Machine Reading and Reasoning with Neural Program Interpreters

- Sebastian Riedel


@riedelcastro

Collaborators

Johannes Welbl

(UCL)

Pontus Stenetorp

(UCL)

Matko Bosnjak

(UCL)

Tim Rocktäschel

(now at Oxford)

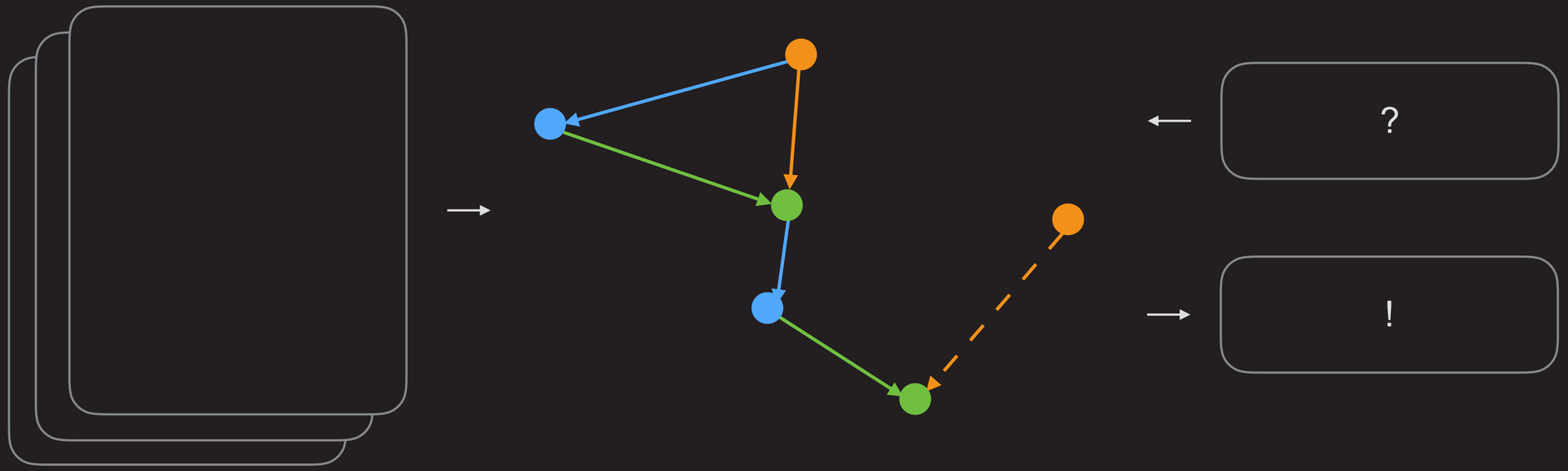
Jason Naradowsky

(now Johns Hopkins University)

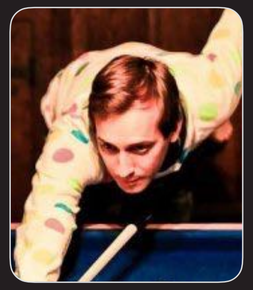
“Should we separate The “Classic” NLP Paradigm language? [Maybe not?]

Convolutional 2D Knowledge Graph Embeddings, **Tim Dettmers et al.** AAAI18, **Mon 11:30-12:30, Room 8**

Chris Manning @AKBC 2013

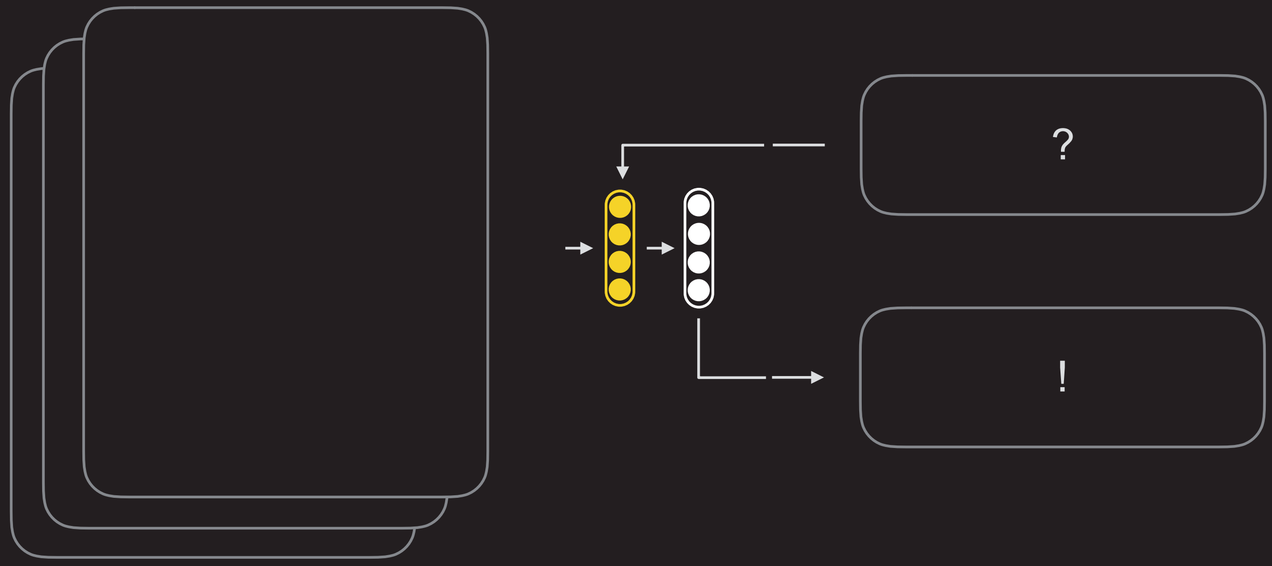


[Language] [Meaning] [Information Need]



End-to-End Reading and Comprehension

(Hermann et. al 2015, Seo et al., 2016, Rajpurkar et al., 2016, Weissenborn 2016…)



[Language] [Information Need]

Machine Reading

What city did Tesla move to in 1880?

Prague

Why was he unable to enroll at the university?

arrived too late to enroll

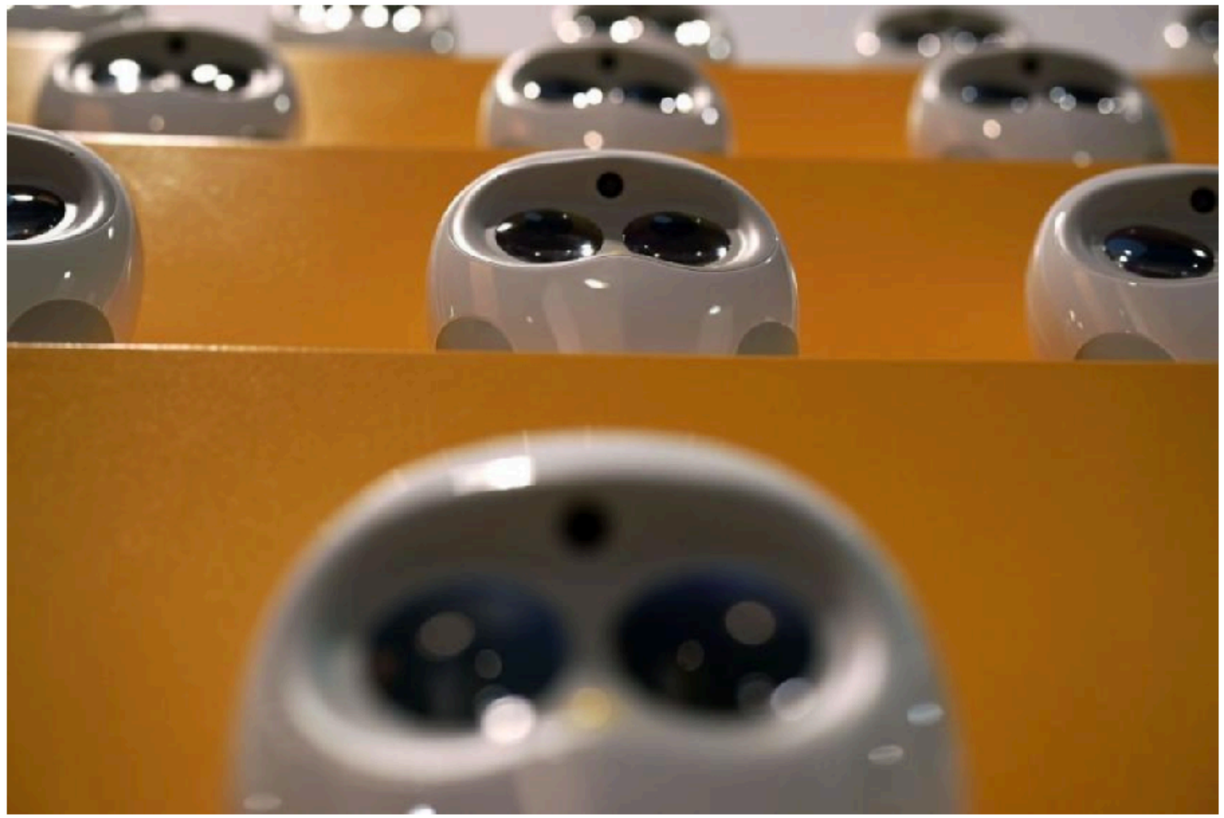
Nicola Tesla

…
In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for **Prague** where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses.

Newsweek

ROBOTS CAN NOW READ BETTER THAN HUMANS, PUTTING MILLIONS OF JOBS AT RISK

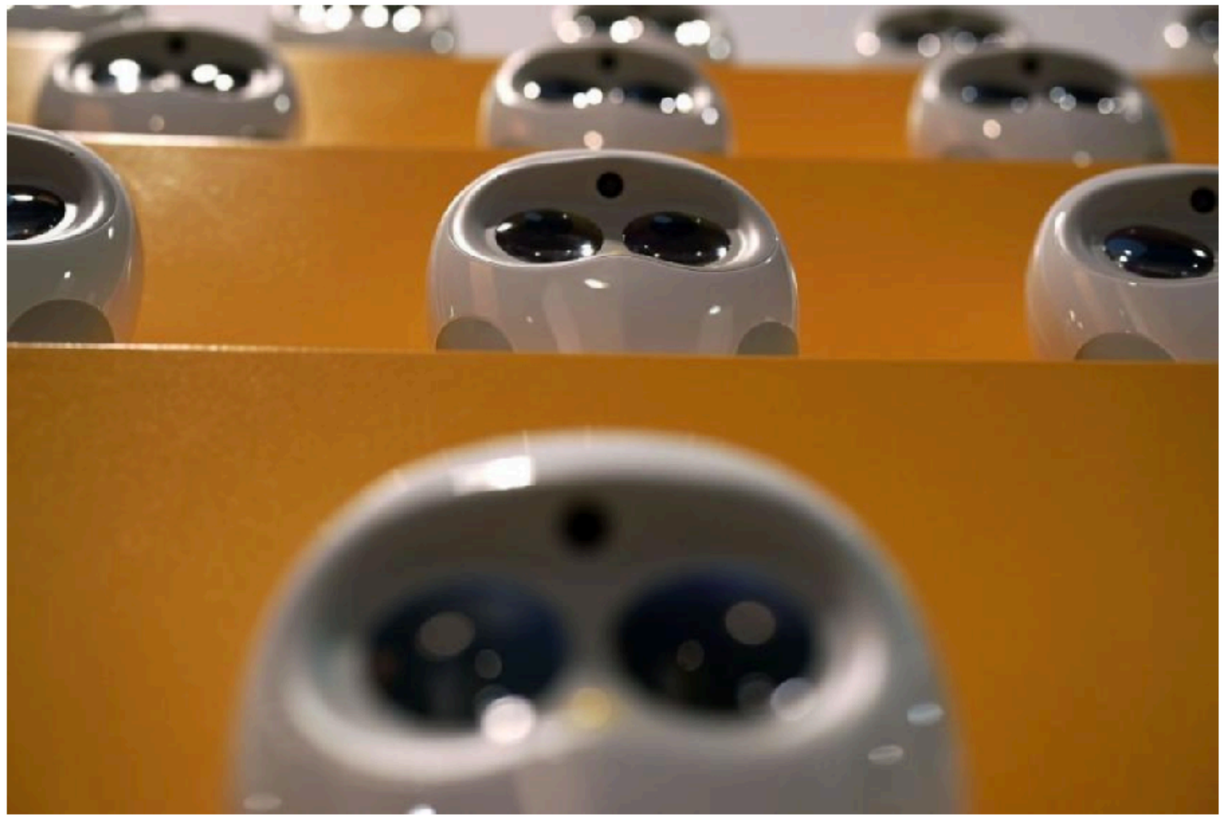
BY ANTHONY CUTHBERTSON ON 1/15/18 AT 8:00 AM



