Stanford

Reasoning with Implicatives

Modular and Compositional Learning for Language Understanding

NALOMA Workshop – Brandeis University

Ignacio **Cases** and Lauri **Karttunen**

Center for the Study of Language and Information (CSLI)
Department of Linguistics – Stanford NLP Group
Stanford University

Stanford

Main collaborators

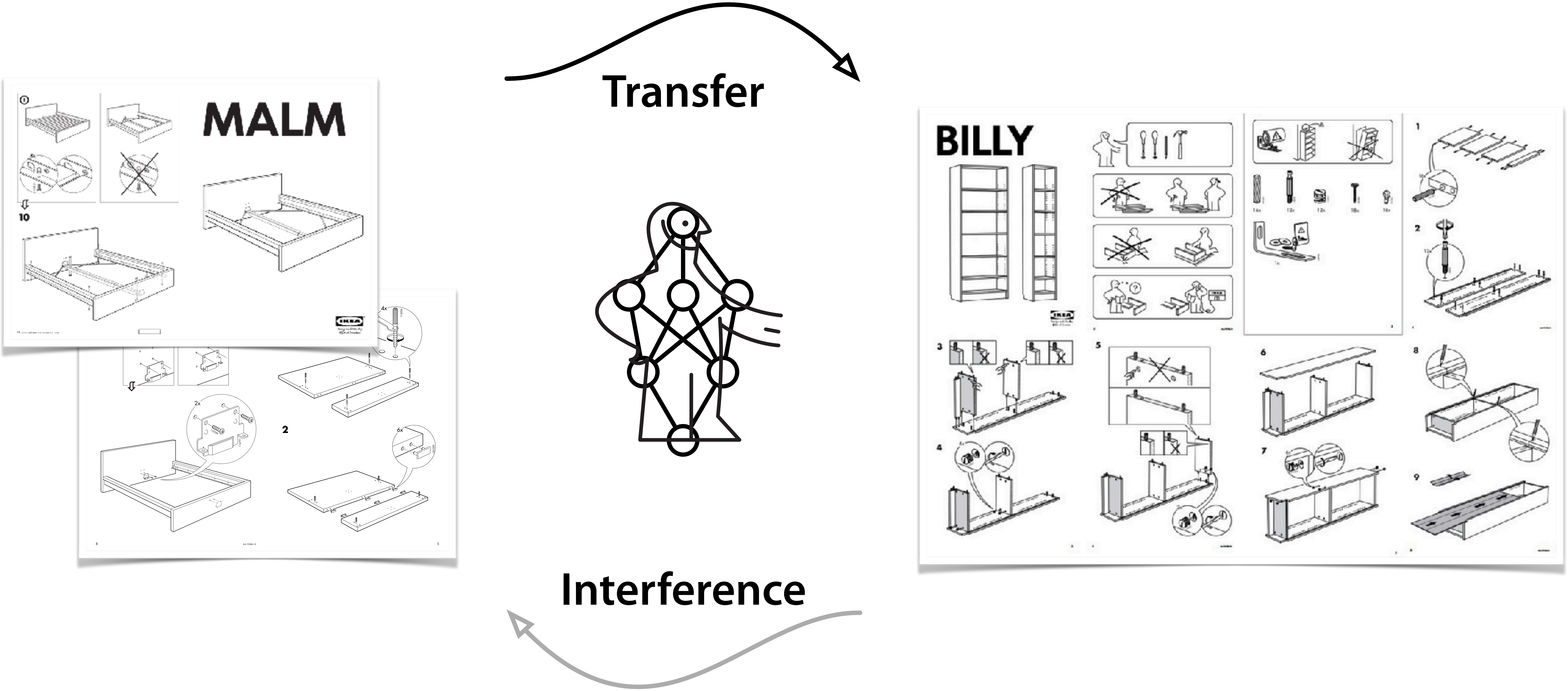
Clemens Rosenbaum form. UMass – Matt Riemer IBM Research

Dan Jurafsky – Chris Potts Stanford

Josh Greene Harvard

Modular and Compositional Learning

Introduction



Interference in Neural Networks is Catastrophic

Introduction

A–B list (List 1)

```
regal     – dax
crazy     – jex
fruitless – qob
entire    – mef
twofold   – vuq
    ...
```

A–C list (List 2)

```
regal     – feh
crazy     – jic
fruitless – xuy
entire    – gyq
twofold   – sij
    ...
```

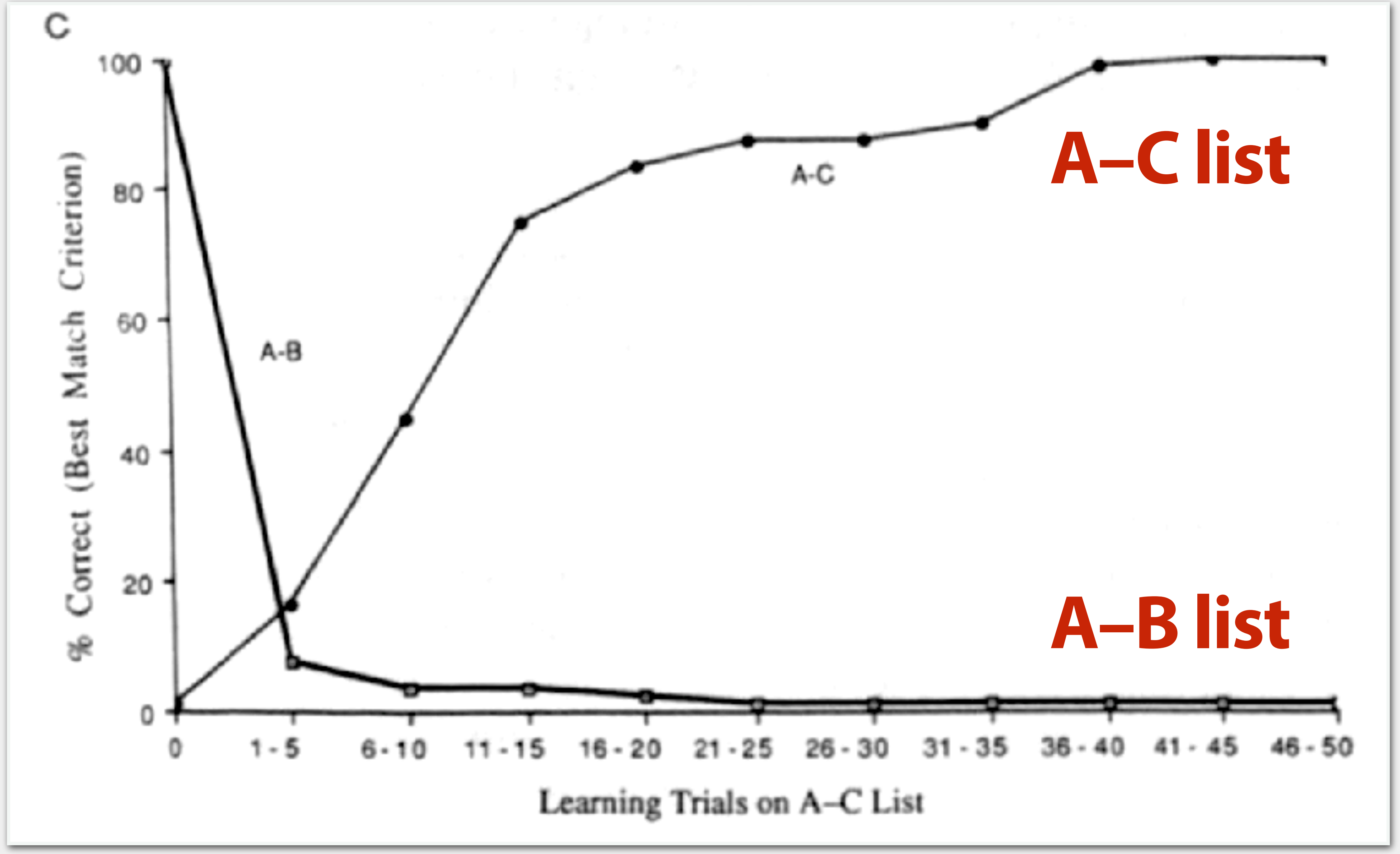
Test on A-B list

```
Q: dax? A: regal
Q: jex? A: crazy
    ...
```

Test on A-C list

```
Q: feh? A: regal
A: jic? A: crazy
    ...
```

In **neural networks**, performance on A–B decreases **catastrophically** when learning A–C



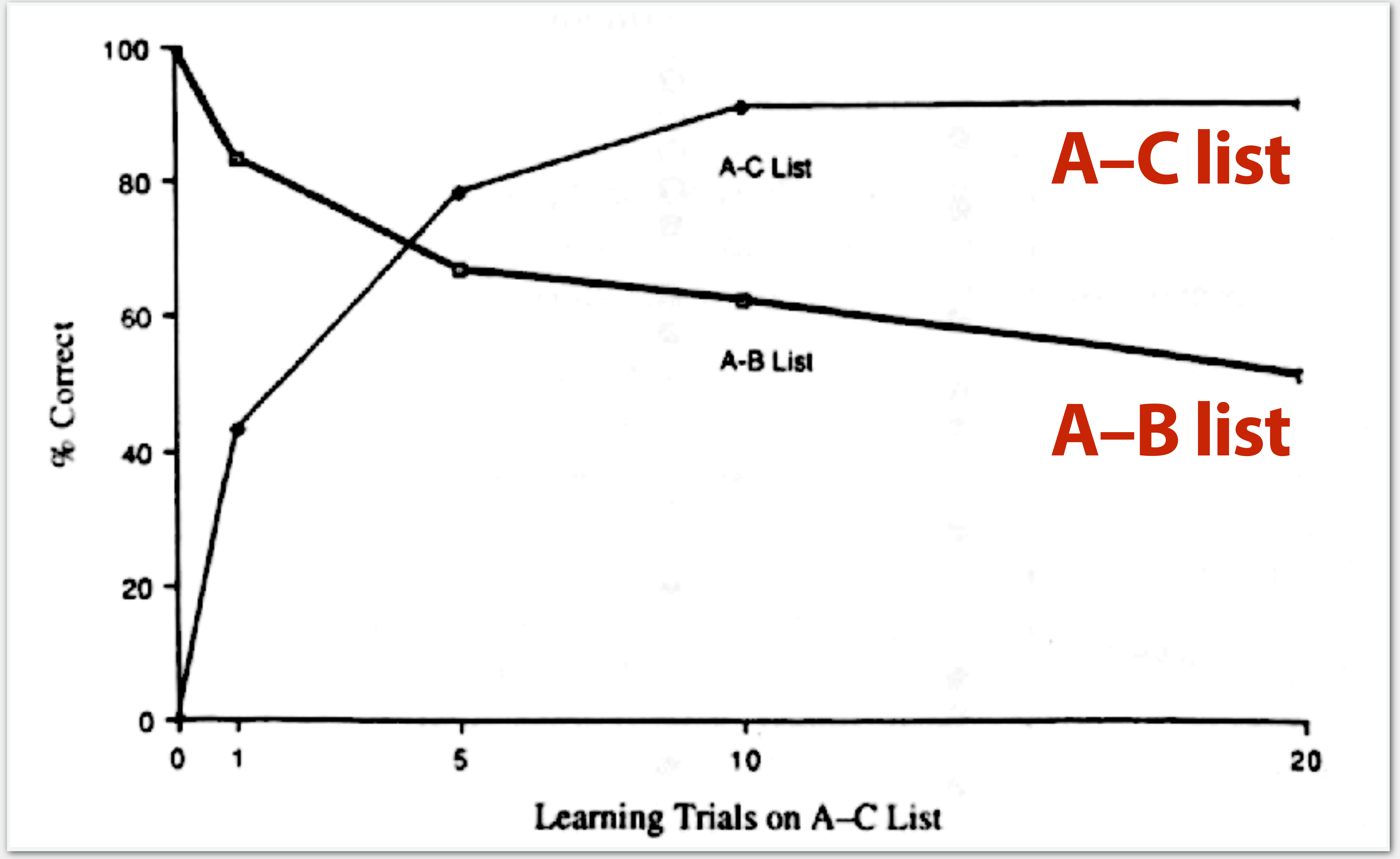
McCloskey and Cohen 1989

Interference in Humans is Gradual

Introduction

In **humans**, performance on A–B decreases **gradually** when learning A–C

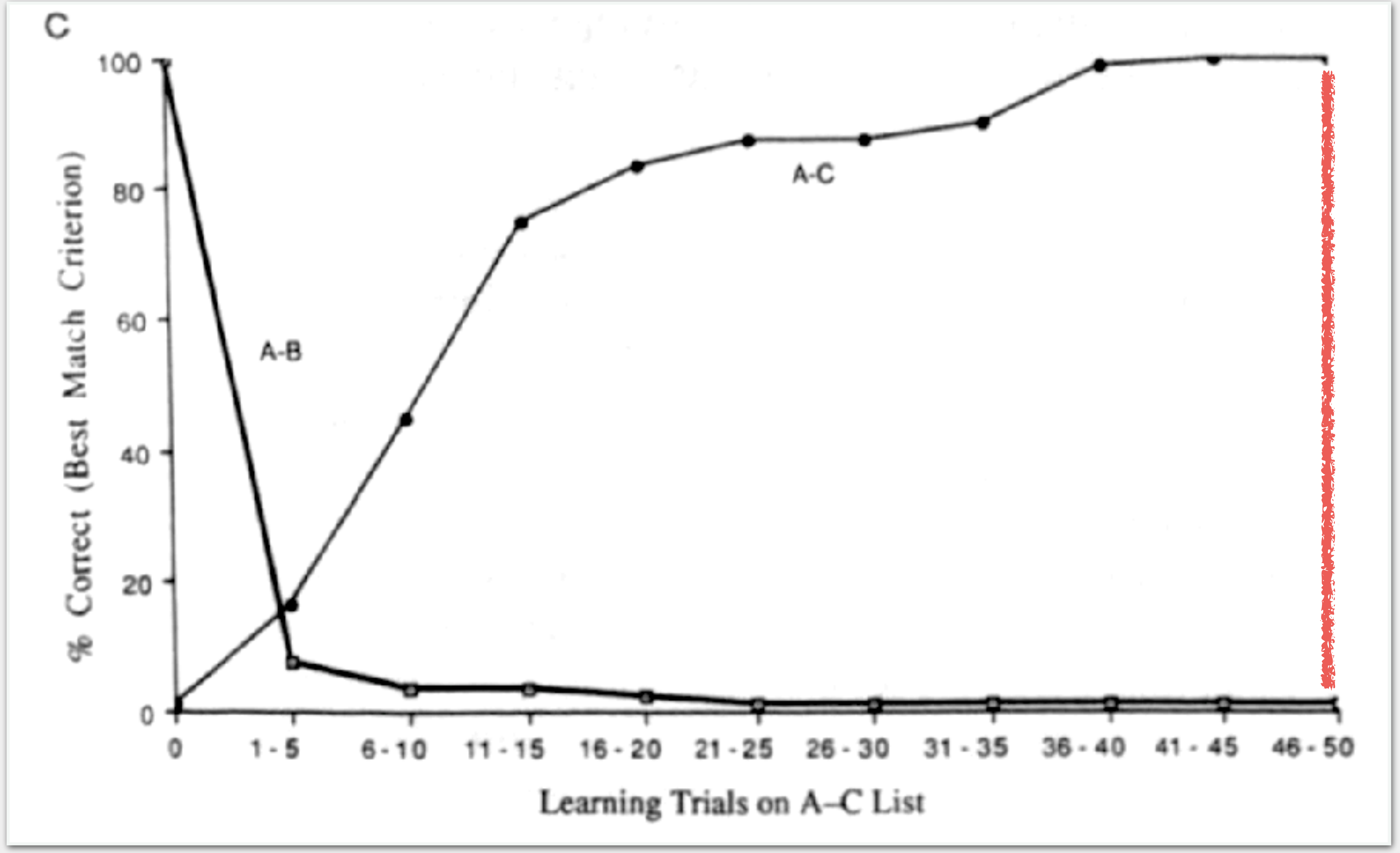
"The results showed that during the learning of List 2 there is a **gradual decrease** in frequency of List 1 responses, the curve being not unlike an extinction curve." (Barnes and Underwood 1959)

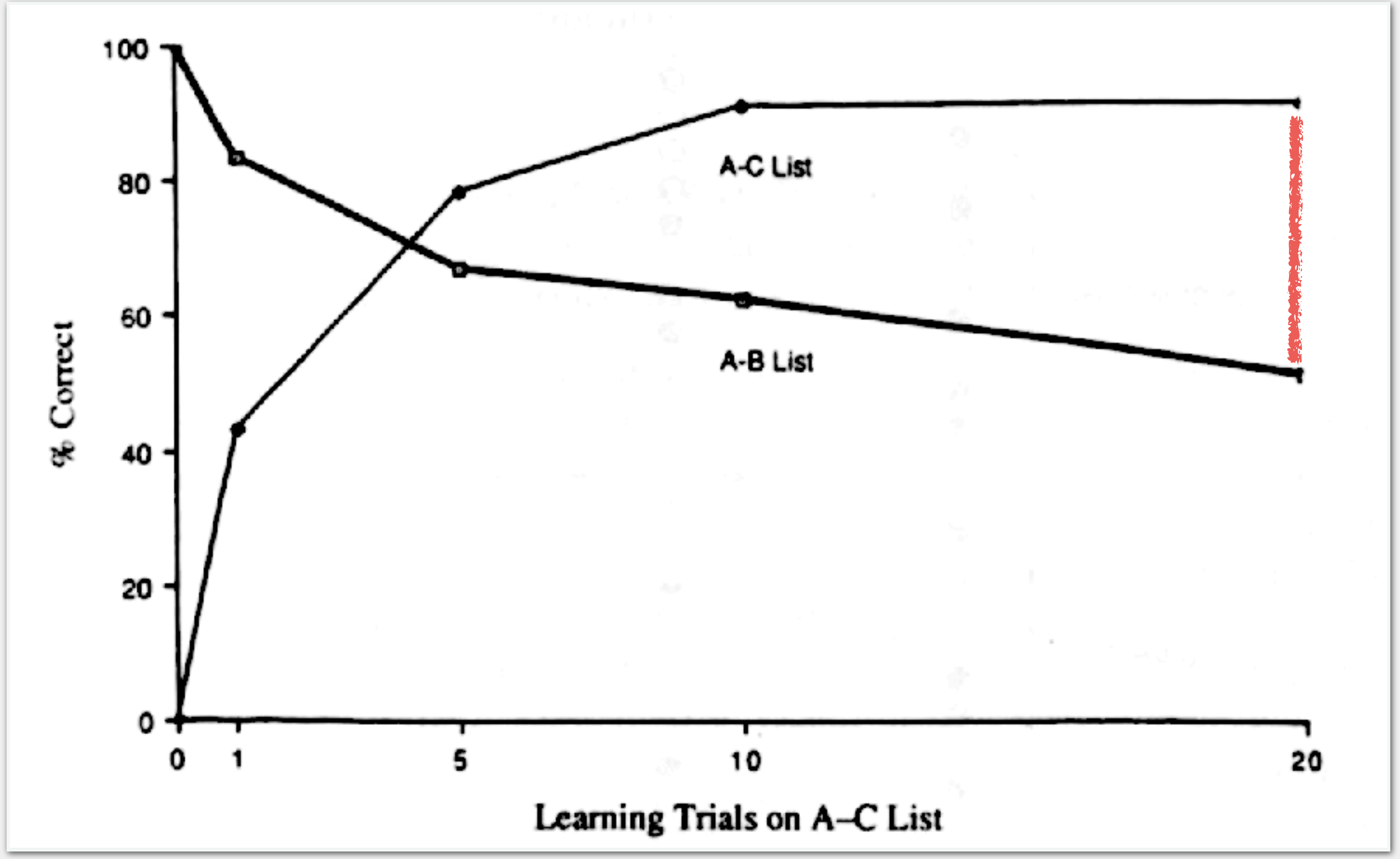


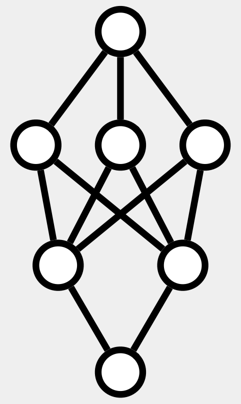
McCloskey and Cohen 1989

Some Degree of Interference is not a *Problem*

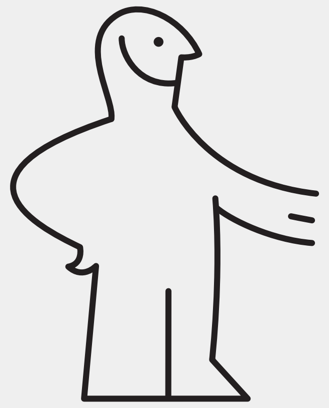
Introduction







Neural Networks



Humans

McCloskey and Cohen 1989

Interference is a Significant Problem and Inevitable

Introduction

"To the extent that one is interested in using connectionist networks to model human learning and memory, this sort of disruption would appear to be a **significant problem**."

