What will it take to solve human cognitive development?

Jay McClelland
Stanford University

Visiting at

DeepMind

Questions

- Where does our knowledge come from?
- How does it evolve over time?
- How do we go from very primitive initial abilities to all of the amazing achievements of human thought?

Questions

- Where does our knowledge come from?
- How does it evolve over time?
- How do we go from very primitive initial abilities to all of the amazing achievements of human thought?

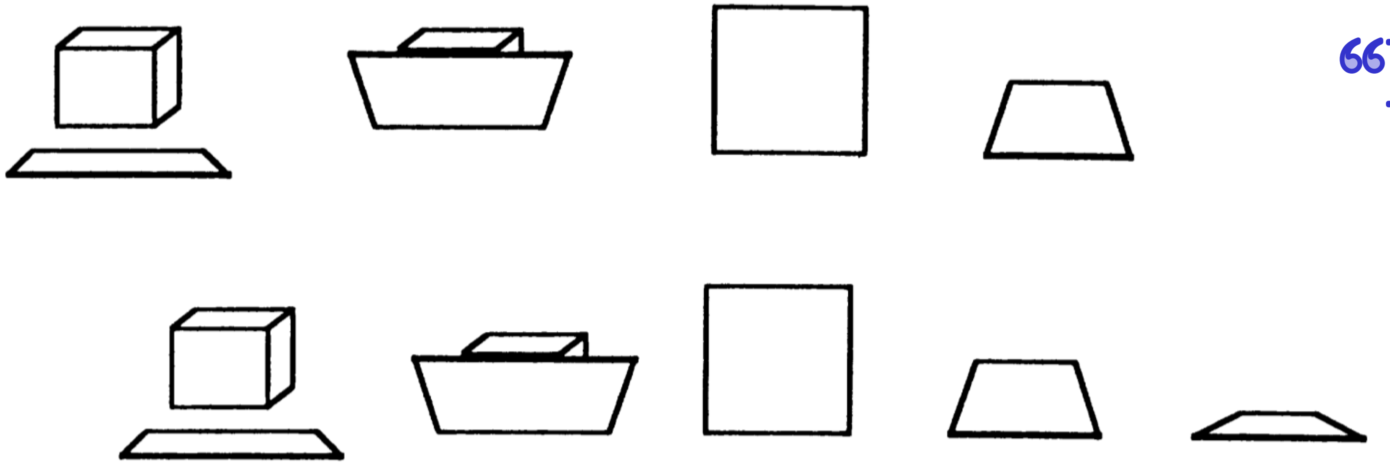
Psychological Review
1992, Vol. 99, No. 4, 605–632

Origins of Knowledge

Elizabeth S. Spelke, Karen Breinlinger, Janet Macomber, and Kristen Jacobson
Cornell University

The authors assert children reason from initial, core principles:

Objects continue to exist even when they are no longer in view

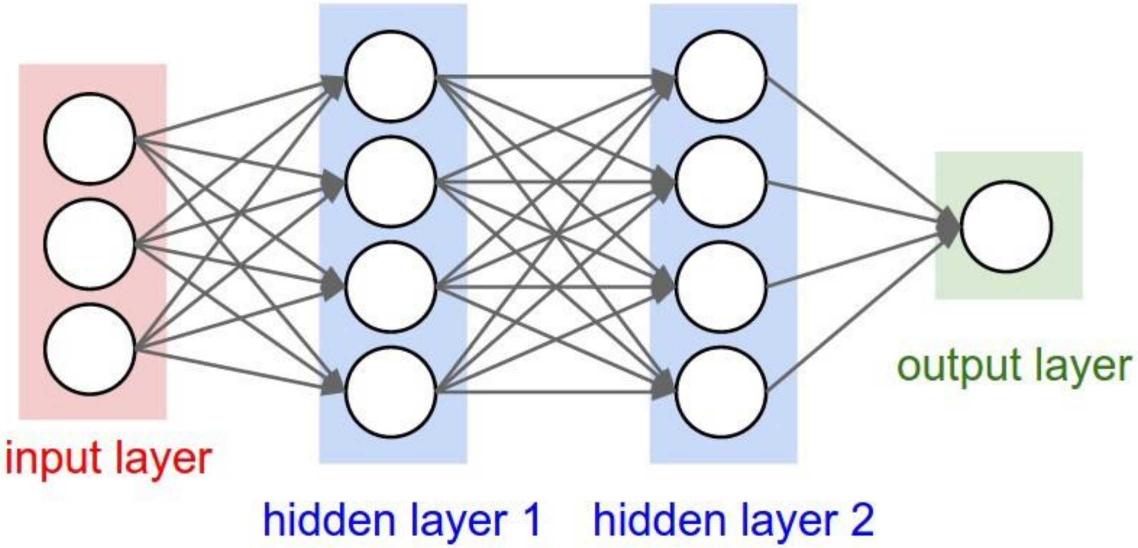


“Ho-Hum”

“Wow!”

My Basic Tenets

- Gradient not discrete
- Connection- not
  Proposition-based
- Learned not built in
- General not specific



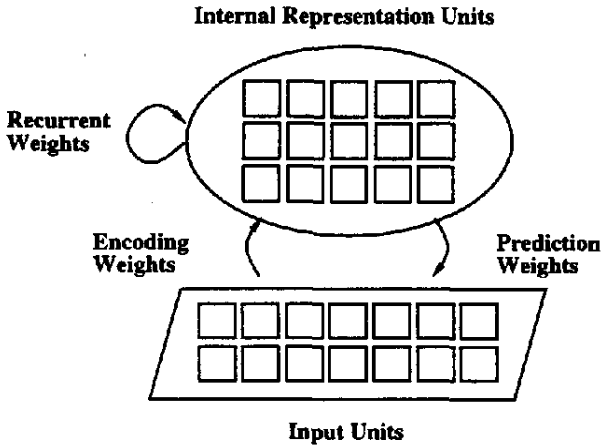
Psychological Review
1997, Vol. 104, No. 4, 686–713

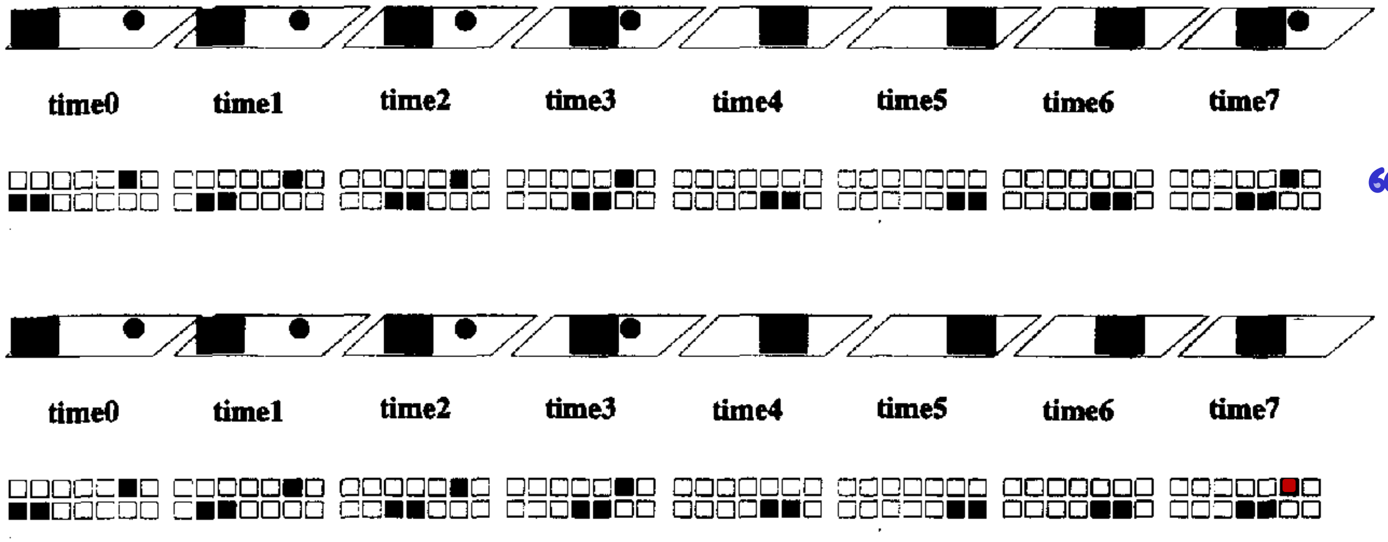
Rethinking Infant Knowledge: Toward an Adaptive Process Account of Successes and Failures in Object Permanence Tasks

Yuko Munakata and James L. McClelland
Carnegie Mellon University and the Center for the Neural Basis of Cognition

Mark H. Johnson
Medical Research Council Cognitive Development Unit

Robert S. Siegler
Carnegie Mellon University





"Ho-Hum"

"Wow!"