Newsweek

DEEP LEARNING KILLED THE LINGUISTS BOSTON SCIENTIST CLAIMS

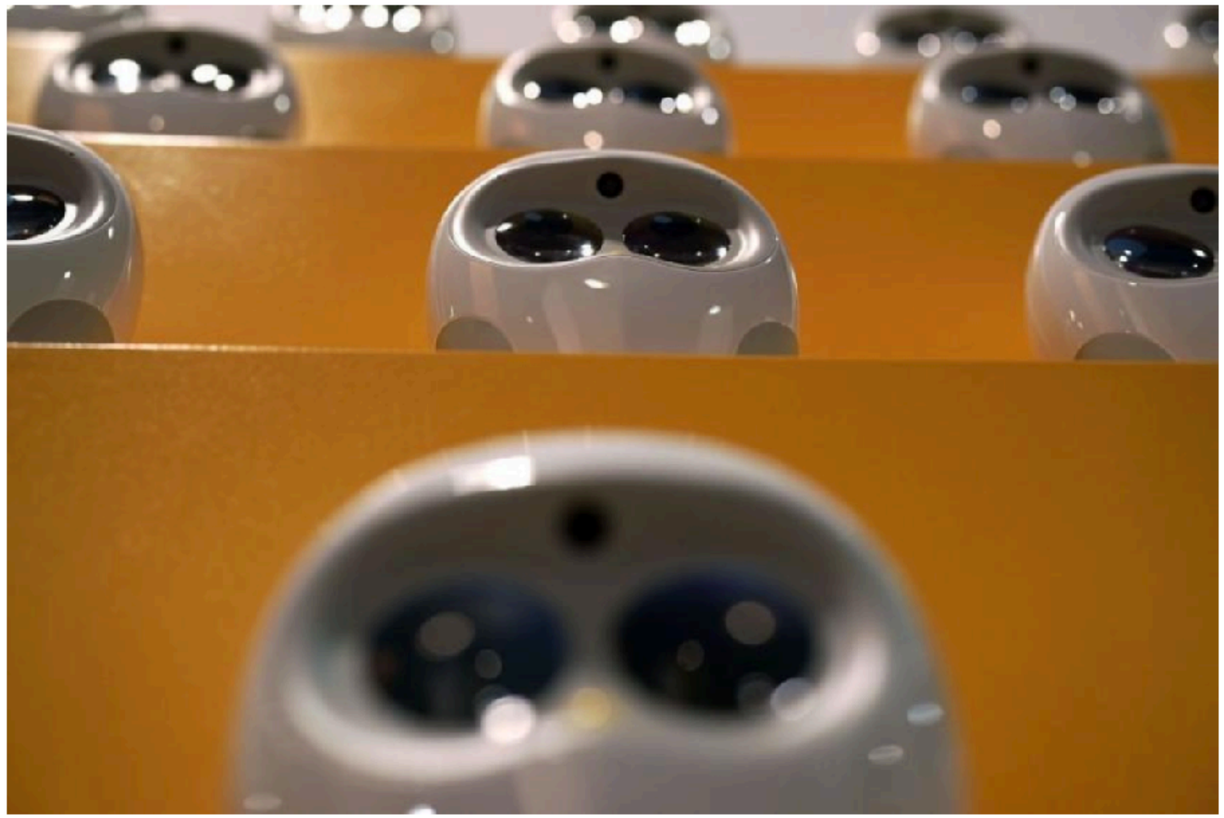
BY ANTHONY CUTHBERTSON ON 1/15/18 AT 8:00 AM



Newsweek

BOSTON SCIENTIST IS COMPLETELY WRONG CLAIMS PHILADELPHIA PROFESSOR

BY ANTHONY CUTHBERTSON ON 1/15/18 AT 8:00 AM



How to read and **reason end-to-end**?

Machine Reading and Reasoning

Which medical specialty deals with pituitary ACTH hypersecretion?

Endocrinology

Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Machine Reading and Reasoning

How many pictures were in each of the albums?

2

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

Can we learn this **end-to-end**?

Matko Bosnjak

Tim Rocktäschel

Jason Naradowsky

Part 1: Learning to Read and Calculate

Machine Reading and Reasoning: Math

How many pictures were in each of the albums?

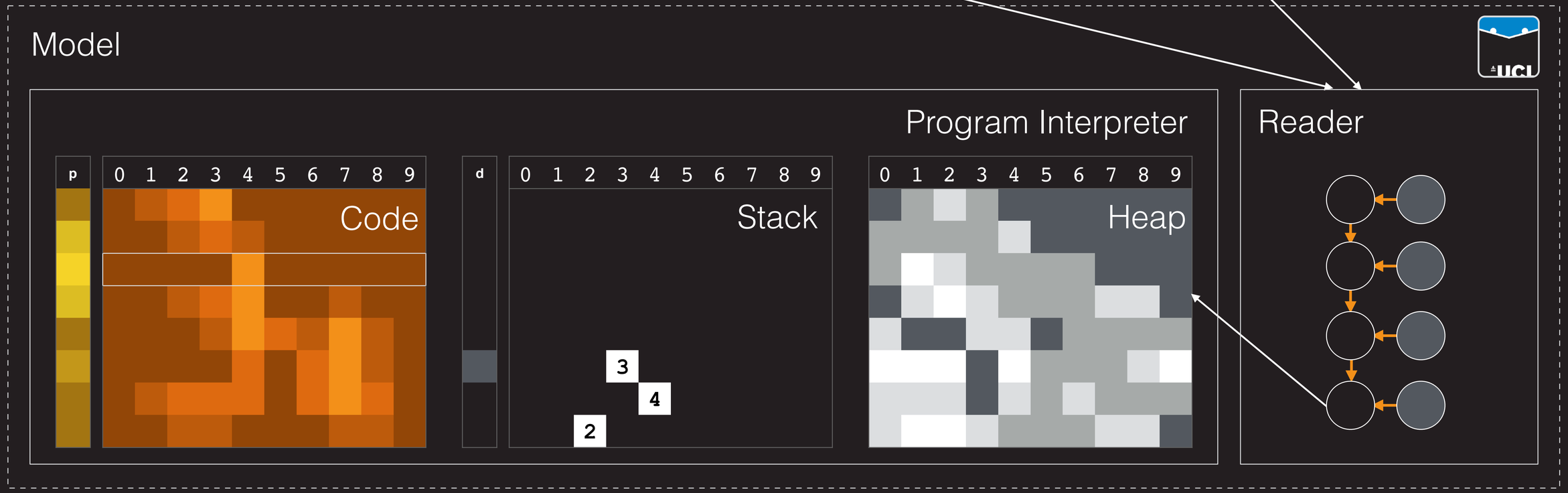
2

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

Differentiable Program Interpreters

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

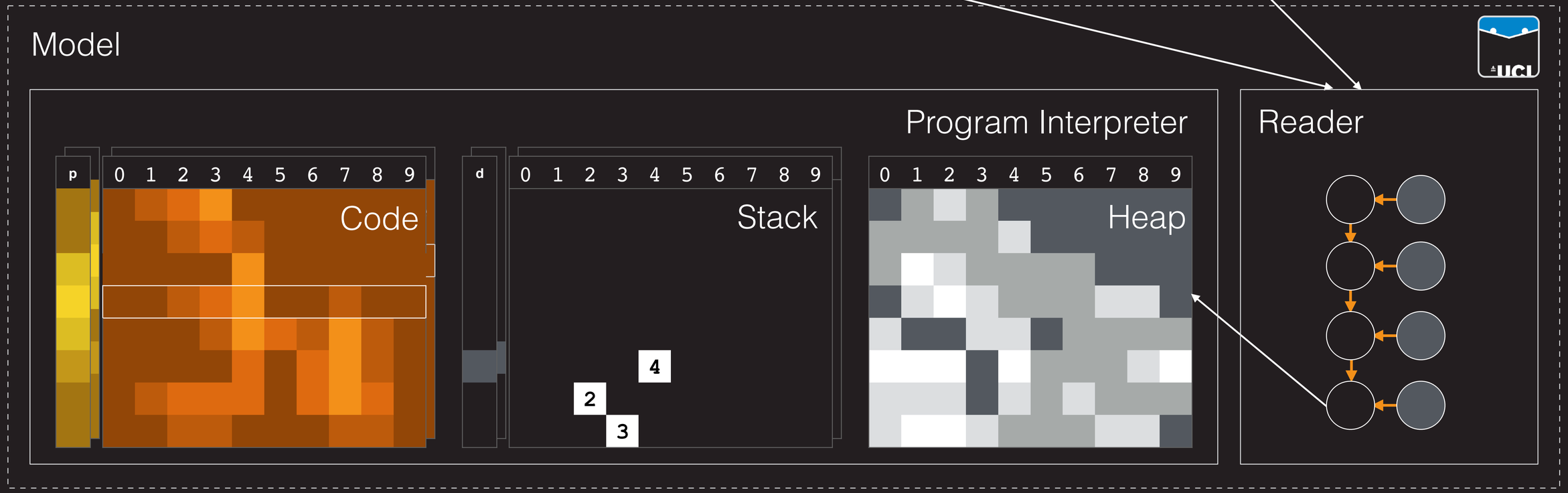


Bosnjak et al. **ICML 2017**

Differentiable Program Interpreters

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.



Bosnjak et al. **ICML 2017**

Differentiable Program Interpreters

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

Model

Program Interpreter

p 0 1 2 3 4 5 6 7 8 9 Code

d 0 1 2 3 4 5 6 7 8 9 Stack

3

6

0 1 2 3 4 5 6 7 8 9 Heap

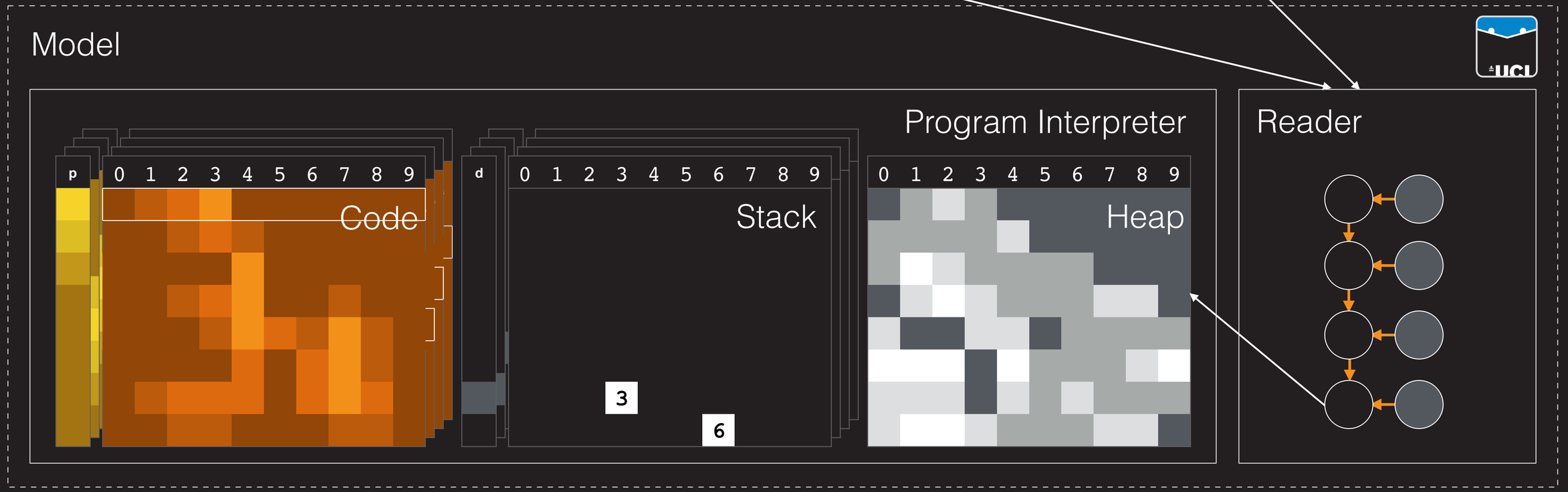
Reader

Bosnjak et al. **ICML 2017**

Differentiable Program Interpreters

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.



