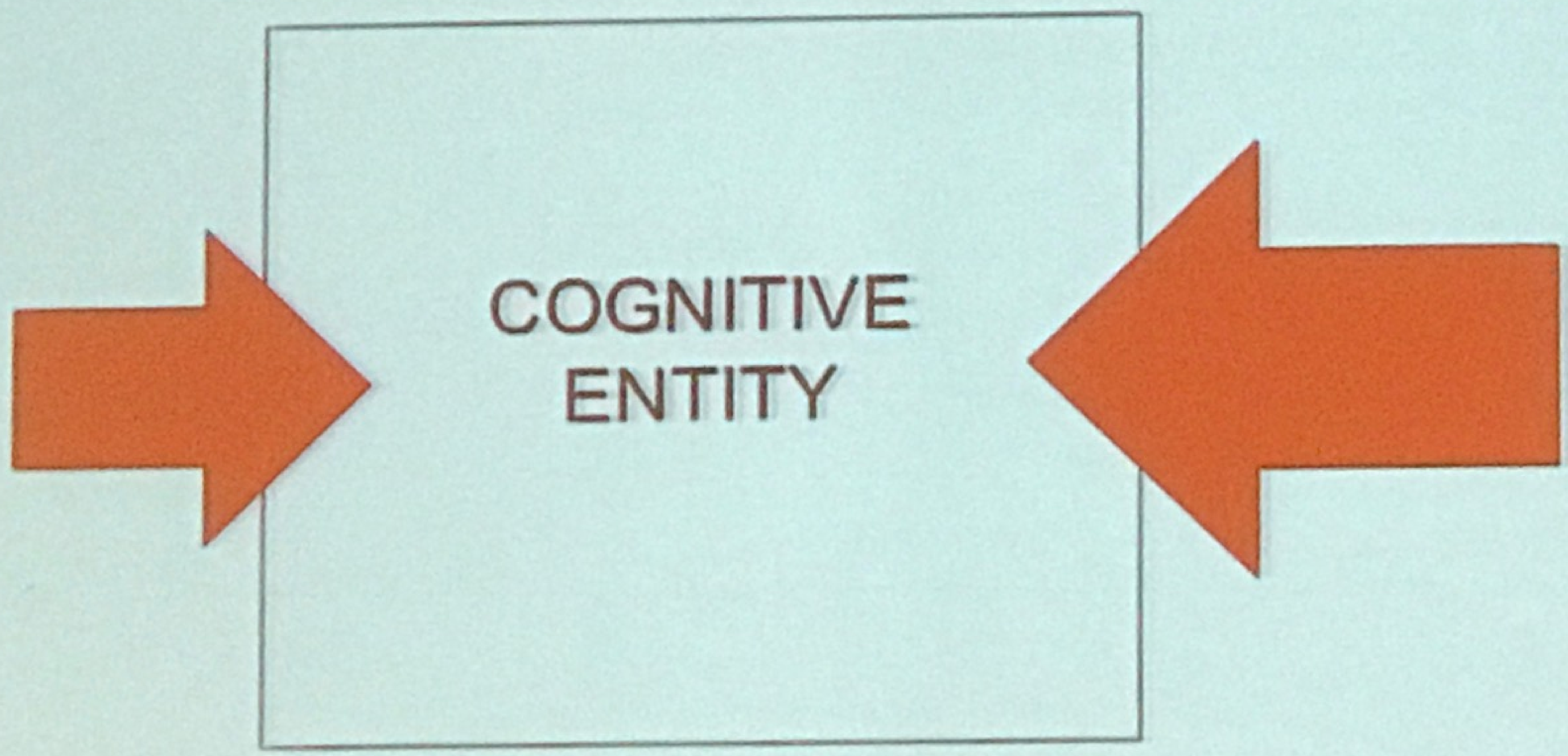
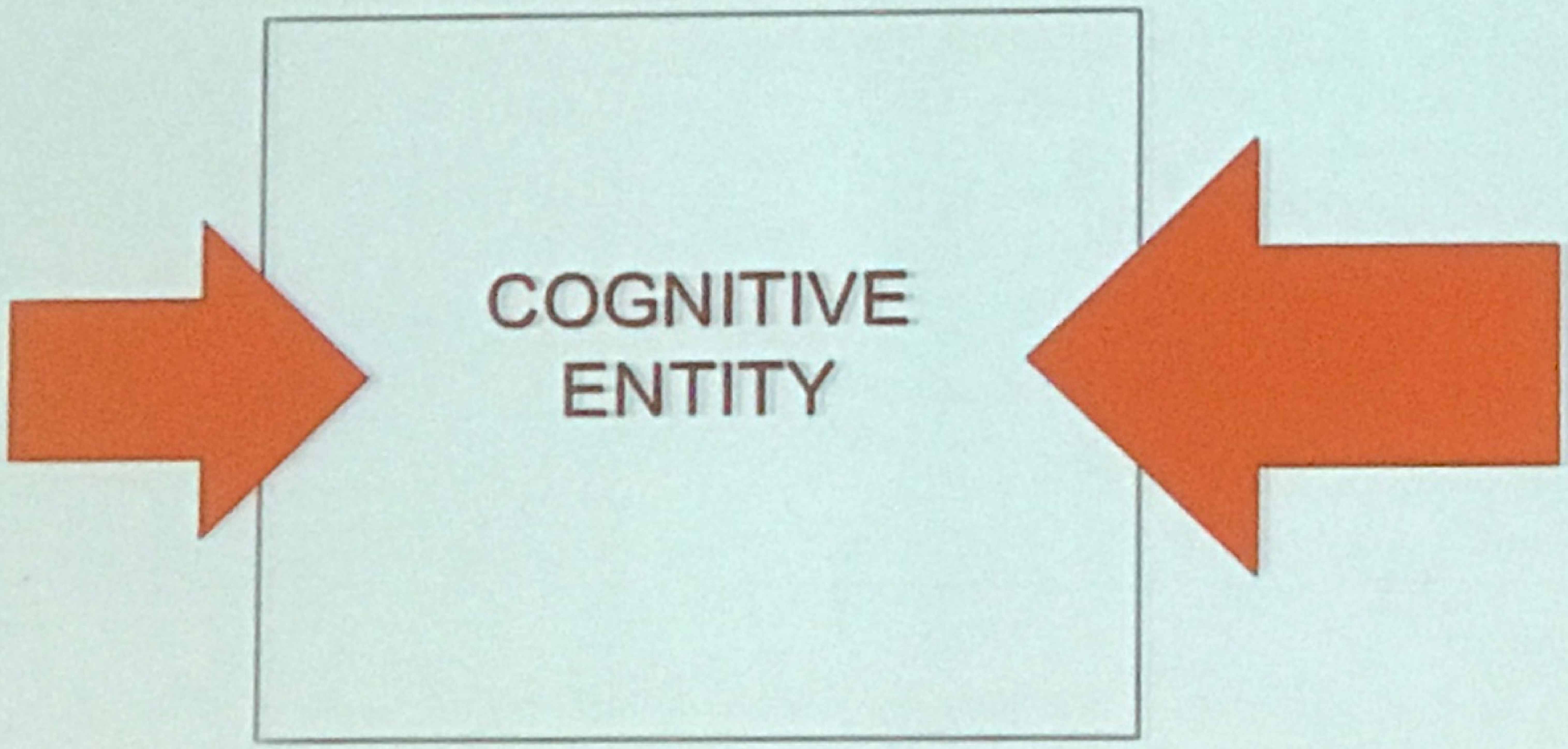
# What Needs to be Added to Machine Learning?

