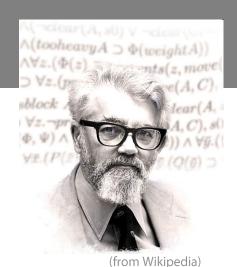
# Artificial Intelligence

Introduction
Physical Symbol Systems

Marco Piastra

# Artificial Intelligence: the origins

# "Artificial Intelligence" (first appearance of the term)



[John McCarthy et al., 1955]

"We propose that a two-month, ten man study of **artificial intelligence** carried out during the summer of 1956 [...]

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. [...]

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture."

# The Physical Symbol System Hypothesis (PSSH)

Newell, A., Simon, H., Computer Science as Empirical Inquiry Symbols and Search, 1976

"A **physical symbol system** consists of a set of entities, called **symbols**, which are physical patterns that can occur as components of another type of entity called an **expression** (or symbol structure).

Thus, a **symbol structure** is composed of a number of **instances** (or tokens) of symbols related in some physical way (such as one token being next to another).

At any instant of time the system will contain a collection of these symbol structures.

Besides these structures, the system also contains a collection of **processes** that operate on *expressions* to produce other *expressions*: processes of creation, modification, reproduction and destruction."

# Symbolic Computation

Solving quadratic equations

$$x^{2} + ax + b = 0$$

$$x^{2} + 2(a/2)x + a^{2}/4 - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} = a^{2}/4 - b$$

$$x = -a/2 \pm (a^{2}/4 - b)^{1/2}$$

A sequence of steps: at each step a transformation rule is applied

Solving quadratic equations

$$x^{2} + ax + b = 0$$

$$x^{2} + 2(a/2)x + a^{2}/4 - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} = a^{2}/4 - b$$

$$x = -a/2 \pm (a^{2}/4 - b)^{1/2}$$

A sequence of steps: at each step a transformation rule is applied

### From start to end

Start point: a premise (i.e. we assume the truth of something)

End point: a conclusion (i.e. we state that something else must <u>also</u> be true)

(both points are decided by us)

Solving quadratic equations

$$x^{2} + ax + b = 0$$

$$x^{2} + 2(a/2)x + a^{2}/4 - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} = a^{2}/4 - b$$

$$x = -a/2 \pm (a^{2}/4 - b)^{1/2}$$

A sequence of steps: at each step a transformation rule is applied

### From start to end

Start point: a premise (i.e. we assume the truth of something)

End point: a conclusion (i.e. we state that something else must <u>also</u> be true)

(both points are decided by us)

Solving quadratic equations

$$x^{2} + ax + b = 0$$

$$x^{2} + 2(a/2)x + a^{2}/4 - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} = a^{2}/4 - b$$

$$x = -a/2 \pm (a^{2}/4 - b)^{1/2}$$

A sequence of steps: at each step a transformation rule is applied

### From start to end

Start point: a premise (i.e. we assume the truth of something)

End point: a conclusion (i.e. we state that something else must <u>also</u> be true)

(both points are decided by us)

### Sequence of steps

Each transformation rule is based on equivalence

Each step must be correct, in this particular sense

Solving quadratic equations

$$x^{2} + ax + b = 0$$

$$x^{2} + 2(a/2)x + a^{2}/4 - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} - a^{2}/4 + b = 0$$

$$(x + a/2)^{2} = a^{2}/4 - b$$

$$x = -a/2 \pm (a^{2}/4 - b)^{1/2}$$

A sequence of steps: at each step a transformation rule is applied

### From start to end

Start point: a premise (i.e. we assume the truth of something)

End point: a conclusion (i.e. we state that something else must also be true)

(both points are decided by us)

### Sequence of steps

Each transformation rule is based on equivalence

Each step must be correct, in this particular sense

### Abstraction and Correctness

Symbols like x, a and b represent any number, equivalences are valid for all of them

What does it mean for the whole sequence to be correct?

# Symbolic Reasoning

# Symbolic descriptions can abstract

 Symbolic descriptions have abstraction capabilities: for instance, many linguistic phenomena are systemic (i.e. their complexity goes beyond a pattern-matching)

Someone may understand English or not.