(McCloskey and Cohen 1989: 123)

"To put it somewhat flippantly, the magnitude of the observed interference makes it seem more like **retrograde amnesia** than retroactive interference."

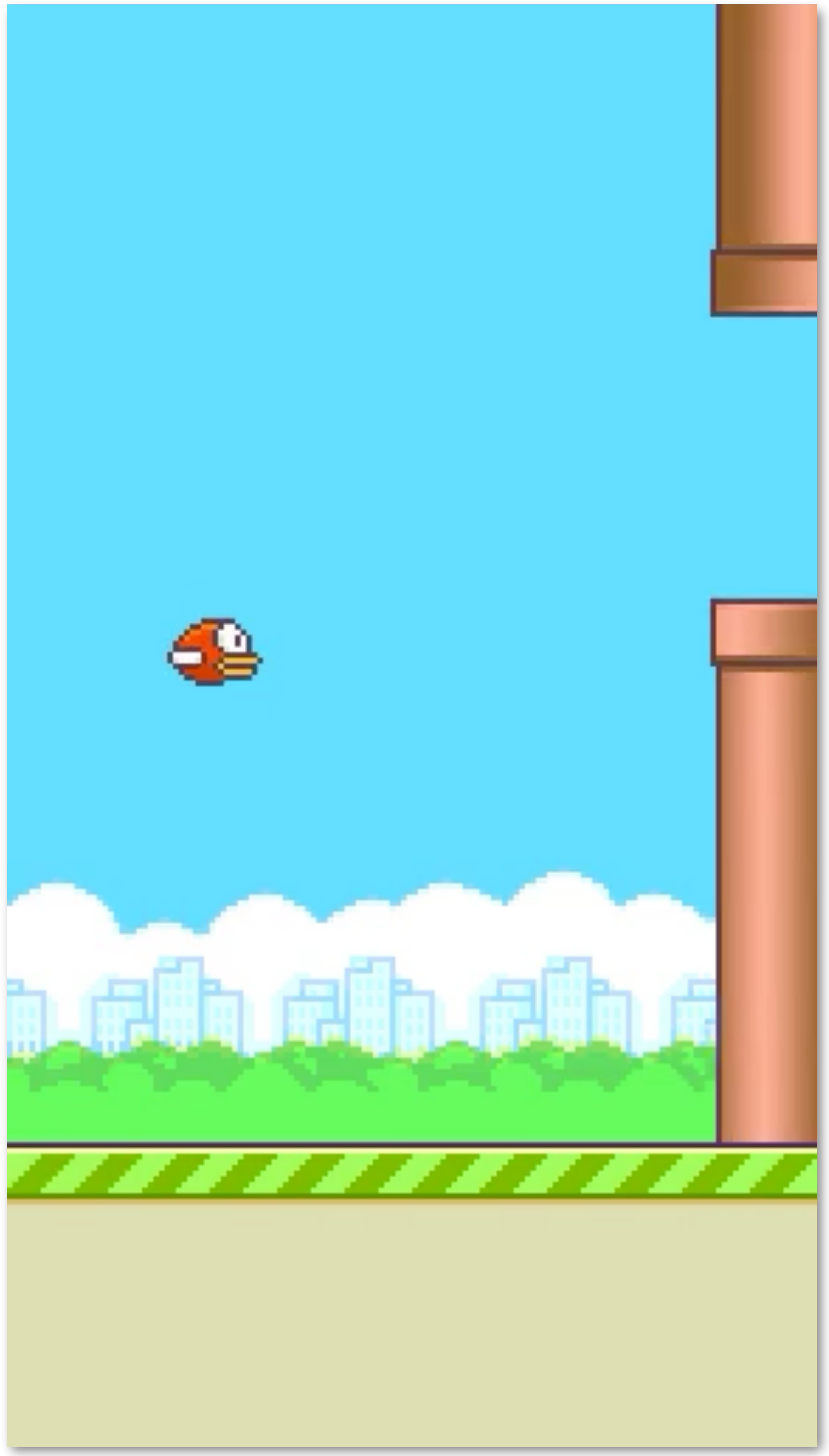
(McCloskey and Cohen 1989: 123)

"[I]t has been widely thought that catastrophic forgetting is an **inevitable feature** of connectionist models."

(Kirckpatrick et al 2017: 3521)

Interference in Continual Learning

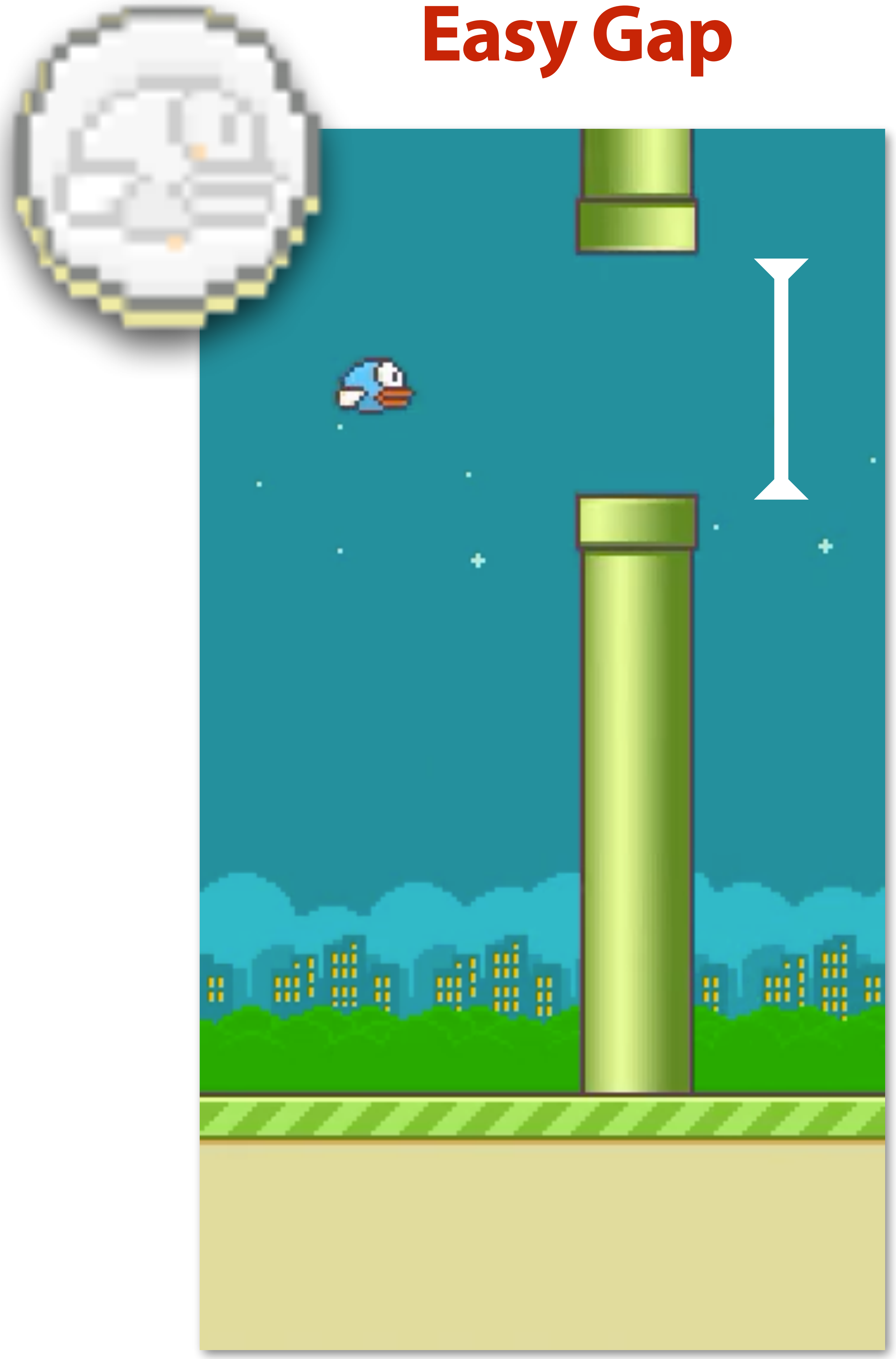
Introduction



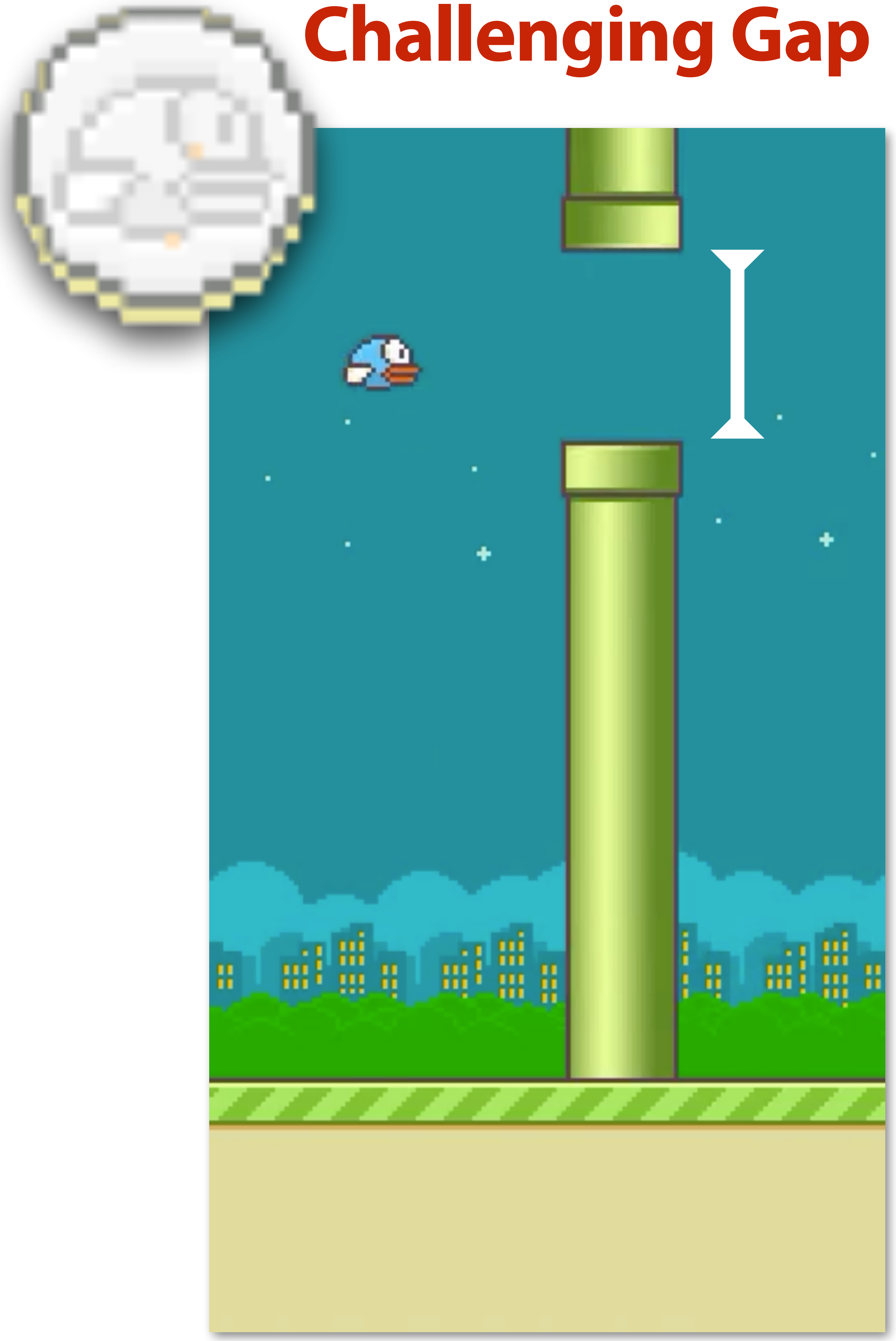
Interference in Continual Learning

Introduction

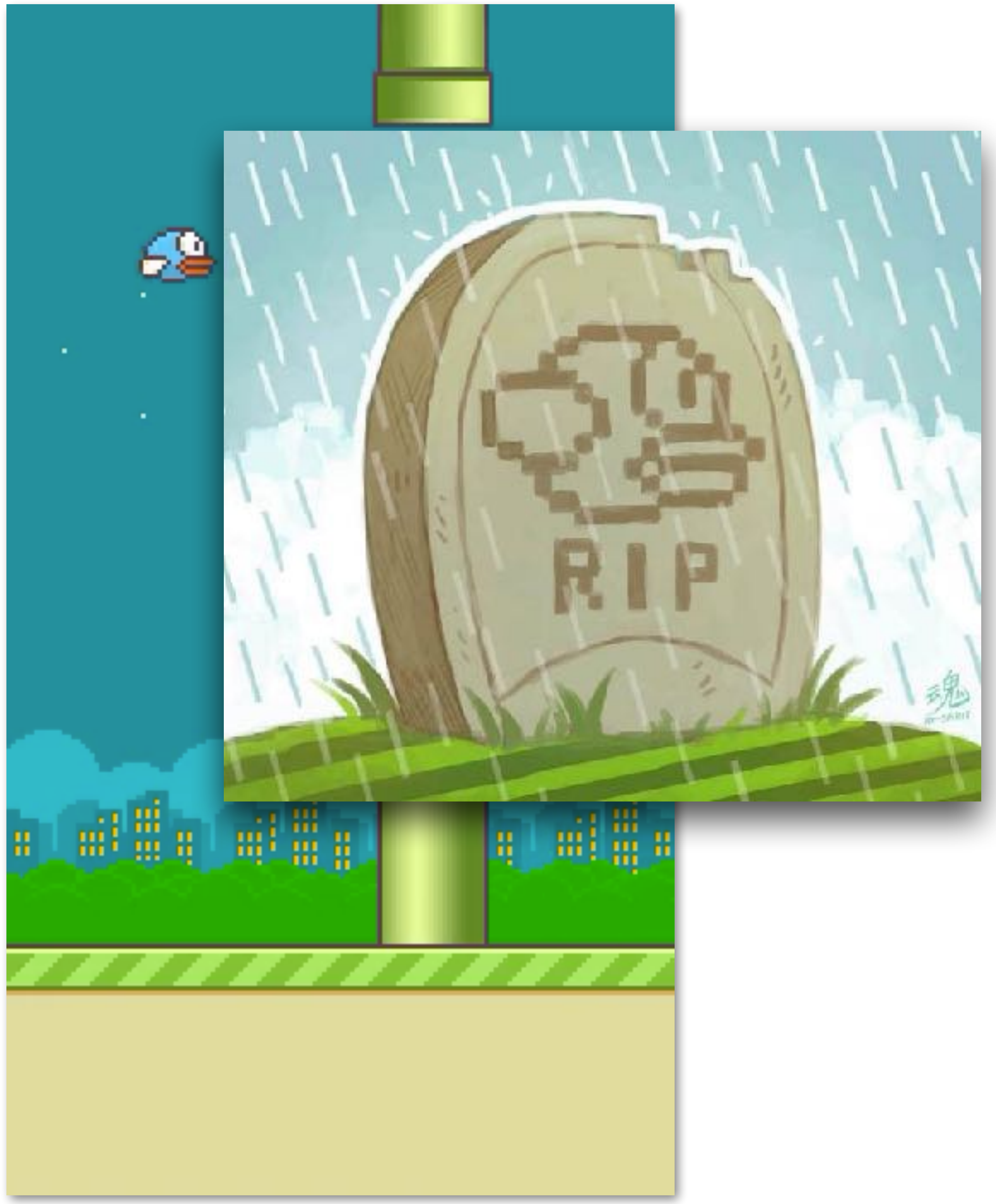
Easy Gap



Challenging Gap

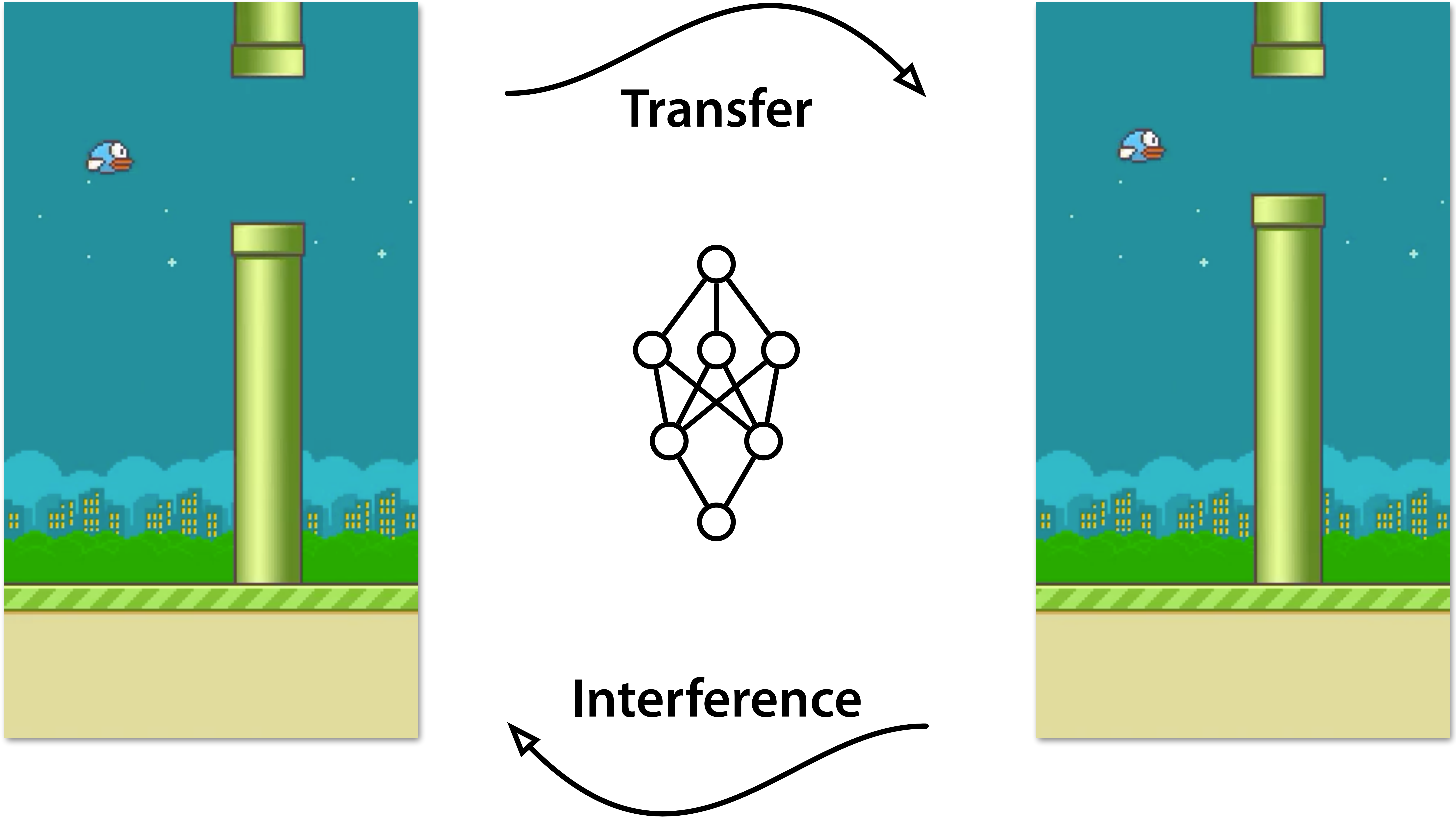


Easy Gap
after learning
Challenging Gap

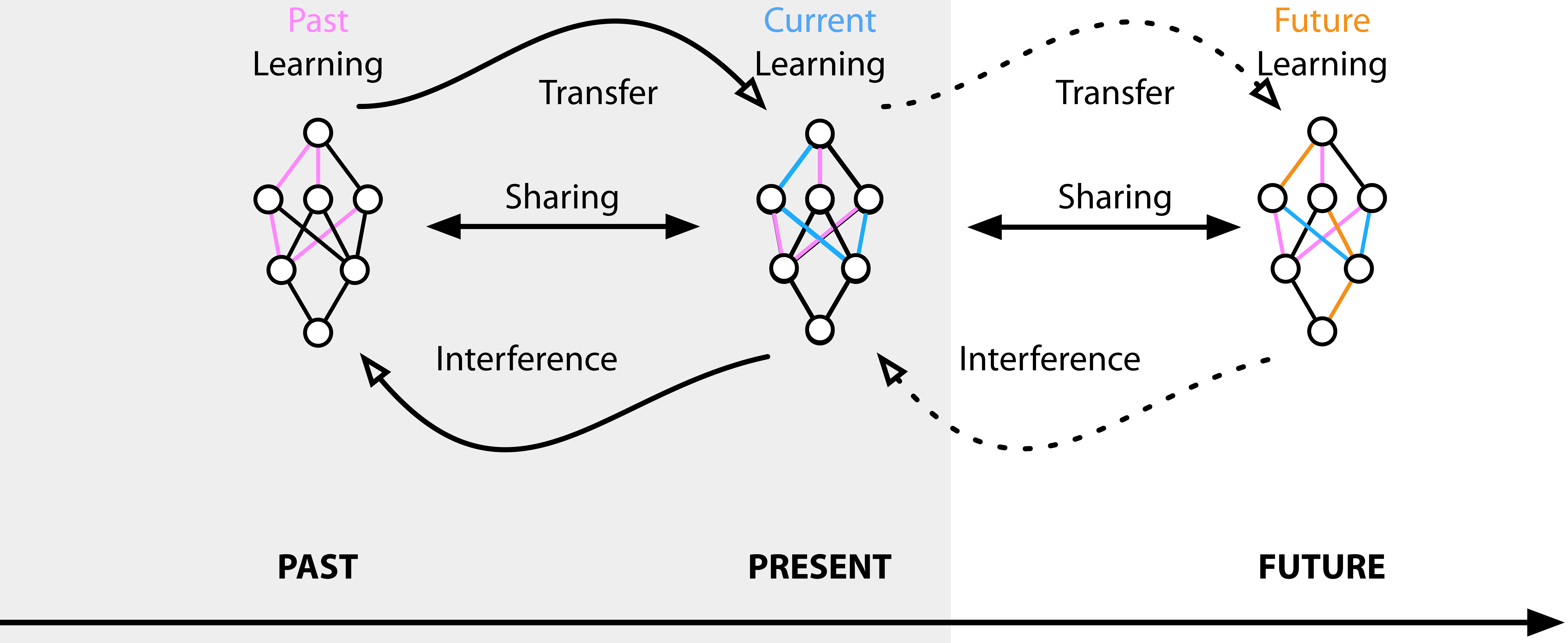


Transfer and Interference in Continual Learning

Introduction



Transfer–Interference Trade-off



Navigating the Transfer–Interference Trade-Off

Learning to Learn without Forgetting

Learning to Learn without Forgetting by Maximizing Transfer and Minimizing Interference

- Introduces the transfer-interference trade off.
- Meta-Experience Replay as a solution to navigate the trade-off.
  - Matt Riemer, Ignacio **Cases**, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro, published in ICLR 2019, NeurIPS 2018 (spotlight talk)

The proposed solution

- **Memory**: incorporates past experiences to make the learned distribution compatible with the past.
- **Learning to learn**: an algorithm that learns to learn to maximize transfer and minimize interference.