Questions

- Where does our knowledge come from?
  - How does it evolve over time?
  - How do we go from very primitive initial abilities to all of the amazing achievements of human thought?

Prediction-error driven learning

Adjust each parameter of the mind to reduce the discrepancy between predicted and observed events

- Could this be the engine that drives cognitive development and the discovery of new knowledge?

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In Morris, R. (Ed)., *Parallel distributed processing: Implications for psychology and neurobiology.* (pp. 8-45). New York: Oxford University Press.

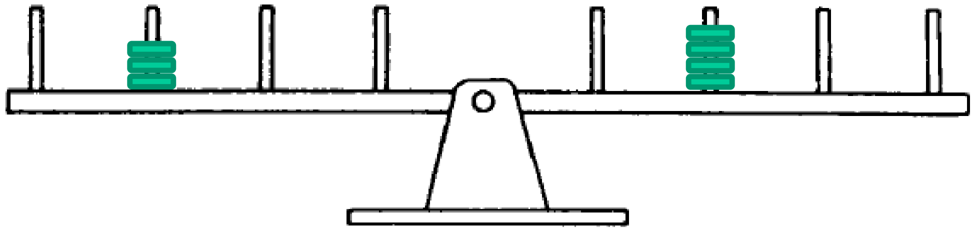
McClelland, J. L. (1994). The interaction of nature and nurture in development: A parallel distributed processing perspective. In P. Bertelson *et al.* (Eds.), *International perspectives on psychological science, Volume 1: Leading themes.* United Kingdom: Erlbaum.

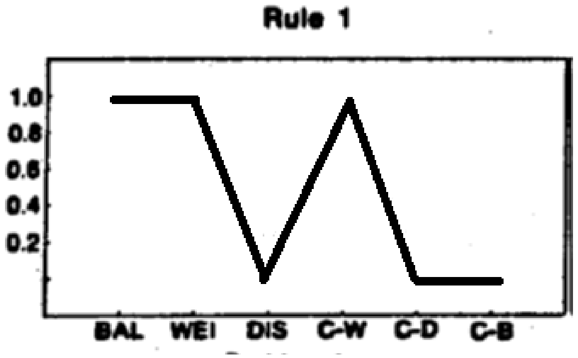
Aspects of Human Developmental Change: A Case Study (McClelland, 1989)

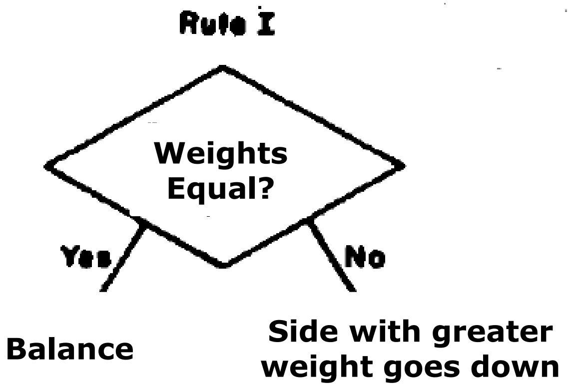
- Stage-like progression:
  - Early incomprehension
  - Systematic errors
  - Gradual progression to approximate intuitive mastery

- Readiness to learn:
  - Gradual developmental change creates conditions for fast learning
  - Consistent new knowledge can be added to existing knowledge easily
  - Inconsistent new knowledge is much harder to learn

Siegler's Balance Scale Task and Rule-Based Analysis







Older children used more complex 'Rules' (III and IV). We focus on I and II today

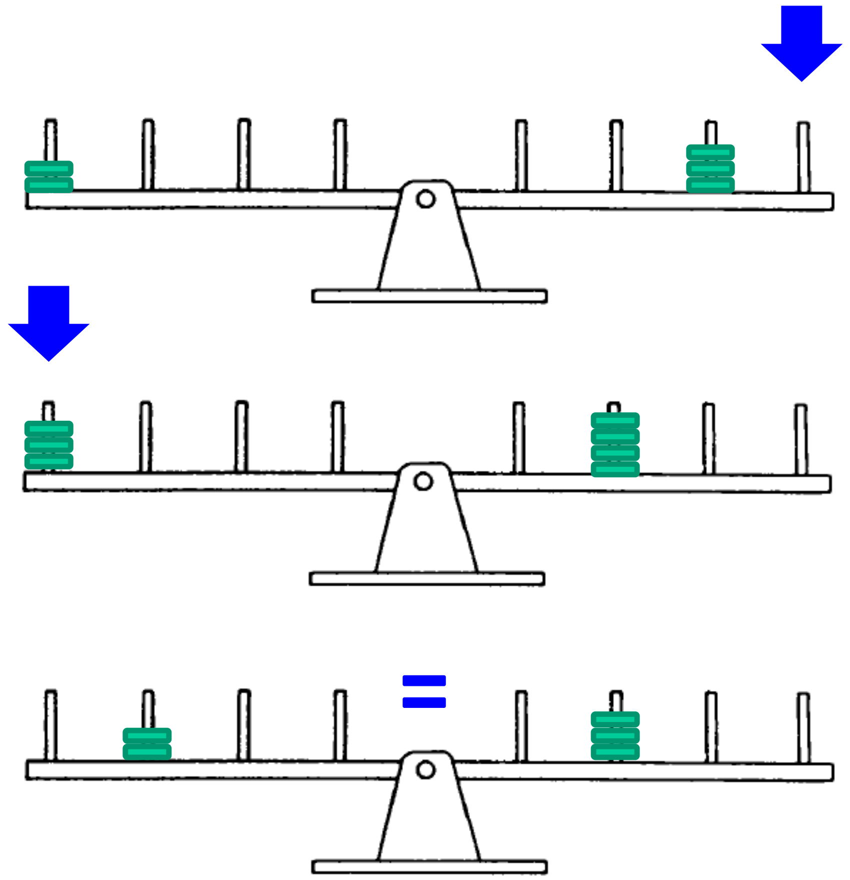
Readiness to Learn Experiment

- Rule 1 children are given 3 types of conflict problems with feedback

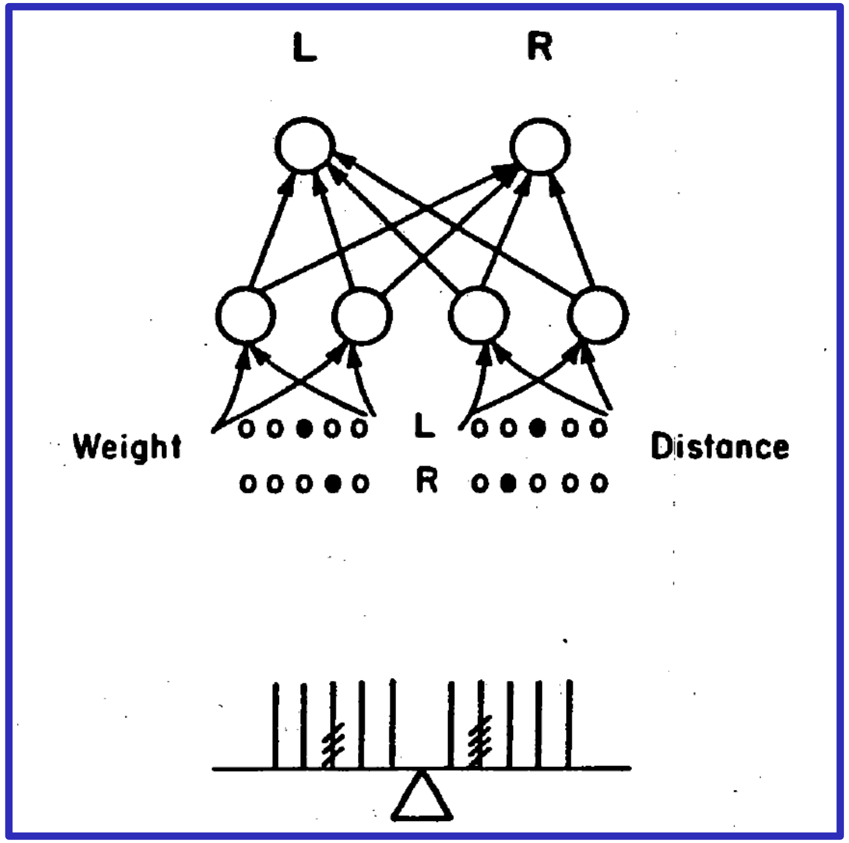
  5 examples of each type
- 5 year olds:

  show no change or regress
- 7-8 year olds:

  progress on to Rule 2



The Balance Scale Model: Setup, training, and testing



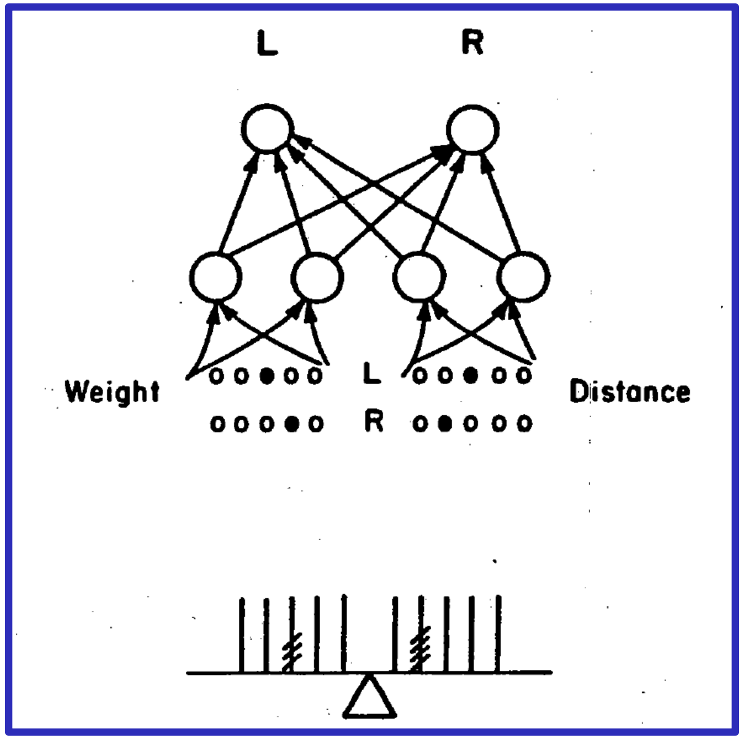
- Training set:
  - All combinations of 2 W’s and 2 D’s
  - W varies more frequently than D
- Test set:
  - Same 24 problems used by Siegler
- Scoring Criterion:
  - 20/24 responses must match a Rule
- Four nets were trained and tested

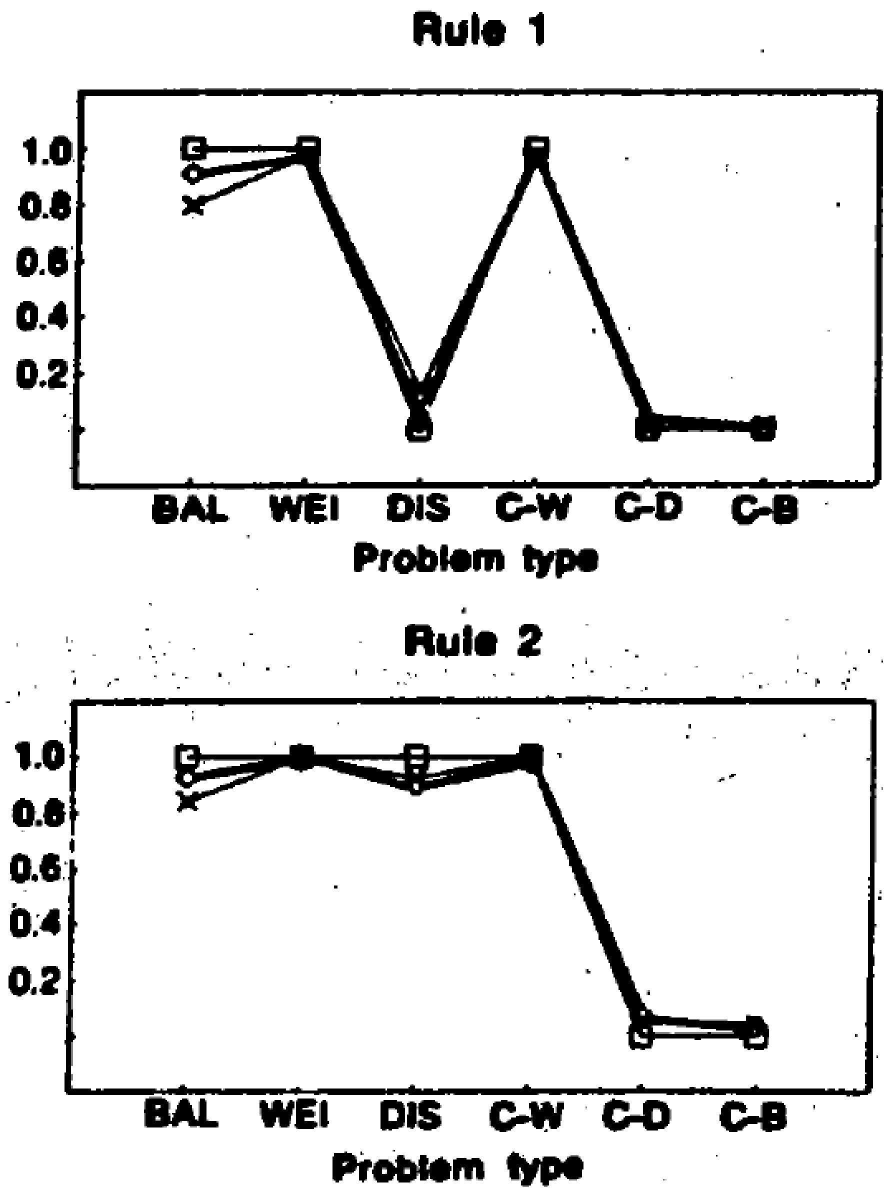
Results

- After an initial delay, networks were consistent with one of the rules 85% of the time (compared to 90% for children)

- One run is shown at right (gap = not rule consistent)