But no human being can understand the sentence:

"Sally **likes** Harry"

without being able to understand:

"Harry **likes** Sally"

or any other sentences of the kind:

"X likes Y"

where X and Y can be any nouns or definite descriptions:

"Ronald's girlfriend likes the cat that Linda brought home yesterday"

(freely adapted from [Fodor e Phylyshyn, 1988])

Many (intuitively) valid arguments obey an abstract schema

All (humans) are (mortals)

All (Greeks) are (humans)

hence

All (Greeks) are (mortals)

Abstract schema:

All H are M

All **G** are **H** hence

All **G** are **M** 

Many (intuitively) valid arguments obey an abstract schema

All (humans) are (mortals)

hence

All (Greeks) are (humans)

All (Greeks) are (mortals)

Abstract schema:

All **H** are **M** 

hence

All **G** are **H** 

All **G** are **M** 

**CAUTION!** 

The position of the line IS relevant: in this case the schema does <u>not</u> work

Many (intuitively) valid arguments obey an abstract schema

All (humans) are (mortals)

All (Greeks) are (mortals)

hence

All (Greeks) are (humans)

Abstract schema:

All H are M

All **G** are **M** hence

All **G** are **H** 

#### **CAUTION!**

The ordering of sentences IS relevant: in this case the schema does <u>not</u> work

Many (intuitively) valid arguments obey an abstract schema

All (humans) are (mortals)

All (Greeks) are (humans) hence

All (Greeks) are (mortals)

Abstract schema:

All H are M

All **G** are **H** hence

All **G** are **M** 

The validity of schemas does not depend on *meaning*:

All (enchanted frogs) are (princesses)

All (princesses) are (young and beautiful) hence

All (enchanted frogs) are (young and beautiful)

Same schema, different impression

# Fallacies (paralogisms)

### Wrong sequence:

All (humans) are (mortals) All (Greeks) are (mortals)

All (Greeks) are (humans)

The last two sentences have been switched

### Referential ambiguities:

(Nothing) is better than (eternal happiness)

(A ham sandwich) is better than (nothing)

(A ham sandwich) is better than (eternal happiness)

### 'Obscure' subtleties (obscure for now):

All (enchanted frogs) are (princesses)

All (enchanted frogs) are (young and beautiful)

There is an (enchanted frog) which is (a young and beautiful princess)

What do we mean by "all"? Do we need at least one specimen to say "all"? In such case, the schema would be valid (in formal logic it is not, as we will see)

# What is the purpose of symbolic logic?

To distinguish correct reasoning from incorrect reasoning

# What is the purpose of symbolic logic?

To distinguish correct reasoning from incorrect reasoning by the formal structure

# Truth and Correctness: the quest for formal semantics

### What is true?

### A world of cats



### Sentences about this world

"Spot likes Tom" and "Tom does not like Spot"

"Tom likes himself"

"Kitty likes Spot" and "Spot likes Kitty"

"Kitty likes herself"

"Felix likes Kitty"

All these sentences are true, in the world above

### What is true?

### A world of cats

	_T_	مکم	o <b>t</b> Kit	tyFel	i)
likes	101			دي. د.	
Tom	Х				
Spot	X		Х		
Kitty		х	х		
Felix			Х		

### Sentences about this world

"Spot likes Felix"

"Spot likes himself"

"Kitty does not like herself"

"Felix likes Spot"

All these sentences are **false**, in the world above

### What is true?

### A world of cats



### Sentences about this world

"Every cat likes a cat" is true

"Every cat likes another cat" is false

"Tom does not like any other cat" is true

"Kitty is liked by every cat" is false

### How to make a sentence true?

### Consider the sentence

"Kitty likes Spot" and "Spot likes Tom"

It can be made true in many different ways "It may be true in many different *possible worlds*" Examples:

	$T_{\Delta t}$	$\infty$ 2 $\infty$	<u> ⊲t⁄it</u>	tyFel	.iv
likes	101			cyi Ci	
Tom					
Spot	X				
Kitty		х			
Felix					

	_T_	<del>mSp</del>	<b>∩t</b> Kit	tvEol	Liv
likes	101			C). C.	
Tom	X				
Spot	X		X		
Kitty		Х	Х		
Felix			X		