Transfer and Interference as an Inner Product

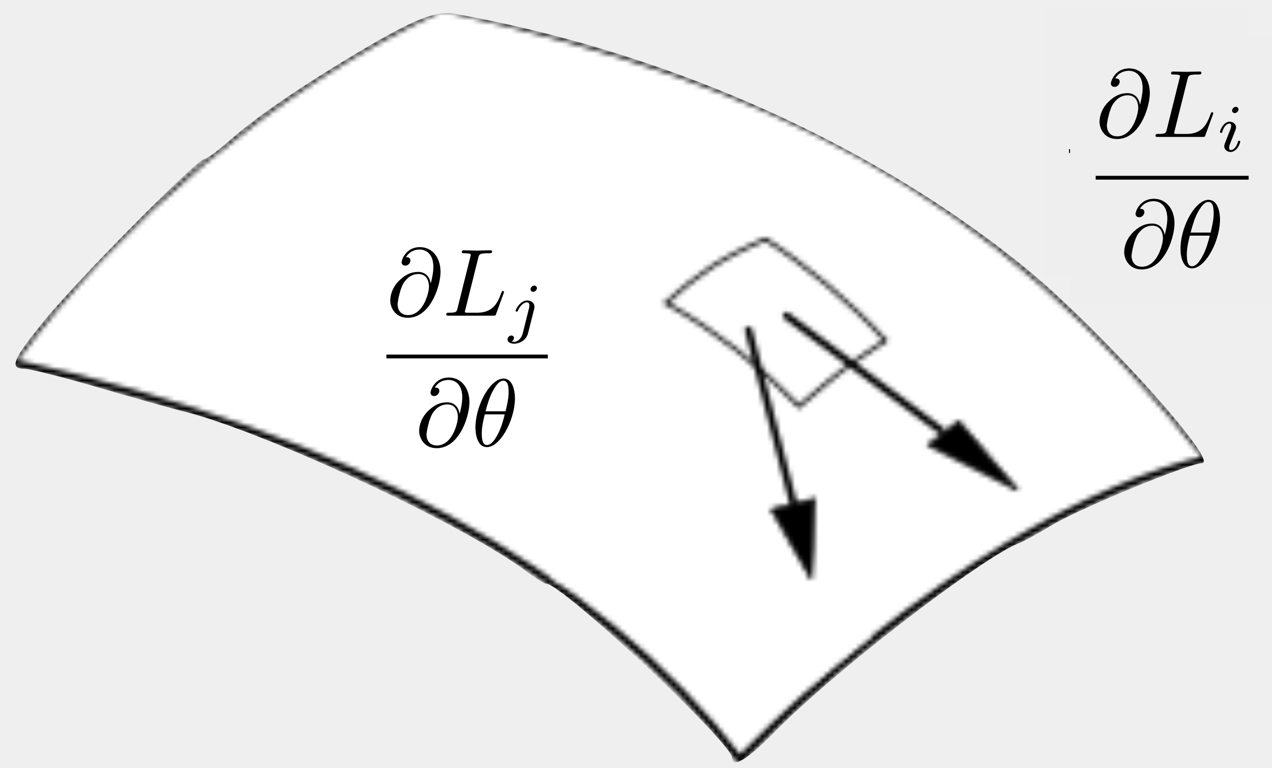
Learning to Learn without Forgetting

For any two data points (x_i,y_i) and (x_j,y_j) we define an instant in time with loss L and parameters :

Transfer

- When we train on one we improve on the other:

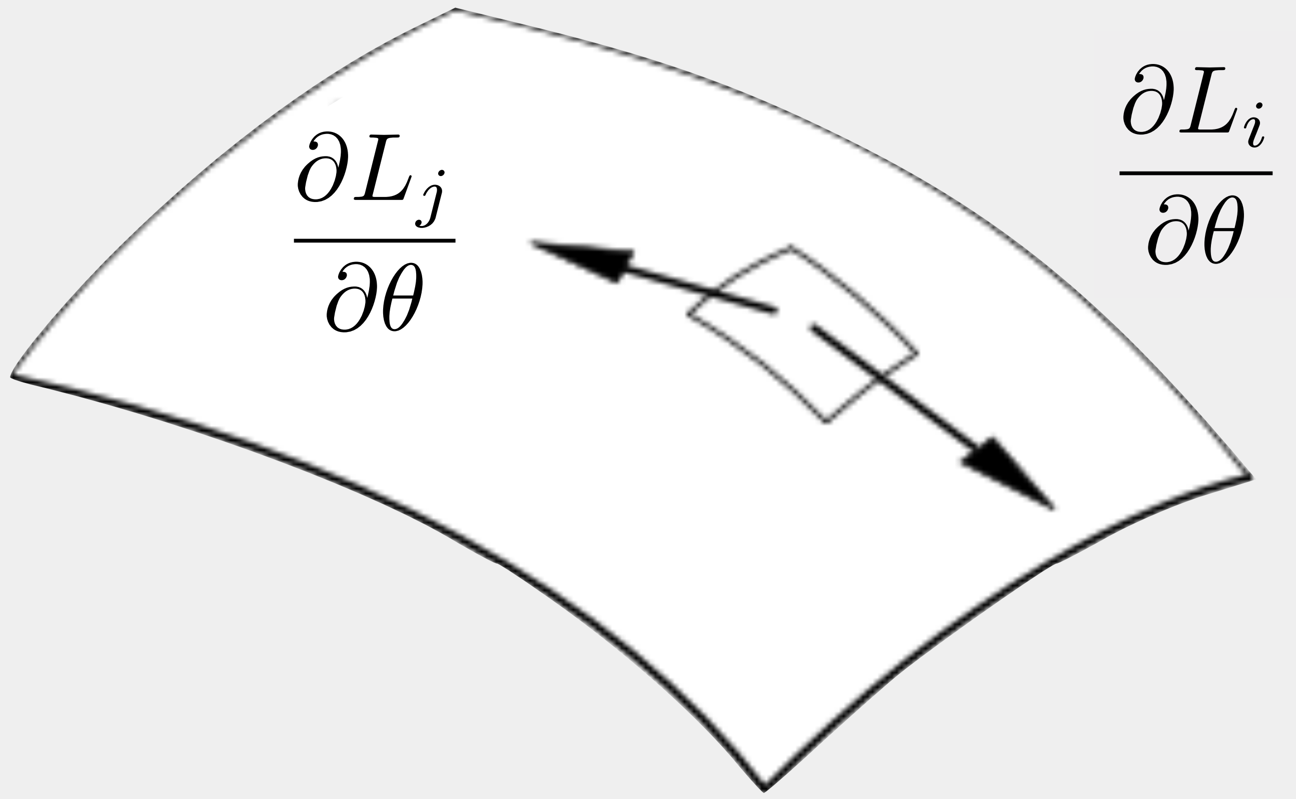
$$\frac{\partial L_i}{\partial \theta} \cdot \frac{\partial L_j}{\partial \theta} > 0$$



Interference

- When we train on one we get worse at the other:

$$\frac{\partial L_i}{\partial \theta} \cdot \frac{\partial L_j}{\partial \theta} < 0$$



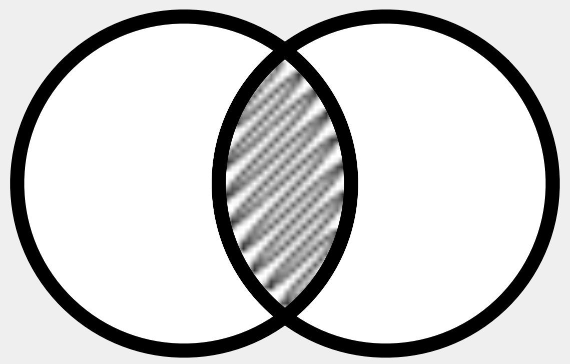
Distribution of the Products

Learning to Learn without Forgetting

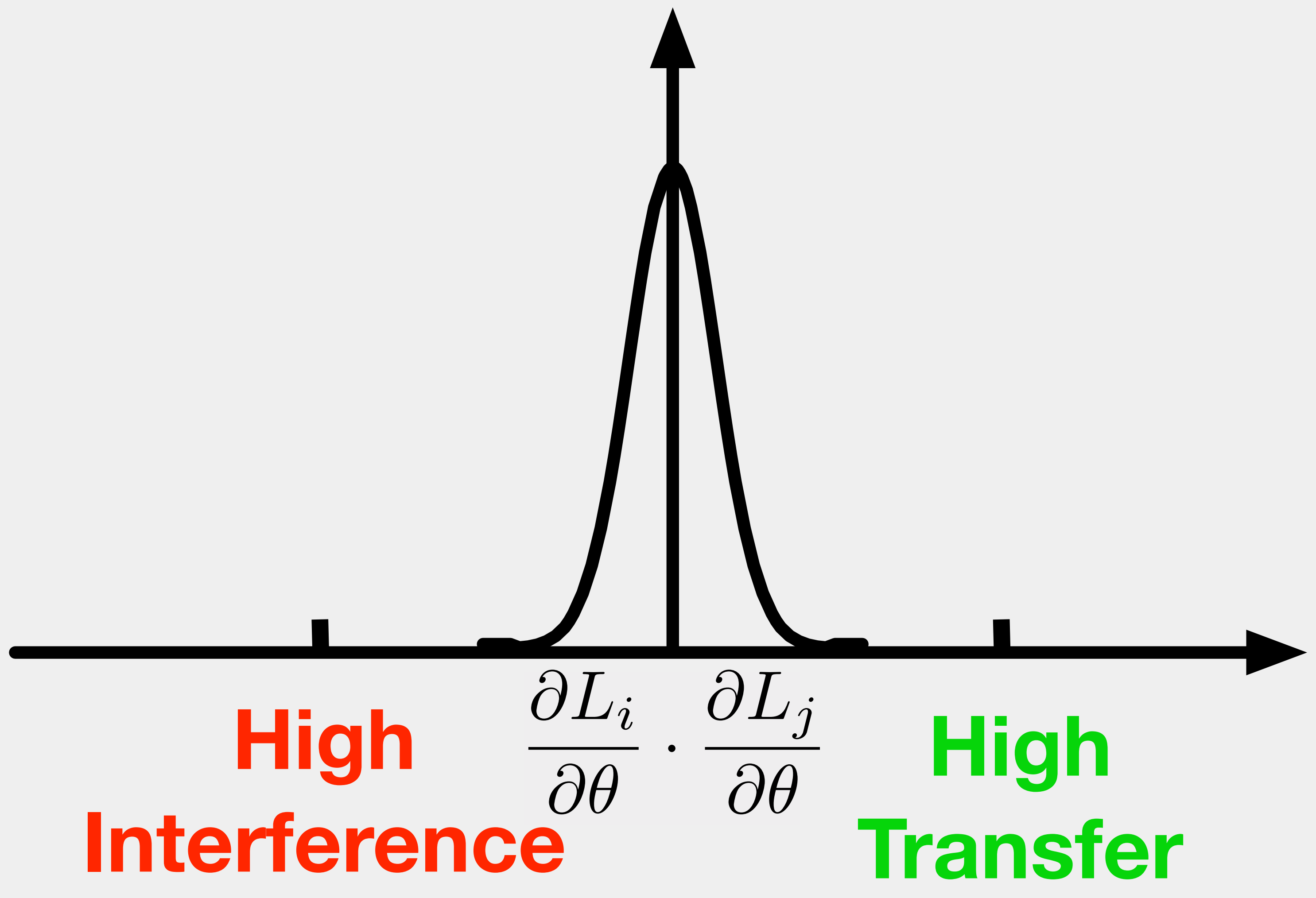
Weight sharing across examples

Possible distribution

Orthogonalized Knowledge

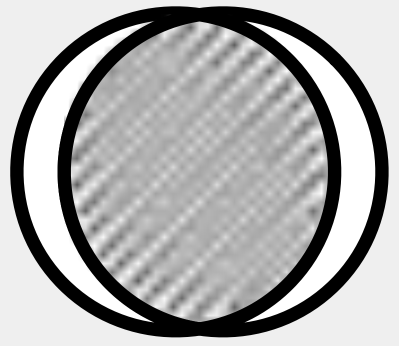


$\theta_i \;\; \theta_j$

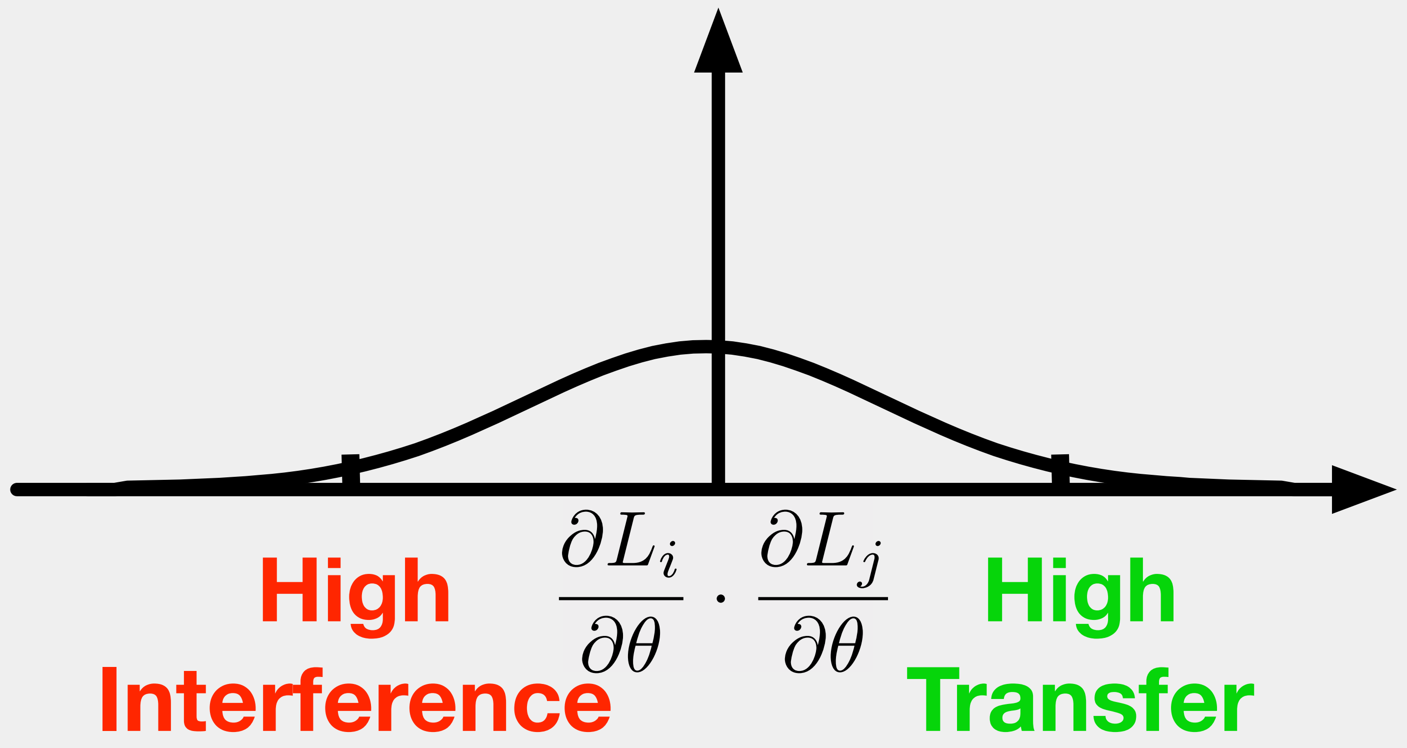


$$[1, 2, 1, 0, 0, 0] \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \\ 1 \\ 2 \end{bmatrix} = 0$$

Compressed Knowledge



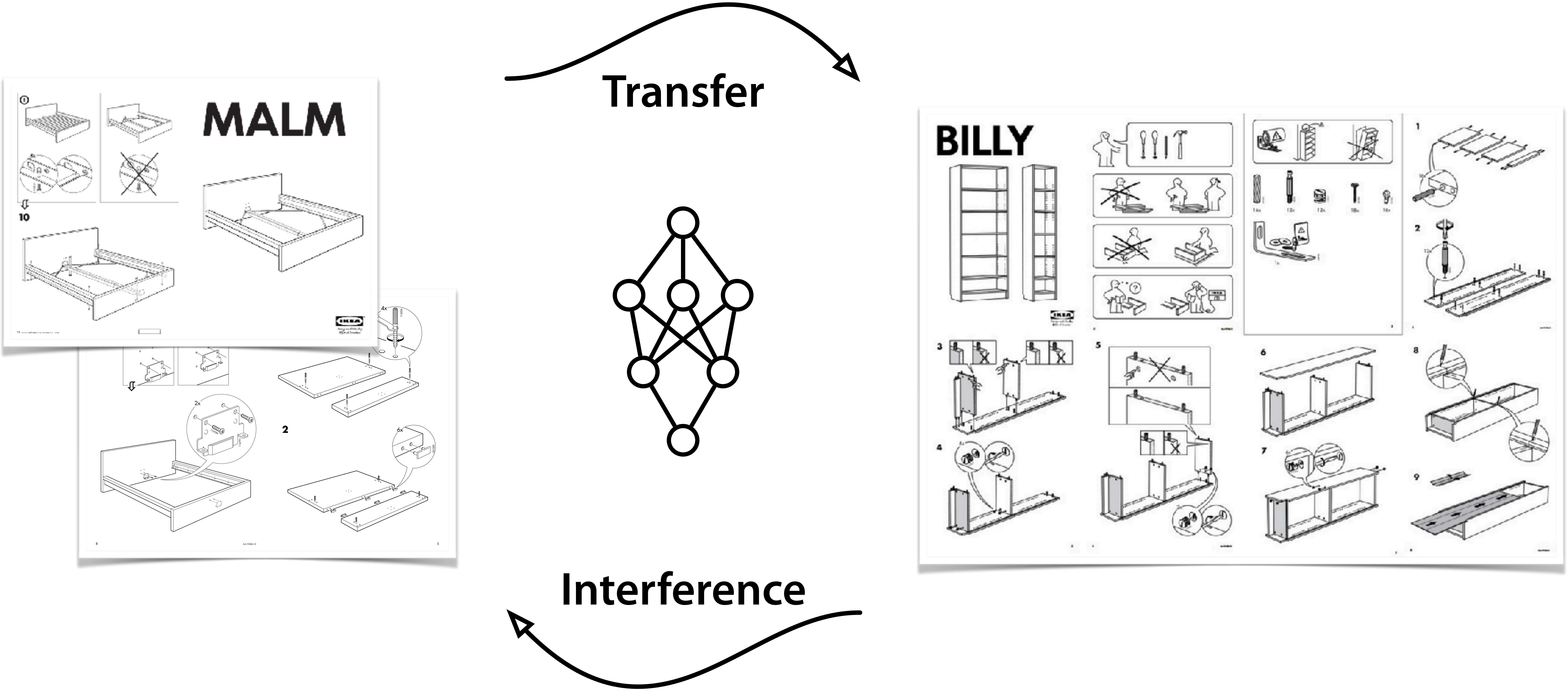
$\theta_i \;\; \theta_j$



$$[1, 2, 1, 1, 1, 1] \cdot \begin{bmatrix} -1 \\ 1 \\ 0 \\ 3 \\ 1 \\ 2 \end{bmatrix} = 7$$