Rules, the model and to human performance

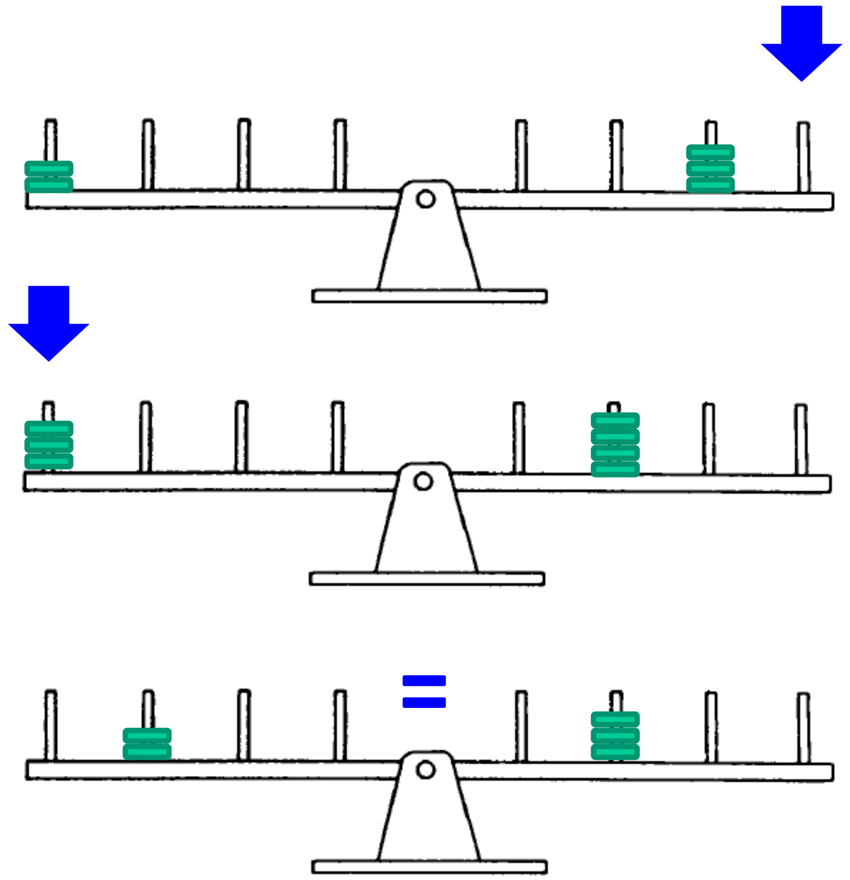




□ Rule
◇ Child
× Model

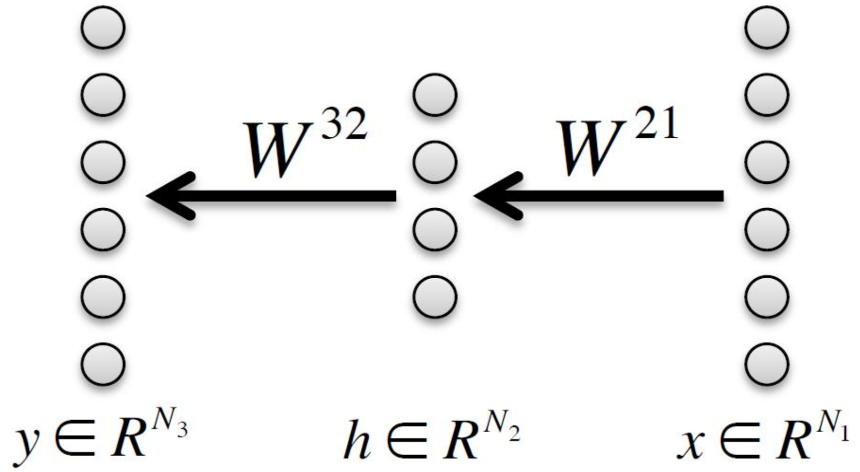
Readiness to Learn: Simulation

- Weights from Run 3 of the model received Siegler's 15 item learning experience
  - Early in the Rule 1 phase
  - Late in the Rule 1 phase
- Early model regresses toward guessing
- Late model progresses to Rule 2



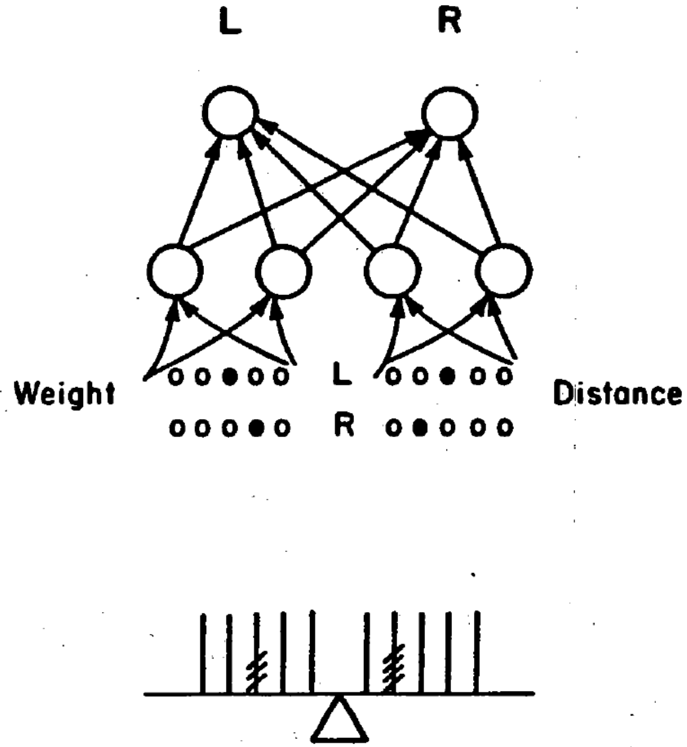
What's going on?

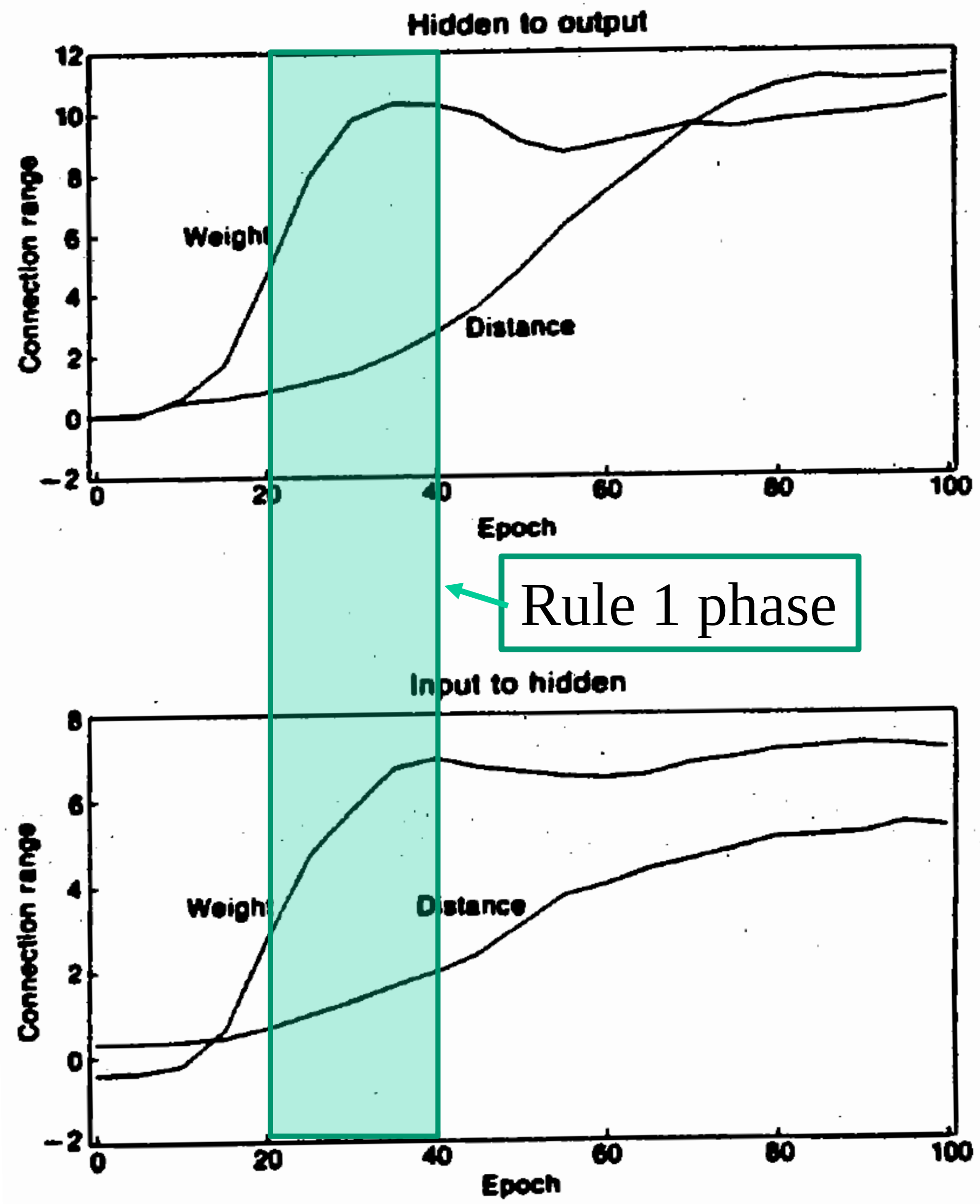
- Learning in a multi-layer network is prior knowledge dependent
- The signals that drive change in each layer of weights depend on the strengths of the connections in all of the other layers of weights
- You have to know something already to be ready to learn



$$\boxed{\Delta W^{21}} = \lambda \boxed{W^{32^T}} \left(y^\mu x^{\mu T} - W^{32} W^{21} x^\mu x^{\mu T} \right)$$