Leslie Valiant
Harvard University

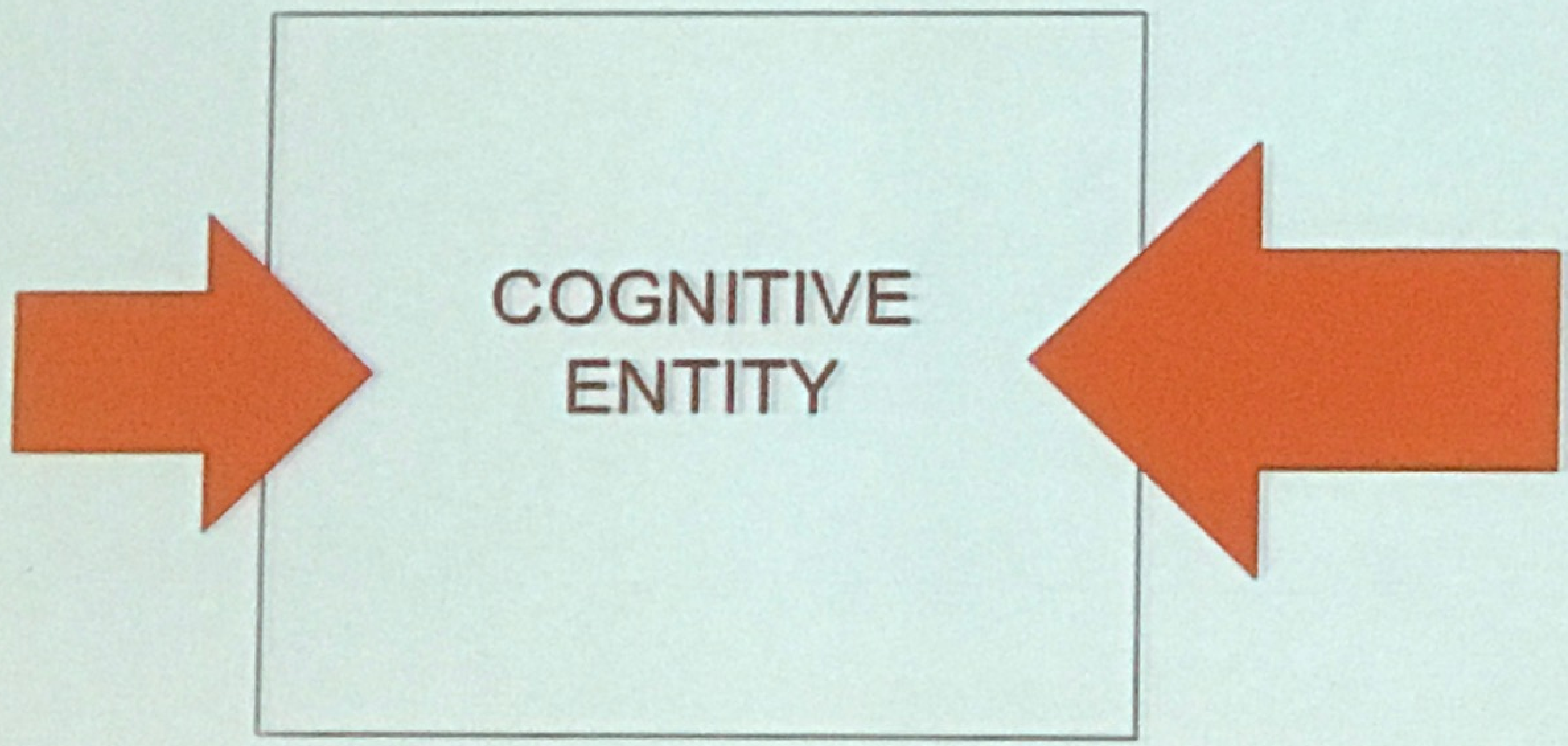
WCC, Poznan, September, 2018


COGNITIVE
ENTITY



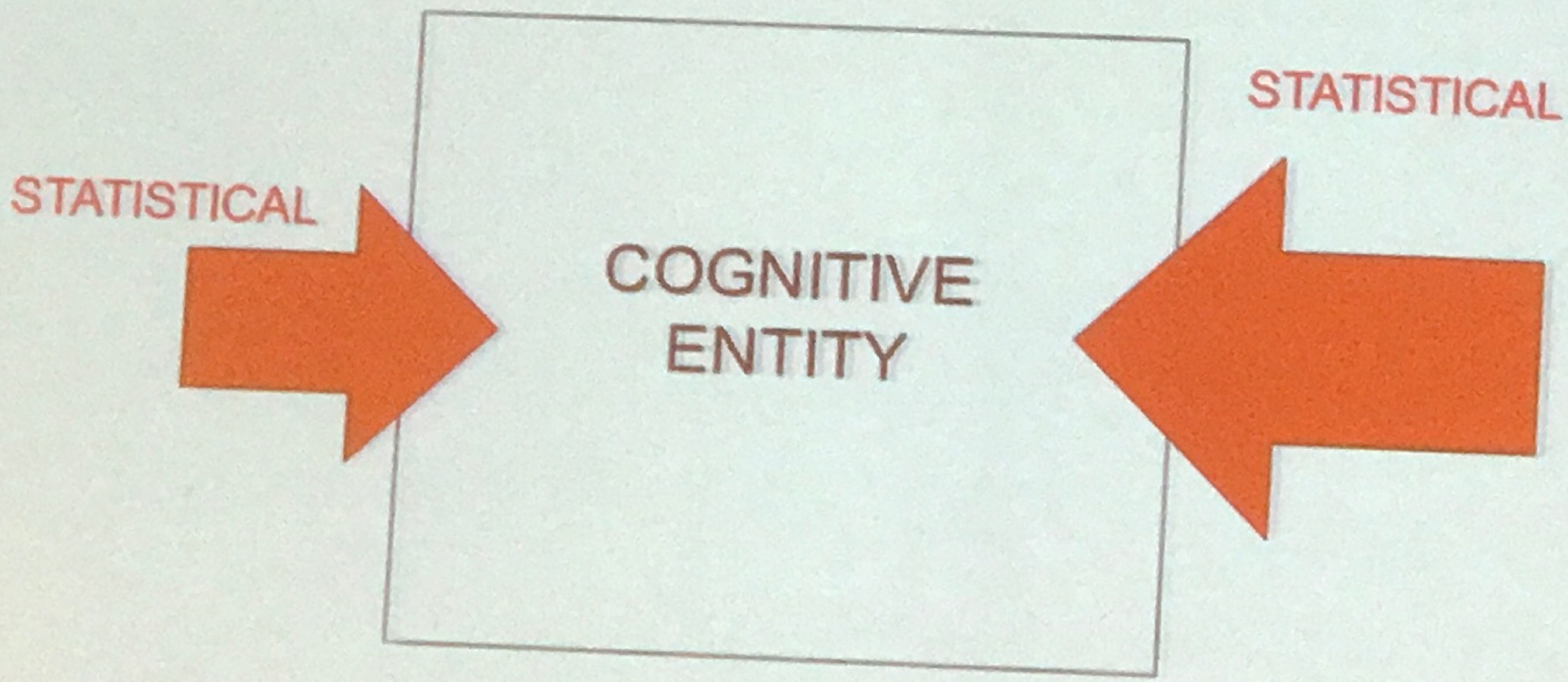

# COGNITION MUST HAVE A THEORY.

SEMANTICS: Correspondence between **State of Cognitive Entity** and **World**




**COGNITION MUST HAVE A THEORY.**

SEMANTICS: Correspondence between **State of Cognitive Entity** and **World**




**COGNITION MUST HAVE A THEORY.**

"All belief comes from Syllogism or Induction" - *Aristotle*

Syllogism ~ Mathematical Logic, Reasoning

Induction ~ Machine Learning Theory

"All belief comes from Syllogism or Induction" - *Aristotle*

Syllogism ~ Mathematical Logic, Reasoning

Induction ~ Machine Learning Theory

**Goals for our Time**

1. Give Learning and Reasoning a common semantics.

“All belief comes from Syllogism or Induction” - *Aristotle*

Syllogism ~ Mathematical Logic, Reasoning

Induction ~ Machine Learning Theory

**Goals for our Time**

1. Give Learning and Reasoning a common semantics.

2. Enable a robust database of commonsense knowledge to be acquired by learning for AI.

Did Aristotle own
a cell phone?

# Word Completion

(From China Daily, 2018)

# Word Completion

(From China Daily, 2018)

Whatever the Year of the Dog holds in store, pet owners will be lavishing more attention than ever on their pooches.

Google: "pet owners" "their pooches" 16,900

# Word Completion

(From China Daily, 2018)

Whatever the Year of the Dog holds in store, pet owners will be lavishing more attention than ever on their pooches.

Google: "pet owners" "their pooches" 16,900

China rises as a maritime powerhouse after snapping up profitable seaport terminals across the world.