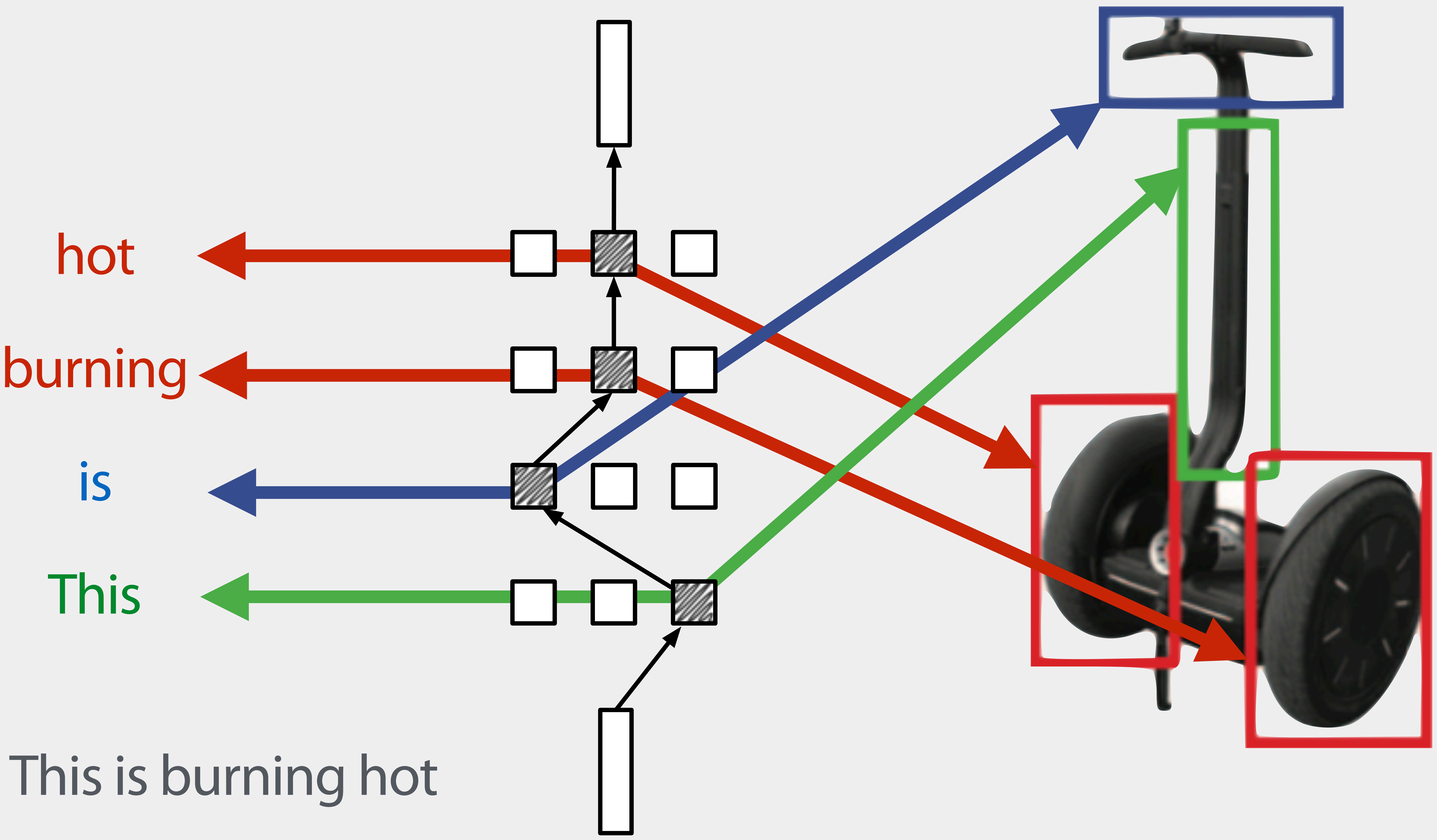
Modular and Compositional Learning

Modular and Compositional Learning for Natural Language Understanding



Modular and Compositional Learning

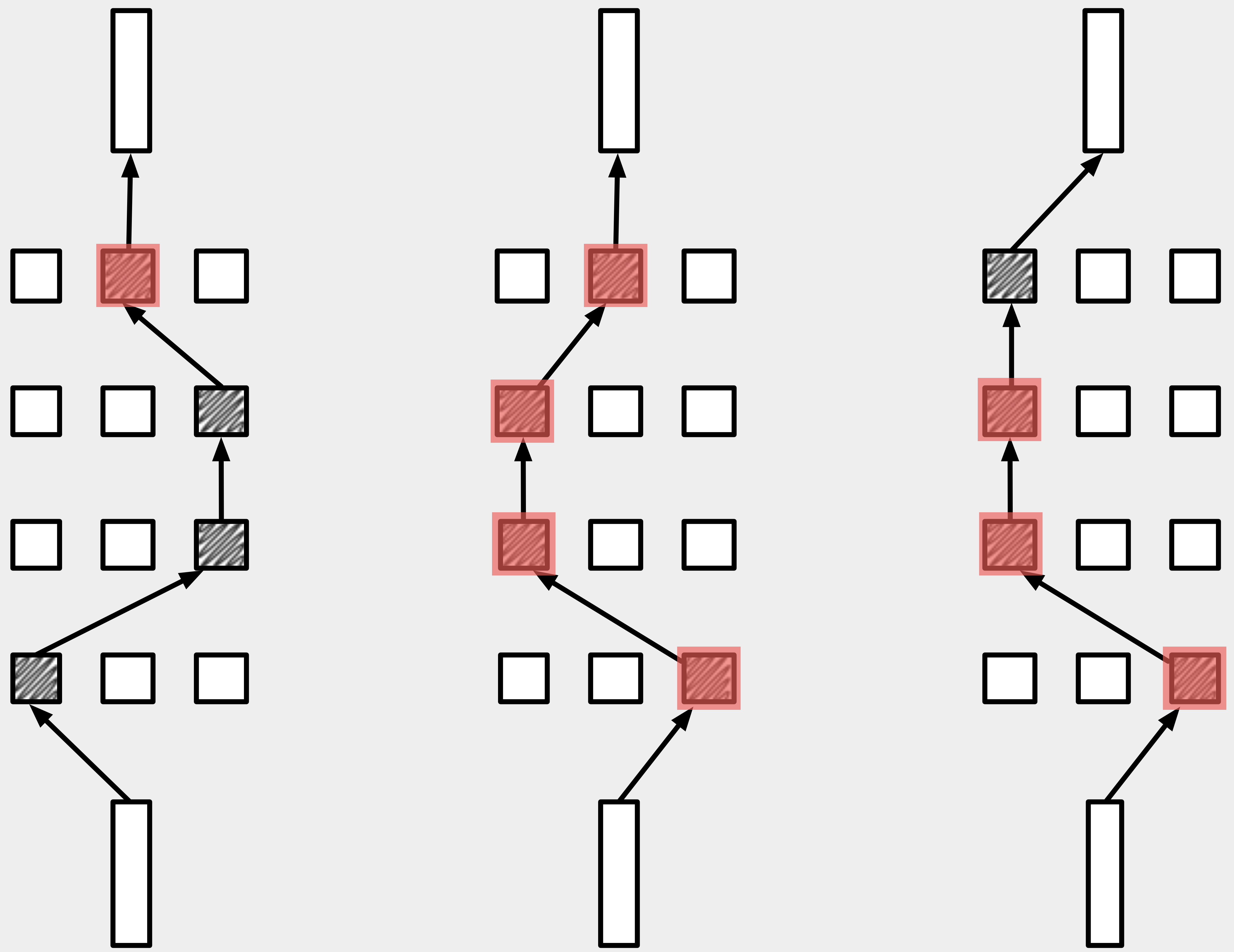
Modular and Compositional Learning for Natural Language Un

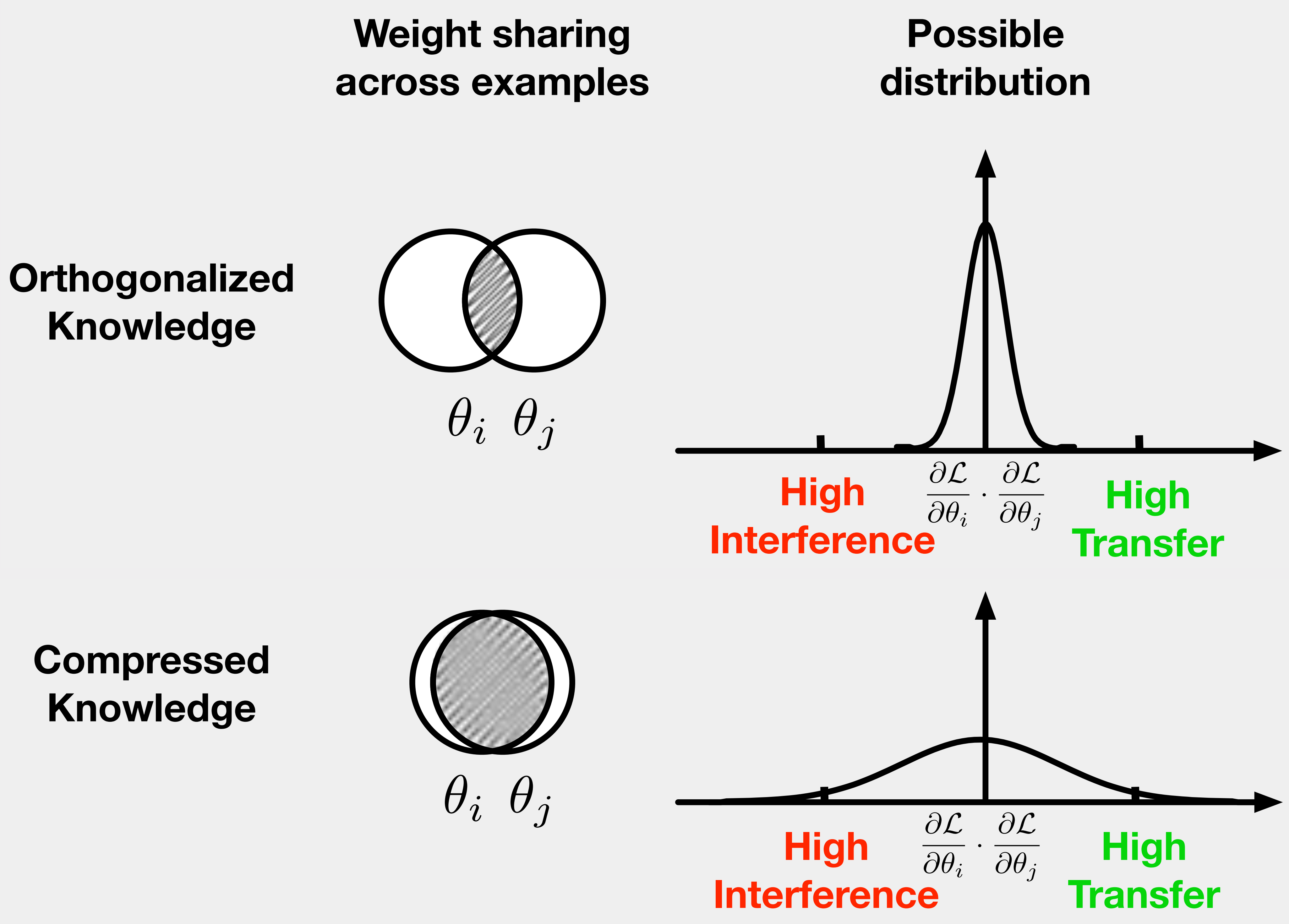


Illustrations from Lake, Salakhutdinov and Tenenbaum 2015

Modular and Compositional Learning

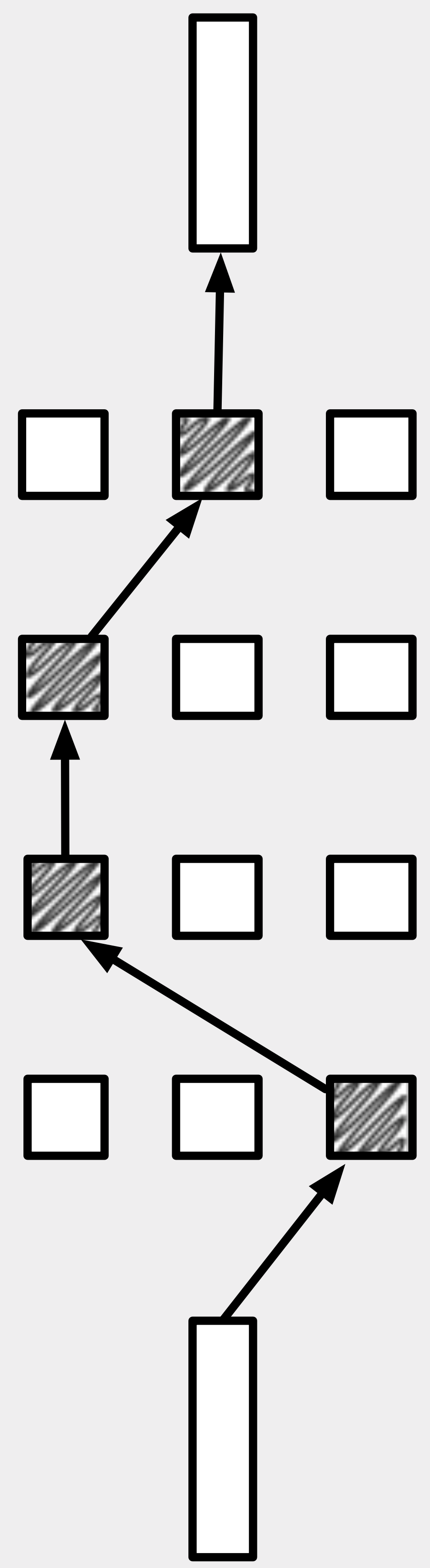
Modular and Compositional Learning for Natural Language Understanding

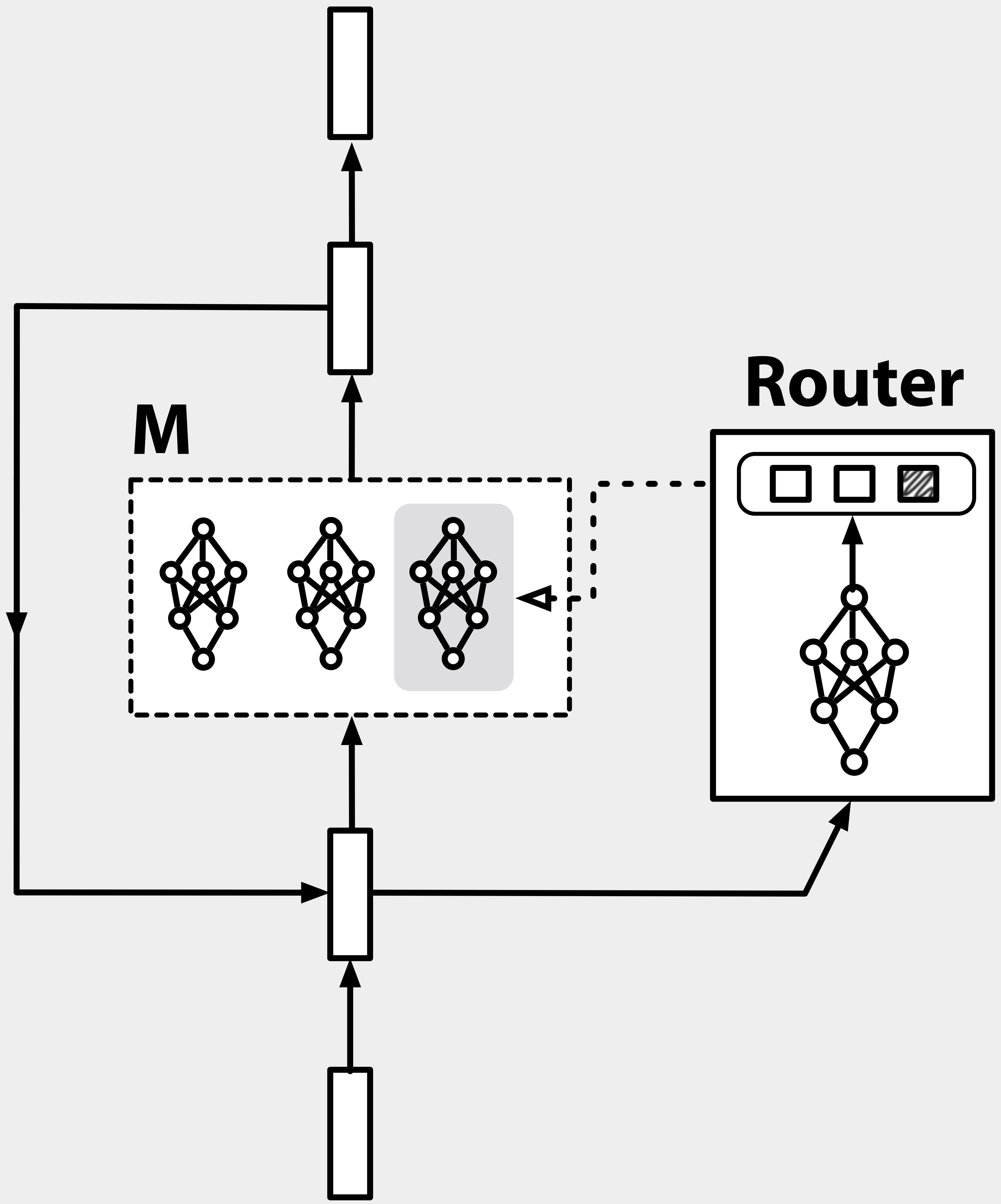




Recursive Routing Networks

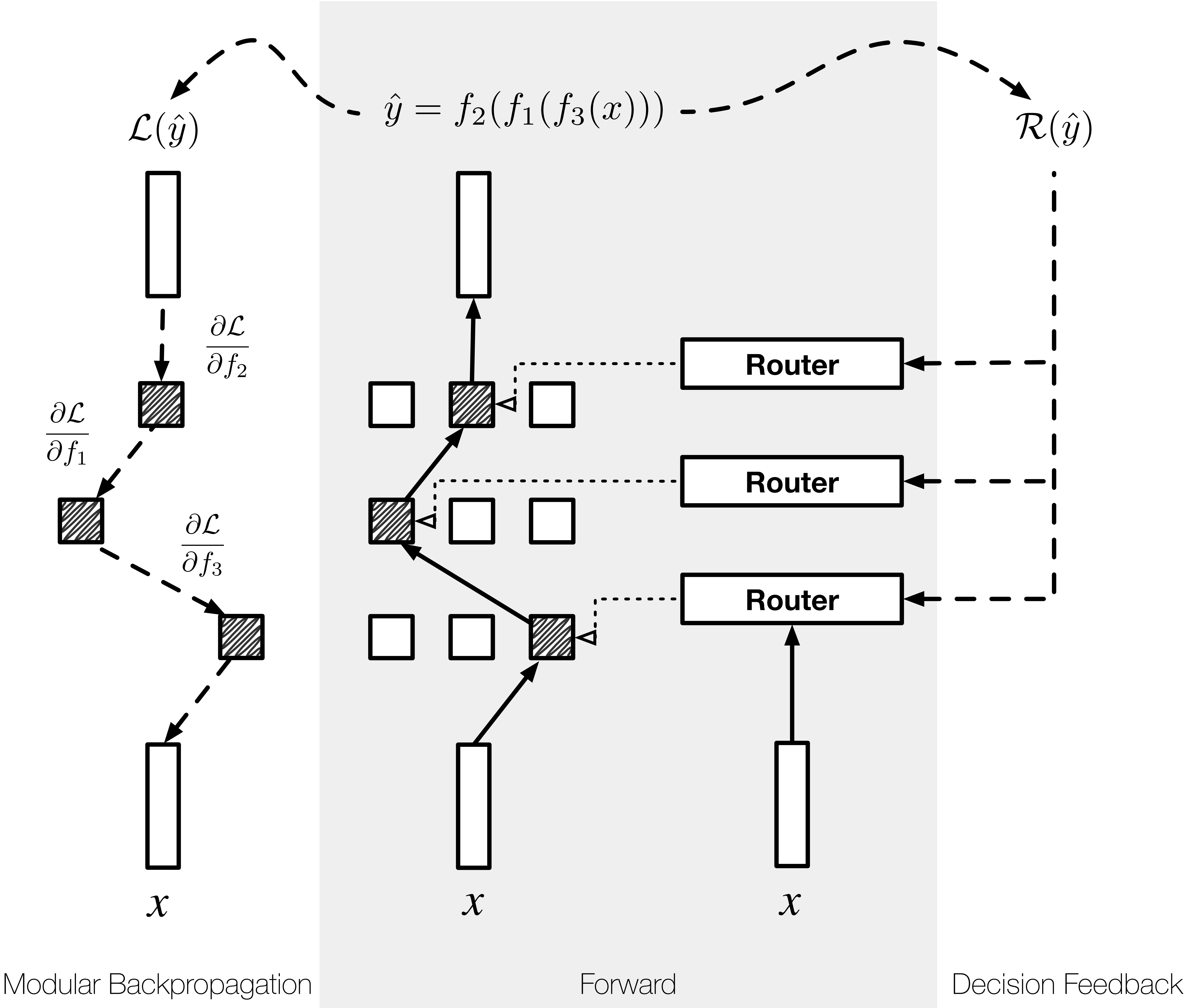
Modular and Compositional Learning for Natural Language Understanding





Training Recursive Routing Networks

Modular and Compositional Learning for Natural Language Understanding



Implicative: manage

Implicative Constructions

Joan managed to solve the problem

entails

Joan solved the problem

contradicts

Joan did not solve the problem

permits

The problem was not about mathematics

Matrix clause

Relation

Complement clause

= neither entails nor contradicts

Implicative: fail

Implicative Constructions

Joan failed to solve the problem

entails

Joan did not solve the problem

contradicts

Joan solved the problem

permits

The problem was about mathematics.