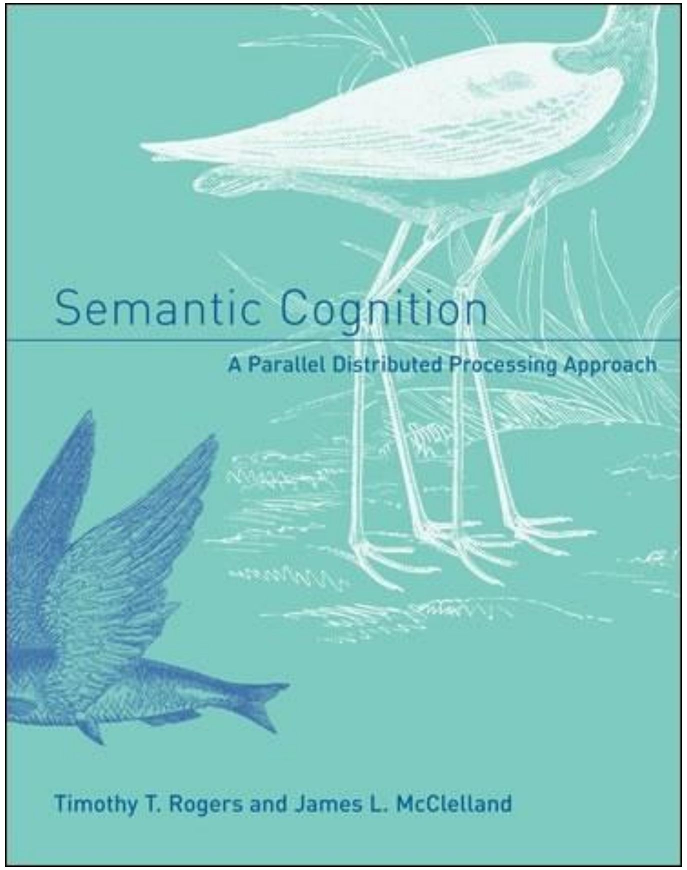
$$\boxed{\Delta W^{32}} = \lambda \left(y^\mu x^{\mu T} - W^{32} W^{21} x^\mu x^{\mu T} \right) \boxed{W^{21^T}}$$


L
R
Weight
L
R
Distance


Hidden to output
Connection range
Weight
Distance
Epoch
Rule 1 phase
Input to hidden
Connection range
Weight
Distance
Epoch

An extended exploration of these ideas

- Human semantic cognition
  - Progressive differentiation
  - U-shaped developmental change
  - Reorganization of conceptual knowledge
  - Acquired domain-specific inductive biases through domain general error-correcting learning



Shortcomings of these models

- Toy vs real problems
- Abstract vs grounded inputs and outputs
- Early-acquired intuitions vs advanced cognitive abilities
- Do not exhibit exploration and discovery

- Do not generalize beyond a specific task space
- Do not exhibit explicit understanding
- Do not benefit from explicit instruction

- Wherein lie the solutions?

Philosophy behind our Approach

1. Learning vs. Handcrafted
2. General vs. Specific
3. Grounded vs. Logic-Based
4. Active vs. Passive

DeepMind

DeepMind provides key aspects of the solution

Real Progress - But Challenges Remain

✓ Real problems
✓ Grounded inputs and outputs
✓ Advanced cognitive abilities
✓ Exhibit exploration and discovery

- Do not generalize beyond a specific task space
- Do not exhibit explicit understanding
- Do not benefit from explicit instruction

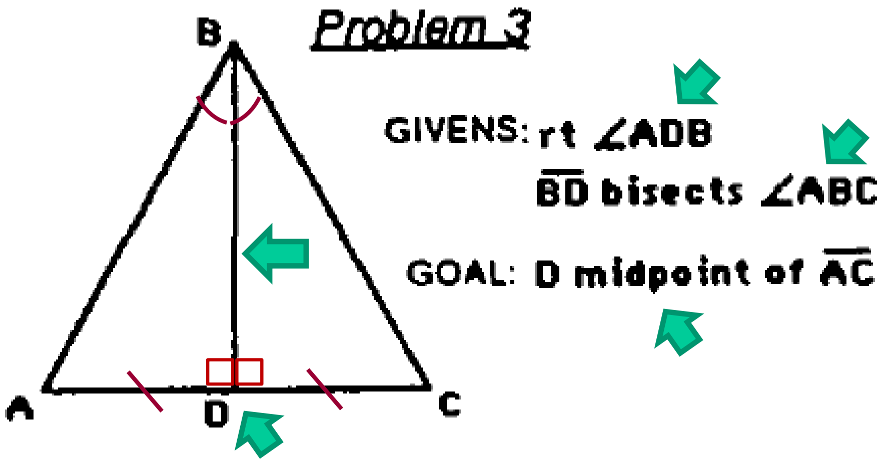
- Wherein lie the solutions?

My Plan

- Focus on human mathematical cognition
  - From number, to arithmetic, algebra, geometry, and beyond
- Apply the principles and resources of DeepMind
- Extend them to address the challenges

Why Mathematical Cognition?

A Human Solves a Geometry Proof Problem



We're given a right angle.
This is a right angle,
perpendicular on both sides
BD bisects angle ABC
And we're done.

We know that *this* is reflexive
We know we have corresponding triangles;
we can determine anything from there
in terms of corresponding parts
And that's what *this* is going to mean…
that *these* are congruent

Koedinger & Anderson, *Cognitive Science*, 1992

Why mathematical cognition?

- Mathematical discoveries are among the highest achievements of human thought
- Mathematical is often hard to learn, yet mastery leads to powerful capabilities
- Some view math as strictly formal, but grounding, intuition and insight play central roles
  - Mathematics can be concretely grounded while still obeying formal rules
  - Grounding facilitates understanding and transfer
- Mathematics includes justification and explanation as well as formal procedures
- Stage-like transitions, and readiness to learn, arise at every step
- The main thread of a mathematics curriculum, from counting to geometry to calculus, appears tractable to explore, since its core contents and grounding structures are circumscribed
- Should complement the existing *Theorem Proving and Maths* effort at DeepMind

What if we could…

- Create a simulated agent that would learn mathematics cumulatively, in a virtual environment, up through the basics of calculus, such that the agent could:
  - Solve novel problems
  - Learn new extensions of its skills quickly through demonstration, explanation and discovery
  - Explain and justify its solutions

Two Conjectures

1. Solving advanced cognitive abilities will depend on building systems that rely on the DeepMind philosophy

   – Learned not Hard-coded
   – General not Specific
   – Grounded not Logic-Based
   – Active not Passive

DeepMind tools essential to this effort

- Powerful stochastic neural networks with memory and attention
  - Gregor *et al.* (2015). DRAW
- Models that learn to read and write from external memory
  - Graves, Wayne *et al.* (2016). DNC
- Models that learn to communicate with others through embodied and situated language
  - Hermann, Hill *et al.* (2017)

- Virtual physical environments in which we can create learning opportunities for simulated learners

Two Conjectures

1. Solving advanced cognitive abilities will depend on building systems that rely on the DeepMind philosophy

   – Learned not Hard-coded
   – General not Specific
   – Grounded not Logic-Based
   – Active not Passive

2. But that alone is not enough

   – Advanced cognitive abilities depend on culturally-constructed tools that leverage human thought

Some technological and conceptual tools for thought

- Writing systems and durable media
  - Chalk and tablet, pencil and paper
  - Electronic documents and editors
- Diagrams and number systems
  - Straight edge and compass,
  - ruler and protractor
  - Place value systems and the abacus
- Inference systems
  - Number and arithmetic,
    algebra and geometry,
    logic and systems of mathematical proof,
    computer programming languages

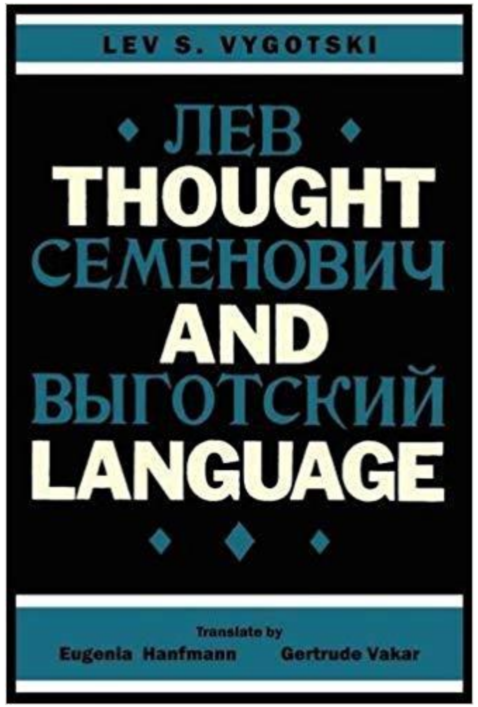
Human External Memory

External Memory
We Compute On

Formal Systems that
Structure Human Thought

Systematic Thought as an Emergent Property
Learning in Advanced Cultures

- Vygotsky:
  - We internalize logical thought as we develop, through language
- Luria:
  - Individuals from non-literate cultures do not engage in hypothetical or deductive reasoning
- Scribner & Cole:
  - People in such cultures appeal to authority and precedent rather than principles or facts