# Word Completion

(From China Daily, 2018)

Whatever the Year of the Dog holds in store, pet owners will be lavishing more attention than ever on their pooches.

Google: "pet owners" "their pooches" 16,900

China rises as a maritime powerhouse after snapping up profitable seaport terminals across the world.

Google: "maritime powerhouse" "seaport terminals" 6

Macao's retail sale up by 23.7 pct in Q2.

# Word Completion

Word Completion has two commonalities with Turing Test:

1. Performance measured quantitatively.
2. Knowledge unrestricted (no microworld.)

# Word Completion

Word Completion has two commonalities with Turing Test:

1. Performance measured quantitatively.
2. Knowledge unrestricted (no microworld.)

Learning Theory approach:

3. Performance relative to particular world (i.e. distribution *D* of natural inputs.)

# Word Completion

Word Completion has two commonalities with Turing Test:

1. Performance measured quantitatively.
2. Knowledge unrestricted (no microworld.)

Learning Theory approach:

3. Performance relative to particular world (i.e. distribution *D* of natural inputs.)
4. Insists on feasible computation and error control.

# Robust Logic

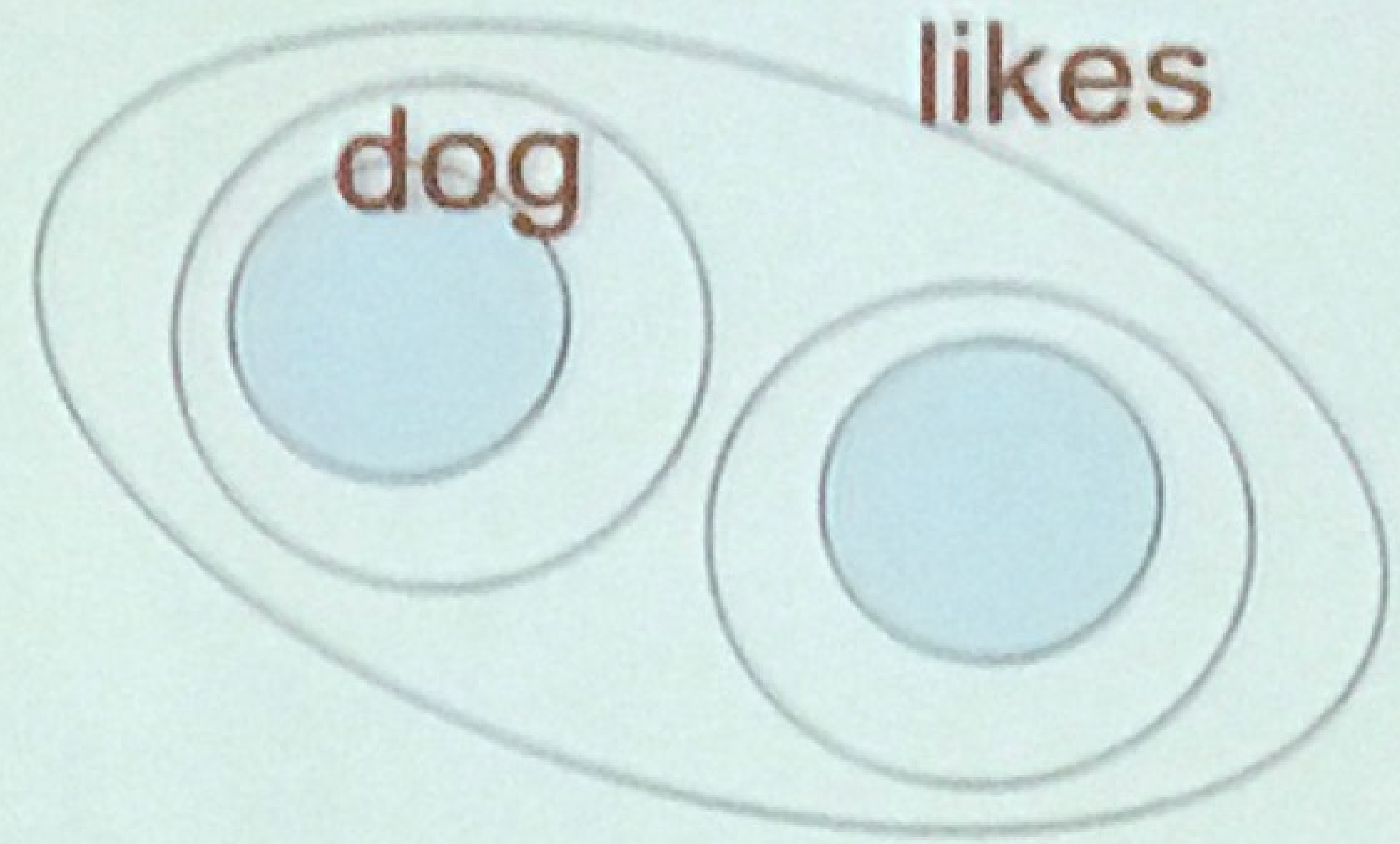
(Semantics for unifying learning and reasoning.)
(Can test on word completion.)