Matrix clause

Relation

Complement clause

= neither entails nor contradicts

Strawson Entailment I

Implicative Constructions

Joan solved the problem

does not entail

Joan managed to solve the problem

because *manage* has a presupposition:

It was difficult for Joan to solve the problem

The entailment only goes in one direction. The sentences in blue are not equivalent for us (as they are in MacCartney's 2009 *NatLog* system).

Strawson entailment (von Fintel)

A entails B just in case A satisfies all the presuppositions of B.

We adopt this notion of entailment.

Strawson Entailment II

Implicative Constructions

Joan did not solve the problem

does not entail

Joan failed to solve the problem

because *fail* has a presupposition:

Joan tried to solve the problem or was expected to solve it

Signatures

Implicative Constructions

John managed to write the paper

entails

John wrote the paper.

+ – ø

1	2

Karttunen 1971

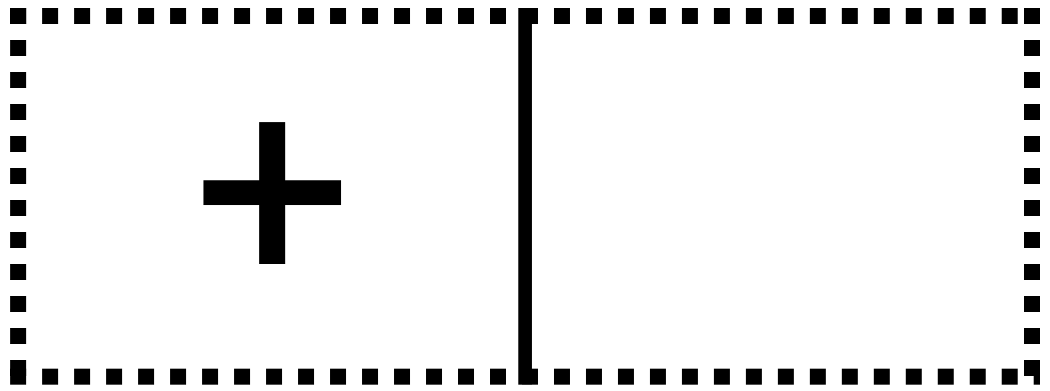
Signatures

Implicative Constructions

John managed to write the paper

entails

John wrote the paper.



Karttunen 1971

Signatures

Implicative Constructions

John didn’t manage to write the paper

entails

John didn’t write the paper.

+	−

Karttunen 1971

Signatures Recap: pos|neg

Implicative Constructions

The pos sign indicates the semantic relation of the matrix sentence to its complement in affirmative environments, the neg sign pertains to negative environments.

+ indicates entailment, – indicates contradiction, o stands for permits

manage	+|–	+	–
fail	–|+	–	+
promise	o|o	ø	ø

Karttunen 1971

Nested Implicatives

Implicative Constructions

Implicatives can be nested:

- John failed to manage to solve the problem
- John managed to fail to solve the problem

both entail

- John did not solve the problem

but they have different presuppositions

What is the difference?

Composition of signatures

Implicative Constructions

manage ∘ fail = manage to fail

+|−   −|+   −|+

fail ∘ manage = fail to manage

−|+   +|−   −|+

promise ∘ manage = promise to manage

o|o   +|−   o|o

manage ∘ promise = manage to promise

+|−   o|o   o|o

fail ∘ promise = fail to promise

−|+   o|o   o|o

Composition of signatures

Implicative Constructions

$$\mathrm{sig}_1 \circ \mathrm{sig}_2 = \mathrm{sig}_2 \qquad \text{if } \mathrm{sig}_1 \text{ is } +|?$$

$$\mathrm{sig}_1 \circ \mathrm{sig}_2 = \mathrm{reverse}(\mathrm{sig}_2) \qquad \text{if } \mathrm{sig}_1 \text{ is } -|?$$

$$\mathrm{sig}_1 \circ \mathrm{sig}_2 = o|o \qquad \text{if } \mathrm{sig}_1 \text{ is } o|?$$

Composition of signatures is associative:

$$(\mathrm{sig}_1 \circ \mathrm{sig}_2) \circ \mathrm{sig}_3 = \mathrm{sig}_1 \circ (\mathrm{sig}_2 \circ \mathrm{sig}_3)$$

not ∘ fail ∘ manage = not fail to manage

$-|+ \quad -|+ \quad +|- \quad +|-$

$+|- \quad -|+$

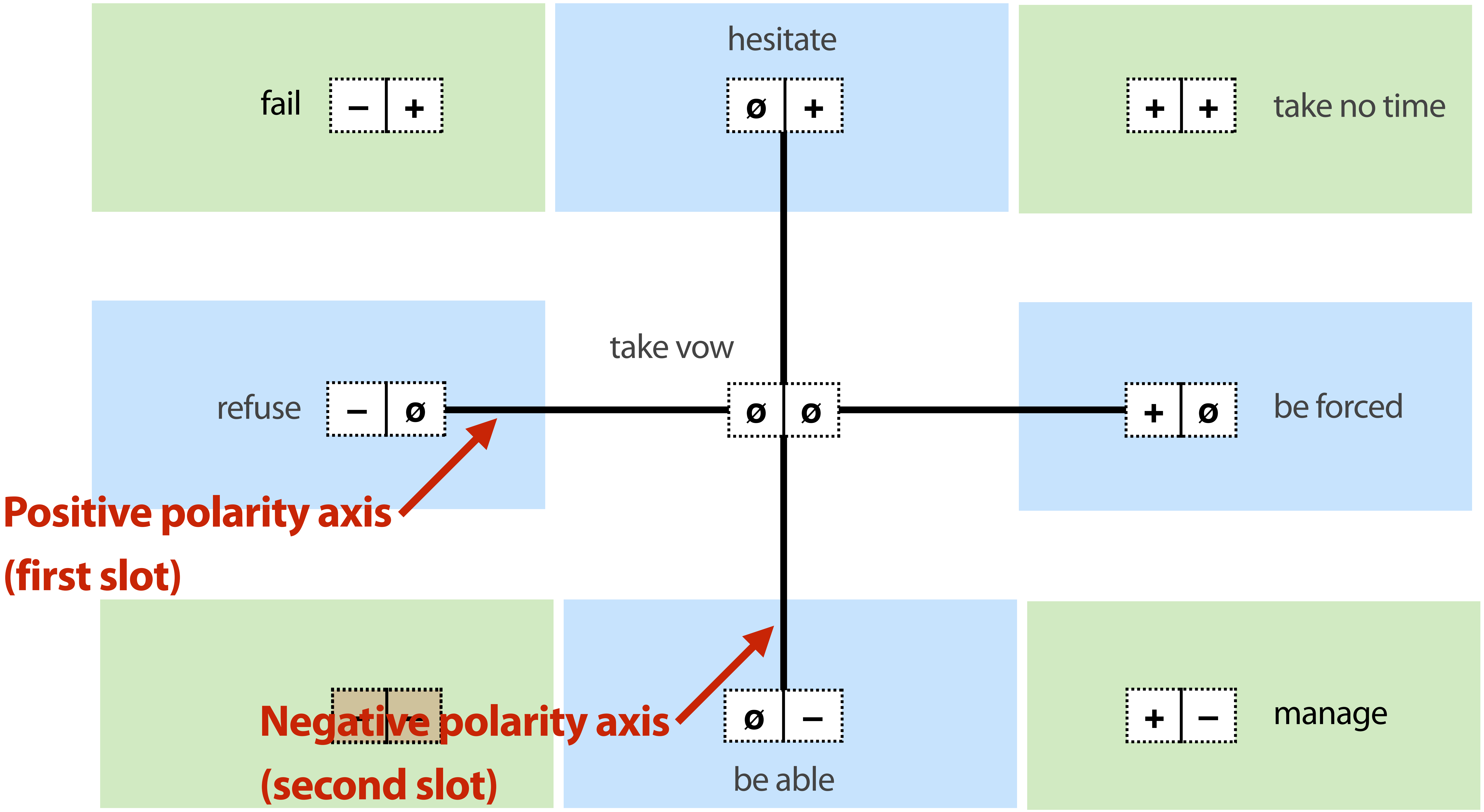
The blue and the red path lead to the same result.

Nairn et al (2006)

Plane of Signatures

Implicative Constructions

Nonimplicatives



Two-way Implicatives

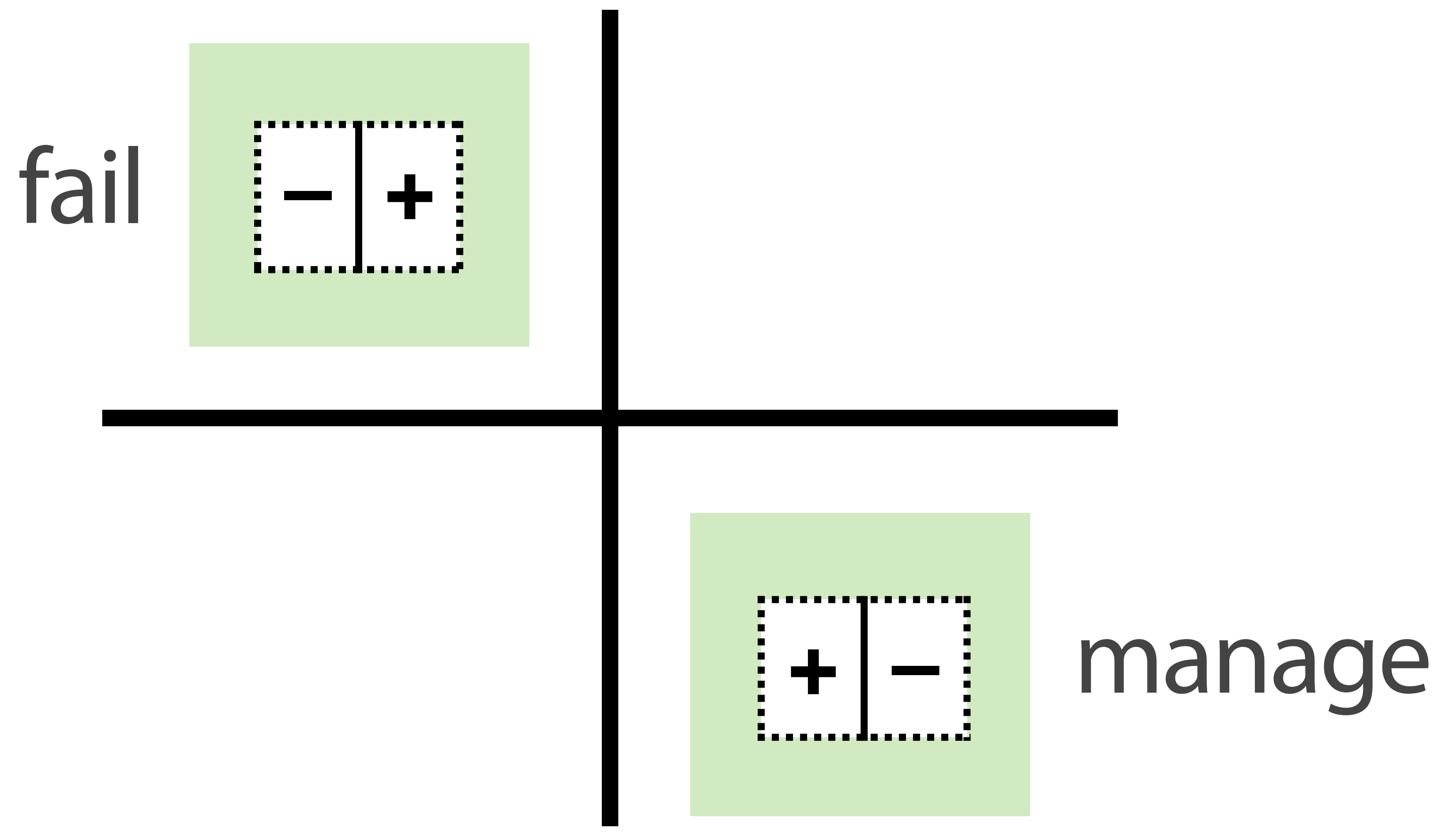
Entailment and Invited Inferences

Two-way implicatives yield an entailment under both positive and negative polarity.

+|− verbs

manage, bother, dare, deign, remember (to), happen, turn out

Preserve polarity



−|+ verbs

fail, neglect, refuse, forget (to)

Reverse polarity

One-way implicatives

Entailment and Invited Inferences

There are four types of implicatives that yield an entailment only under one polarity.

The entailments of able and force are polarity-preserving, refuse and hesitate reverse the polarity.

For the polarities associated with non-zero signature, we have for example

- Ann was not able to speak. $\sqsubset$ Ann didn't speak. o|–
- Ann was forced to speak. $\sqsubset$ Ann spoke. +|o
- Ann refused to speak. $\sqsubset$ Ann didn't speak. –|o
- Ann didn't hesitate to speak. $\sqsubset$ Ann spoke. o|+

With the polarity corresponding to zero-signature there is no entailment but there may be a suggestion, an **invited inference**.

Invited inferences

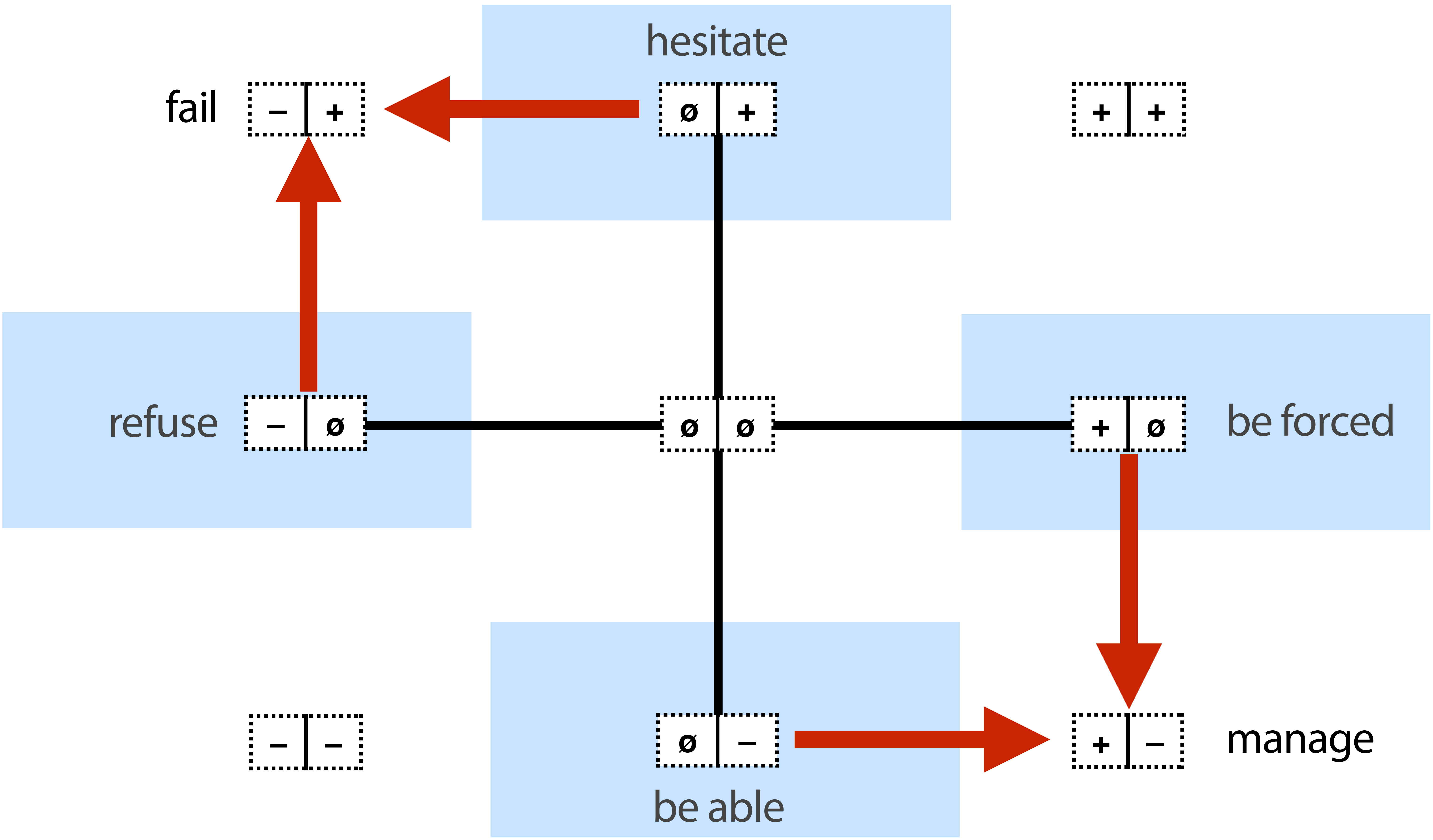
Entailment and Invited Inferences

Ann was able to speak.
⇝ Ann spoke. (+)|–

Ann was not forced to speak.
⇝ Ann didn't speak. +|(–)

Ann did not refuse to speak.
⇝ Ann spoke. –|(+)