Bosnjak et al. **ICML 2017**

Differentiable Program Interpreters

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

Model

Program Interpreter

p 0 1 2 3 4 5 6 7 8 9 Code

d 0 1 2 3 4 5 6 7 8 9 Stack

2

0 1 2 3 4 5 6 7 8 9 Heap

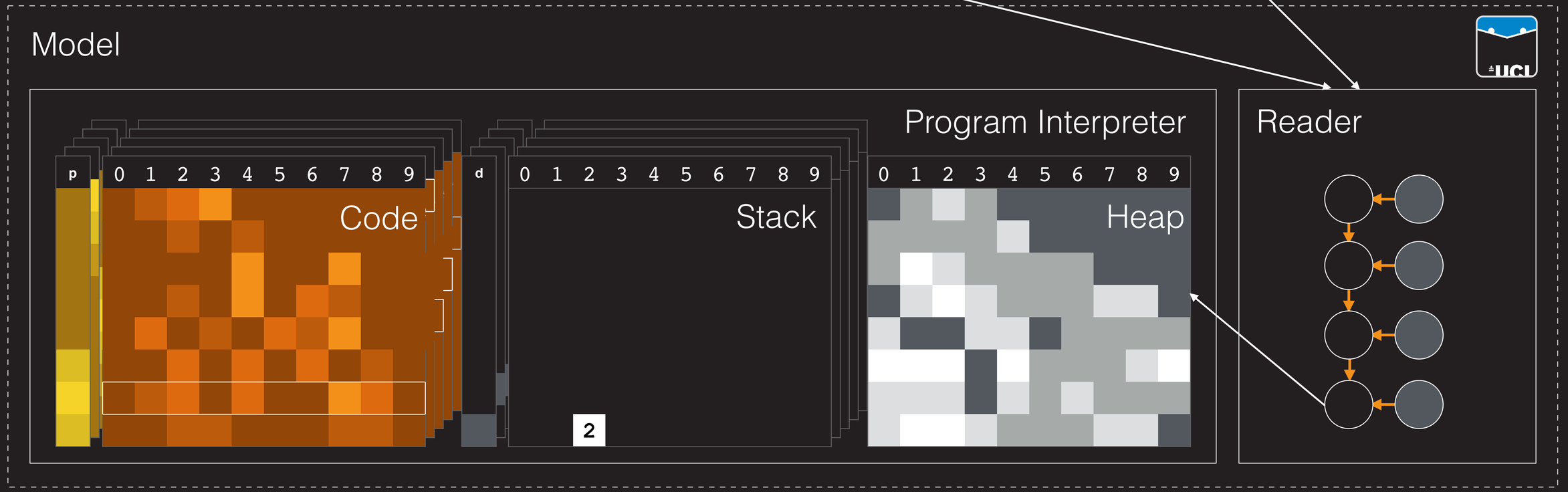
Reader

Bosnjak et al. **ICML 2017**

Training

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

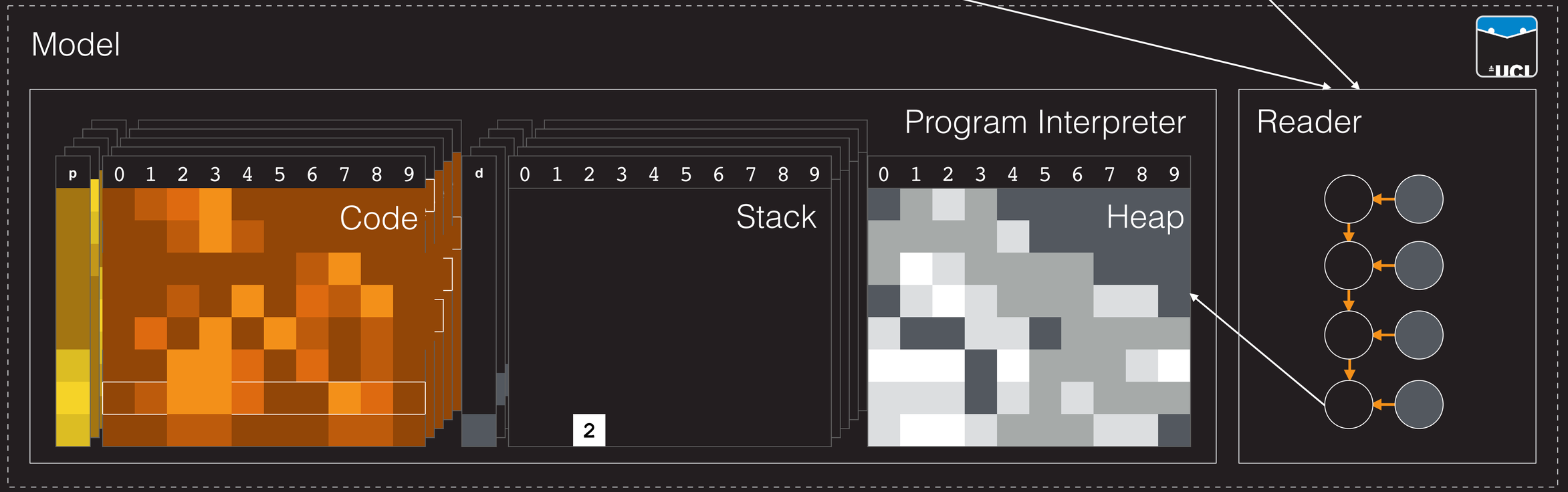


Bosnjak et al. **ICML 2017**

Training

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

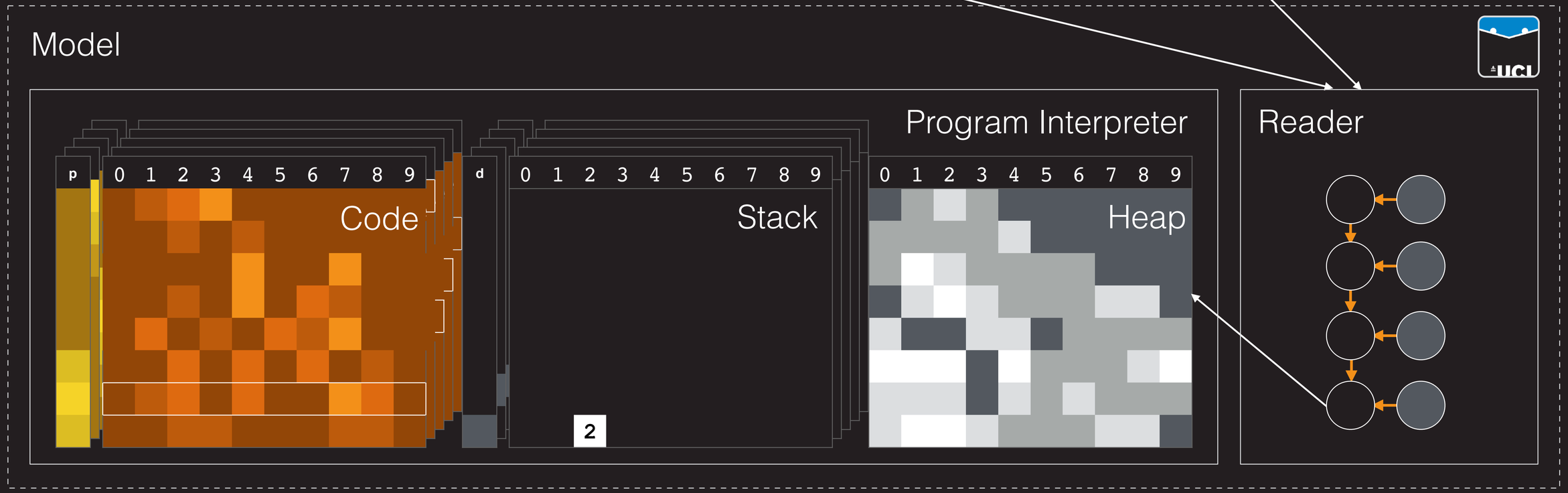


Bosnjak et al. **ICML 2017**

Training

How many pictures were in each of the albums?

Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.

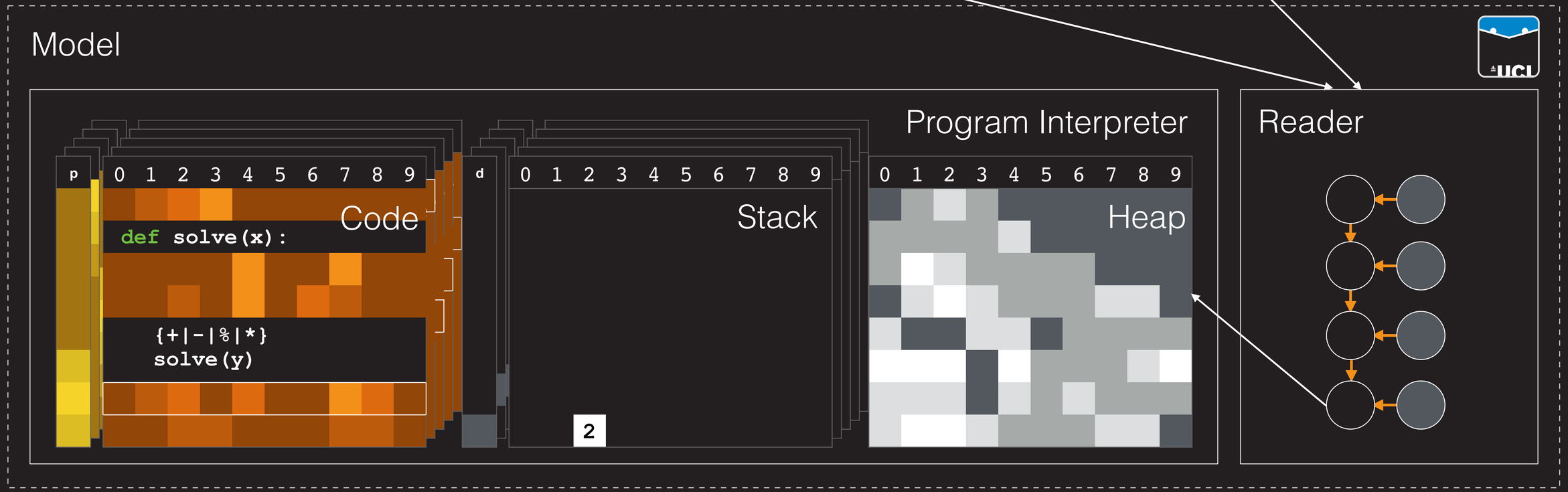


Bosnjak et al. **ICML 2017**

Training

How many pictures were in each of the albums?

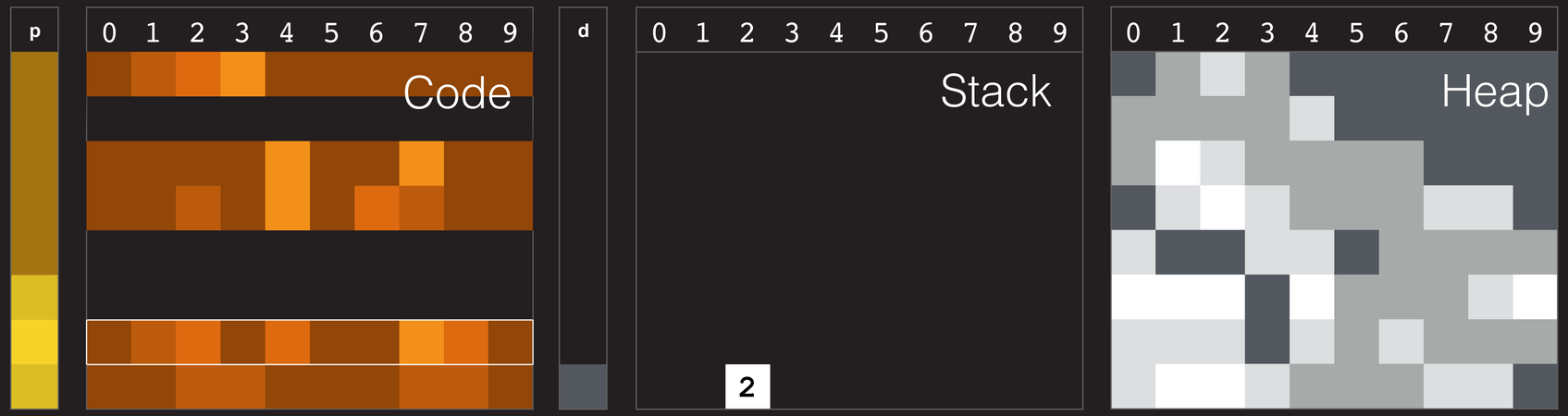
Isabel uploaded **2 pictures from her phone** and **4 from her camera** to facebook. She sorted the pics into **3 different albums** with the same amount of pics in each album.