*An Idea From Cognitive Science:*

Working Memory ≈ Mind's eye

WCC 2018
POZNAŃ
THE 24th IFIP
World Computer Congress

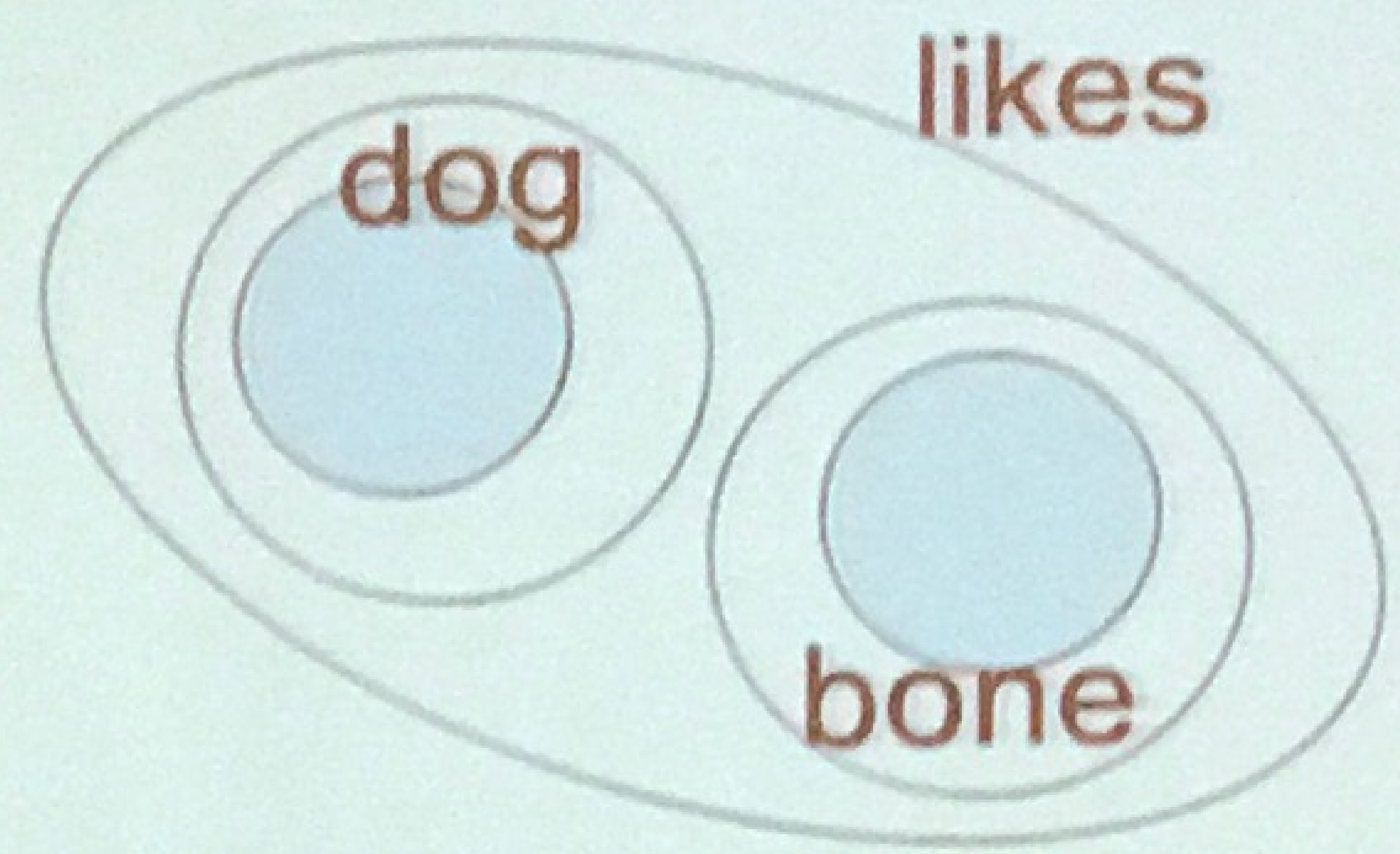
# A Scene




## A Logical Rule:

$\exists x_1 Dog(x_1) \wedge Likes(x_1,x_2) \rightarrow Bone(x_2)$

# A Scene




## A Logical Rule:

$$\exists x_1 \mathrm{Dog}(x_1) \wedge \mathrm{Likes}(x_1, x_2) \rightarrow \mathrm{Bone}(x_2)$$

## A Learned Rule:

"Expression from learnable class" $\equiv \mathrm{Bone}(x_2)$


Computers have to ACQUIRE knowledge, and then ACT on it.
Type A
Type B
Accept
Labeled
Examples
?
Classifies
Unlabeled
Examples
ifip
PTI
ODDZIAŁ WIELKOPOLSKI
WCC
2018
POZNAŃ
THE 24th IFIP
World Computer Congress
ifip
PAN
PTI


SUPERVISED LEARNING
Computers have to ACQUIRE knowledge, and then ACT on it.
Type A
Type B
Accept
Labeled
Examples
?
Classifies
Unlabeled
Examples
There is a science of this.
ifip
PAN
PTI
ODDZIAŁ WIELKOPOLSKI
WCC
2018
POZNAŃ
THE 24th IFIP
World Computer Congress
ifip
PAN
PTI
ODDZIAŁ WIELKOPOLSKI

# SUPERVISED LEARNING

Computers have to ACQUIRE knowledge, and then ACT on it.