Ann hesitated to speak.
⇝ Ann didn't speak. (–)|+

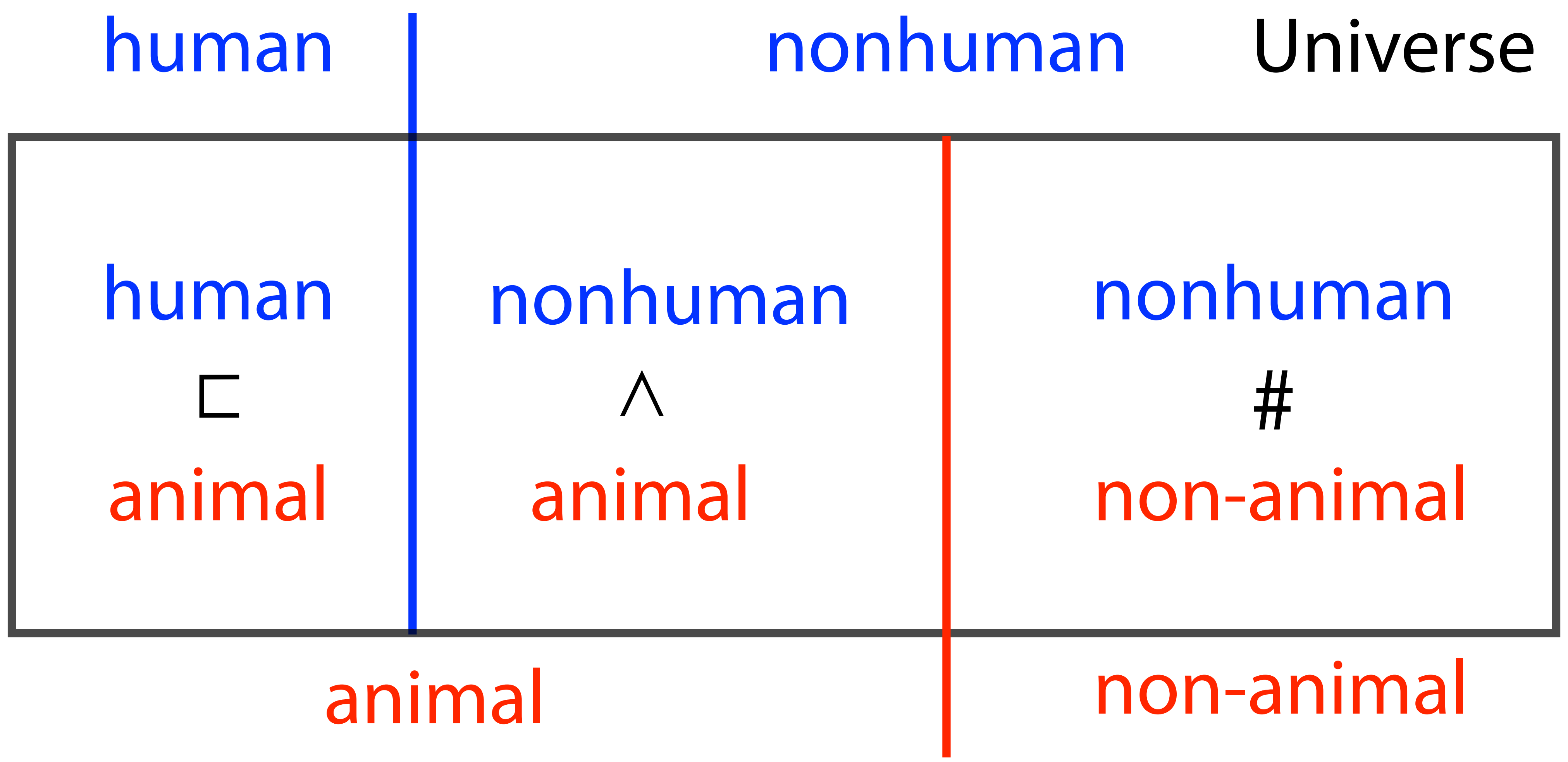


animal $\smile$ nonhuman

Entailment and Invited Inferences

$$X \smile Y \;\equiv\; X \cap Y \neq \varnothing \;\wedge\; X \cup Y = U$$

where U is the Universe of Discourse



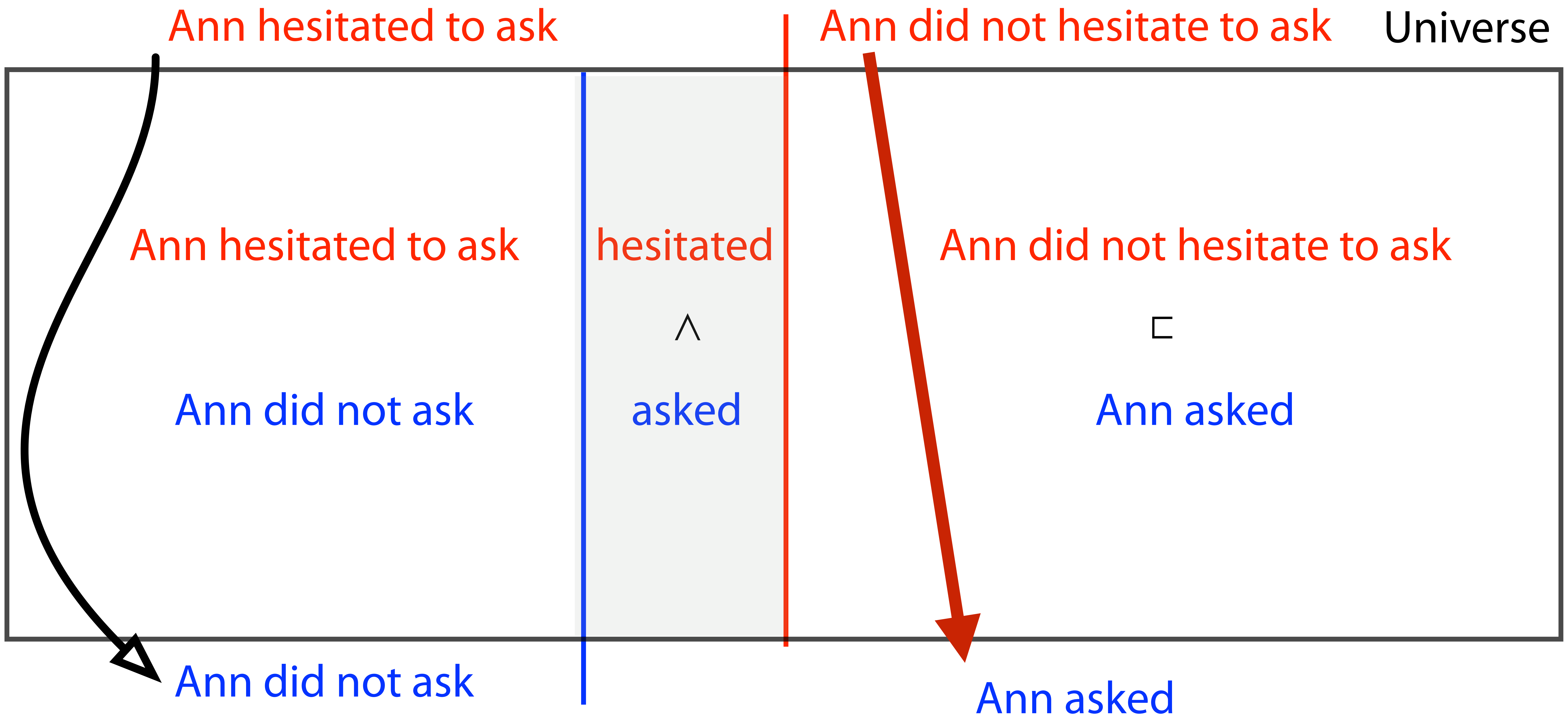
You are an animal!

Why is this an insult?

One way implicatives: o|+

Entailment and Invited Inferences

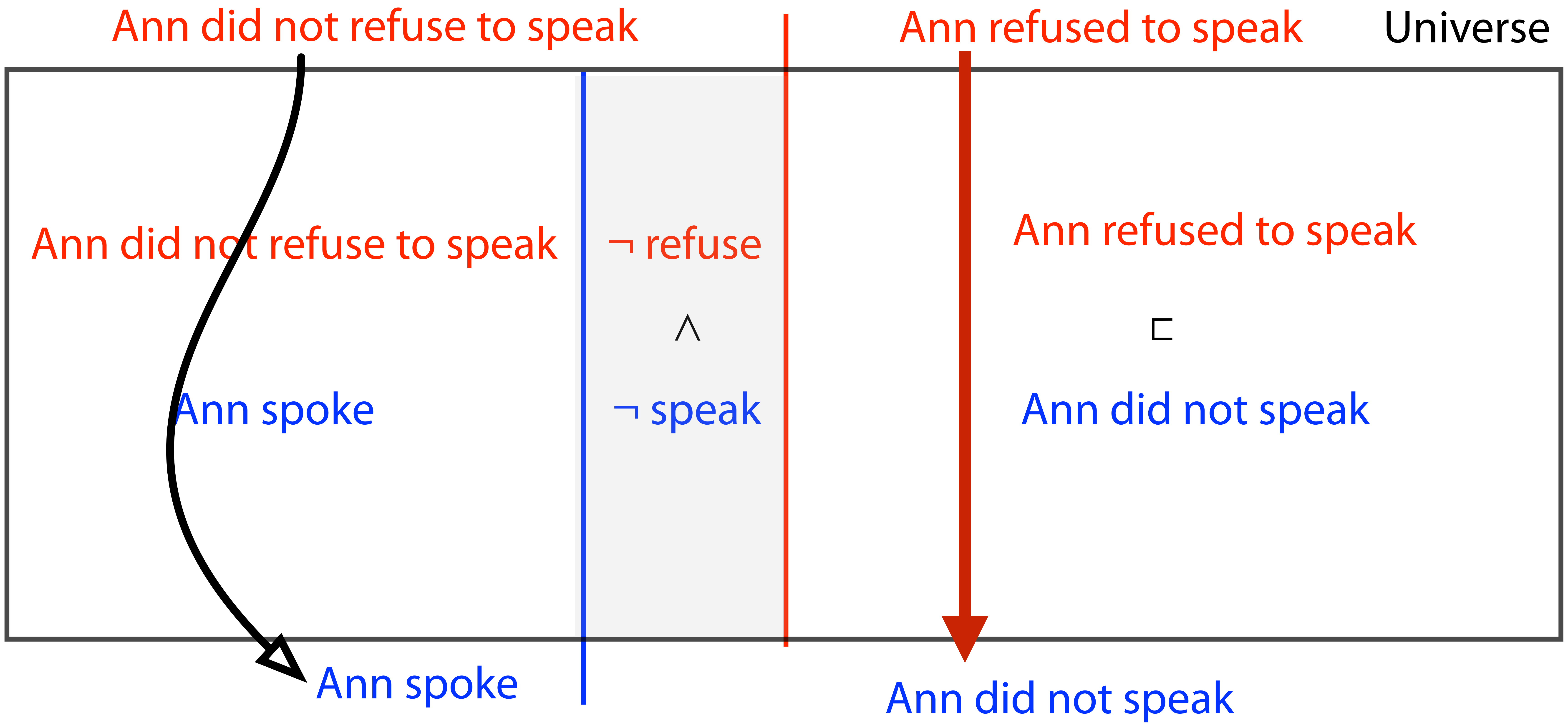
Ann hesitated to ask ᴗ Ann asked



One way implicatives: –|o

Entailment and Invited Inferences

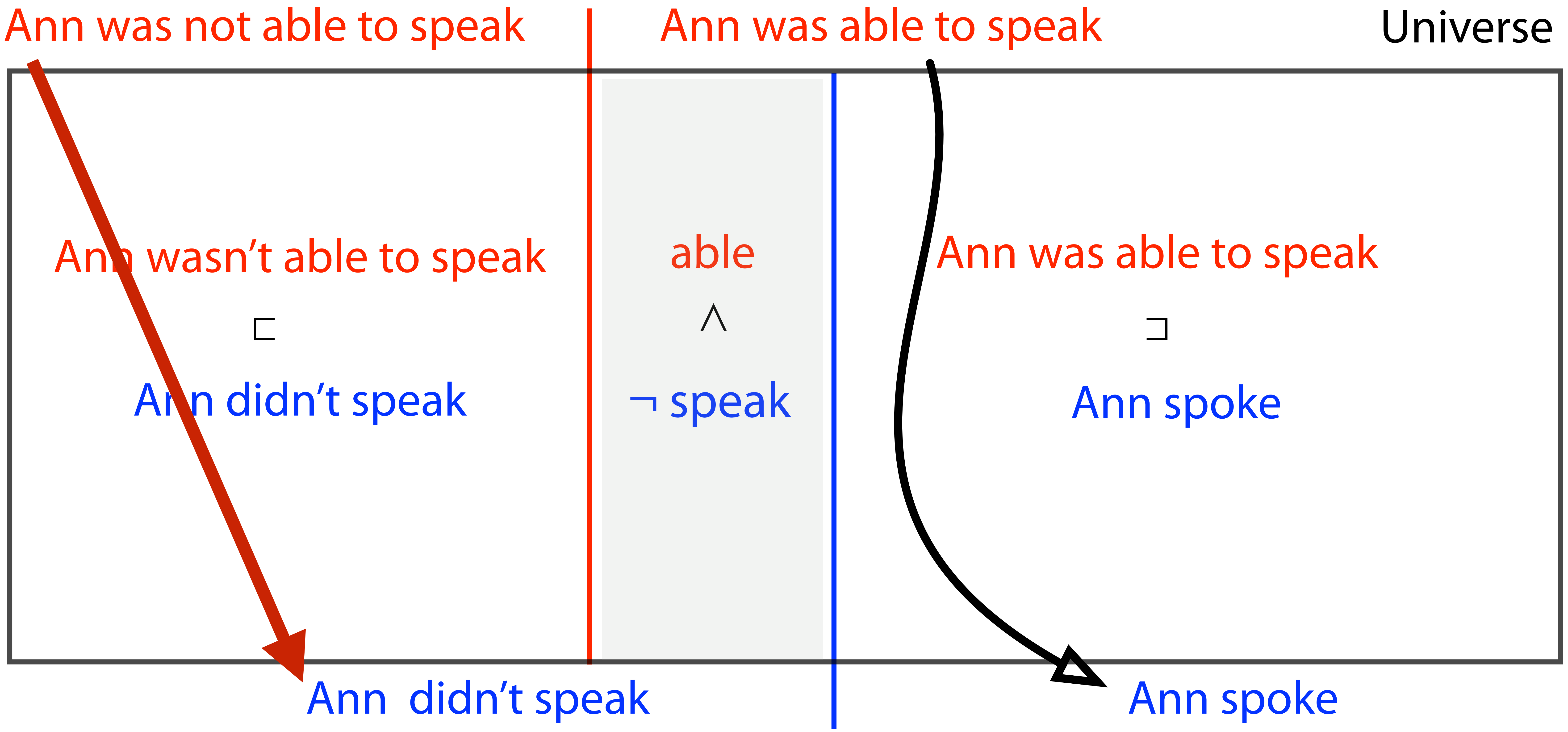
Ann did not refuse to speak ᴗ Ann did not speak



One way implicatives: o|–

Entailment and Invited Inferences

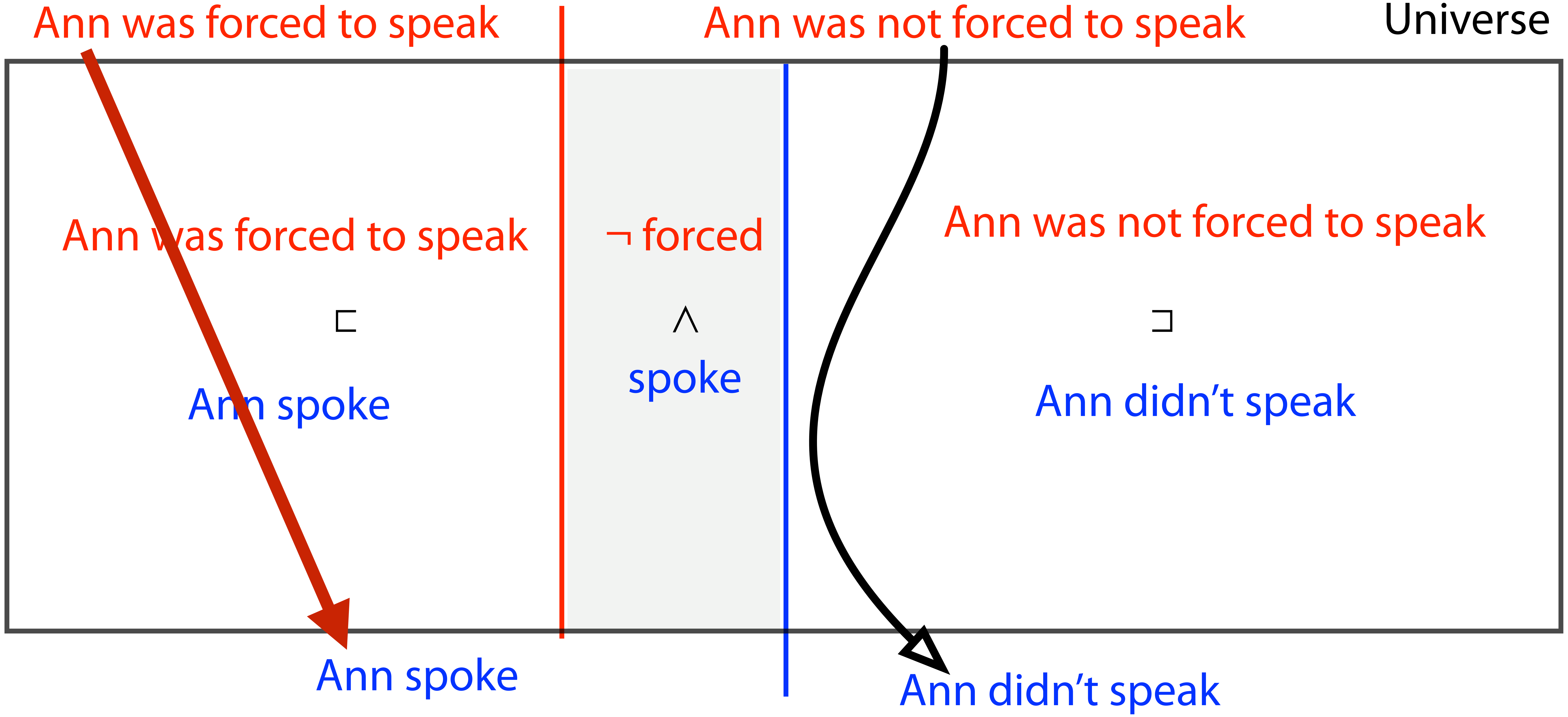
Ann was able to speak ᴗ Ann didn't speak



One way implicatives: +|o

Entailment and Invited Inferences

Ann was not forced to speak ∪ Ann spoke



Explaining the invited inferences

Entailment and Invited Inferences

As shown in the previous diagrams all the one-way implicatives are in a cover relation:

- Ann was able to speak. $\cup$ Ann did not speak. invites Ann spoke.
- Ann did not refuse to speak. $\cup$ Ann did not speak. invites Ann spoke.
- Ann was not forced to speak. $\cup$ Ann spoke. invites Ann didn't speak.
- Ann hesitated to speak. $\cup$ Ann spoke. invites Ann didn't speak.

In all cases the invited inference is the negation of the corresponding statement in the right, from the intersection of the two propositions.

Probabilistic Signatures

Entailment and Invited Inferences

- be able .9|–
  - ⋆ behaves a (+)|–
  - ⋆ ~90% of the cases

- prevent –|.7
  - ⋆ behaves as (–)|+
  - ⋆ ~70% of the cases

Phrasal two-way implicatives

Reasoning with Implicatives

+|−

use an asset, opportunity

I used the money to buy shoes and food.

Randy didn’t use the opportunity to toot his own horn.

−|+

waste an asset

I wasted the money to buy a game that I cannot play.

I’m glad I didn’t waste 90 minutes to see this film.

waste an opportunity

Mr. Spitzer wasted the opportunity to drive a harder bargain.

She didn't waste the chance to smile back at him.

fail an obligation

The Avatar failed his duty to bring peace to a broken world.

Orlando didn't neglect his duty to escort the dead.

Phrasal one-way implicatives

Reasoning with Implicatives

–|o

lack opportunity

She lost the chance to qualify for the final.

o|–

have ability

The defendant had no ability to pay the fine.

make effort

I have made no effort to check the accuracy of this blog.

o|+

show hesitation

She did not have any hesitation to don the role of a seductress.

Fonseka displayed no reluctance to carry out his orders.

Phrasal Implicatives: Most Common Verbs

Reasoning with Implicatives

<table>
<tr><th>VERB FAMILY</th><th>NOUN FAMILY</th><th>IMPLICATIVE SIGNATURE</th></tr>
<tr><td>HAVE</td><td>ABILITY | OPPORTUNITY
COURAGE | WISDOM</td><td>o|–
+|–</td></tr>
<tr><td>LACK</td><td>ABILITY | OPPORTUNITY</td><td>–|o</td></tr>
<tr><td>MAKE</td><td>EFFORT</td><td>o|–</td></tr>
<tr><td>MEET
FAIL</td><td>OBLIGATION</td><td>+|–
–|+</td></tr>
<tr><td>SHOW</td><td>HESITATION</td><td>o|+</td></tr>
<tr><td>TAKE</td><td>ASSET | EFFORT</td><td>+|–</td></tr>
<tr><td>USE</td><td>ASSET | OPPORTUNITY</td><td>+|–</td></tr>
<tr><td>WASTE</td><td>ASSET
OPPORTUNITY</td><td>+|–
–|+</td></tr>
</table>

Verb families

Reasoning with Implicatives

FAIL	fail, neglect
HAVE	get, have, possess
LACK	discard, give up, lack, lose, miss, throw away
MAKE	do, make, undertake
MEET	acquit, do, fulfill, meet, perform (OBLIGATION)
SHOW	have, show, display
TAKE	grab, seize, snap, snatch, take
USE	expend, exploit, use, utilize
WASTE	drop, squander, waste

A Corpus of Implicatives

Stanford Corpus of Implicative Constructions

Attested around 1000 implicative constructions (DARPA FAUST Project) (Karttunen et al. 2007)

Stanford Corpus of Implicatives (SCI)

- Consists of ~11K natural language triplets with a premise, hypothesis, and label.
- Human-generated triplets from seed sentences collected from Google Books and the web.
- 90+ implicative constructions.
- Additional 20+ nested implicatives in a extra set of the corpus.
- Four different subsets:
  - Joint/disjoint/mismatch,
  - Nested.

