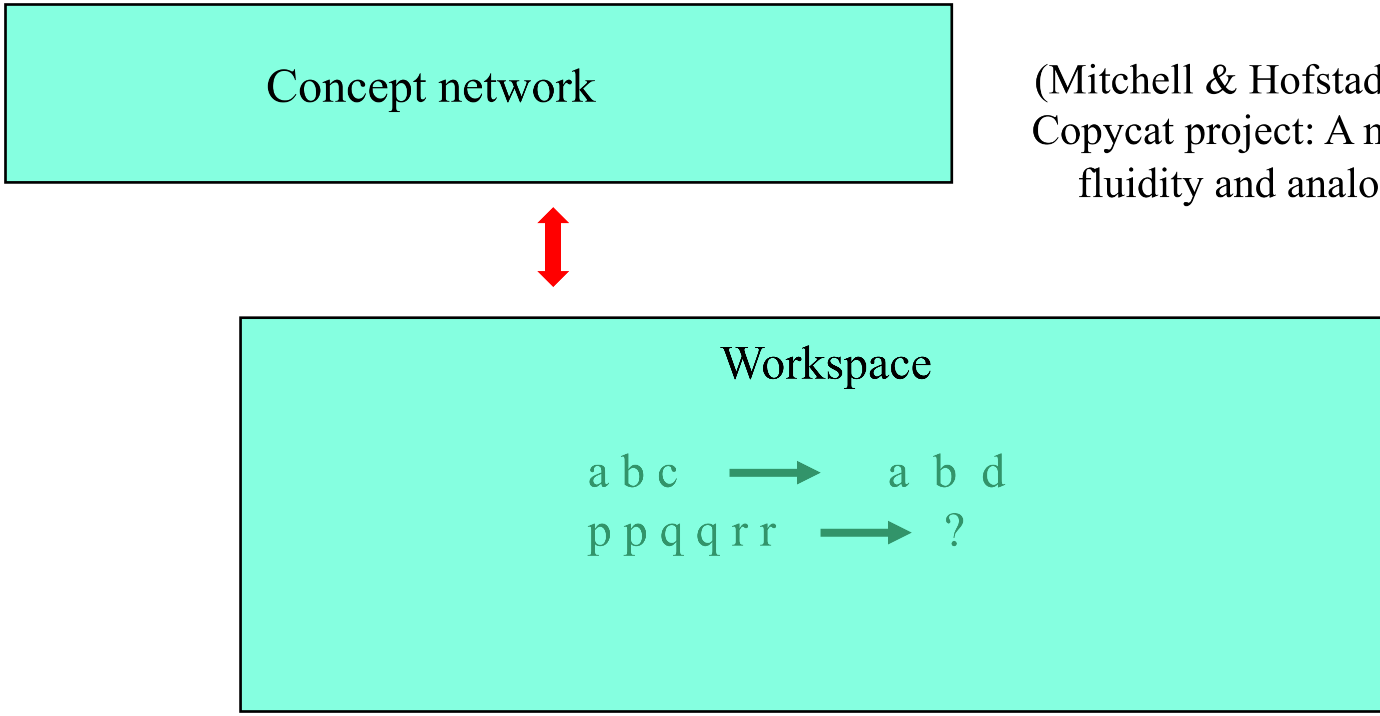
Workspace

a b c → a b d

p p q q r r → ?

