

# COGNITIVE PSYCHOLOGY AND

## BIG, TINY + GIGANTIC DATA

RAY DOUGHERTY

XLP TSINGHUA

NEW YORK UNIVERSITY

rcd2@nyu.edu

rayidou@icloud.com

Feb 1, 2015

### 1. The differences

#### BIG DATA

usually business

#'s (Numbers)

QUANTITY HOW MUCH  
MANY

STATISTICS

FIBONACCI

USE EXISTING

- IDEAS

AMAZON

#### TINY ATOMS

#### GIGANTIC BIO-

LINGUISTIC  
LANGUAGE  
BIOLOGY

CERN

ALL LIFE  
coll  
organ  
organism  
(BRENNER)

# RESEARCH FUNDING

2

BIG DATA \$  
BUSINESS  
SELF

Billions

GOVT \$ \$ \$  
CERN  
BOSON GENOME  
COMPUTER LOGICS

Zillions  
TAXPAYERS

DATA

IS NOT INFORMATION

DECISIONS

ARCHITECTURE  
FORMAT  
NOTATIONS  
QUANTITY  
(Numbers)  
PATTERN  
(Geometry)

TYPES

Animal  
Insect  
BIRD  
MACHINE  
NORMAL/ABNORMAL

3

PASCAL (1663)

DATA from PAST & FUTURE

$$\text{1) HARRY + TESS } 300 + 500 \\ = 1000$$

(2) FLIP COIN

First to get 10 H or T wins.

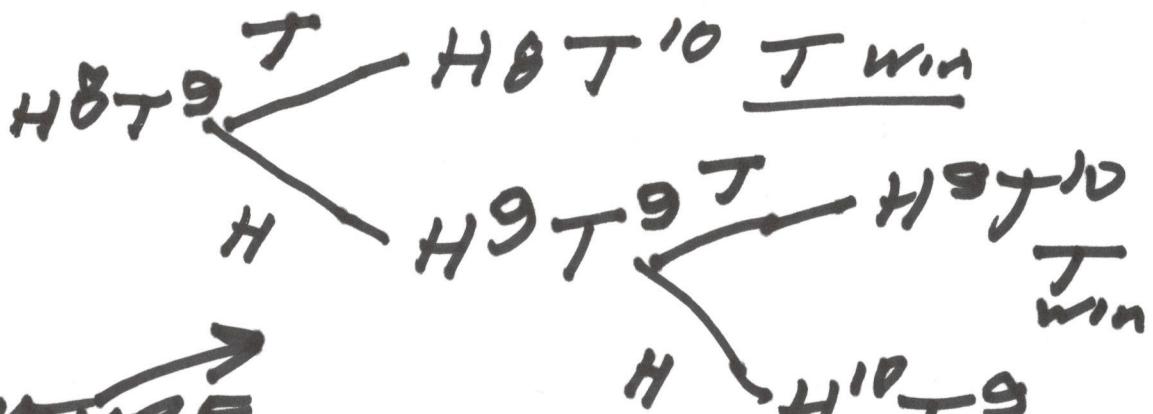
	PAST	NOW
17	H <sup>10</sup> T <sup>7</sup>	T <sup>9</sup>
Times H <sup>7</sup> T <sup>10</sup>		H <sup>8</sup>

POLICE RAID GAME → Ø  
DIVIDE \$1000 FAIRLY

PAST (1) 50/50

(2) T = 9/17 H = 8/17 T = 9 wins  
H = 8 wins

(3) 17 = TOSSES



ON FUTURE

H is  $\frac{1}{3}$  \$333  
T is  $\frac{2}{3}$  \$677

NATURE DOES NOT CALCULATE  
BUT COMPUTES.  
CALCUL ≠ COMPUT

## NOTATION DEFINITIONS

### CHOMSKY

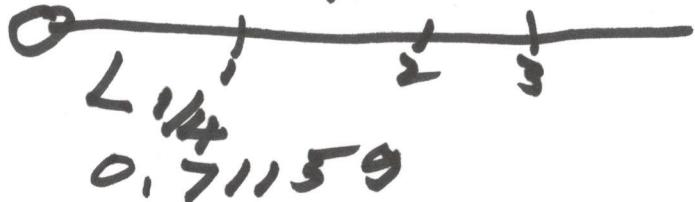
A sentence is a string of elements taken from an alphabet (symbols)  
A language is a set of sentences  
A grammar defines a language.  
A linguistic theory (meta-grammar) defines "possible grammars".