	<u>_</u> T_	<del>mSp</del>	<b>∆t</b> Kit	tıEal	iv
likes	101		Juli	cyi Ci	
Tom	X	Х	X	X	
Spot	X	x	X	x	
Kitty	X	X	Х	X	
Felix	X	X	X	X	

### How to make a sentence true?

### Consider the sentence

"Kitty likes Spot" and "Kitty does not like Spot"

There is no way to *make it true*"There is no *possible world* where this can be true"

### Three sentences

- 1) "Every cat that likes Kitty likes Spot as well"
- 2) "Tom likes Kitty"
- 3) "Tom likes Spot"

There is no way to make true sentences 1) and 2) without making true sentence 3) as well... (just give it a try...)

likes	TOI	пЗро	rtKiti	yFei	ľ
Tom					I
Spot	х				l
Kitty		х			l
Felix		х	х		

likes	Toi	n Spo	otKiti	yFel	İ×
Tom					
Spot	х	х	х		
Kitty		х	х		
Felix		х	х		

likes	Toi	n Spo	rtKiti	yFel
Tom	х			
Spot			х	
Kitty		х		
Felix				х

likes	Toi	пэрс	ikiti	yrei
Tom				
Spot	х	х	х	х
Kitty			х	
Felix		х	х	

likes	Ťoi	пSр	otKiti	yFel
Tom				
Spot				
Kitty			х	
Felix				

likes	Tol	n Spo	ιKiti	yFel
Tom	х	х	х	
Spot	х			
Kitty		х		
Felix		х	х	

		_			İ×
likes	101	n Spo	PUNIU	уге	^
Tom		х	х		
Spot	х	х	х		
Kitty	х	х			
Felix					

	Tor	пЗро	τΚίτ	yFel
likes				
Tom	х		х	
Spot	х			
Kitty		х		
Felix		х		х

likes	Tol	пЅро	ρίΚΙτι	yFel
Tom			х	
Spot	х			
Kitty		х		
Felix		х	х	

likes	Toi	n <del>Sp</del> o	tKiti	yFel
Tom	х	х	х	х
Spot		х		
Kitty			х	
Felix				х

### Three sentences

- 1) "Every cat that likes Kitty likes Spot as well"
- 2) "Tom likes Kitty"
- 3) "Tom likes Spot"

There is no way to make true sentences 1) and 2) without making true sentence 3) as well... (just give it a try...)

#### Sentence 1) is true in these worlds

	Tor	п Ѕрс	σιKit	vFel	İ×
likes				,,	ĺ
Tom					
Spot	х				
Kitty		х			
Felix		х	х		

	To	ii Siii	n Kin	yFel	İ١
likes	5	пэрс	Perkie	y r Ci	ľ
Tom					
Spot	х	х	х		ĺ
Kitty		х	х		ĺ
Felix		х	х		ĺ

likes	Toi	n Spo	rtKiti	yFel	Х
Tom	х				
Spot			х		
Kitty		х			
Felix				х	

1		Tor	пSр	πKit	vFeli
	likes				, y . c .
	Tom				
	Spot	х	х	х	х
	Kitty			х	
	Felix		х	х	

	_	_		
likes	101	пэрс	JUNIU	уге
Tom				
Spot				
Kitty			х	
Felix				

likes	Toi	n Spo	νίΚίτι	yFel
Tom	х	х	х	
Spot	х			
Kitty		х		
Felix		х	х	

likes	Tol	n <del>S</del> po	tKiti	yFel
Tom		х	х	
Spot	х	х	х	
Kitty	х	х		
Felix				

likes	Toi	n <del>Sp</del> o	ιΚiti	yFel
Tom	х		х	
Spot	х			
Kitty		х		
Felix		х		х

likes	Toi	n Spo	iKiti	yFei	X
Tom			х		
Spot	х				
Kitty		х			
Felix		х	х		

likes	Toi	n <del>Sp</del> o	tKiti	yFel
Tom	х	х	х	х
Spot		х		
Kitty			х	
Felix				х

### Three sentences

- 1) "Every cat that likes Kitty likes Spot as well"
- 2) "Tom likes Kitty"
- 3) "Tom likes Spot"

There is no way to make true sentences 1) and 2) without making true sentence 3) as well... (just give it a try...)