Type A 

Type B 

Accept Labeled Examples

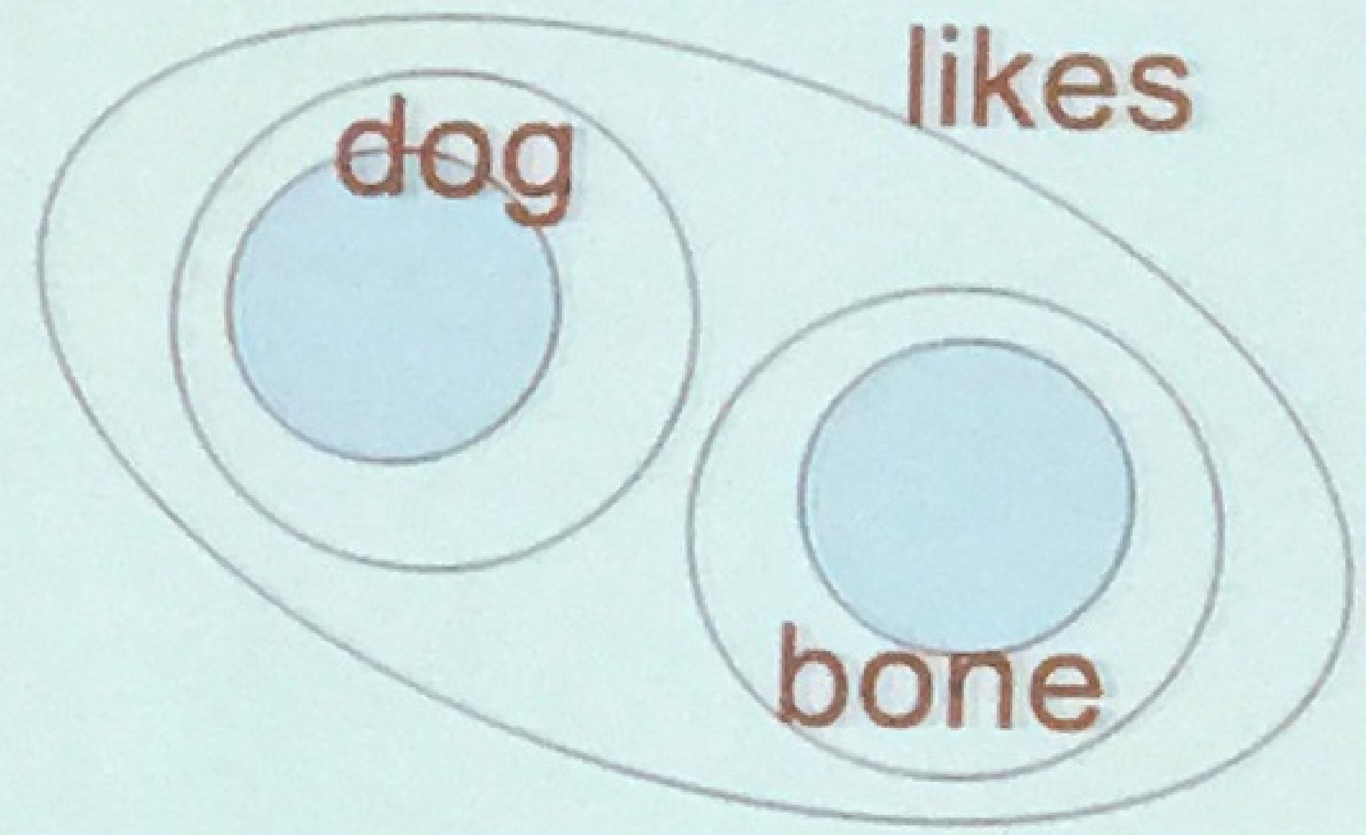
? 

Classifies Unlabeled Examples

There is a science of this.


Prediciton
Error
Decreases as fast as $N^{-\beta}$
$N$ = #computational steps
To learn to error 1/100 need only $N = 100^{1/\beta}$
for *some* constant, such as $\beta = 0.5$.
ifip
PAN
PTI
ODDZIAŁ WIELKOPOLSKI
WCC
2018
POZNAŃ
THE 24th IFIP
World Computer Congress
ifip
PAN
PTI
ODDZIAŁ WIELKOPOLSKI

# Robust Logic Feature 2:
## Quantifiers are Local




### A Logical Rule:

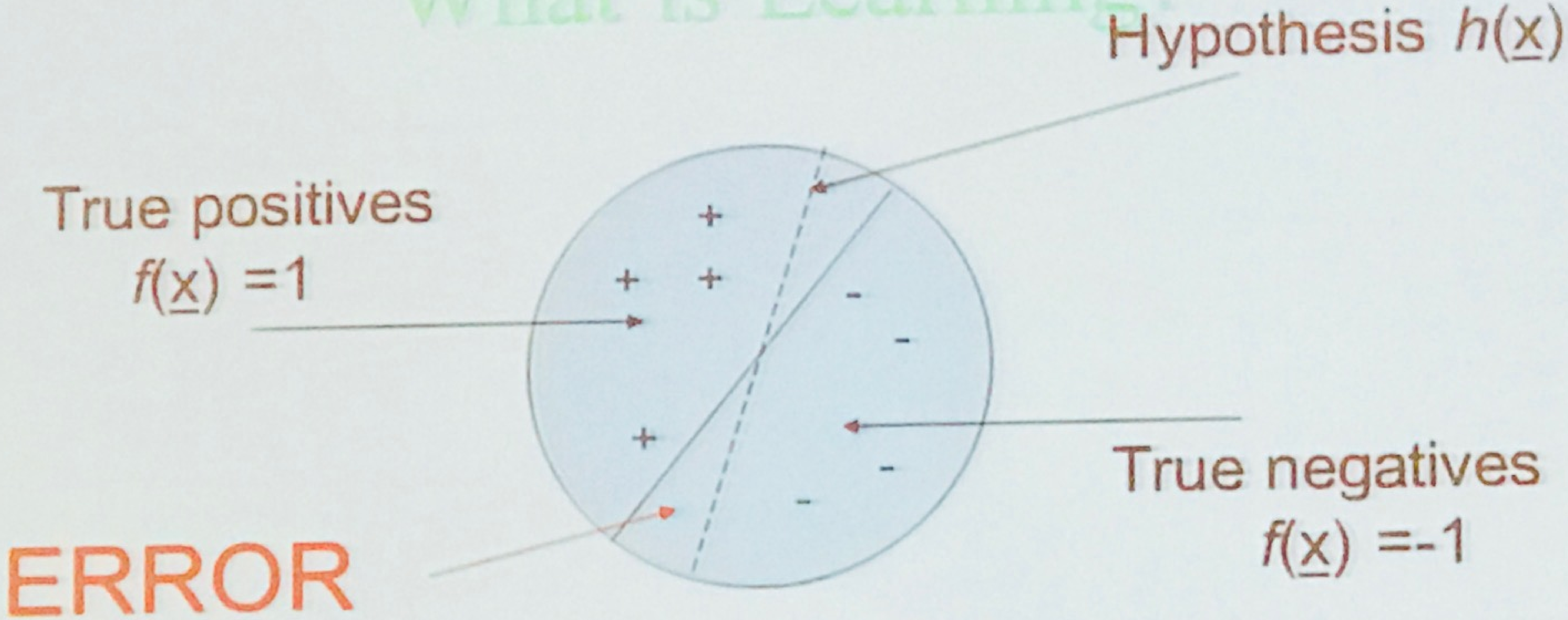
$$\exists x_1 Dog(x_1) \wedge Likes(x_1,x_2) \rightarrow Bone(x_2)$$