## SUPER COMPLEX(HARD) FACTS BASIS OF ALL WORK TODAY.

### FACT A

The Fibonacci Type numbers (quantity) have a contradiction

A) Turing A number is the unique name of a single unique point on a line.



B) Fibonacci type numbers (you learned in school)  $7 \times 3 = 21$

$$1) \quad x = 0.99999 \dots, \quad 5+6=11$$

$$2) \quad 10x = 9.99999 \dots,$$

$$2-1) \quad 9x = 9,000000$$

$$(3) \quad x = 9/9 = 1$$

~~But~~ BUT (NS says  $x = 0.999\dots$ ,

(3) NS says  $x = 1$   
 $\therefore$  Contradiction

## META LANGUAGE

## COMPUTATION >> CALCULATION

S	DATA	INFORMATION
H	anything	MAKE DECISION
A	in some	1) YES/NO
N	notation	2) which (choice)
N		3) probability

P  $11+17+20+31+162+413+1 = 662$

Q  $11 \pm 17 \pm 20 \pm 31 \pm 162 \pm 413 \pm 1 = 224$   
 $\uparrow$  which set of +/ -

6

## CALCULATION

### BRUTE FORCE

## META LANGUAGE COMPUTATION

### HEURISTIC PROGRAMMING (XLP)

$$\text{ODD} \pm \text{ODD} = \text{EVEN}$$

$$\text{EVEN} \pm \text{EVEN} = \text{EVEN}$$

$$\text{EVEN} \pm \text{ODD} = \text{ODD}$$

So  $P + Q$  can not be true

since the result is even but the #s contain 5 ODD #s

## META LANGUAGE

USE PROPERTIES OF  
the SET + Architecture of DATA

# (7)

## TURING'S GREAT DISCOVERY 1935

(Age 23)

There are ~~the~~ two sizes of infinity (CANTOR)

1) Countable, 1 2 3 4 5 6  
integers

(2) # of points on line  $0 \rightarrow 1$   
(# of decimals)

PROOF (in Chomsky's notation)

ALL INTEGERS	ALL DECIMALS
1	0. <u>7</u> 596283
2	0, <u>1</u> 928635
3	0, <u>2</u> 2 <u>2</u> 5,87
4	0, <u>8</u> 3 <u>8</u> <u>2</u> 16
5	0,751 <u>8</u> <u>3</u> 0
6	
7	
8	

Grammar - Select  
DIAGONAL

0.79223..... Forever.

↗ This # is NOT IN LIST

Turing Showed 100% certainly  
The # of problems one can  
define in computation/

= the # of decimals  
calculation

BUT

The # of Solutions (answers)  
is = the integers

HENCE

There are an uncountable  
infinity (decimals) of problems  
but only a countable infinity  
of programs that can do  
anything interesting.

(9)

## The Collatz $3N+1$ problem

TAKE ANY integer N

if N is even divide it by 2

if N is odd multiply by  
3 AND ADD 1

12 → 6 → 3 → 16 → 8 → 4 → 2 → 1

DOES EVERY INTEGER  
GO BACK TO 1?

Eg.) # of even & odd #s.

~~CHOMSKY~~ HUMAN  
LANGUAGE  
IS NOT  
LINEAR

The men want to go to NY.  
~~the men want to go to NY~~ → NY

The men want may to go to NY  
who do the men want to go to NY?

~~the men want to go to NY~~ → NY      MY LIEE

10 \$

FOR MUCH DATA  
CALCULATE Numbers of the old  
grade school work

For some DATA in which a  
number is a point on a line  
Something else is needed  
computation (Turing)

FOR SOME DATA (Chomsky)  
HUMAN LANGUAGE we  
need symbolic ~~or~~ computational  
(Natural Language Processing)

FOR SOME Biological  
Big/Genetic DATA  
we need 3 D geometry  
Turing Mottling Thompson  
Spots Spirals  
Bumps