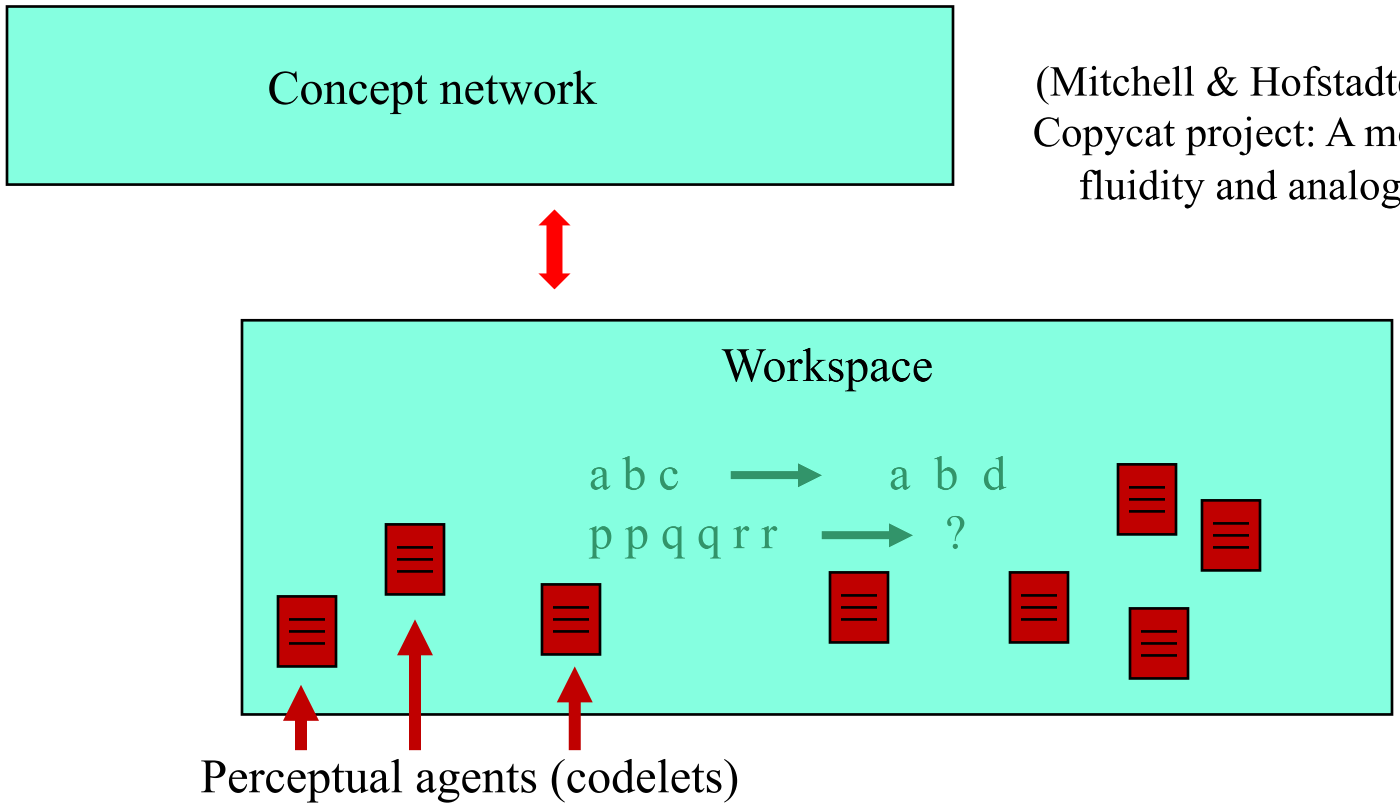
Copycat Architecture

(Mitchell & Hofstadter, 1995, "The Copycat project: A model of mental fluidity and analogy-making")