### A Learned Rule:

"Expression from learnable class" $\equiv Bone(x_2)$

# Robust Logic Feature 4:

## What is Learning?




PAC Learning

1. Examples from *arbitrary distribution D*, but testing from same.
2. Errors can be reduced *arbitrarily* with more examples and more computation, and with only polynomial cost in these.

Robust Logic Feature 5:
What is world D?
ifip
PAN
PTI
ODDZIAŁ WIELKOPOLSKI
WCC
2018
POZNAŃ
THE 24th IFIP
World Computer Congress
POLISH
INFORMATION
PROCESSING
SOCIETY
WIELKOPOLSKA BRANCH

# Robust Logic Feature 5:

## What is world D?

Examples in Mind's Eye!

# Robust Logic Feature 5:

## What is world D?

Examples arise from ground truth vectors:

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, \ldots\ldots, x_{1000})$.

But this is about **incomplete specification**.

An example is (1,*,0,0,*,1,0,*,*,*, ….. ,*)

D over {0,1,*}.

# Robust Logic Feature 5:

## What is world D?

Examples arise from ground truth vectors:

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, \ldots\ldots, x_{1000})$.

But this is about **incomplete specification**.

An example is (1,*,0,0,*,1,0,*,*,*, ….. ,*)

D over {0,1,*}.

E.g. bird = 1, penguin = * → fly = 1.

(No problem with nonmonotonicity.)

# Robust Logic Feature 5:

## What is world D?

Examples arise from ground truth vectors:

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, \ldots\ldots, x_{1000})$.

But this is about **incomplete specification**.

An example is (1,*,0,0,*,1,0,*,*,*, ….. ,*)

D over {0,1,*}.

E.g. bird = 1, penguin = * → fly = 1.

(No problem with nonmonotonicity.)

To predict $x_6$ apply rules with entry $x_6$ **obscured**:

e.g. (1,*,0,0,*,?,0,*,*,*, ….. ,*).

ifip

WCC

ifip

# Robust Logic Feature 6: Is Not Hierarchical Learning Difficult?

e.g. to learn $F(x_1, x_2)$, and $G(F(x_1, x_2), x_3)$ while F is inaccurately known.

# Robust Logic Feature 6: Is Not Hierarchical Learning Difficult?

e.g. to learn $F(x_1, x_2)$, and $G(F(x_1, x_2), x_3)$ while F is inaccurately known.

Consider that examples come with correct labels. (As in learning math from teacher.)

# Upshot

Will learn rules like

# Upshot

# Upshot

Will learn rules like:

[illegible]

using the most appropriate supervised learning algorithm, [illegible] [illegible] each word in dictionary or [illegible], and chain these for reasoning.

# The Experiment (L. Michael & L.V.)

500,000 sentences from Wall Street Journal

Use mechanical parser to find subject/object etc.

(Some programmed knowledge: Wordnet dictionary)

# The Experiment (L. Michael & L.V.)

500,000 sentences from Wall Street Journal

Use mechanical parser to find subject/object etc.

(Some programmed knowledge: Wordnet dictionary)

Learn rules from single sentences to predict unstated things.

# The Experiment (L. Michael & L.V.)

500,000 sentences from Wall Street Journal

Use mechanical parser to find subject/object etc.