Bosnjak et al. **ICML 2017**

Zoom in: State Transitions

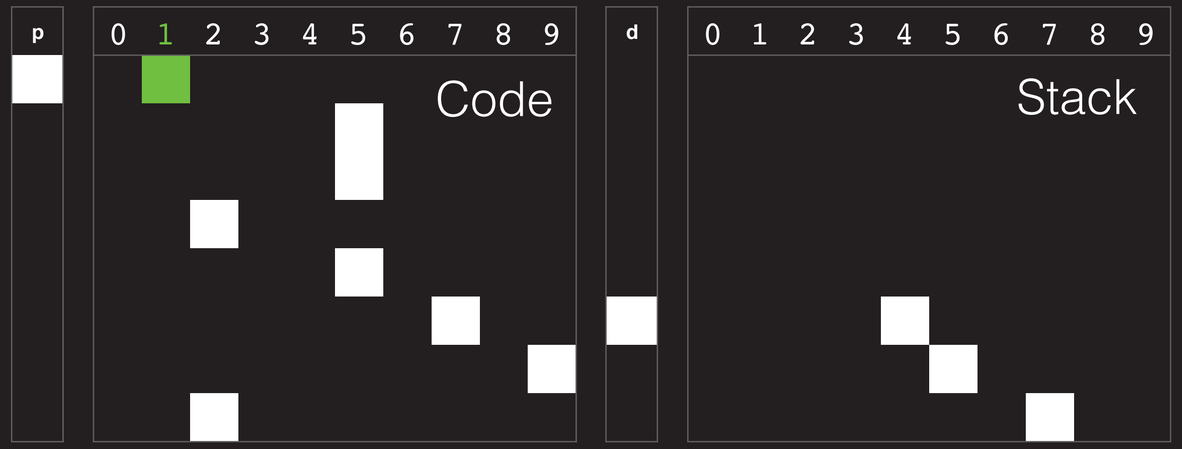
h_{t-1}



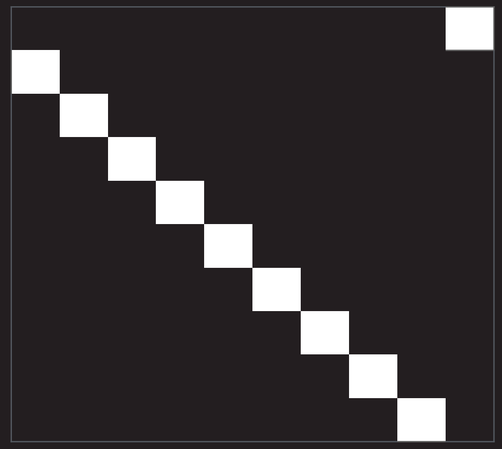
h_t

Zoom in: Pop Operation

Pop



Circular Shift Matrix

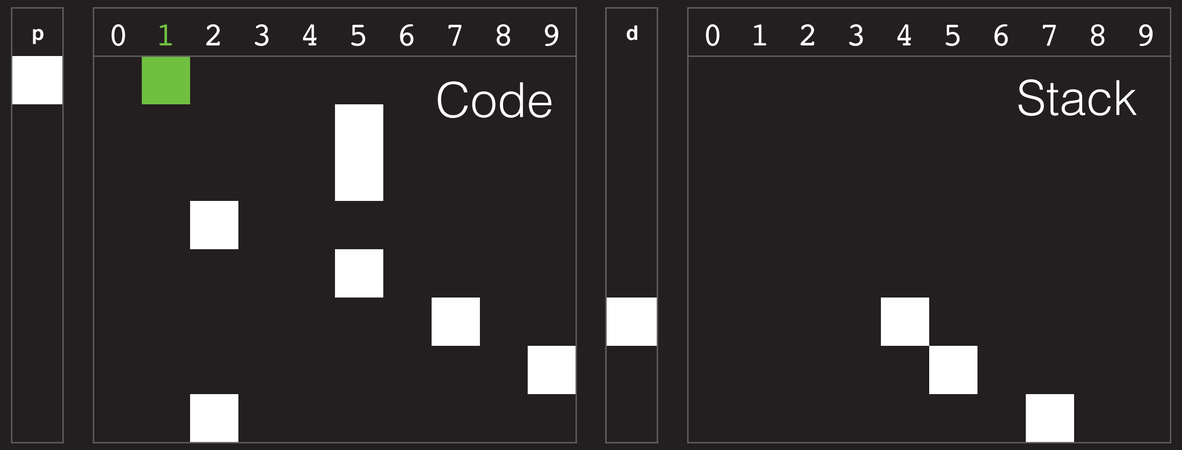


h_{t-1}

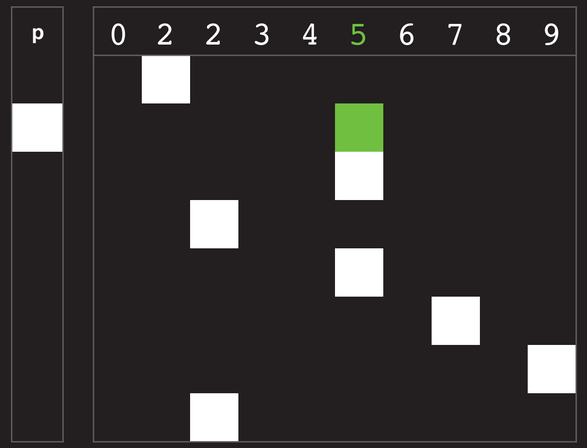
h_t

Zoom in: Pop Operation

Pop

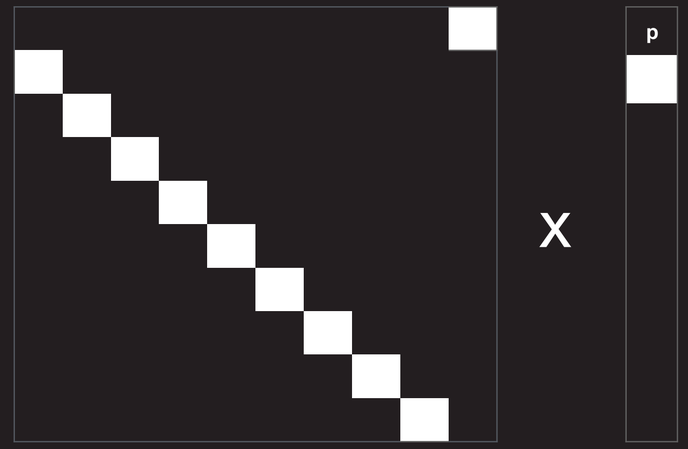


h_{t-1}



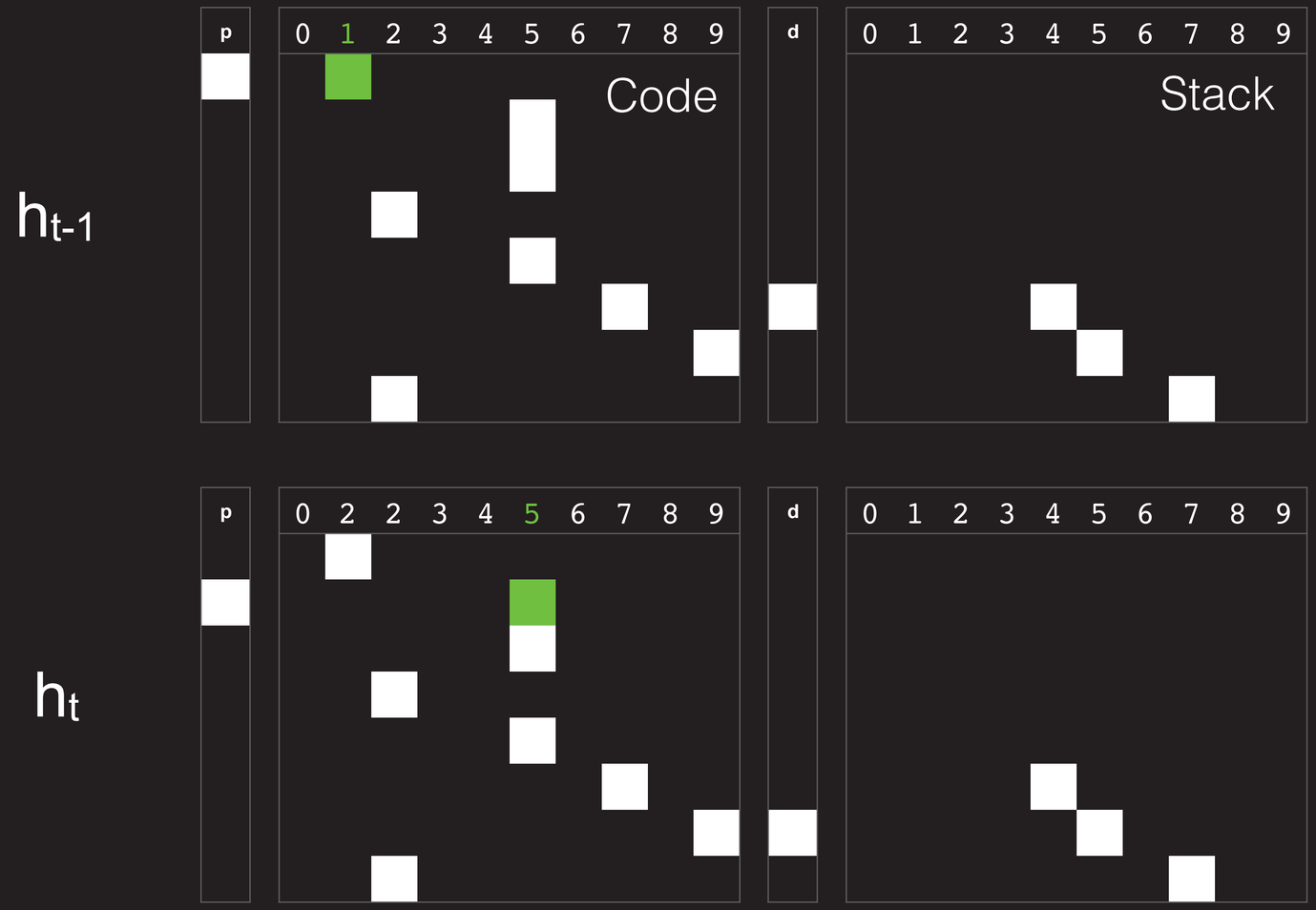
h_t

Circular Shift Matrix

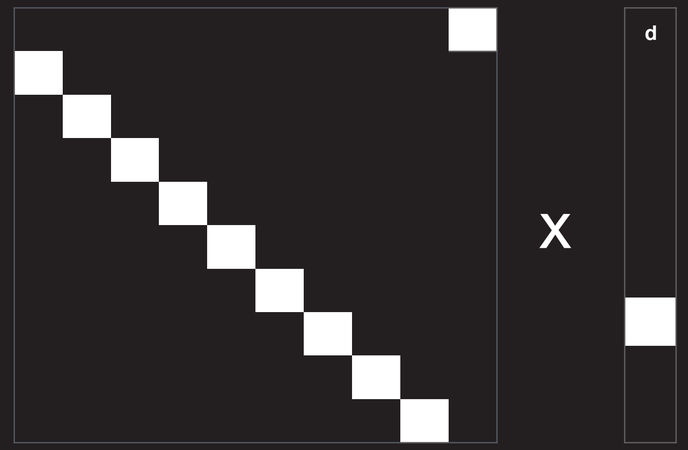


Zoom in: Pop Operation

Pop



Circular Shift Matrix

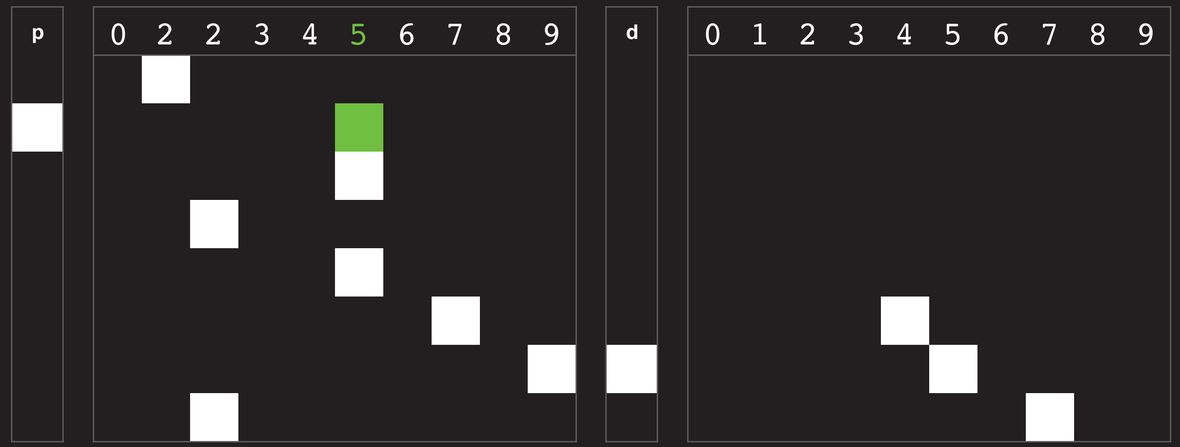


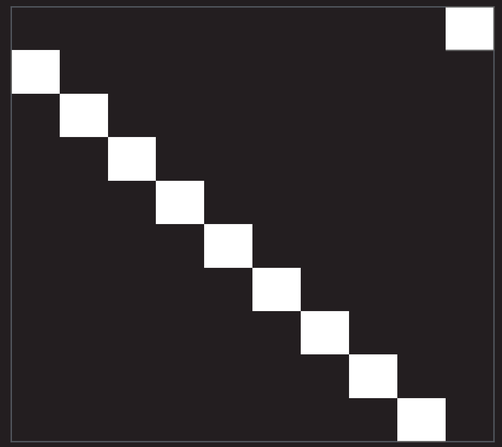
Zoom in: Add 1 Operation

+1

Circular Shift Matrix

h_{t-1}



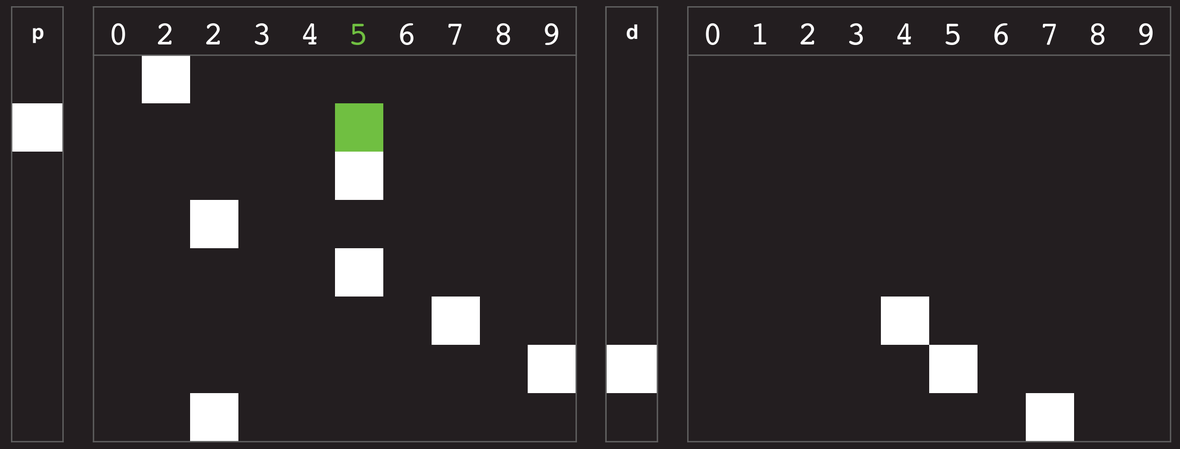


h_t

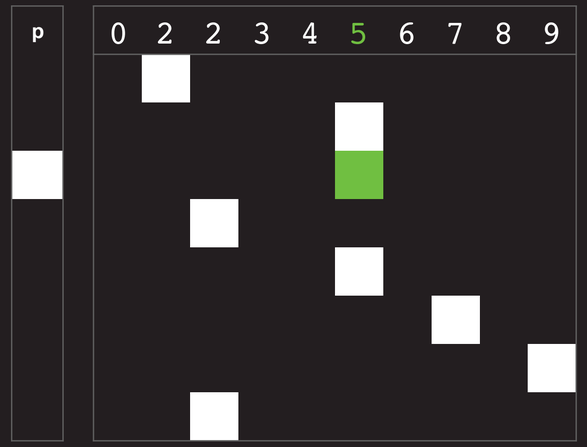
Zoom in: Add 1 Operation

+1

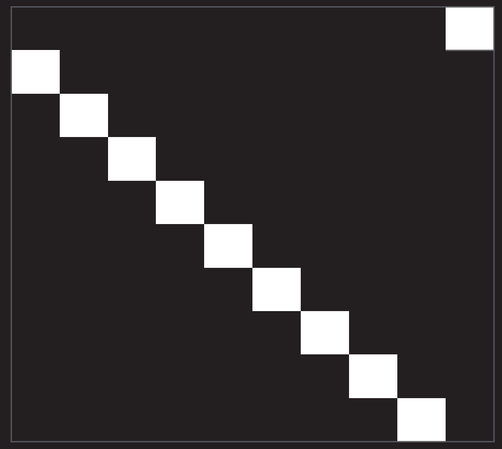
h_{t-1}