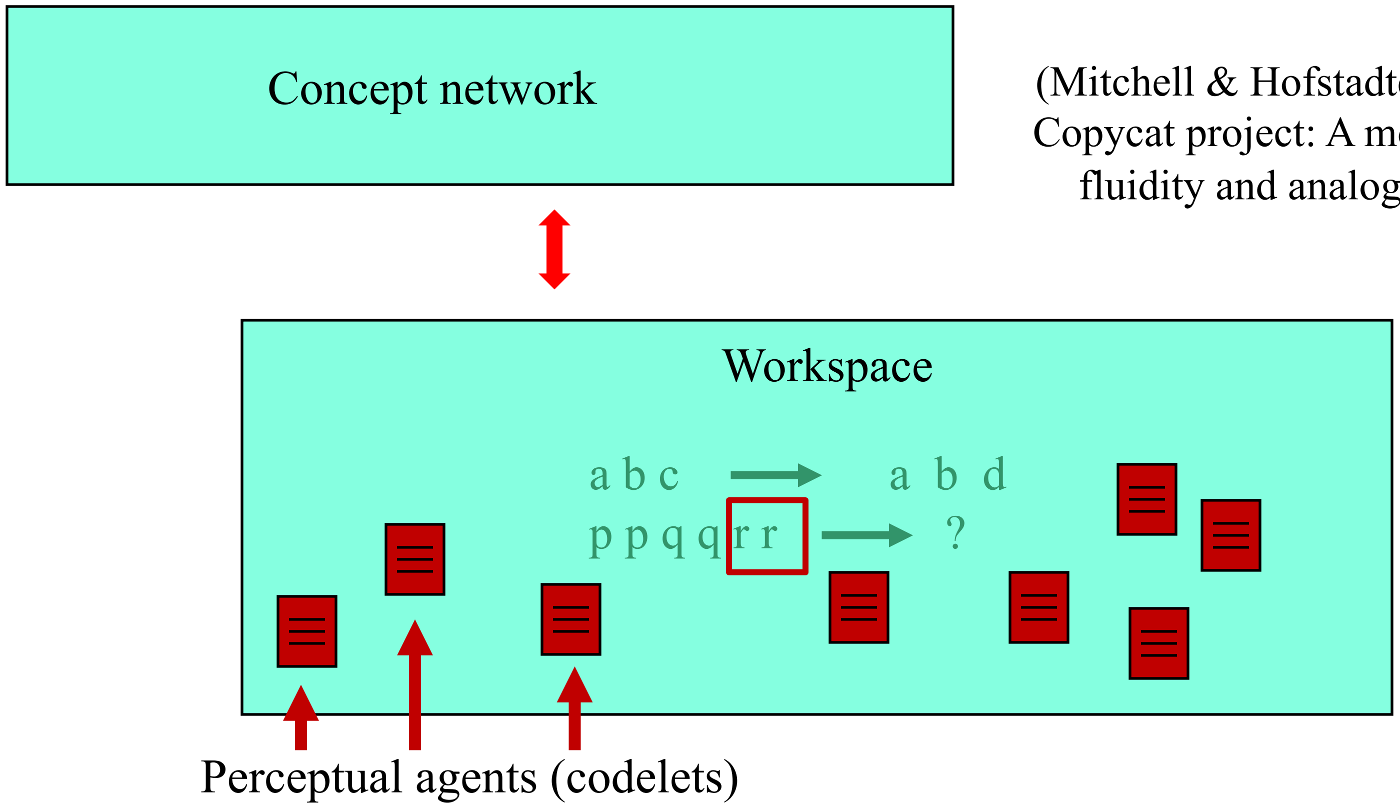
Copycat Architecture

(Mitchell & Hofstadter, 1995, “The Copycat project: A model of mental fluidity and analogy-making”)