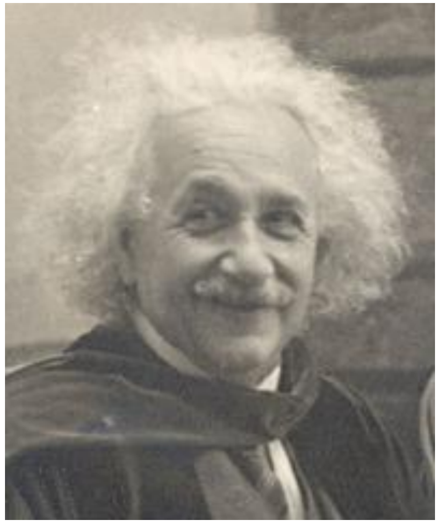
Cognitive Consequences of Formal and Informal Education

Sylvia Scribner and Michael Cole

Science, New Series, Vol. 182, (Nov. 9, 1973), pp. 553-559

My bottom line: the basis of insight and discovery

- Intuition and acquired systematic thinking abilities are required for deep understanding and discovery
- Powerful computer-based mathematical processing systems exist but
  - They were programmed by humans
  - They lack insight and intuition
- Modeling the human case will give us insight into how to build artificial systems that incorporate these human-like qualities



Einstein in
Academic Regalia

Understanding the Counting Numbers: A Sudden Discovery or a Gradually Emerging Cognitive Skill?

- Reciting the count list
  - Often learned early, but without apparent meaning
- Answering “how many?”
  - first for sets of 1, then up to 2, then up to 3 or 4 items
- A sudden discovery?
  - If a child tends to succeed for 5, they’ll tend to succeed with larger numbers (though errors continue to occur)
- However the child may fail many other tasks
  - Give-a-number task
  - Which is more task
  - The ‘remove an object and replace it with another’ task

Understanding the Counting Numbers: A Sudden Discovery or a Gradually Emerging Cognitive Skill?

- We take [such findings] to suggest that the development of the logical underpinnings of number knowledge is advanced by [gradual] increases in children's overall experience with numbers.

- Davidson, Eng & Barner, *Cognition,* 2012

Can a Recurrent Neural Network Learn to Count Things?

Mengting Fang, Zhenglong Zhou, Sharon Y. Chen, James L. McClelland

PDP Lab Stanford University

Questions

- **Competence:** Can a recurrent neural network architecture that can move its center of attention across a series of 'glimpses' learn to count the number of blobs in a display?

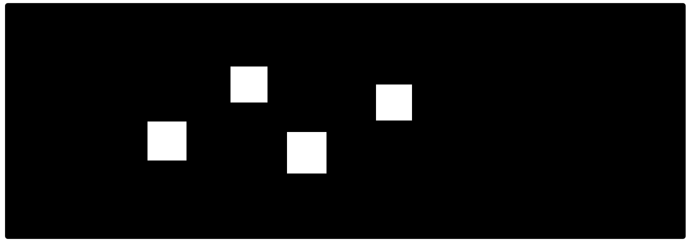
- **Learning Condition:** Can prior and concurrent learning in related tasks help the network learn to count?

- **Development:** Does this architecture allow us to capture features of the developmental progression of counting performance seen in children?

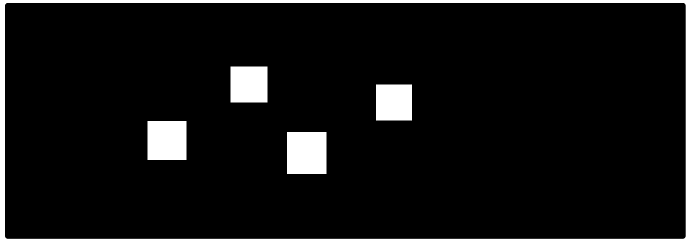
Two skills and their co-ordination

- Recite the count list (blank display)
  - Produce the numbers 1-15 in order then activate the ‘done’ unit

- Touch the blobs (1-15 blobs)
  - Start from left, ‘touch’ each blob once then stop



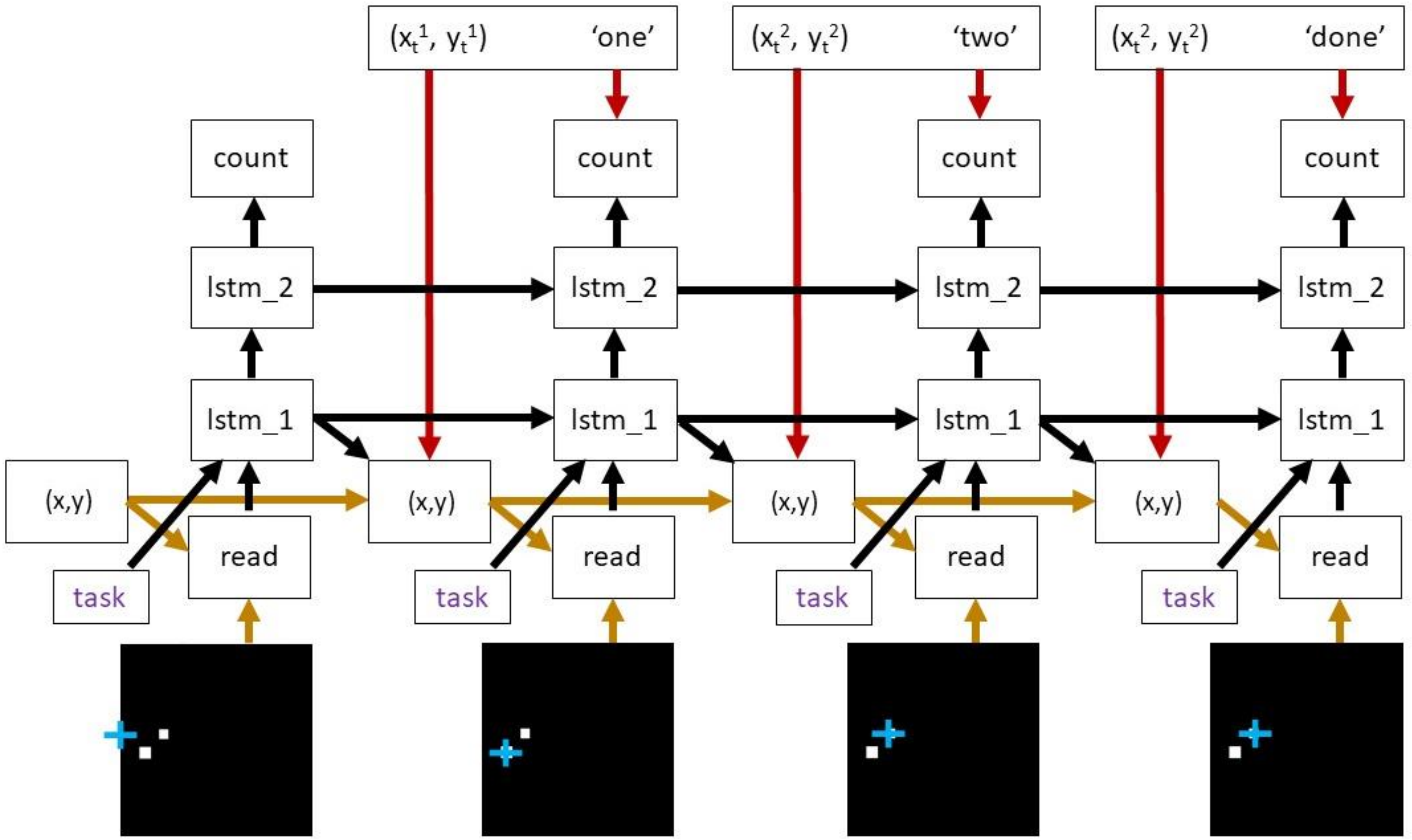
- Touch-and-count the blobs (1-15 blobs)
  - Start from left, ‘touch’ each blob once then stop
  - Produce the numbers 1-N in order then activate the ‘done’ unit