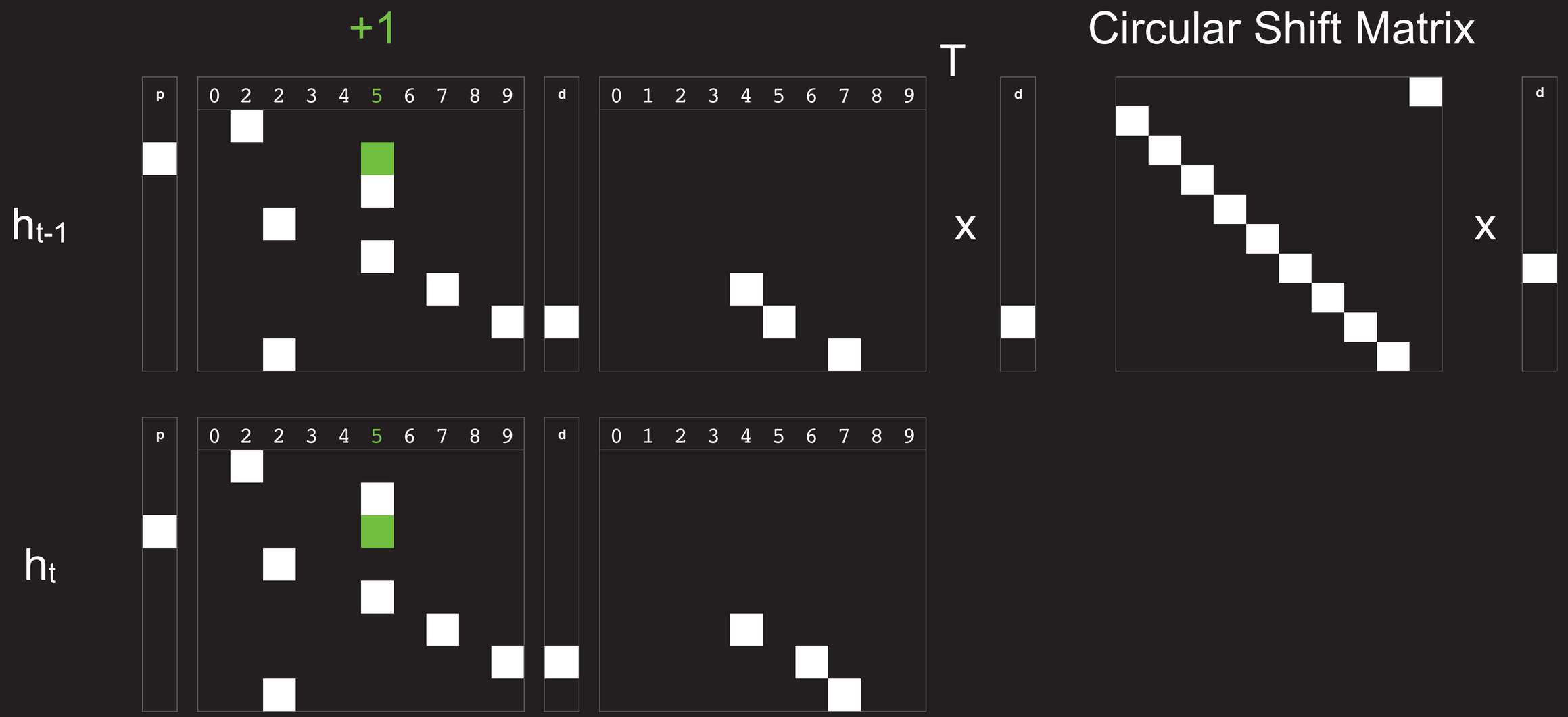
h_t



Circular Shift Matrix



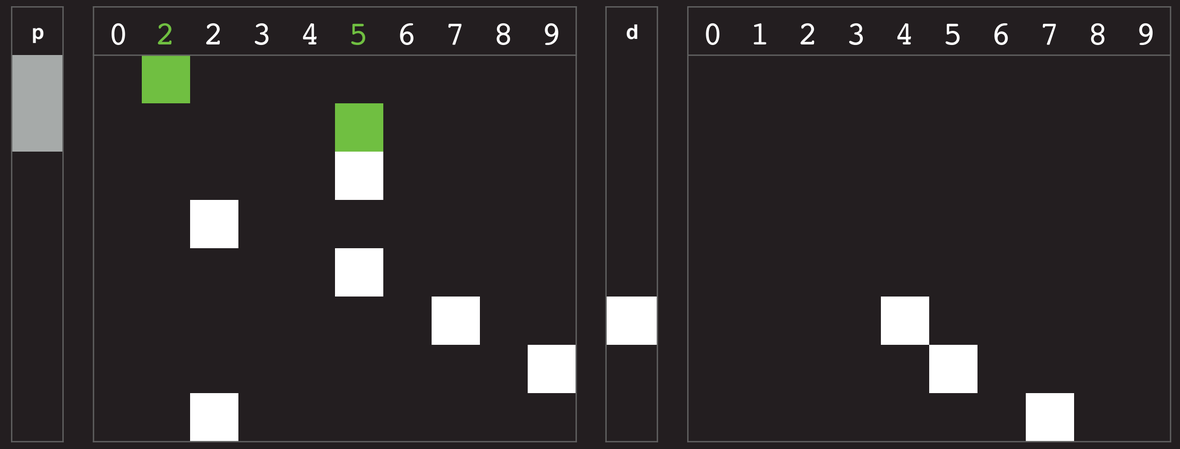
Zoom in: Add 1 Operation



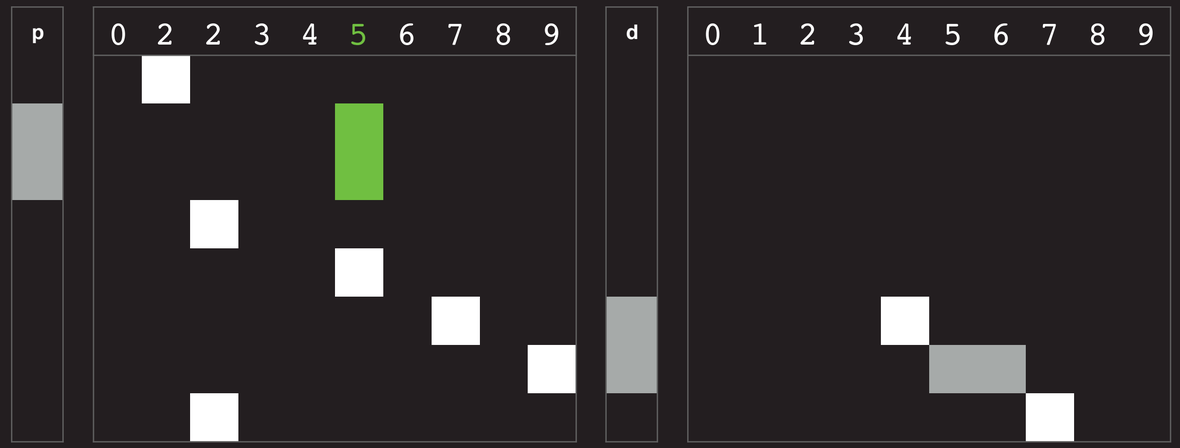
Code Pointer Uncertainty

Pop +1

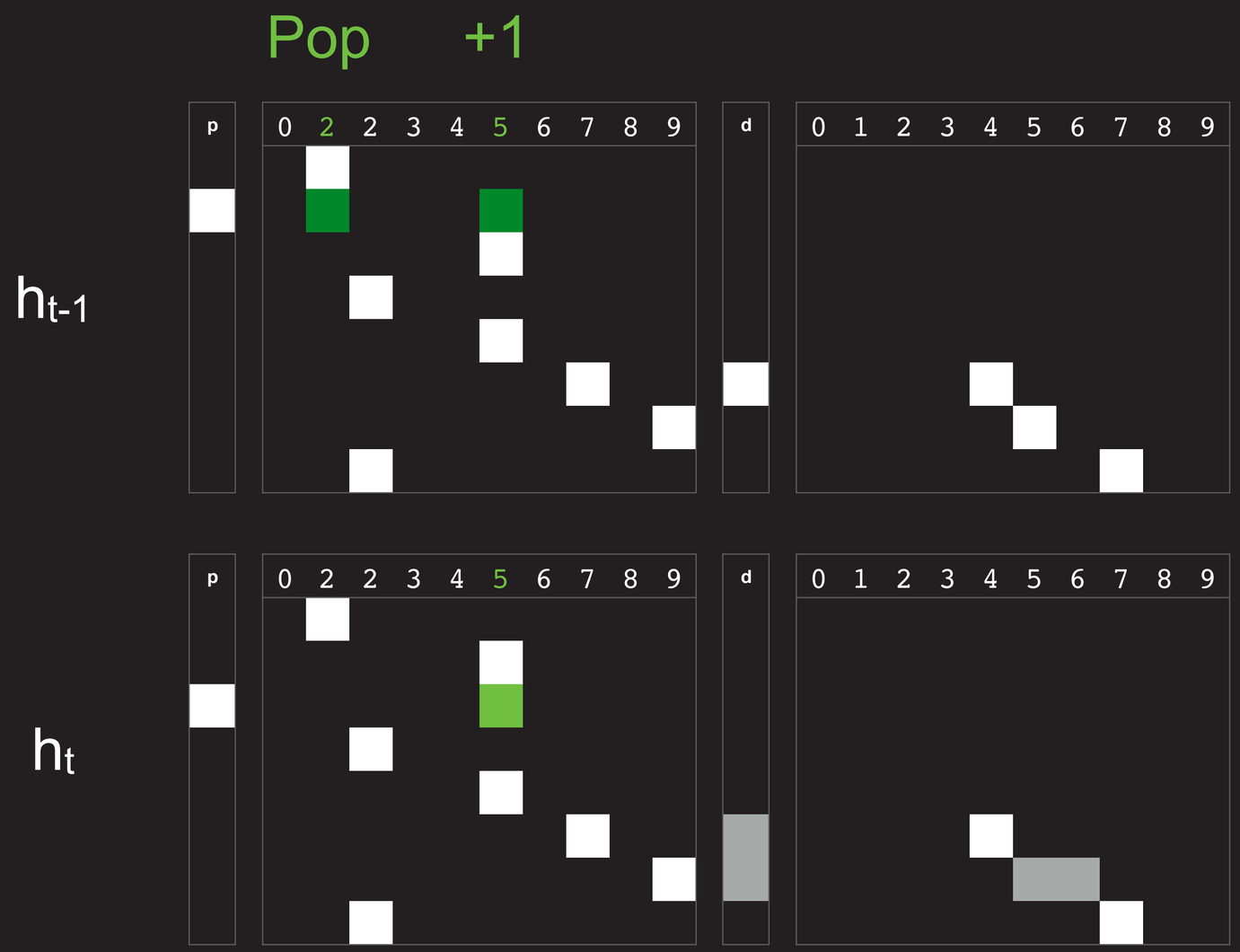
h_{t-1}



h_t



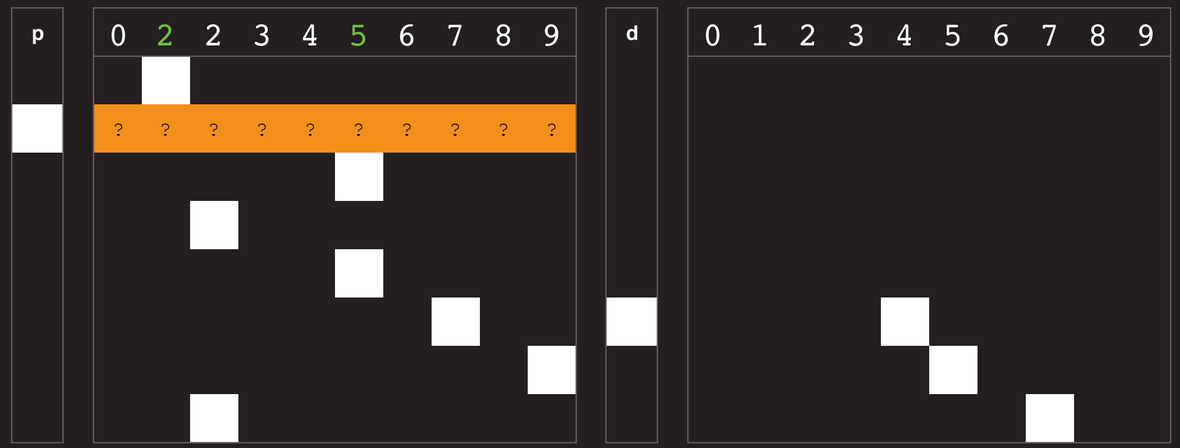
Code Uncertainty



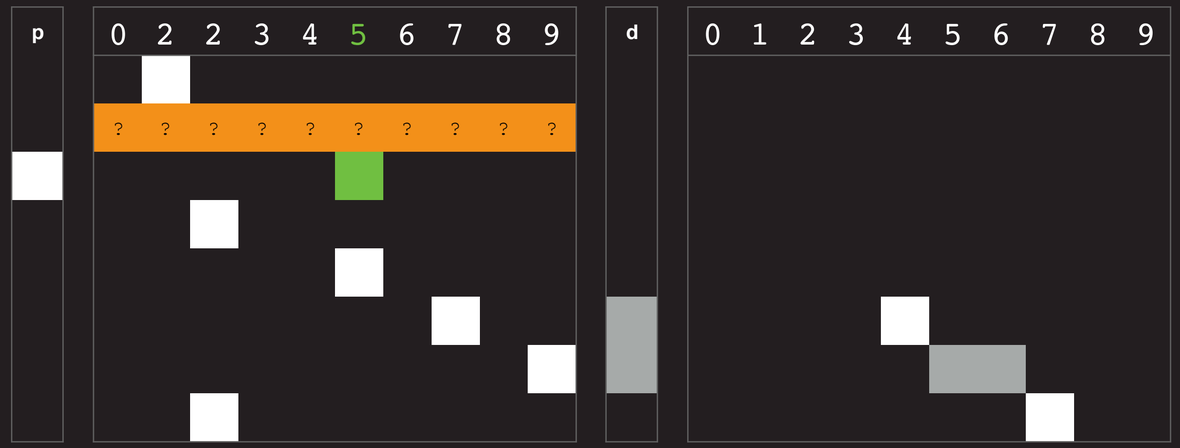
Code Learning

Pop +1

h_{t-1}

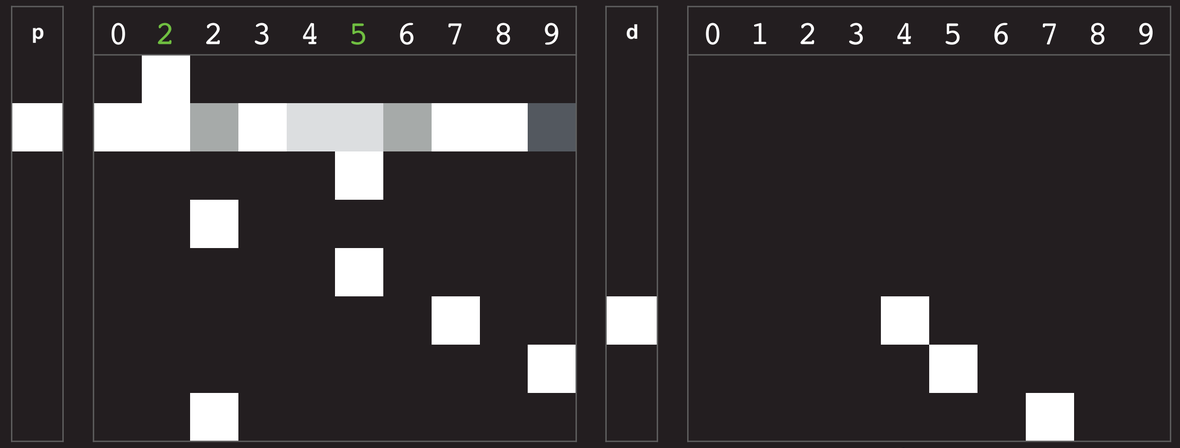


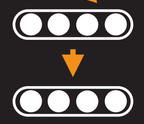
h_t



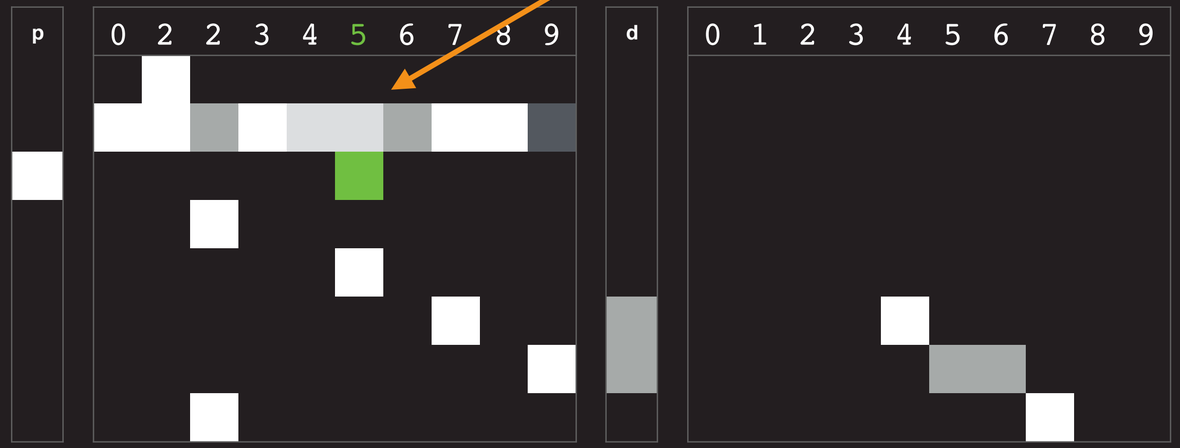
Dynamic Code

h_{t-1}



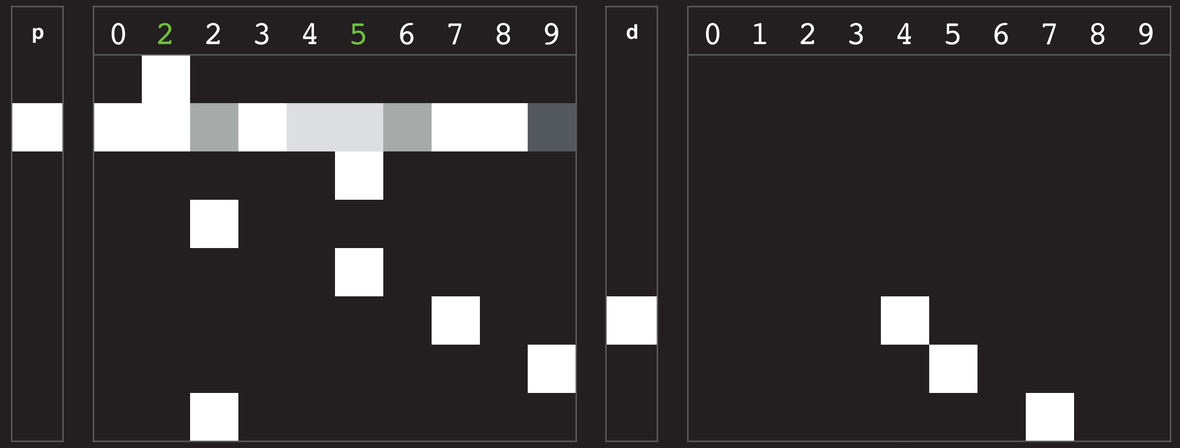


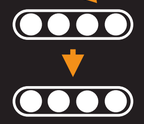
h_t



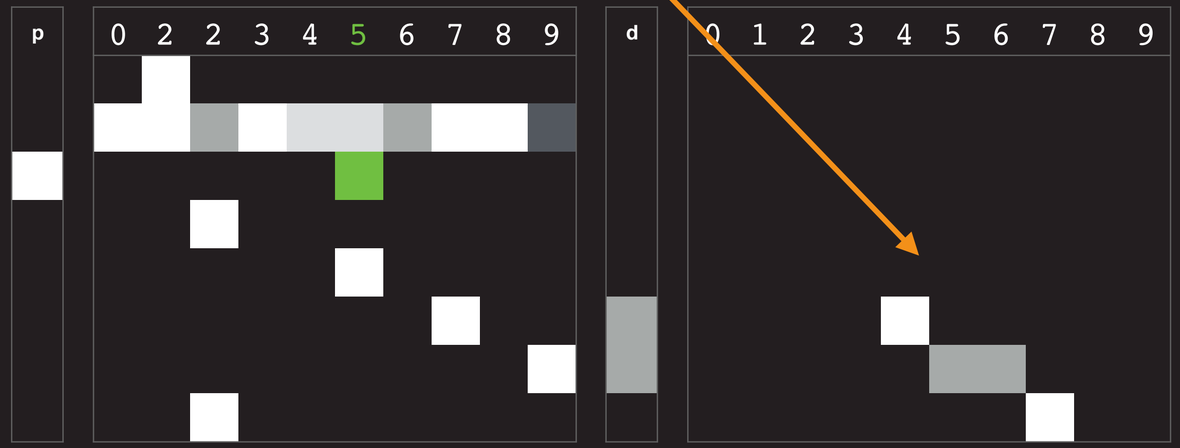
Direct Data Manipulation

h_{t-1}





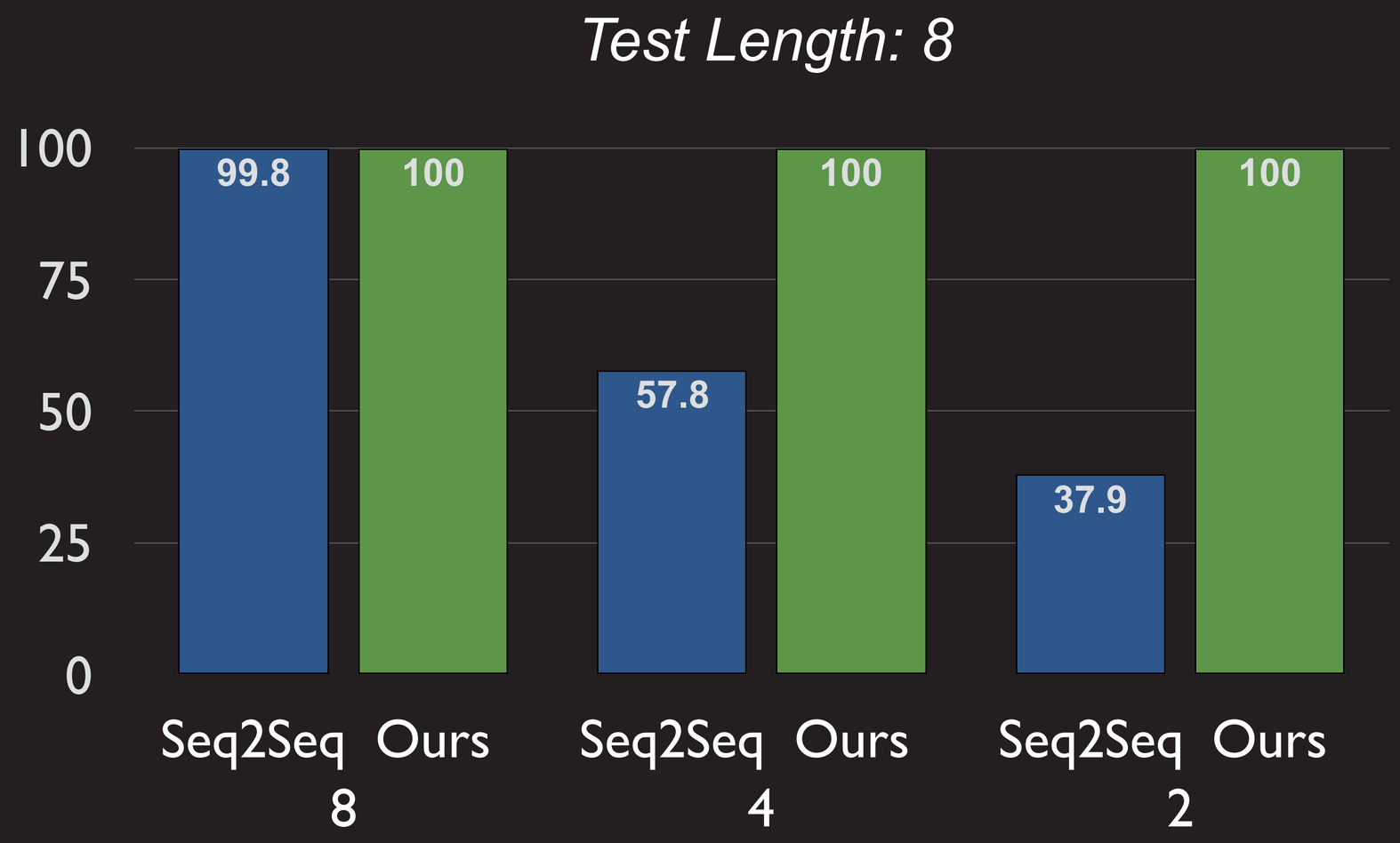
h_t



Related Work

- Program Synthesis (Manna & Waldinger 71, Koza 92, Nordin 97, …)