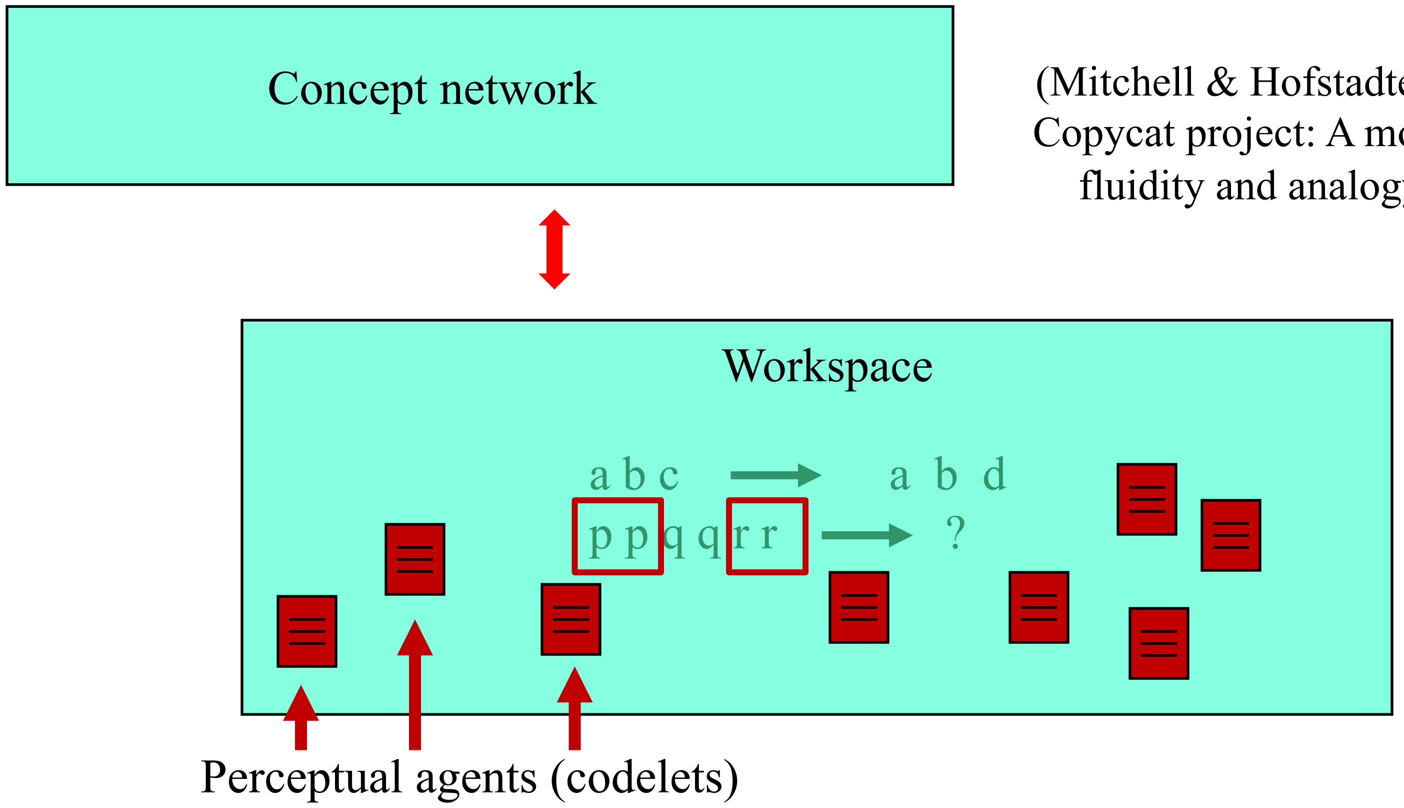
Copycat Architecture

(Mitchell & Hofstadter, 1995, “The Copycat project: A model of mental fluidity and analogy-making”)