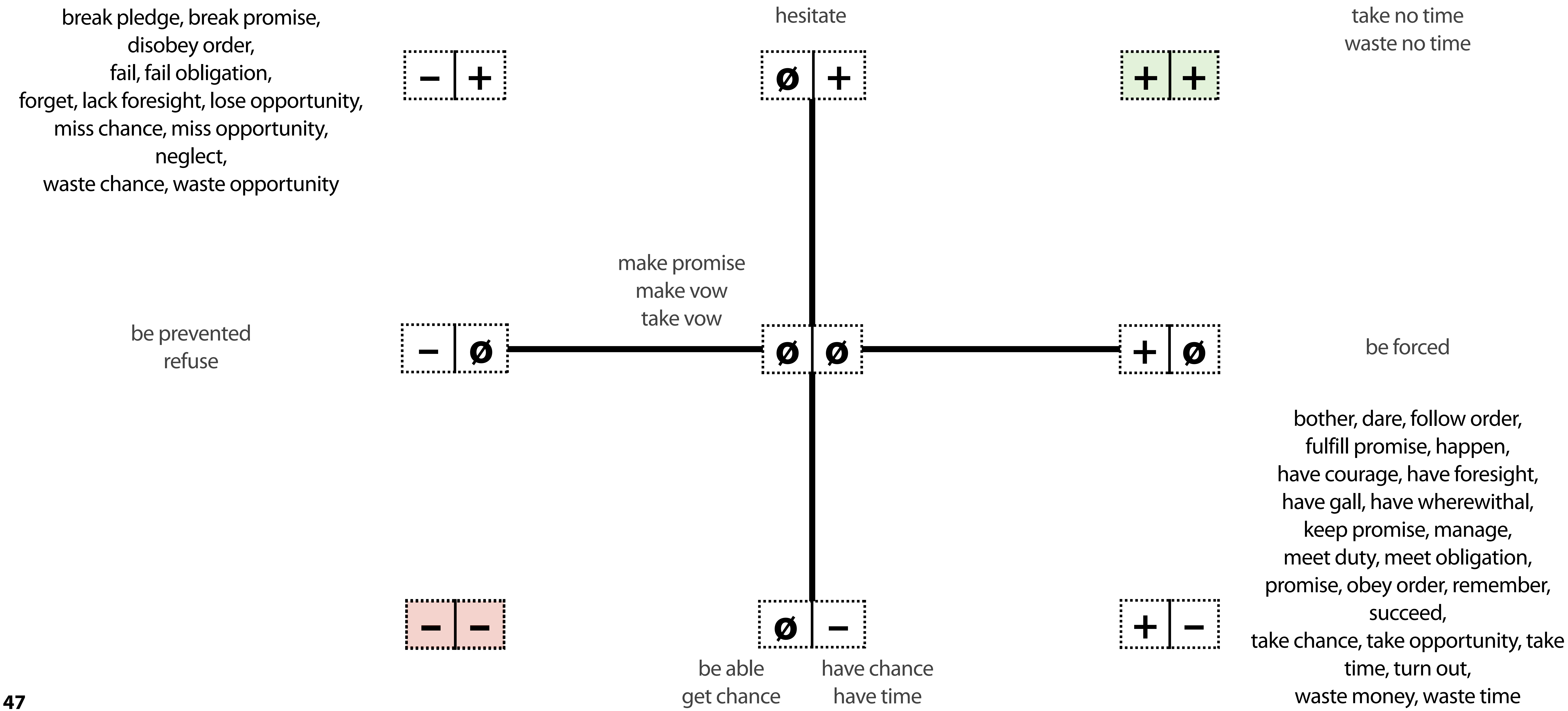
Implicative **signatures**

- Signatures provide high quality information that can be used as meta-information for NLI.

http://nlp.stanford.edu/projects/sci

Signatures in the Corpus of Implicatives

Stanford Corpus of Implicative Constructions



Desiderata I

Reasoning with Implicatives

1. Basics.

- The model should have a good test performance on sentences containing implicative constructions it has seen in training.

2. Generalization.

The model should able to generalize in the way people do.

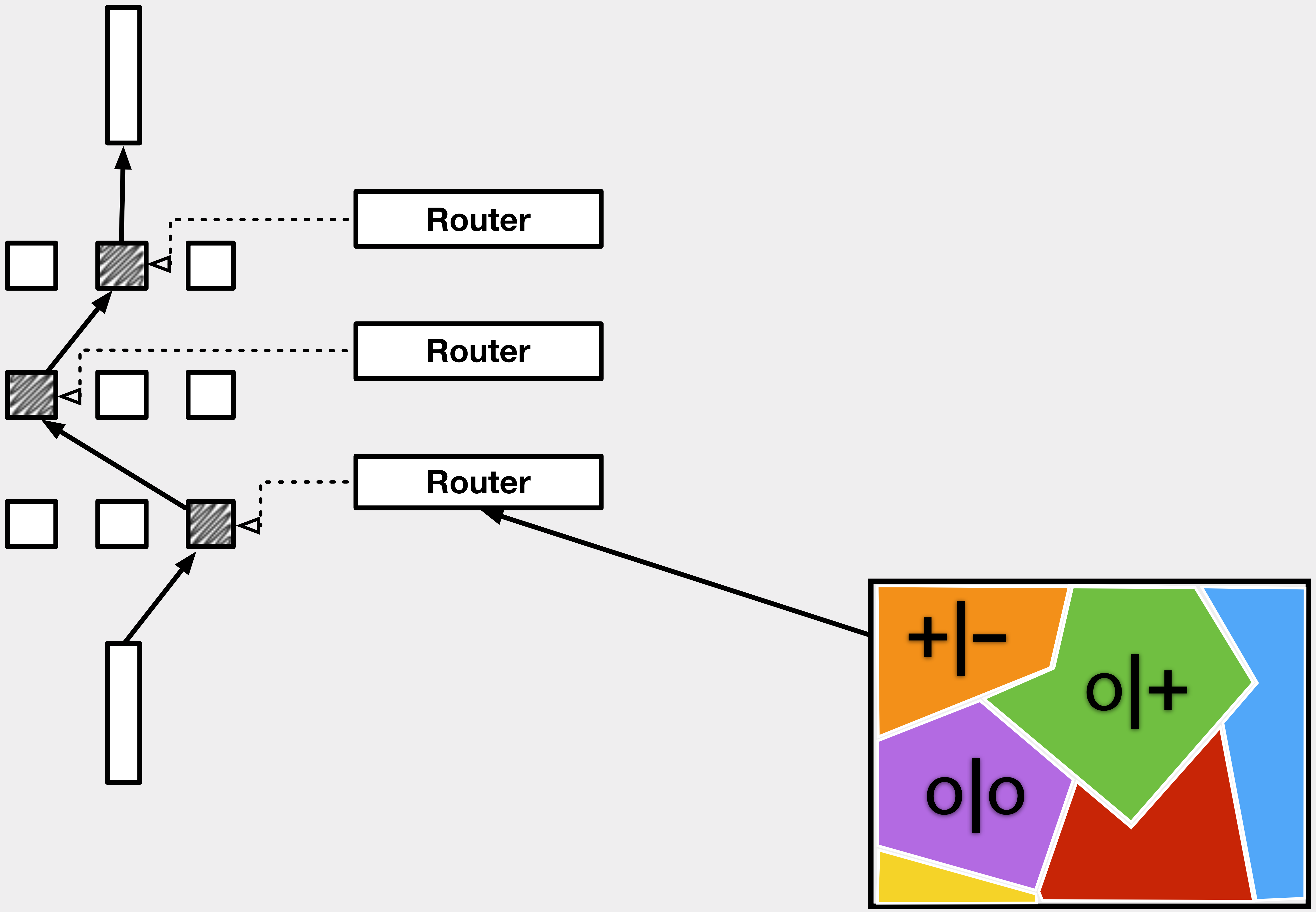
- Unseen lexical items
- Longer sequences