(Some programmed knowledge: Wordnet dictionary)

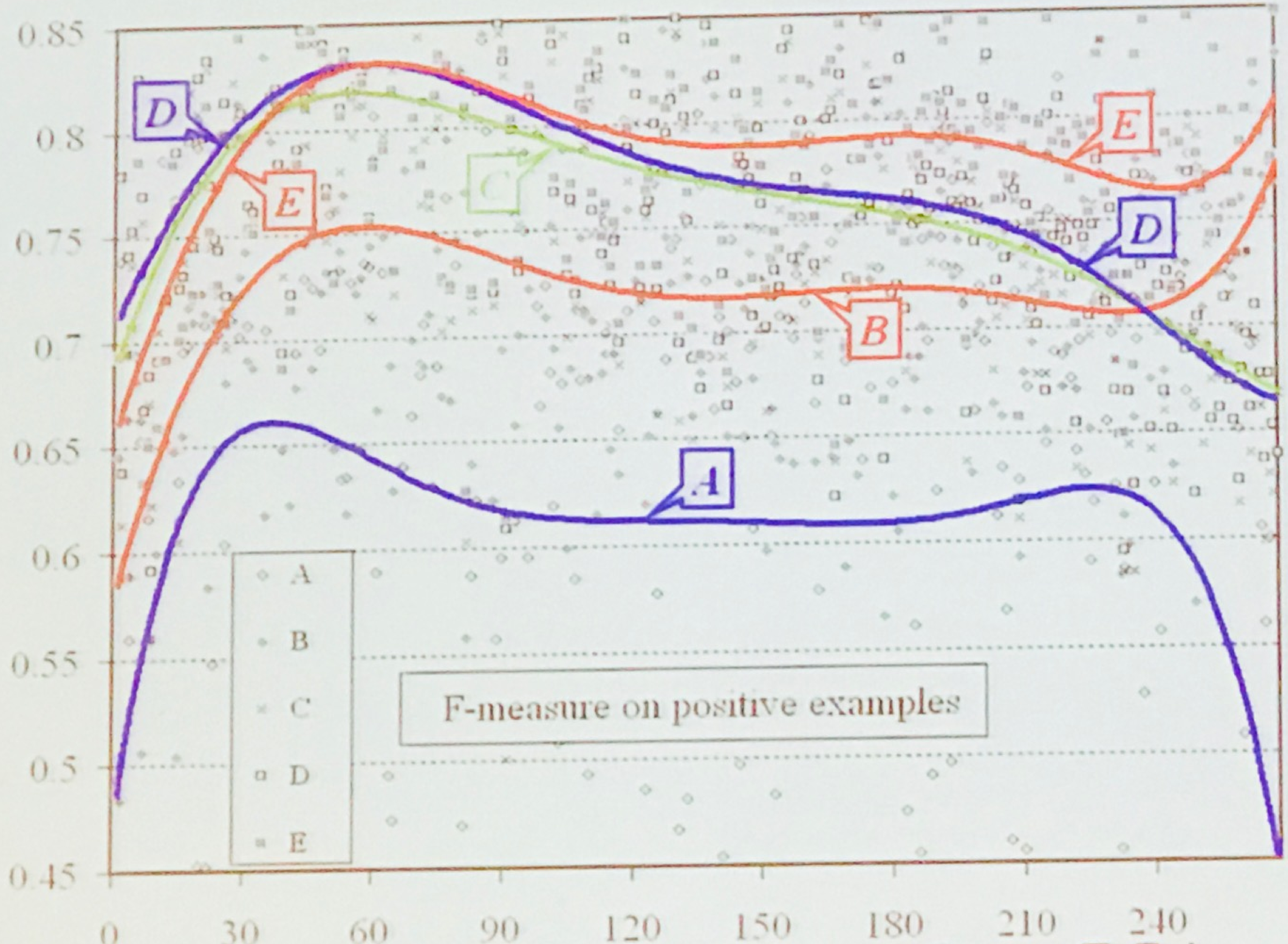
Learn rules from single sentences to predict unstated things.

See how much better chaining these is in predicting missing word, than a baseline "syntactic" method.

# The Experiment (L. Michael & L.V.)

Learned rule are *linear inequalities on compound features*: e.g.

"1.09 * (∃ x: (x lowers y) ∧ (bargain x)) +
0.99 * (∃ x: (x lowers y) ∧ (competition x)) + 0.51 *
(∃ x: (y lowers x) ∧ (demand x)) +
………… > 1"  ≡  (price y).




D = without chaining. E = with chaining. Order E-D increasing

# Robust Logic

**Examples**: Scenes on $t$ objects.

**RHS of rule**: e.g. $R(x_3, x_4)$.

**LHS of rule**: Linear inequality over compound features. (or any other learnable class.)

**Compound features**, local quantifiers

e.g. $\exists x_1 \forall x_2 A(x_1, x_2, x_3) B(x_1, x_4)$.

**Propositional learning algorithm**: to inherit good quantitative characteristics, such as error resilience, attribute efficiency.

e.g. Winnow (Littlestone), deep learning.

# Desirable Properties of Robust Logic

**Theorems**: Rules learnable; chaining of rules applied to one example is polynomial time, sound and complete. (All in pac-semantics.)

# Desirable Properties of Robust Logic

**Theorems:** Rules learnable; chaining of rules applied to one example is polynomial time, sound and complete. (All in pac-semantics.)

**Restriction:** relations of constant arity.

**Otherwise:** run time, error bounds polynomial in rule sizes, rule numbers, number *t* of objects in scenes.

**Here all of:** learning, reasoning, multi-objec

# A Challenge: Developing Good Teaching Materials

"grey elephant":

"pink elephant":

"white elephant":

# A Challenge: Developing Good Teaching Materials

| | (Google 5-11-18) |
|---|---|
| "grey elephant": | 1,370,000 |
| "pink elephant": | 546,000 |
| "white elephant": | 707,000 |

# Challenges for AI that are addressed by Robust Logic

**Infusing machines with general knowledge so that robust reasoning on it is possible.**

- Learning enables system to stay "approximately consistent".
- Learning as method of choice for knowledge that is nowhere defined.
- Principled reasoning on learned knowledge.