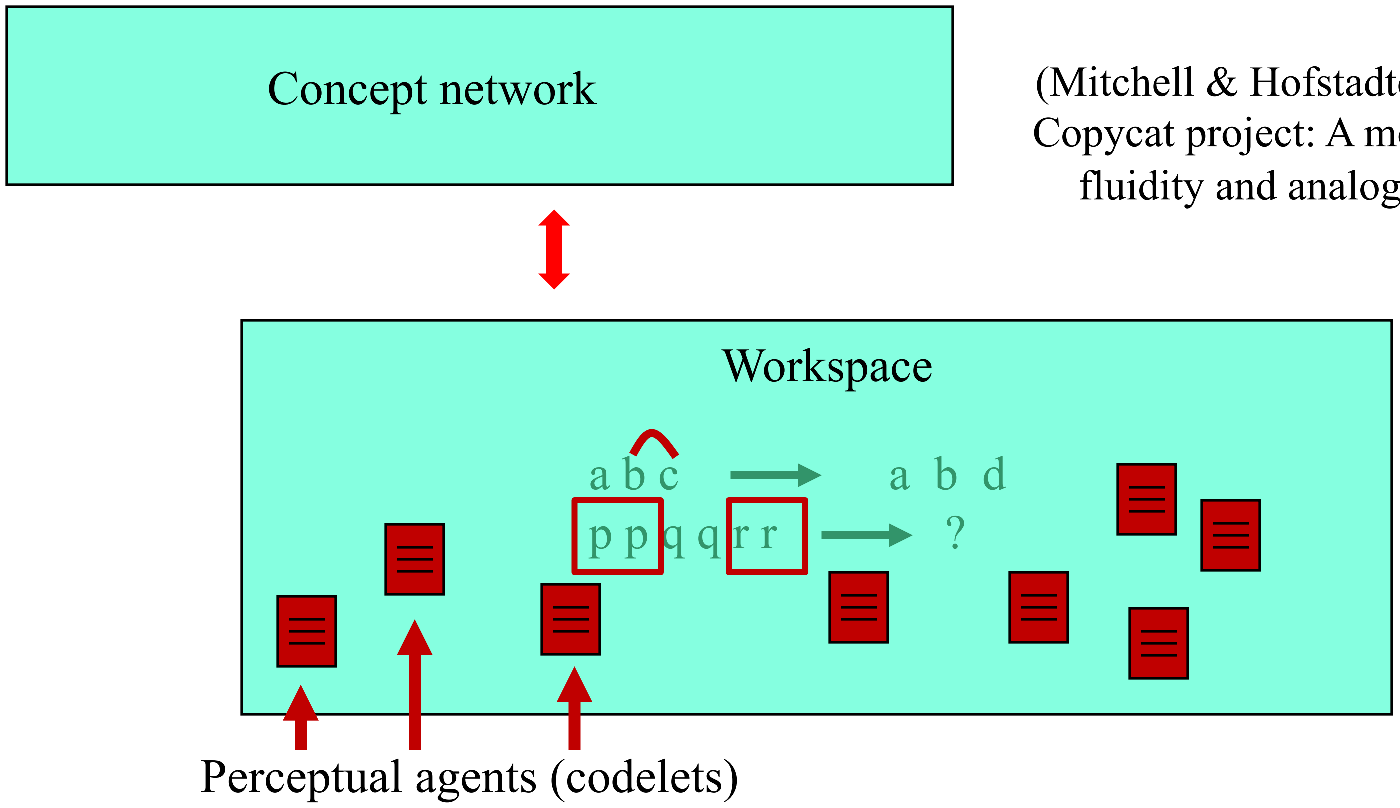
Copycat Architecture

(Mitchell & Hofstadter, 1995, “The Copycat project: A model of mental fluidity and analogy-making”)