  - We learn via SGD, allow dynamic code outside the host language
- Probabilistic Programming (Goodman 08, Pfeffer 01, Milch 05, De Raedt 07)
  - We use procedural language, *discriminative*, easy to integrate into end-to-end neural architectures
- Neural Program Induction (Graves et al. 14, Reed & Freitas 15, …)
  - We enable program sketches, host language is proper 3rd generation language

Results on Learning To Sort



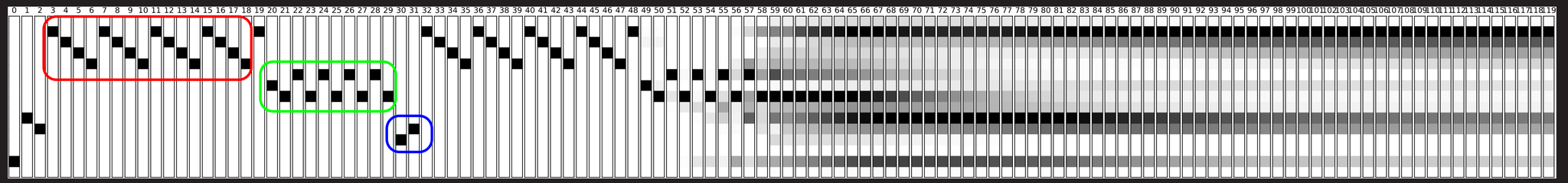

```
def bubble(list):
  if len(list) == 1:
    return list
  else:
    return
    slot1(bubble(slot2(list))

def sort(list):
  for i in range(0,len(list):
     list = bubble(list)
  return list
```

Bosnjak et al. **ICML 2017**

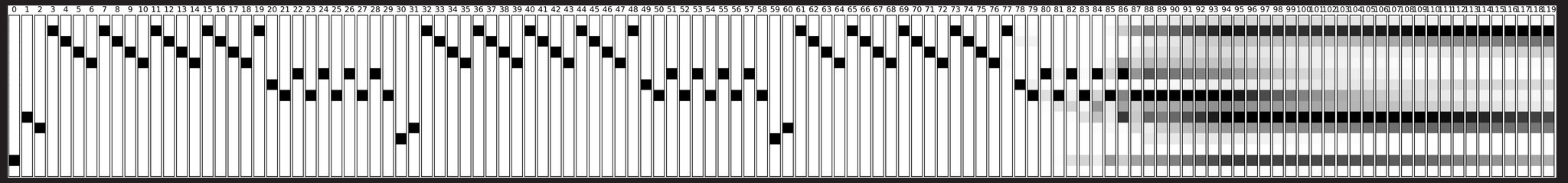
Program Trace During Learning

Early Stage of Training



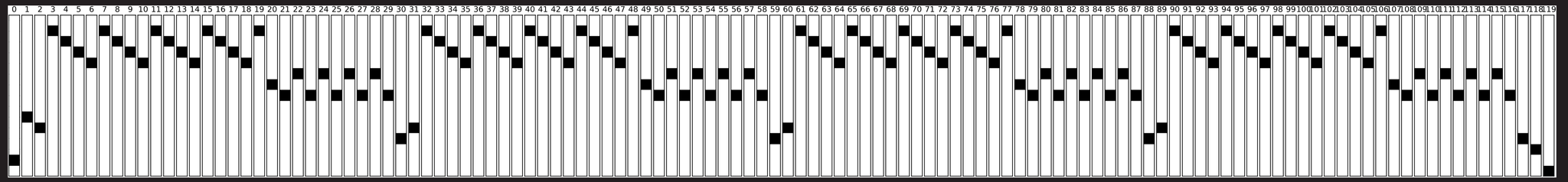
(a) Program Counter trace in early stages of training.