DRAW: A Recurrent Neural Network For Image Generation

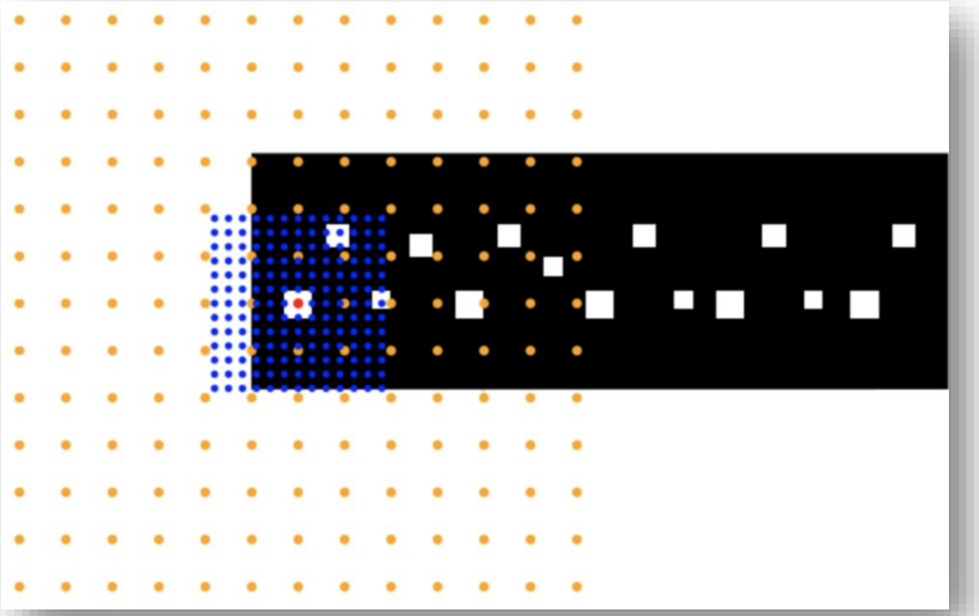
Karol Gregor KAROLG@GOOGLE.COM
Ivo Danihelka DANIHELKA@GOOGLE.COM
Alex Graves GRAVESA@GOOGLE.COM
Danilo Jimenez Rezende DANILOR@GOOGLE.COM
Daan Wierstra WIERSTRA@GOOGLE.COM

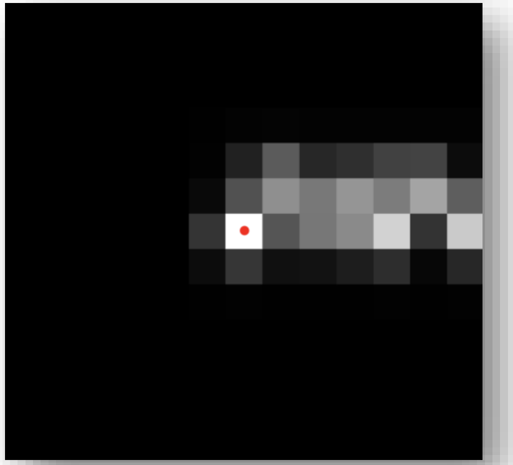
The Differentiable Recurrent Attentional Counting Model Learning the Count-the-Blobs Task



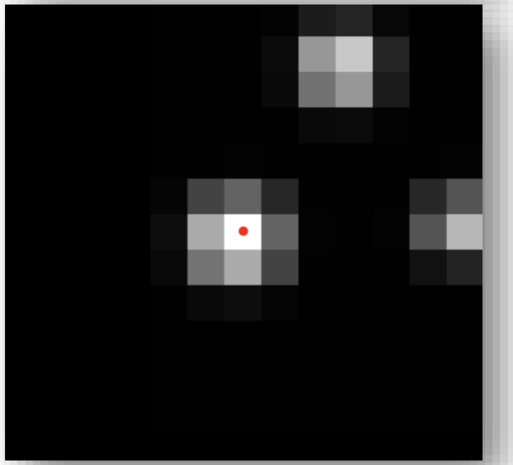
Two layer attentional window (~Mnih *et al*, 2014)

- Two sets of 13 x 13 evenly spaced Gaussian filters, one approximating a fovea (with higher resolution) and one approximating peripheral vision (with lower resolution).





peripheral output

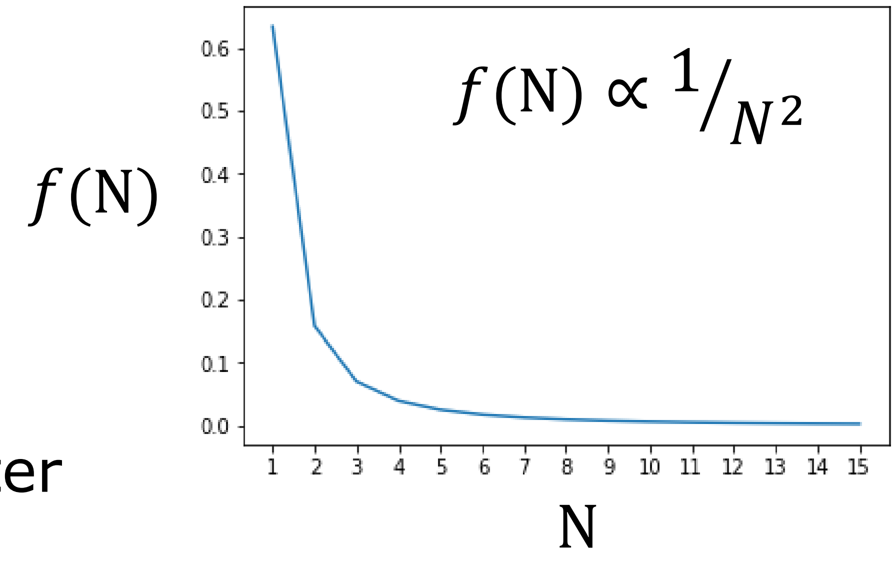


fovea output

Training Regimes

- One task regime
  - Train on the combined ‘touch and count the blobs’ task only
- Three task regime
  - Interleave training on all three tasks
- Touch first, then three tasks
  - Learn to touch til 90% correct for up to 10 items
  - Then interleave training on all three tasks

Zipfian training frequency distribution



- Corpus and scene analyses show we encounter small N’s far more often than larger N’s
- We used a Corpus-based frequency distribution (Piantadosi, 2016) in the touch the blobs and the count-and-touch the blobs tasks
  - Only 10% of trials contain 5 or more blobs
  - The network counts 8 blobs only 10 times per 1,000, higher N’s even less often

Testing

- Focus on the touch-and-count task for numbers from 1-9
- All networks were tested on 500 examples of each N from after every 500 training examples

Results

All three networks in each of the three learning conditions achieved **perfect** performance, without making any subsequent errors

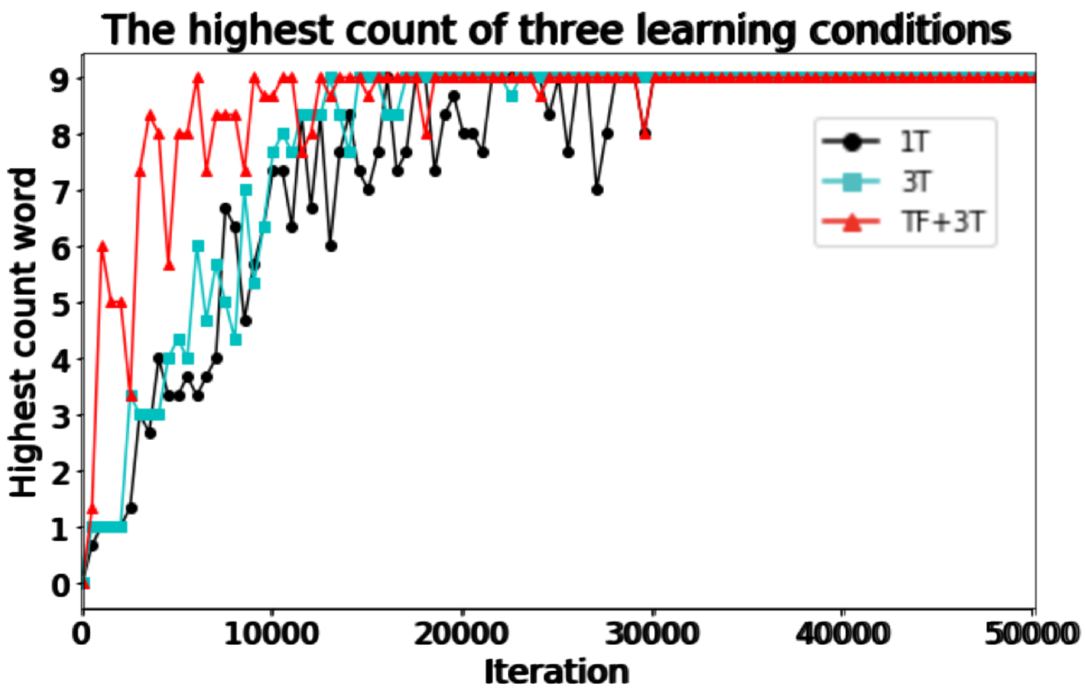
Count Performance		Perfect Iteration	Subsequent Errors
1T	Run 1	27,500	0
	Run 2	32,000	0
	Run 3	50,000	0
3T	Run 1	17,000	0
	Run 2	23,000	0
	Run 3	17,000	0
TF+3T	Run 1	30,000	0
	Run 2	12,500	0
	Run 3	30,000	0