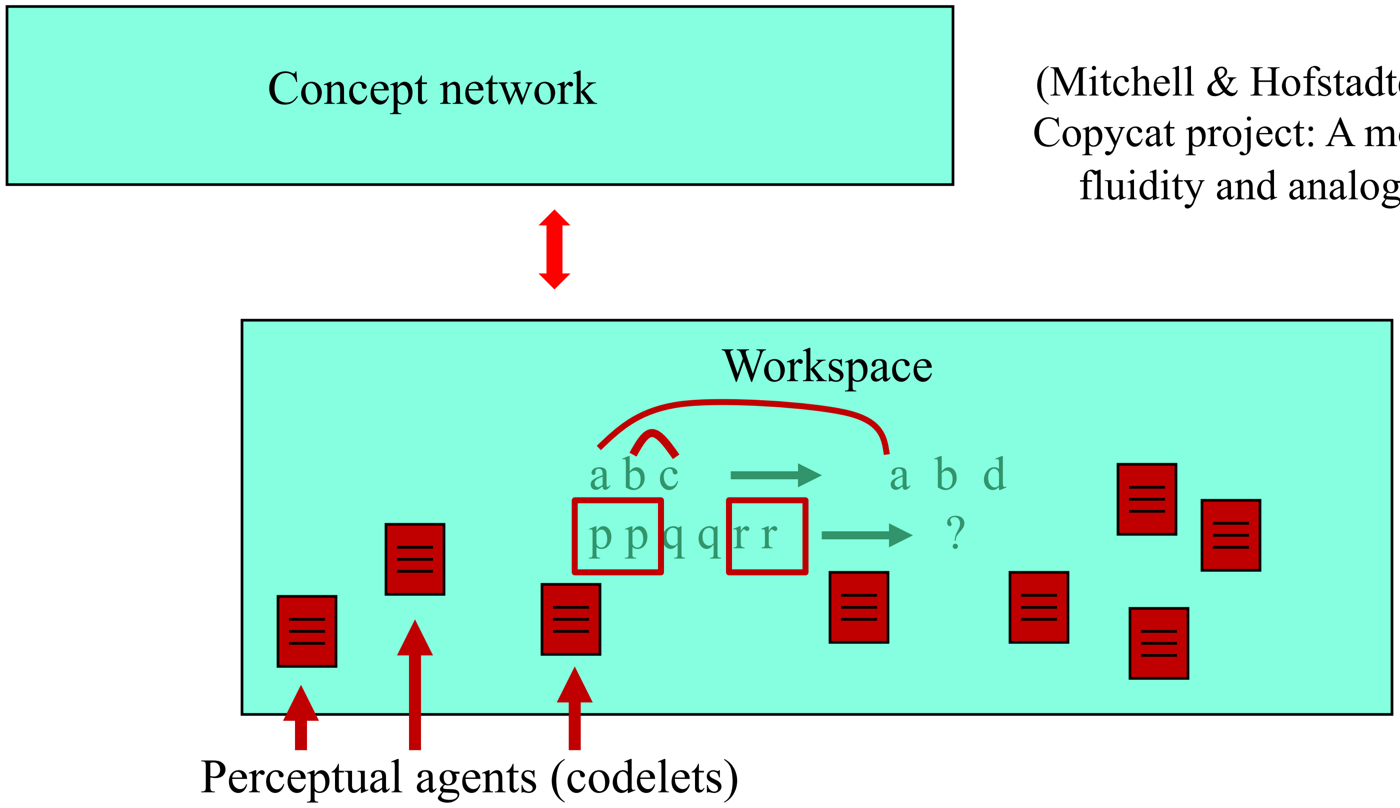
Copycat Architecture

(Mitchell & Hofstadter, 1995, “The Copycat project: A model of mental fluidity and analogy-making”)