Middle of Training



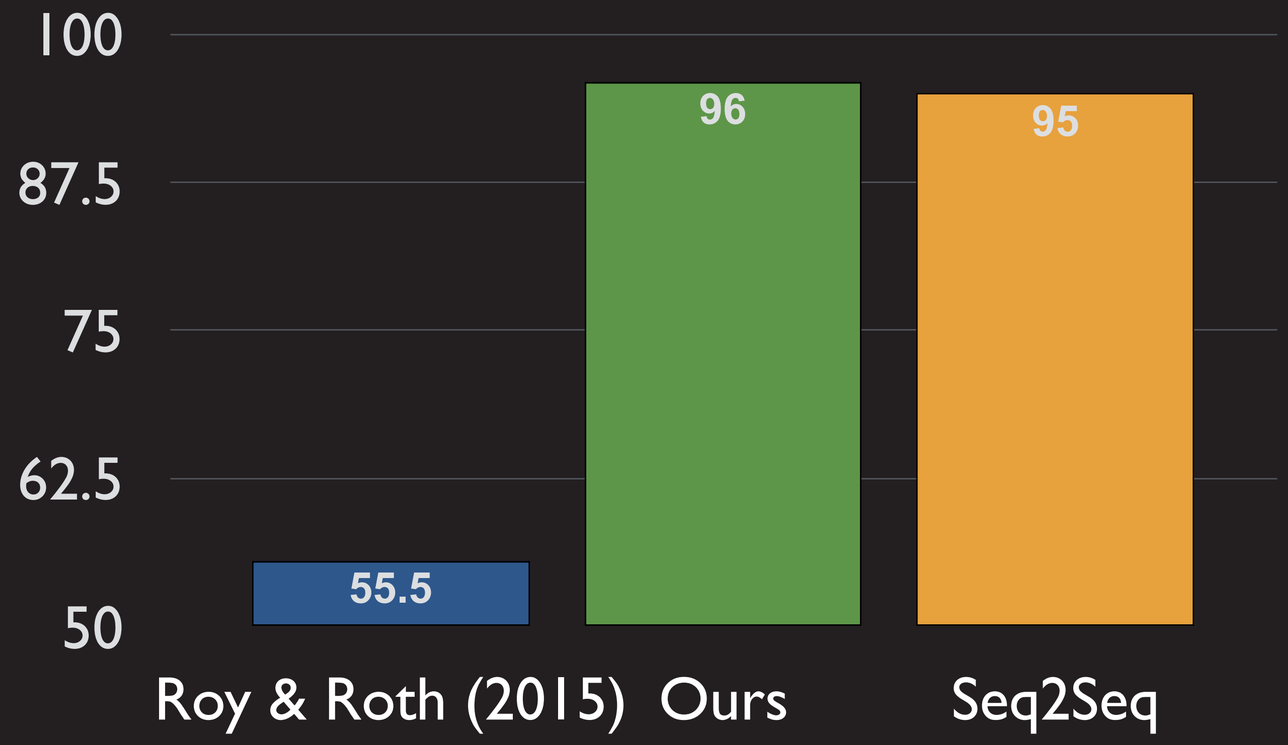
(b) Program Counter trace in the middle of training.

End of training



(c) Program Counter trace at the end of training.

Results on Math Problems (Accuracy)



Seq2Seq solves simpler problem

How many pictures were in each of the albums?

(2 + 4) / 3

Bosnjak et al. **ICML 2017**

Limitations

- Continuous relaxations difficult to train
  - Hard to learn long programs
  - Hard to learn with recursive function calls
- Need better gradients in presence of discrete variables

Tim Rocktäschel

Part 2: Learning to Aggregate

Neural Theorem Provers

Which medical specialty deals with pituitary ACTH hypersecretion?

Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Neural Theorem Provers

Which medical specialty deals with pituitary ACTH hypersecretion?

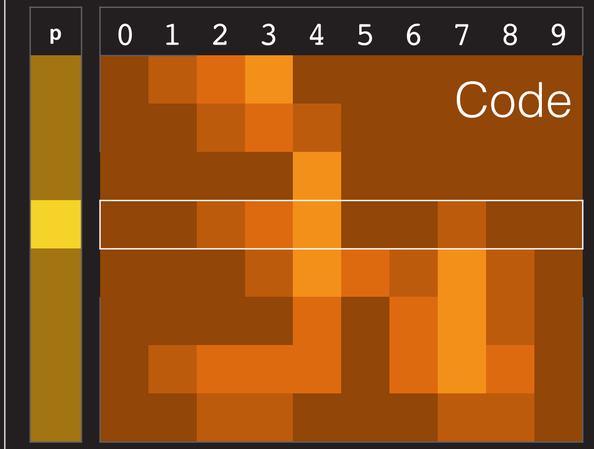
Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Agent

Program Interpreter



Question

Which specialty deals with pituitary ACTH hypersecretion?

Reader

Rocktäschel and Riedel **NIPS 2017**

Neural Theorem Provers

Which medical specialty deals with pituitary ACTH hypersecretion?

Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Agent

Program Interpreter

p | 0 1 2 3 4 5 6 7 8 9

Code

Question

What is ACTH hypersecretion created by?

Reader

Rocktäschel and Riedel **NIPS 2017**

Neural Theorem Provers

Which medical specialty deals with pituitary ACTH hypersecretion?

Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Agent

Program Interpreter

p 0 1 2 3 4 5 6 7 8 9

Code

Question

What is the pituitary a part of?

Reader

Rocktäschel and Riedel **NIPS 2017**

Neural Theorem Provers

Which medical specialty deals with pituitary ACTH hypersecretion?

Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Agent

Program Interpreter

p 0 1 2 3 4 5 6 7 8 9

Code

Question

What topic covers the endocrine system?

Reader

Rocktäschel and Riedel **NIPS 2017**

Vectors Correspond to Interpretable Rules

Which medical specialty deals with pituitary ACTH hypersecretion?

Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Agent

Program Interpreter

p 0 1 2 3 4 5 6 7 8 9

Code

```
X deals with Y if
  Y produced by Z
  Z is a part of U
  X deals with U
```

Question

What topic covers the endocrine system?

Reader

Catch: Currently only Works on Relational Data

dealsWith(**hypersecretion**, X)

createdBy(**hypersecretion, anterior pituitary**)

partOf(**endocrine system**, **anterior pituitary**)

dealsWith(**endocrine system**, **endocrinology**)

Agent

Program Interpreter

p 0 1 2 3 4 5 6 7 8 9

Code

```
X deals with Y if
  Y produced by Z
  Z is a part of U
  X deals with U
```

Question

covers(**endocrine system**, **X**)

Reader

Rocktäschel and Riedel **NIPS 2017**

Catch: Currently only Works on Relational Data

dealsWith(**hypersecretion**, X)

createdBy(**hypersecretion, anterior pituitary**)

partOf(**endocrine system**, **anterior pituitary**)

dealsWith(**endocrine system**, **endocrinology**)

Agent

*Amounts to Differentiable Version of the **Backward Chaining** algorithm used in Prolog*

Supports Soft Unification

dealsWith(**hypersecretion**, X)

createdBy(**hypersecretion, anterior pituitary**)

partOf(**endocrine system**, **anterior pituitary**)

dealsWith(**endocrine system**, **endocrinology**)

Agent

Program Interpreter

p 0 1 2 3 4 5 6 7 8 9

Code

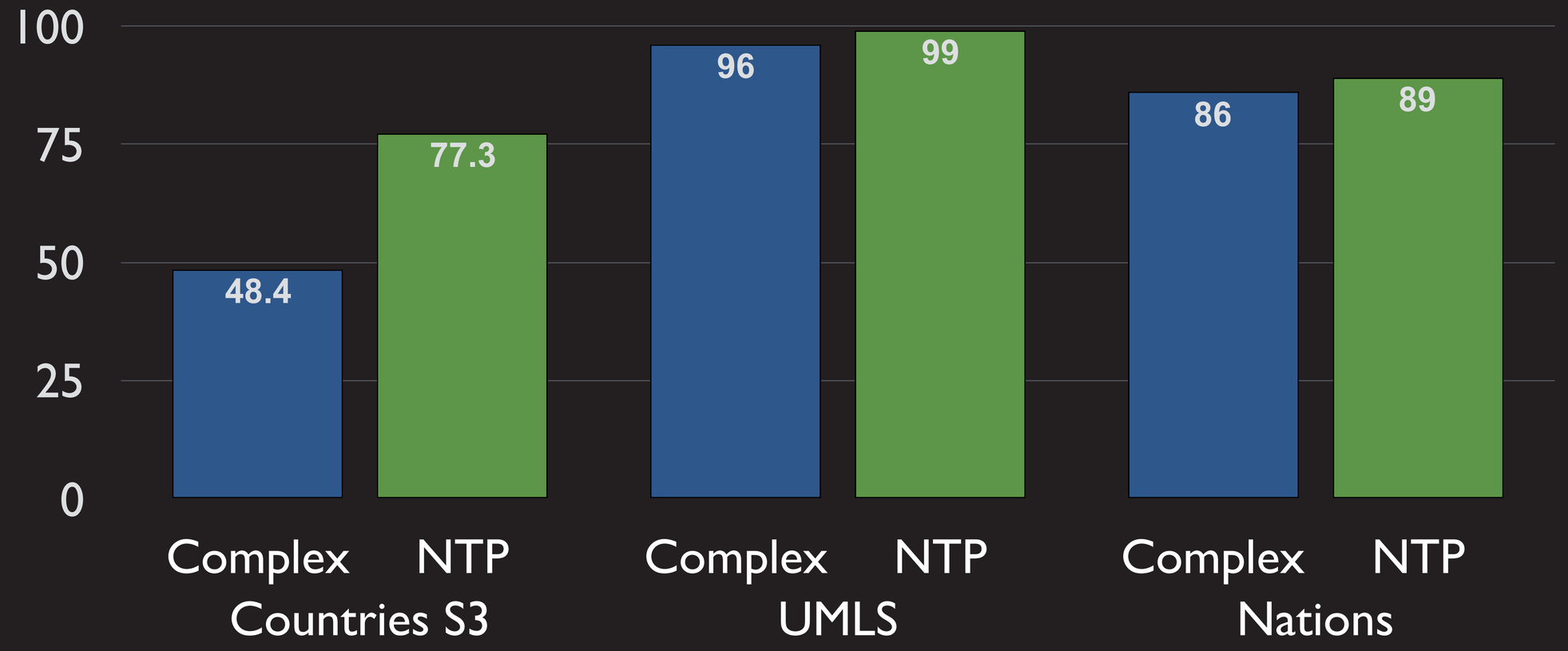
```
X deals with Y if
  Y produced by Z
  Z is a part of U
  X deals with U
```

Question

covers(**endocrine system**, **X**)

Reader

Results on Benchmark (Rank of Correct Answer)



*Comparable or Better than Baselines, and **interpretable***

Rocktäschel and Riedel **NIPS 2017**

Interpretability: Learnt Rules

- if X is located in Y and Y is located in Z then X is located in Z
- if X expels diplomats of Y then X shows negative behaviour towards Y
- if X interacts with Y and Y interacts with Z then X interacts with Z

Related Work

- **Probabilistic Logic Programming**: IBAL (Pfeffer, 2001), BLOG (Milch et al., 2005), Markov Logic Networks (Richardson and Domingos, 2006), ProbLog (De Raedt et al., 2007), BLP (Kersting, De Raedt, 2007)
- **Inductive Logic Programming**: Plotkin (1970), Shapiro (1991), Muggleton (1991), De Raedt (1999) . . . Statistical Predicate Invention (Kok and Domingos, 2007)
- **Neural-symbolic Connectionism**
  - *Propositional* rules: EBL-ANN (Shavlik and Towell, 1989), KBANN (Towell and Shavlik, 1994), C-LIP (Garcez and Zaverucha, 1999)
  - First-order inference (*no training of symbol representations*): Unification Neural Networks (Holldobler, 1990; Komendantskaya 2011), SHRUTI (Shastri, 1992), Neural Prolog (Ding, 1995), CLIP++ (Franca et al. 2014), Lifted Relational Networks (Sourek et al. 2015), TensorLog (Cohen 46)

Limitations

- Scalability
  - Currently only works for KBs with < 10k facts
  - Small proof depth
- Still requires relational representation

Johannes Welbl

Pontus Stenetorp

Part 3: A Read & Reason Dataset


QAngaroo
Reading Comprehension with Multiple Hops
Paper
Dataset Download
Dataset Distribution under CC BY-SA 3.0

A Single Instance

What is the nationality of Jamie Burnett?