3. Composition of signatures.

- Nested implicatives

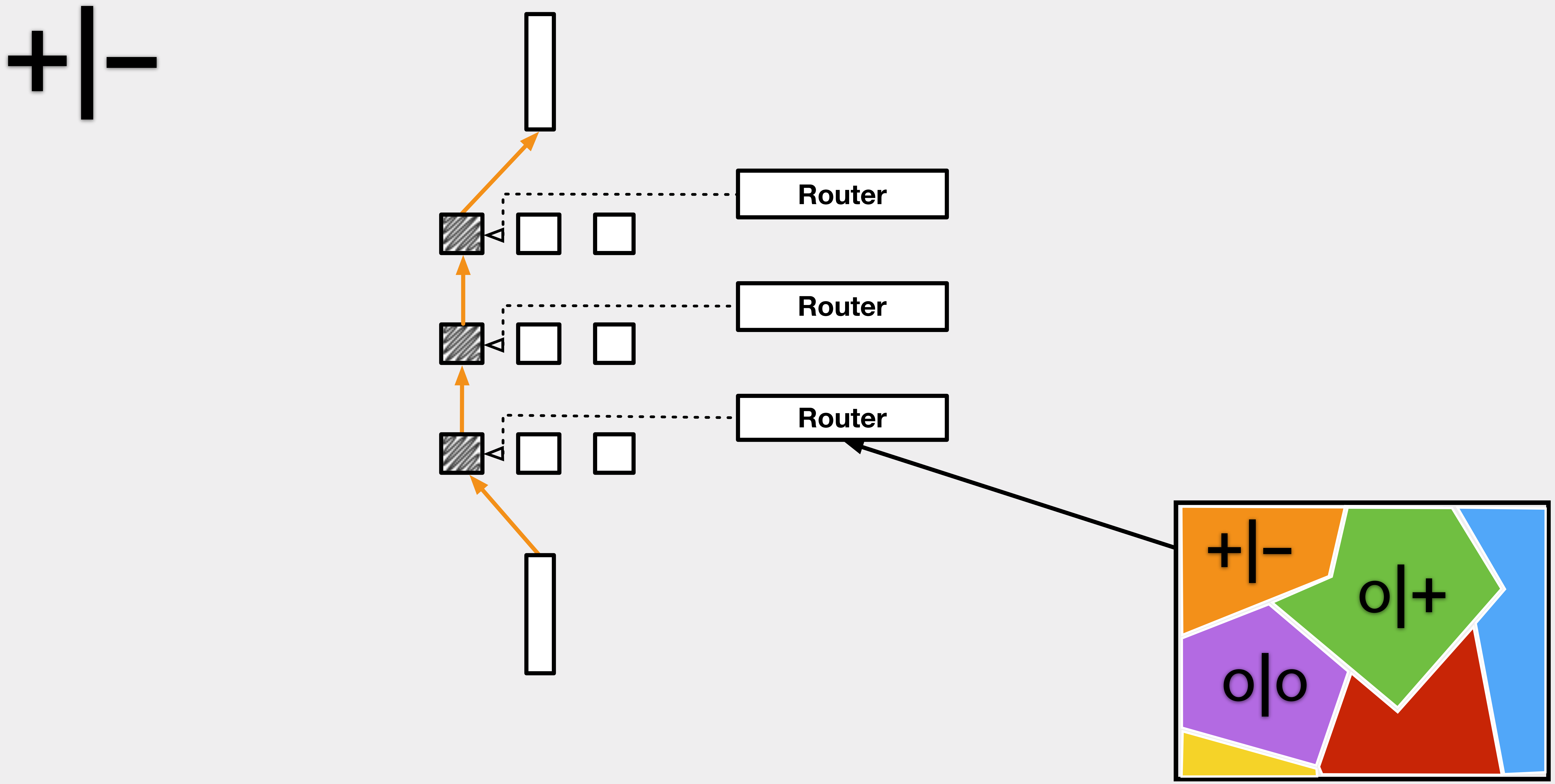
Signatures as Meta-Information

Modular and Compositional Learning for Natural Language Understanding



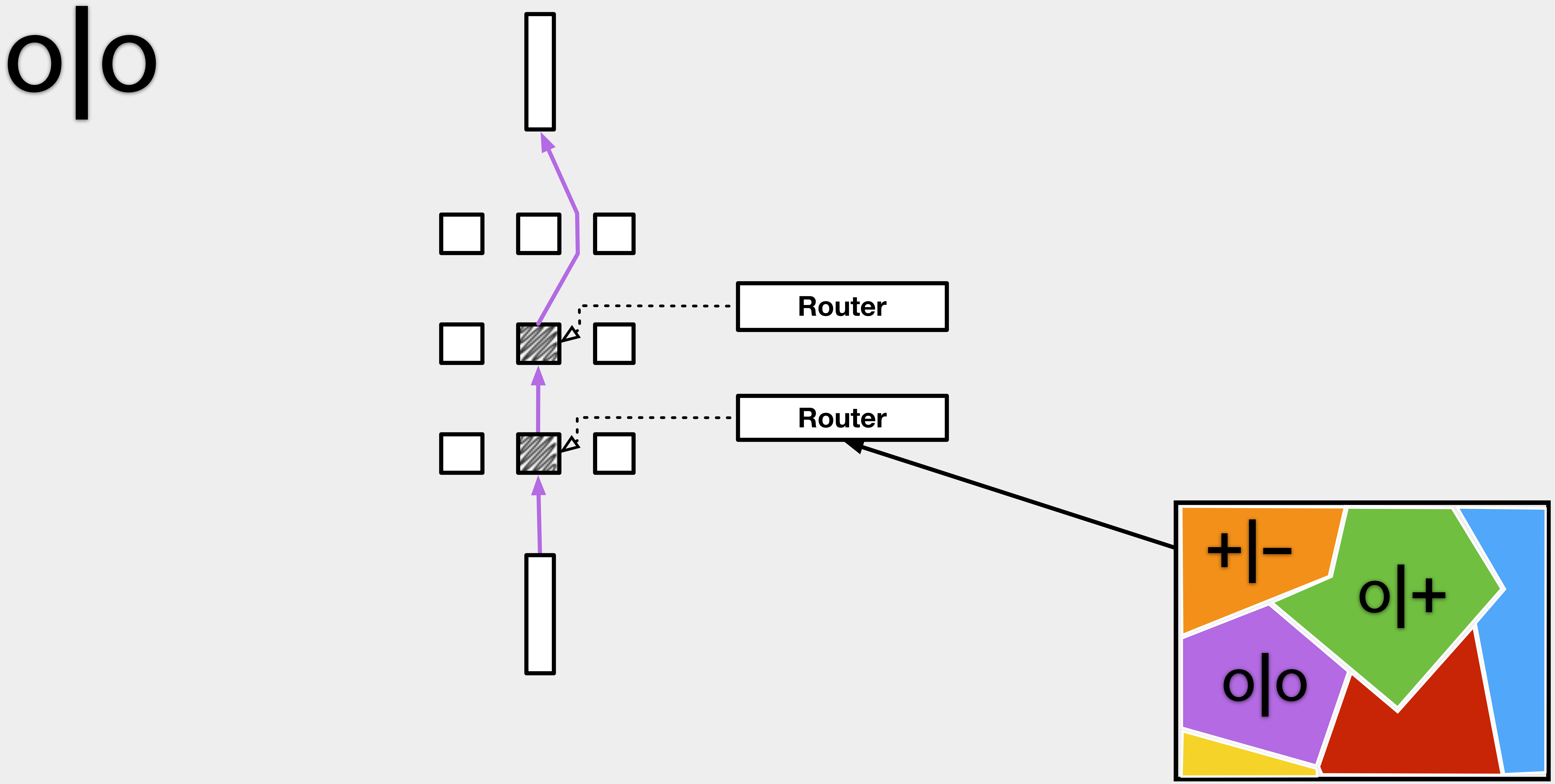
Signatures as Meta-Information

Modular and Compositional Learning for Natural Language Understanding



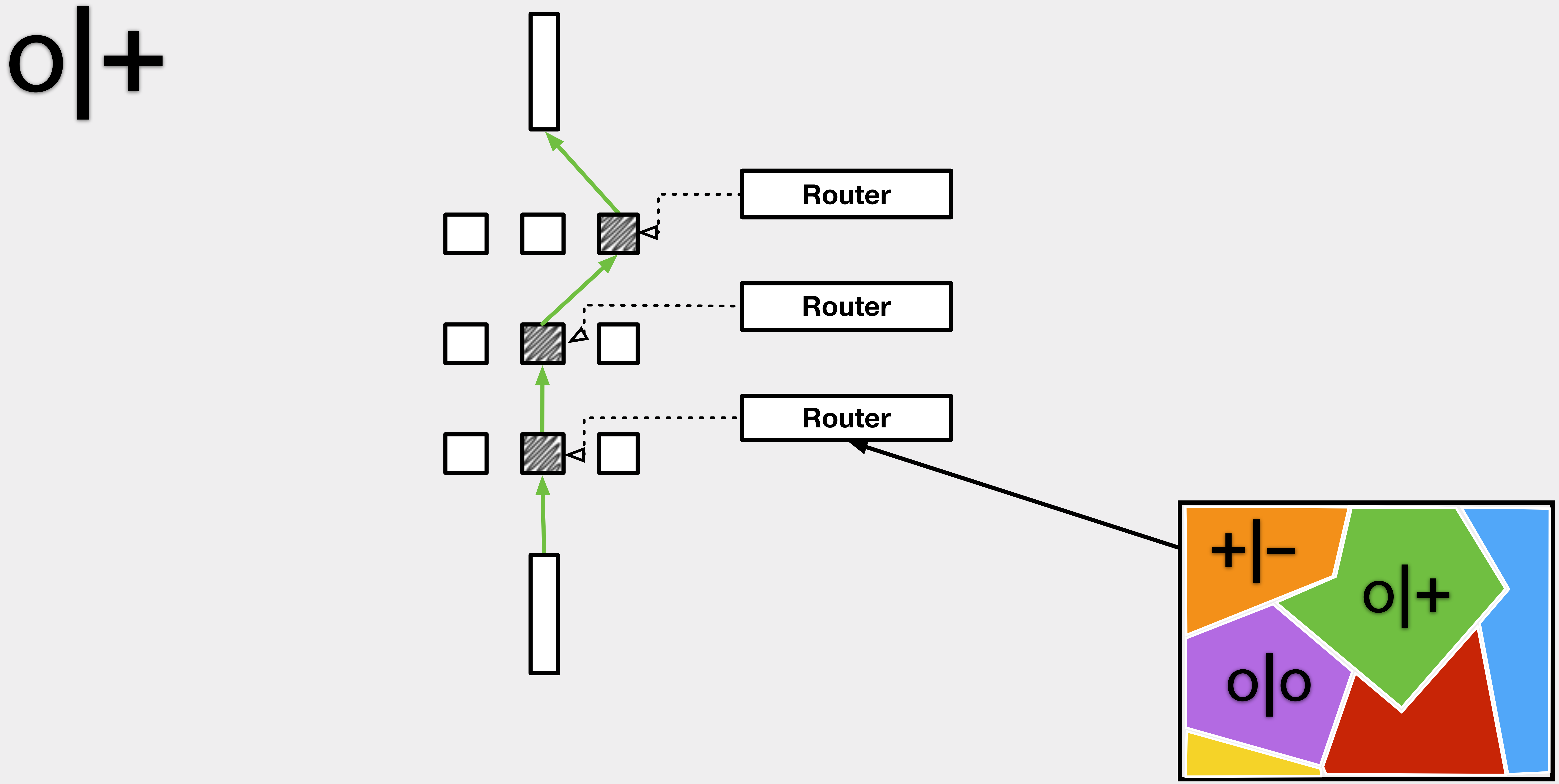
Signatures as Meta-Information

Modular and Compositional Learning for Natural Language Understanding

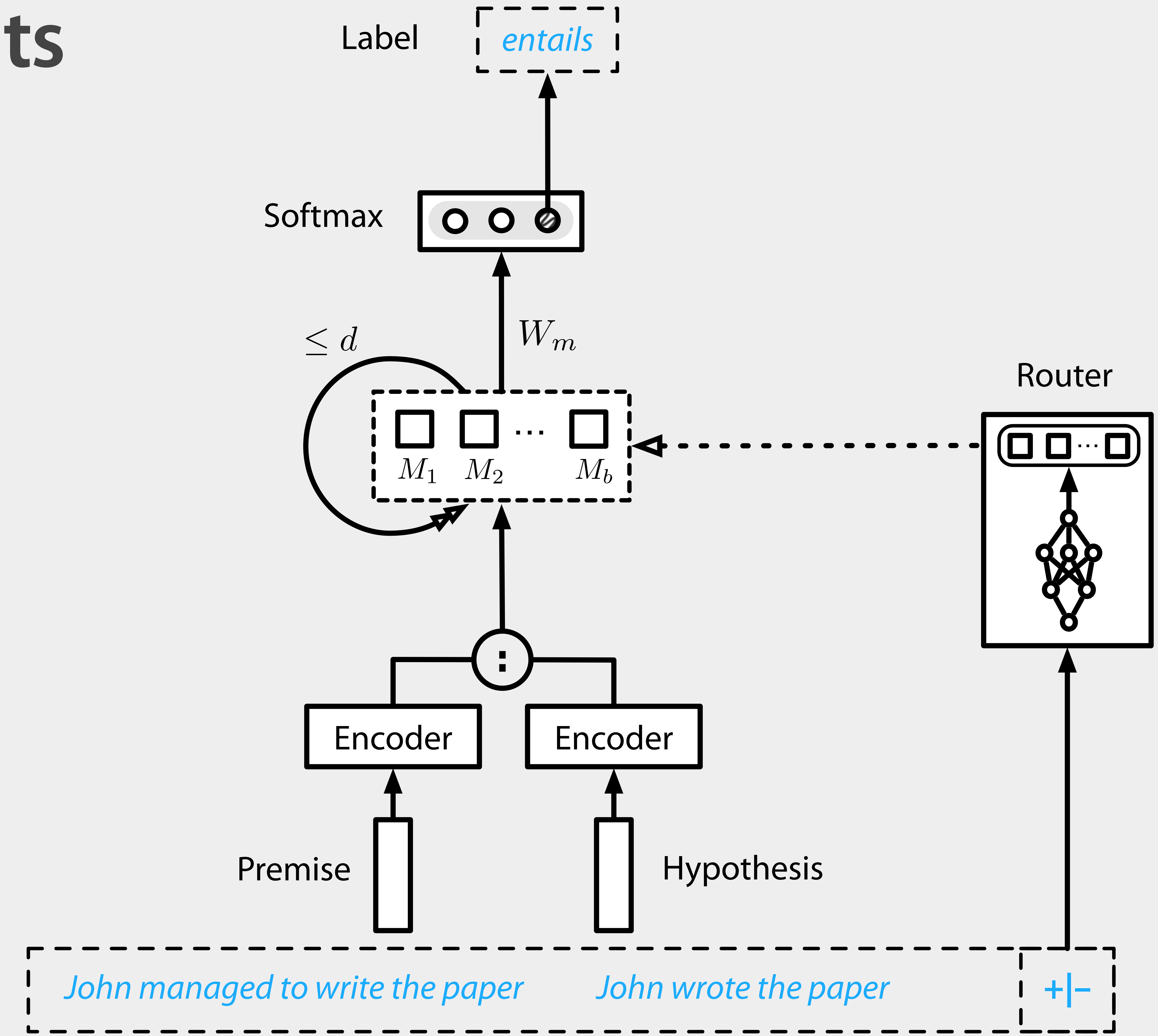


Signatures as Meta-Information

Modular and Compositional Learning for Natural Language Understanding



Experiments



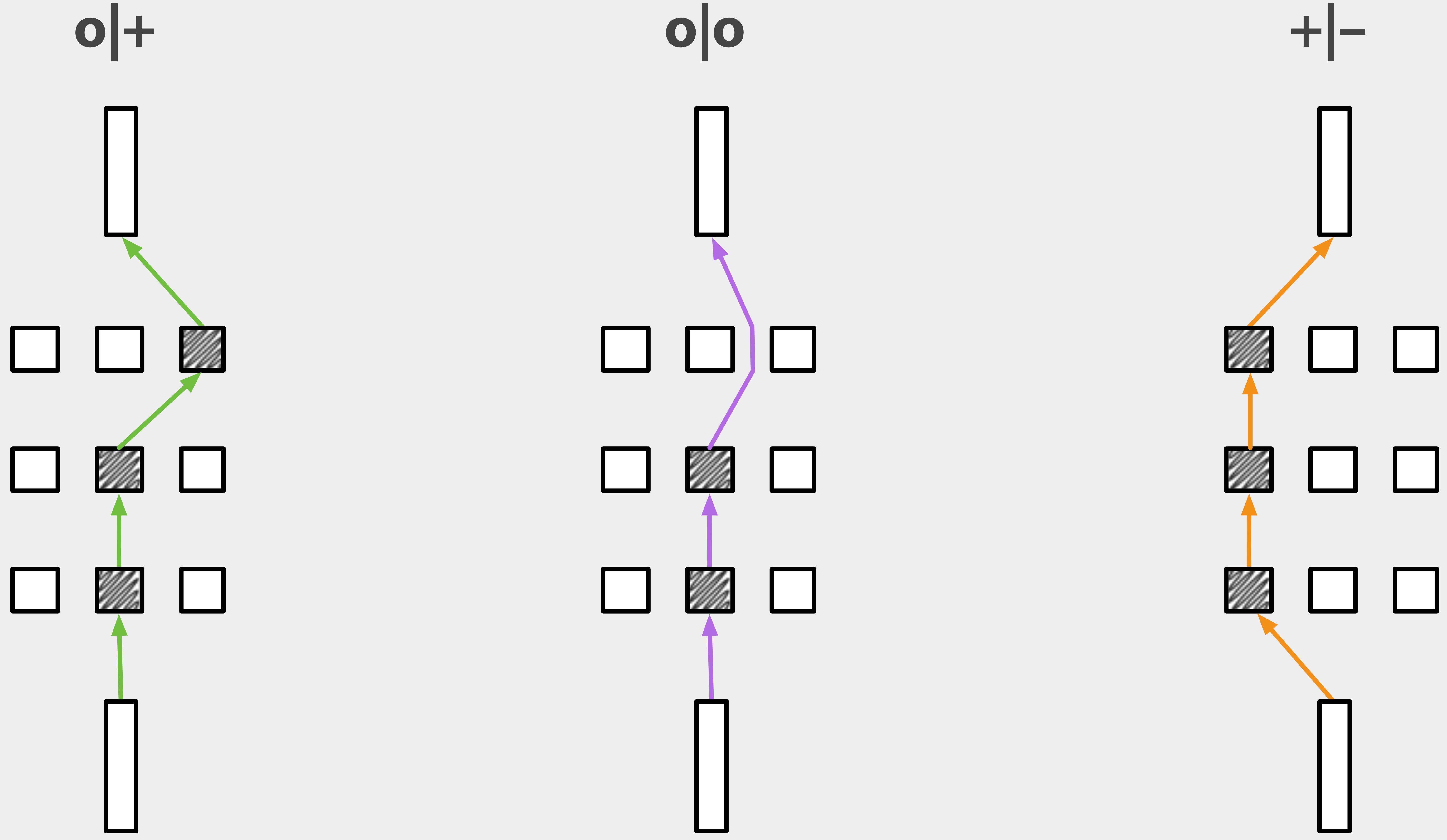
Results SCI

Modular and Compositional Learning for Natural Language Understanding

		SCI			
Embedding	Routing	joint	disjoint	mismatch	nested*
Glove	None	57.26±0.18	55.68±0.47	53.41±0.86	50.98±0.92
Glove	WP+D	**75.56**±0.77	**74.87**±0.49	**71.08**±0.52	**75.43**±0.29

Navigating the Transfer–Interference Trade-off

Modular and Compositional Learning for Natural Language Understanding

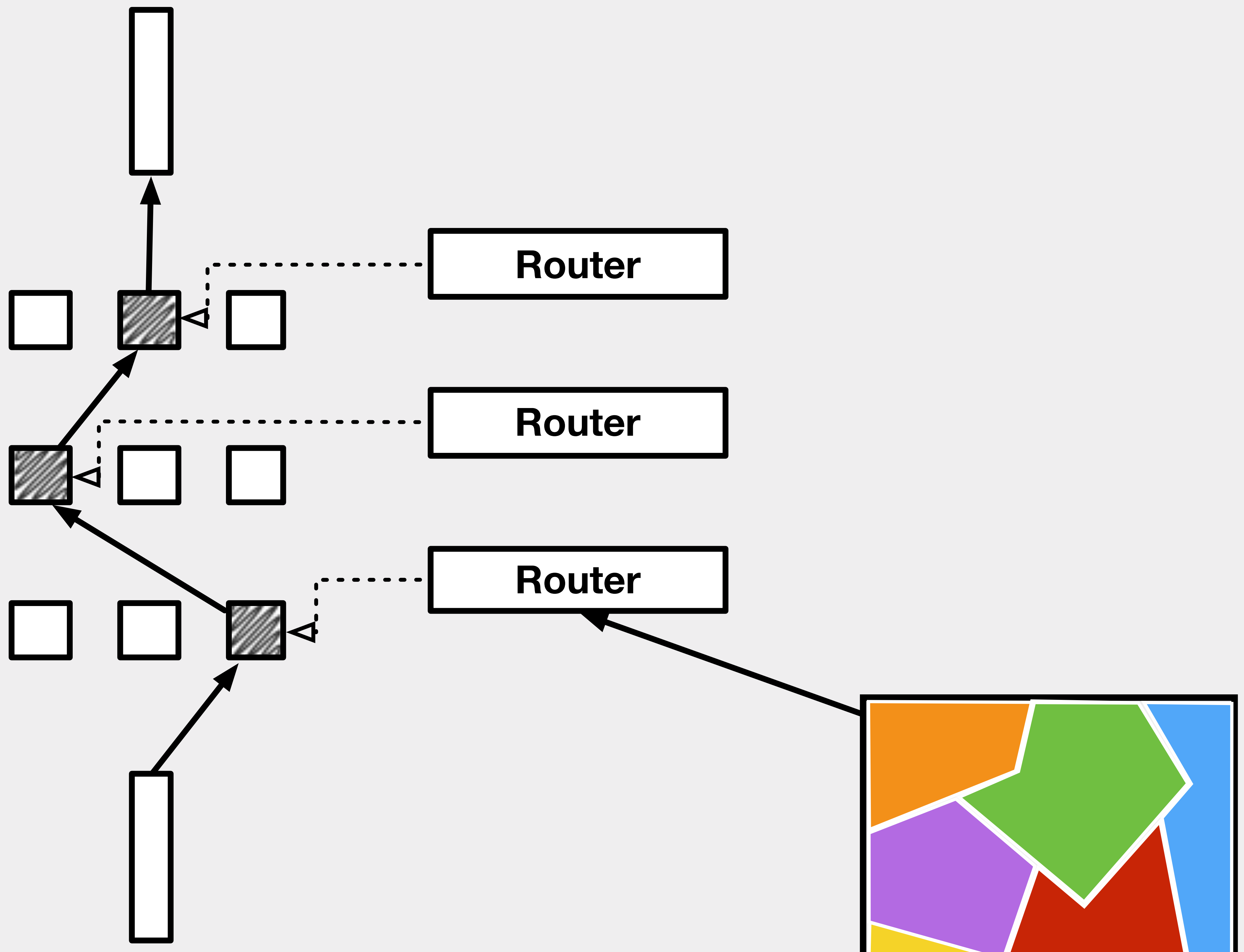


take vow

take opportunity

From Meta-information to Dispatching

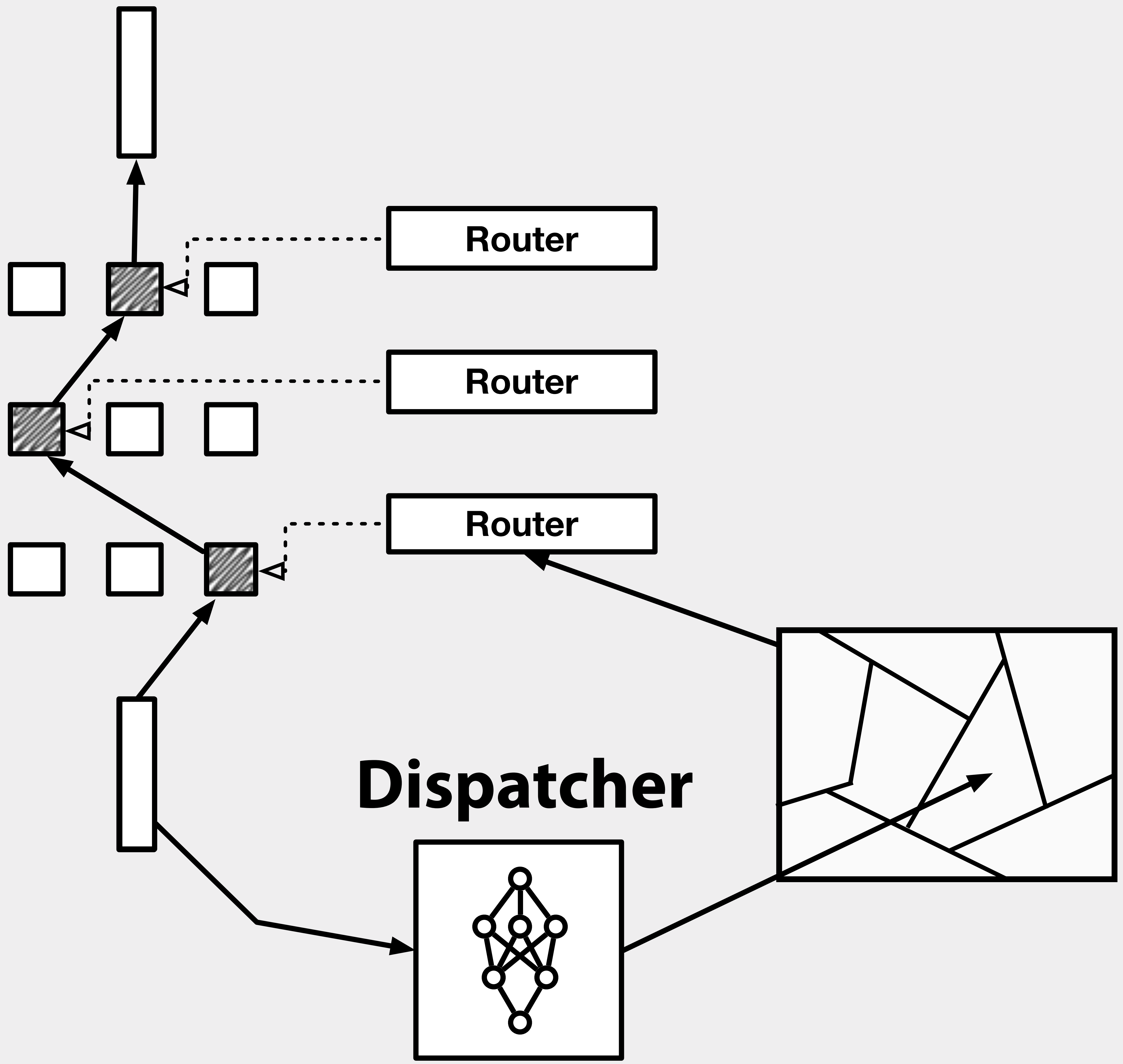
Modular and Compositional Learning for Natural Language Understanding



signatures
as **meta-information**

From Meta-information to Dispatching

Modular and Compositional Learning for Natural Language Understanding

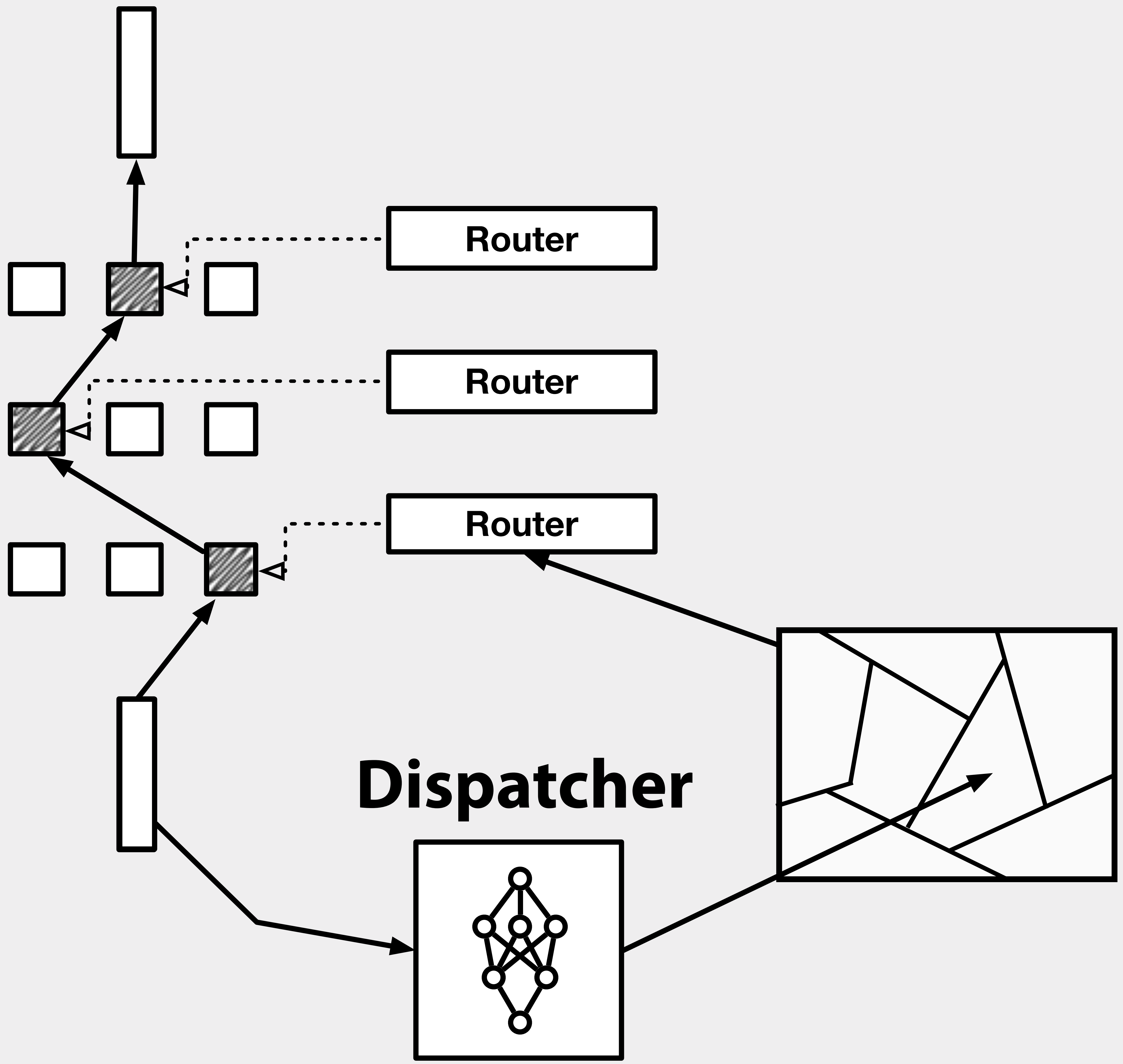


can we train an RRN
so that it does **not rely on meta-information at all?**

?

From Meta-information to Dispatching

Modular and Compositional Learning for Natural Language Understanding



Results SCI

Modular and Compositional Learning for Natural Language Understanding

Routing	SCI joint	SCI disjoint	SCI mismatch	SCI nested*
None	57.26±0.18	55.68±0.47	53.41±0.86	50.98±0.92
WP+D	**75.56**±0.77	**74.87**±0.49	**71.08**±0.52	**75.43**±0.29
Basic Dispatch	**72.7**	**65.05**	67.74	80.35
AE Dispatch	71.4	64.13	**67.7**	**83.24**

Summary

Modular and Compositional Learning for Language Understanding

Transfer–Interference Trade-off

- Introduces the transfer-interference trade off.
- Meta-Experience Replay as a solution to navigate the trade-off.
- Unification of continual learning and meta-learning.
- Matt Riemer, Ignacio **Cases**, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro. *ICLR 2019*.

Recursive Routing Networks

- Recursive models that learn to compose modules.
- Flexibly incorporated in other models.
- Leverage meta-information.
- End-to-end learning.

Summary

Modular and Compositional Learning for Language Understanding

Stanford Corpus of Implicatives

- Natural Language Inference dataset of implicative constructions with ~11K examples.
- 90+ implicative constructions with compositional *signatures*.
- Identified a new type of implicative signature.
- Ignacio **Cases**, Clemens Rosenbaum, Matthew Riemer, Atticus Geiger, Tim Klinger, Alex Tamkin, Olivia Li, Sandhini Agarwal, Joshua D. Greene, Dan Jurafsky, Christopher Potts, and Lauri **Karttunen**. *NAACL 2019*.

Dispatching Routing Networks

- Relaxation in the need of meta-information / no need for meta-information.
- Full end-to-end training with three new types of dispatching mechanisms.
- Online dispatching with different objectives.
- Ignacio **Cases***and Clemens Rosenbaum*, and Matthew Riemer, Atticus Geiger, Lauri **Karttunen**, Joshua D. Greene, Dan Jurafsky, and Christopher Potts. *Stanford NLP Tech Report 2019-1*.

Desiderata II

Reasoning with Implicatives

4. Negation. Every statement contradicts its negation.

***5.* Symmetry of contradiction**. Whenever A contradicts B, B contradicts A as well, provided that A does not have presuppositions that B does not have. (“Strawson entailment”.)

6. Reflexivity and transitivity of entailment. Every statement entails itself.

7. Distant contradiction. If A entails C and B entails not C, will the model be able to conclude that A and B contradict each other?

THANKS!