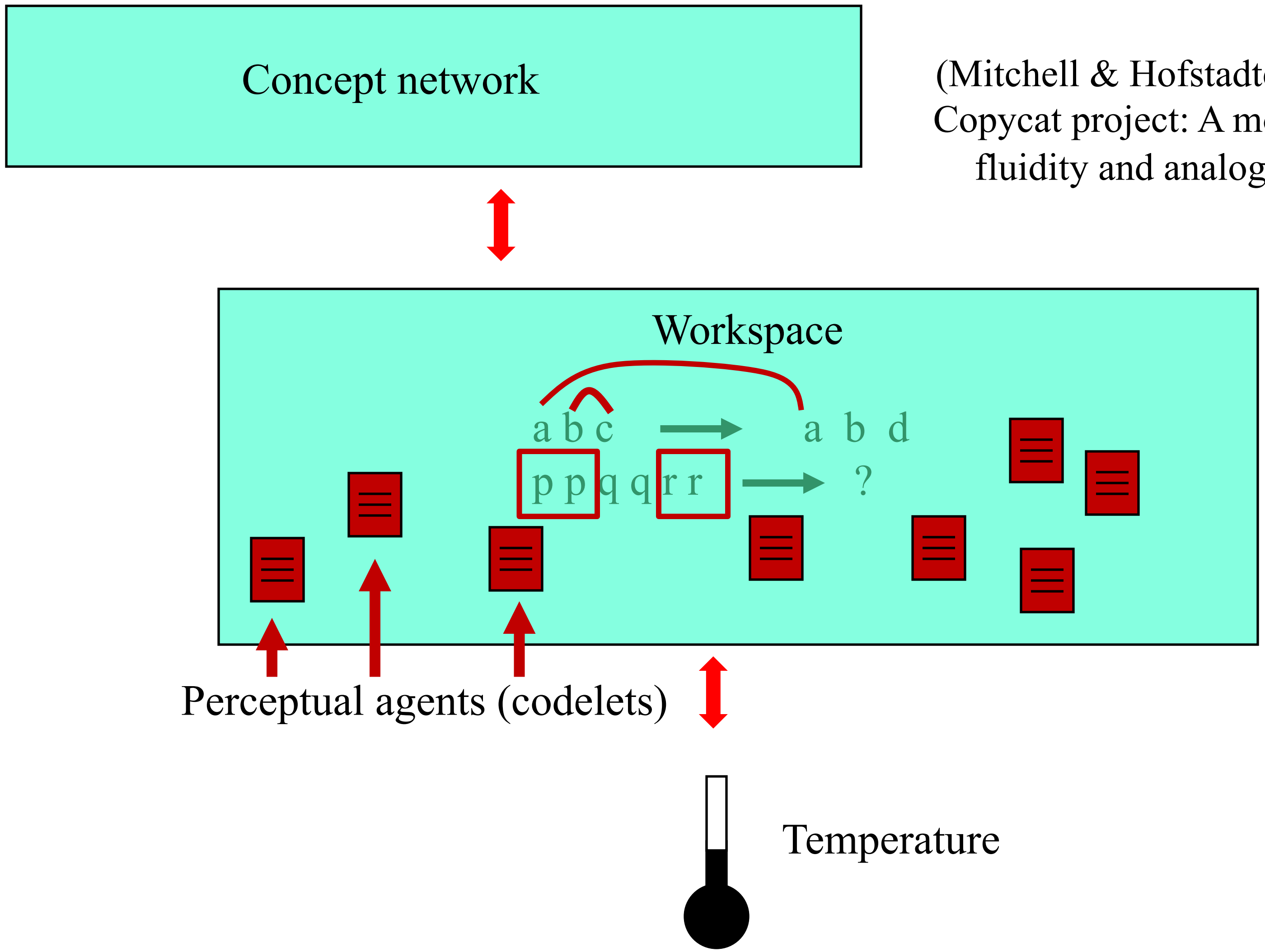
Copycat Architecture

(Mitchell & Hofstadter, 1995, "The Copycat project: A model of mental fluidity and analogy-making")