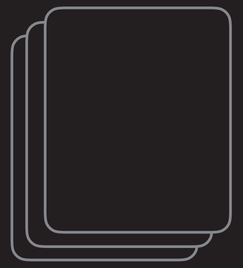
- Candidates:
  - Scotland (correct)
  - China (incorrect)
  - …

Jamie Burnett (born 16 September 1975) is a professional snooker player from Hamilton, South Lanarkshire…He began the 2014/2015 season with a quarter-final showing at the Yixing Open…

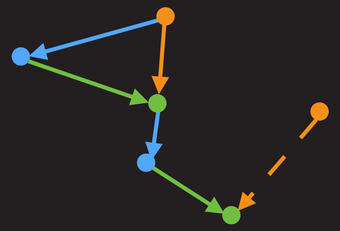
…Hamilton is a town in South Lanarkshire, in the central Lowlands of Scotland…

The Yixing Open was a professional minor-ranking snooker tournament that took place at the Yixing Sports Centre in Yixing, China.

Dataset Construction Method



Unlabelled Text



Knowledge Base

Dataset Construction Method

Multihop Dataset

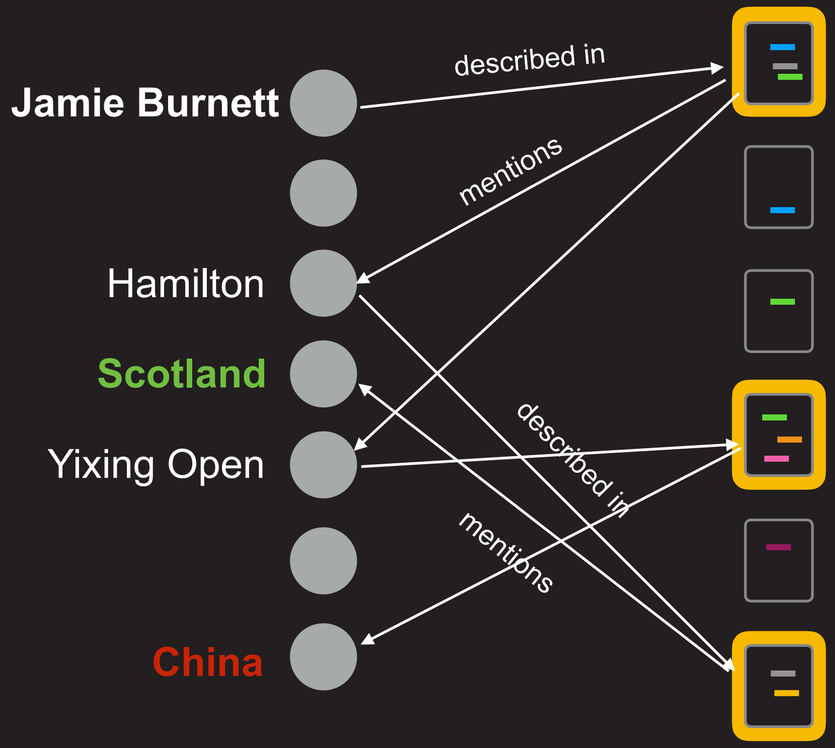
[Conditionally Accepted to TACL]

Dataset Construction Method

KB Triple

Jamie Burnett, citizenship, Scotland

Entities **Documents**



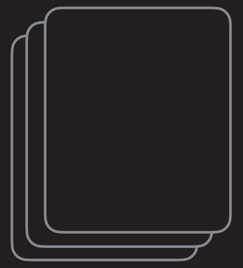
Instance

What is the nationality of Jamie Burnett?

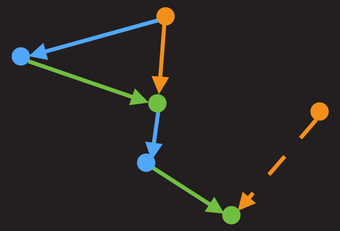
Scotland

China

Dataset Construction Method



Unlabelled Text

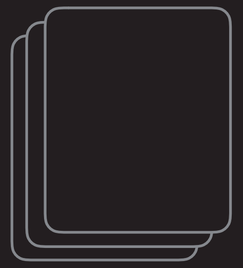


Knowledge Base

Dataset Construction Method

Multihop Dataset

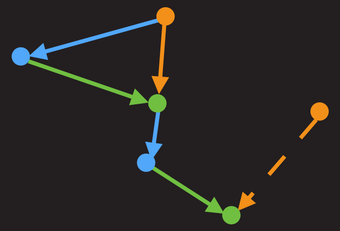
Dataset Construction Method



Unlabelled Text

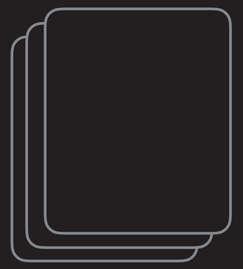
Dataset Construction Method

WikiHop



Knowledge Base

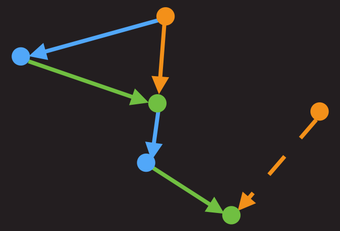
Dataset Construction Method



Unlabelled Text

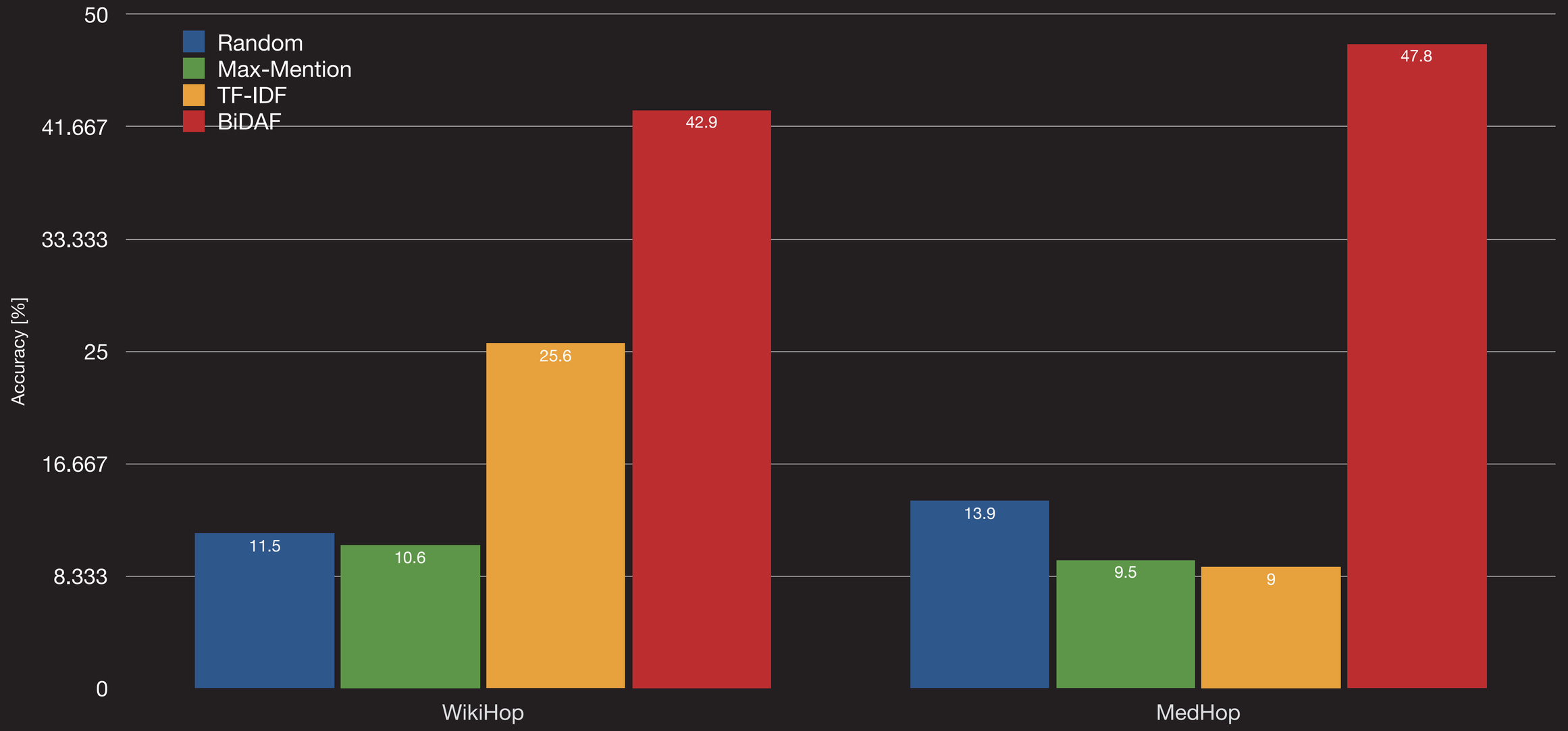
Dataset Construction Method

MedHop



Knowledge Base

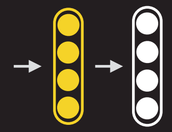
Baseline Results



Reduction to Traditional Machine Comprehension

Which medical specialty deals with pituitary ACTH hypersecretion?

Endocrinology



Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

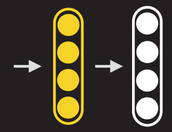
A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Reduction to Traditional Machine Comprehension

Which medical specialty deals with pituitary ACTH hypersecretion?

Endocrinology



Pituitary ACTH hypersecretion ... is a form of hyperpituitarism characterized by an abnormally high level of ACTH produced by the **anterior pituitary** …

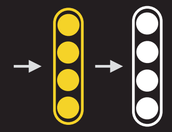
A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Do Neural Reading Models Aggregate?

Which medical specialty deals with pituitary ACTH hypersecretion?

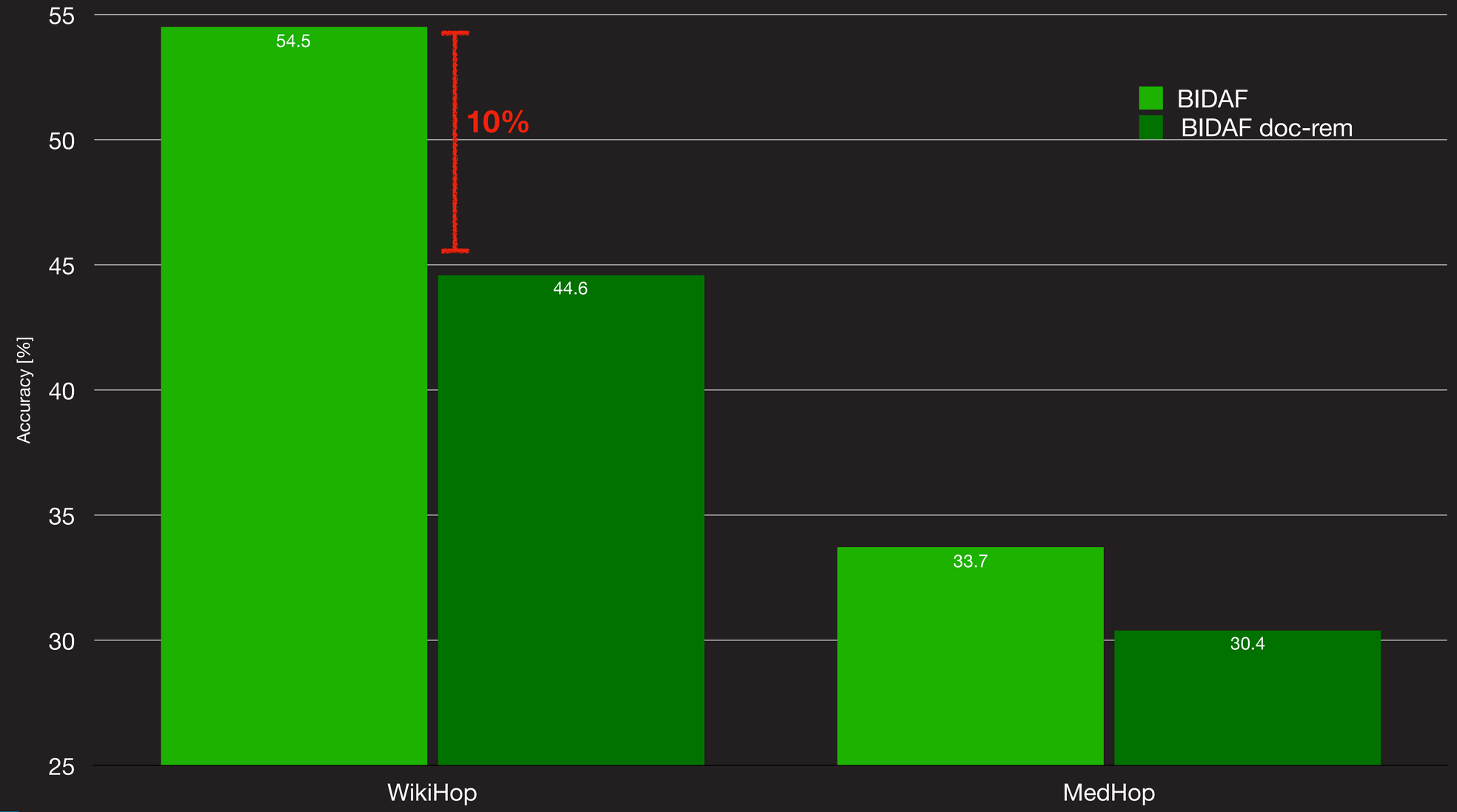
Endocrinology



A major organ of the **endocrine system**, the **anterior pituitary** is the glandular, anterior lobe that ...

The endocrine system is ... ... The field of study dealing with the **endocrine system** and its disorders is **endocrinology**, a branch of internal medicine.

Removing Relevant Documents, Keep Answer Documents



Conclusion

- Great Progress in End-to-End Reading Comprehension
- **Reasoning** (aggregation, calculation etc.) end-to-end is still very challenging
- Our Approaches
  - create datasets
  - cast reasoning as **program learning** and **execution**
  - are **end-to-end differentiable** (can be trained on downstream loss)
  - are inspired and tied to **traditional symbolic formalisms** (Forth, Prolog/Datalog)
  - are learnt models are **interpretable**
  - allow injection of **prior knowledge**