#### Sentence 1) is true in these worlds

	Tor	п Ѕрс	otKiti	vFei	İ×
likes				,	
Tom					
Spot	х				
Kitty		х			
Felix		х	х		

likes	Toi	пЅро	tKiti	yFel	×
Tom					
Spot	х	х	х		
Kitty		х	х		
Felix		х	х		

likes	Toi	n <del>Sp</del> o	tKiti	y <del>Fel</del>
Tom	х			
Spot			х	
Kitty		х		
Felix				х

likes	Toi	пSр	otKiti	yFel
Tom				
Spot	х	х	х	х
Kitty			х	
Felix		х	х	

likes	10	пэрс	JUNIU	yı ei
Tom				
Spot				
Kitty			х	
Felix				

likes	Toi	n Spo	iKiti	yFei
Tom	х	х	х	
Spot	х			
Kitty		х		
Felix		х	х	

	Tor	יעל ח	νίΚία	yrei
likes				,
Tom		х	х	
Spot	х	х	х	
Kitty	х	х		
Felix		·		

likes	Toi	п Ѕрс	νίΚίτι	yFel	Х
Tom	х		х		
Spot	х				
Kitty		х			
Felix		х		х	

ı		Tor	пбро	ρίΚΙι	vEel	İX
	likes	. 0.	., .,		,y . c.	
	Tom			х		
	Spot	х				
	Kitty		х			
	Felix		х	х		

I	likes	Tor	пЅрс	τΚίτι	yFel
	likes				
	Tom	Х	х	Х	х
	Spot		х		
	Kitty			х	
	Felix				х

Sentence 2) is true in these worlds

### Three sentences

- 1) "Every cat that likes Kitty likes Spot as well"
- 2) "Tom likes Kitty"
- 3) "Tom likes Spot"

There is no way to make true sentences 1) and 2) without making true sentence 3) as well... (just give it a try)

#### Sentence 1) is true in these worlds

		Tom SpotKittyFellx Tom S							пЅрс	ριΚίτι	y Feli	× I		Tor	п Ѕрс	ιKiti	yFel	Х		Tor	пЗр	ρίΚίτ	yFel	X		Toi	пЅро	ptKiti	<del>y Feli</del> x
	likes						likes						likes						likes						likes				
	Tom					ĺ	Tom						Tom	х					Tom						Tom				
	Spot	х					Spot	х	х	х			Spot			х			Spot	х	х	х	х		Spot				
	Kitty		х				Kitty		х	х			Kitty		х				Kitty			х			Kitty			х	
	Felix		х	х			Felix		х	х			Felix				х		Felix		х	х			Felix				
F																													
	likes	Toi	n <del>Sp</del> o	tΚiti	yFel	×	likes	Toi	n <del>Sp</del> o	tKiti	<del>y Feli</del> x	×	likes	Toi	n Spo	tKlu	yFel	×	likes	Toi	n Spo	tKiti	yFel	x	likes	Tor	n <del>Sp</del> o	otKiti	<del>y Feli</del> x
	Tom	х	х	х		l	Tom		х	х			Tom	х	х	х			Tom			х			Tom	х	х	х	х
3)	is there i	n×tŀ	ies	e w	orlo	s	Spot	х	х	х			Spot	х					Spot	х					Spot		х		
	Kitty		х				Kitty	х	х				Kitty		х				Kitty		х				Kitty			х	
- 1	_ I:					I	Felix						Felix		х		х		Felix		х	х			Felix				х
	Felix		Х	Х		I	I CIIX						I Clix		^		^	Ш	1 Clist		L ^_	^			I Clix				_^_

#### Three sentences

- 1) "Every cat that likes Kitty likes Spot as well"
- 2) "Tom likes Kitty"
- 3) "Tom likes Spot"

There is no way to make true sentences 1) and 2) without making true sentence 3) as well... (just give it a try)

Questions:

Was this just coincidence?
How many possible worlds
must we consider in order to be sure?