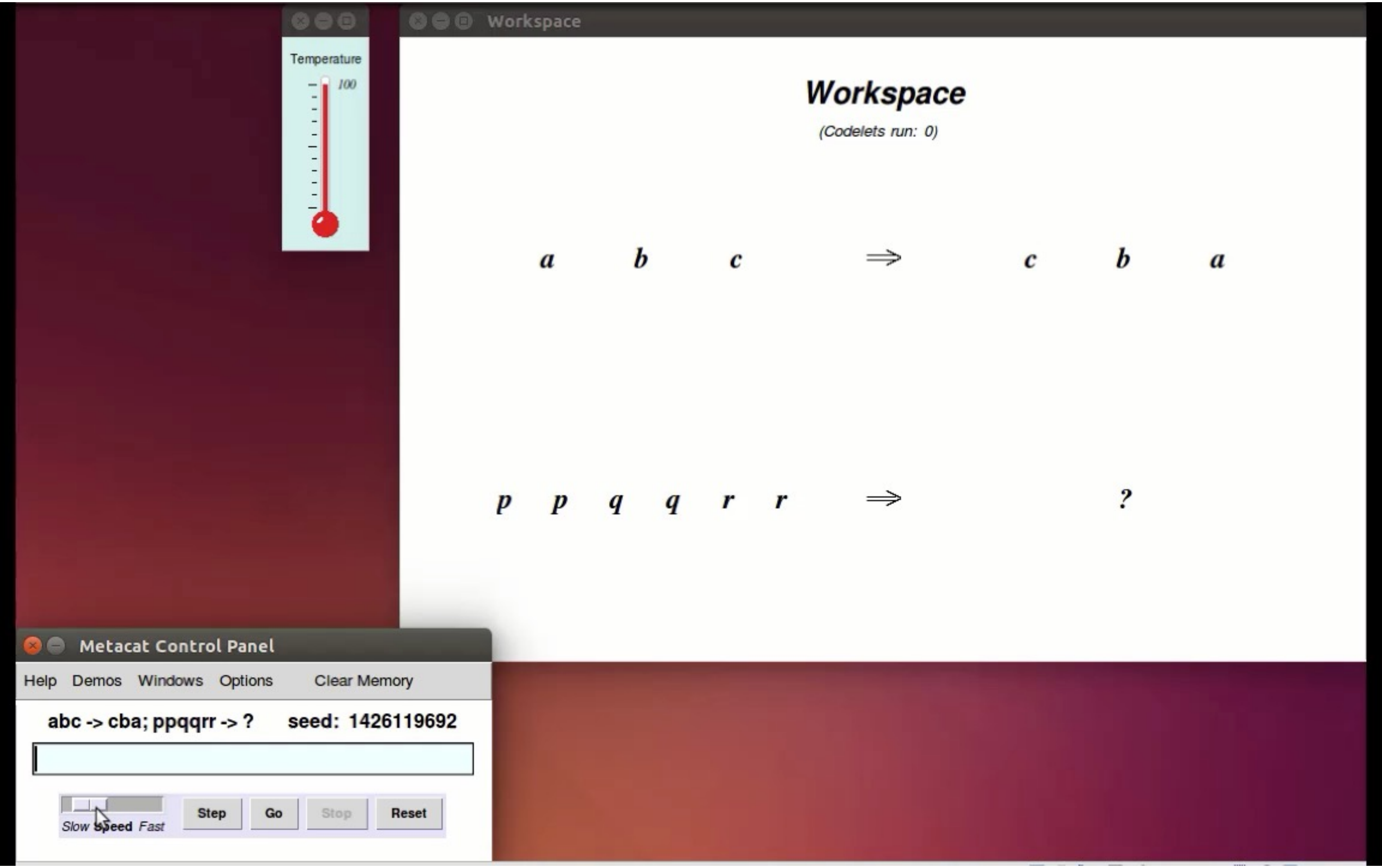


Copycat Architecture

(Mitchell & Hofstadter, 1995, “The Copycat project: A model of mental fluidity and analogy-making”)



Copycat (Metacat) demo



Some important ideas from Copycat

- Modeling analogy-making—and other “high-level” cognitive processes—as **perception**, where a representation is **actively** built up over time

Some important ideas from Copycat

- Modeling analogy-making—and other "high-level" cognitive processes—as **perception**, where a representation is **actively** built up over time

- Perception unfolds **dynamically**, continually integrating symbolic/subsymbolic and top down/bottom-up processes

Some important ideas from Copycat

- Modeling analogy-making—and other "high-level" cognitive processes—as **perception**, where a representation is **actively** built up over time

- Perception unfolds **dynamically**, continually integrating symbolic/subsymbolic and top down/bottom-up processes

- Continual integration of **prior knowledge** with bottom-up perceptions and perceived context

Copycat
Limitations

Copycat
Limitations

- Copycat's architecture is too ad hoc

Copycat
Limitations

- Copycat's architecture is too ad hoc
- Not clear how general the architecture is

Copycat
Limitations

- Copycat's architecture is too ad hoc
- Not clear how general the architecture is
- How to form new concepts beyond what is given in its prior conceptual repertoire?

How to make progress on abstraction and analogy in AI?

How to make progress
on abstraction and analogy in AI?

- Need common suite of challenging tasks

How to make progress on abstraction and analogy in AI?

- Need common suite of challenging tasks
- Advantage of idealized domains:

How to make progress on abstraction and analogy in AI?

- Need common suite of challenging tasks
- Advantage of idealized domains:
  - We can be explicit about what prior knowledge and assumptions are needed for each task domain.

How to make progress on abstraction and analogy in AI?

- Need common suite of challenging tasks

- Advantage of idealized domains:
  - We can be explicit about what prior knowledge and assumptions are needed for each task domain.

  - By avoiding language-based tasks, we can avoid anthropomorphizing what a system has achieved.

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation “test sets”).

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation “test sets”).

- Evaluation should be based on several factors, including:

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation "test sets").

- Evaluation should be based on several factors, including:
  - accuracy on hidden examples

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation “test sets”).

- Evaluation should be based on several factors, including:
  - accuracy on hidden examples
  - performance across different tasks (generality)

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation "test sets").

- Evaluation should be based on several factors, including:
  - accuracy on hidden examples
  - performance across different tasks (generality)
  - ability to generate (as opposed to simply recognize) solutions

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation “test sets”).

- Evaluation should be based on several factors, including:
  - accuracy on hidden examples
  - performance across different tasks (generality)
  - ability to generate (as opposed to simply recognize) solutions
  - ability to abstract with only a few examples

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation “test sets”).

- Evaluation should be based on several factors, including:
  - accuracy on hidden examples
  - performance across different tasks (generality)
  - ability to generate (as opposed to simply recognize) solutions
  - ability to abstract with only a few examples
  - robustness to modifications in tasks

- AI methods should be evaluated on hidden human-created examples that periodically change (no static evaluation "test sets").

- Evaluation should be based on several factors, including:
  - accuracy on hidden examples
  - performance across different tasks (generality)
  - ability to generate (as opposed to simply recognize) solutions
  - ability to abstract with only a few examples
  - robustness to modifications in tasks
  - scalability to more complex examples of tasks

Thank you for listening!