Results

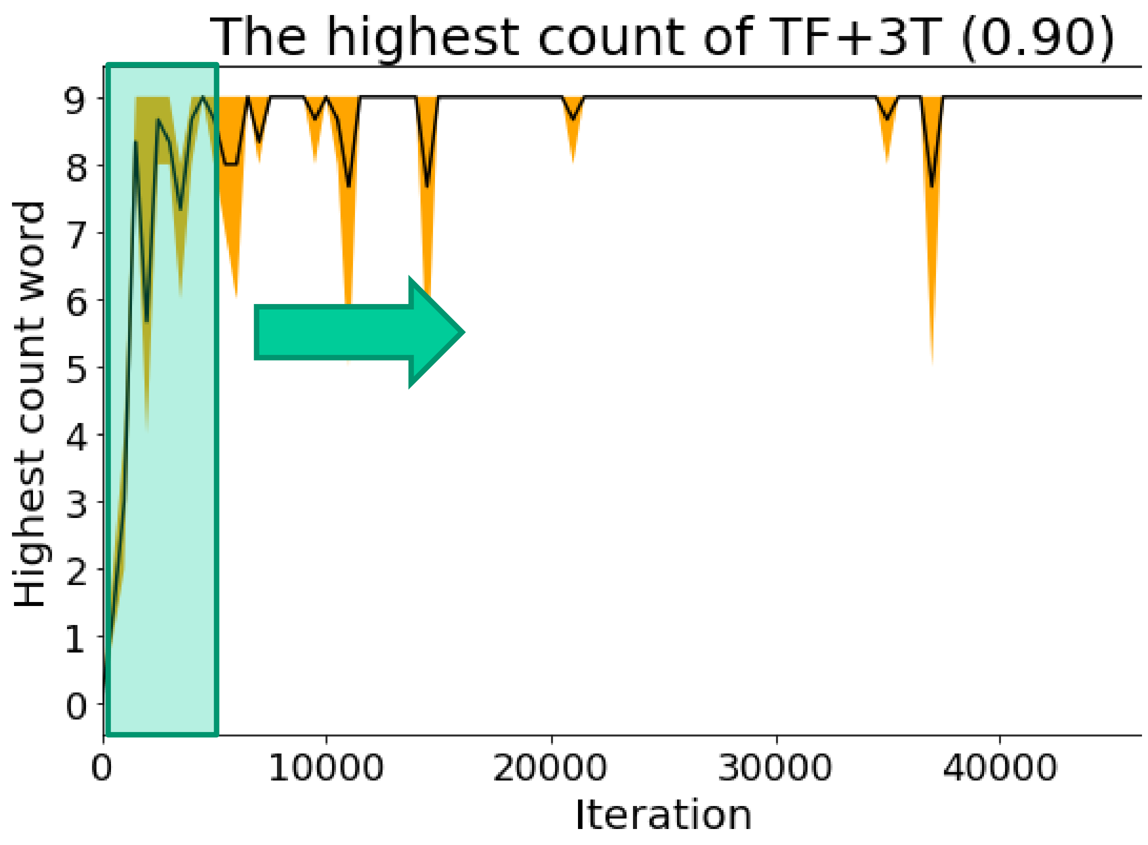
- **Relation between Learning to Touch and Learning to Count**

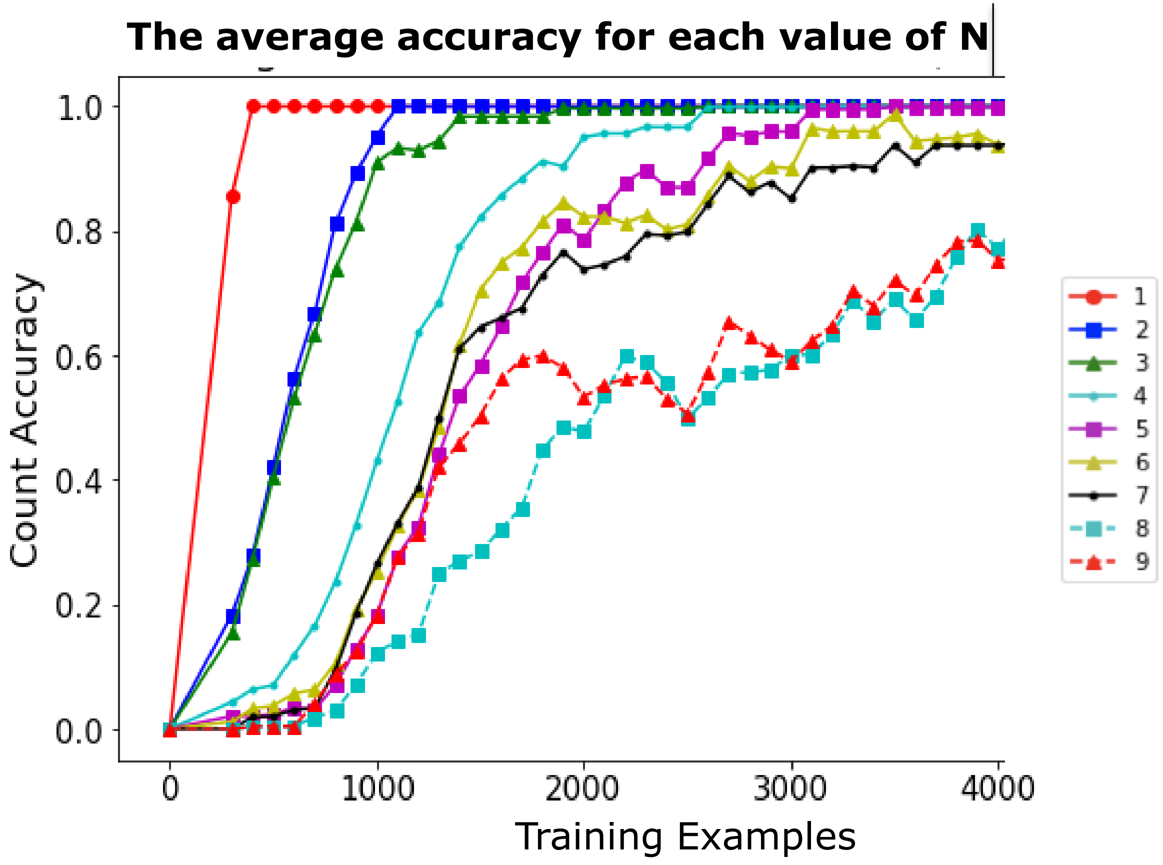
— Learning to count occurred most quickly in the **TF+3T** condition.

More experience, and earlier ability, in pointing helps the network master the ‘count the blobs’ task more quickly.



Work in Progress: Our latest TF+3T networks count 6 and 7 as well as they count 5





Next steps

- Integrate with an interactive training and testing environment
- Integrate with RL to allow exploration as well as imitation learning
- Incorporate a wider range of number tasks
  - Estimate numerosity
  - Give N items
  - Determine equivalence
  - Explore effects of transformations, variations in count orders
  - …
- Integrate language input and output to allow task instructions, question
- Extend to addition, arithmetic, algebra, geometry …

There will be many challenges on the road ahead

- I hope to engage with many of you as we try to address them

Thanks for Listening!

Many Thanks for Listening!

A Distant Goal

- Understand how expert mathematicians acquire the ability to engage in abstract mathematical reasoning
- It is remains to be understood how humans acquire the ability to think at such a level
  - They may have to learn to do it over and over as they master more and more advanced domains
  - But I think they can at least achieve a readiness to progress to the next level quickly
- I look forward to thinking and working on these questions with you.