Mind and Data Science

Herman Y Sutarto

East Continent Research Center

ECF Extension Course Filsafat
Universitas Katolik Parahyangan

DARE TO SLACK

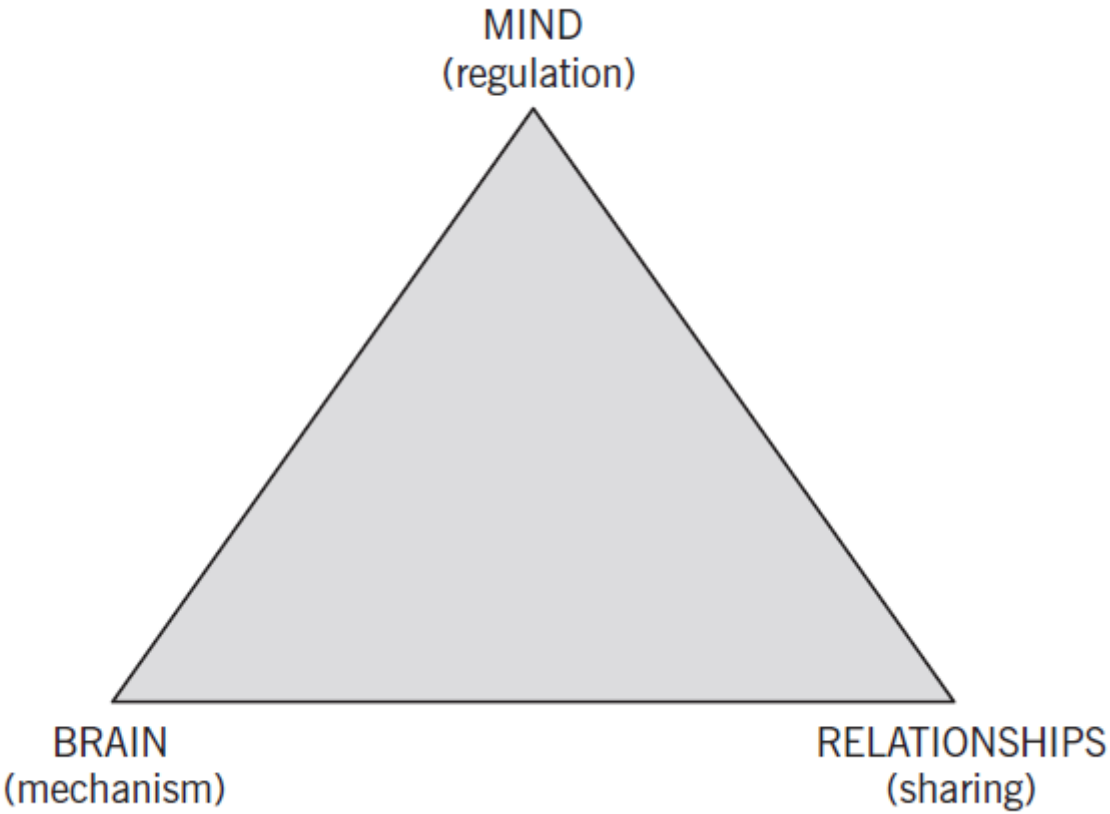
WHEN BIRDS FLY IN THE RIGHT FORMATION, THEY NEED ONLY EXERT HALF THE EFFORT.
EVEN IN NATURE, TEAMWORK RESULTS IN COLLECTIVE LAZINESS.

www.despair.com

Video

Emergence

- Novel behavior
- Properties of the whole
- **Cannot be predicted** from properties of the components that make up the system


MIND
(regulation)
BRAIN
(mechanism)
RELATIONSHIPS
(sharing)

The mind is an embodied and relational process that regulates the flow of energy and information.

The emergence of consciousness may be intimately related to the development of memory.

Memory is not a static thing, but an active set of processes.

Our internal experiences are constructive processes.

Experiences can shape not only what energy and information enters the mind, but also how the mind processes that information.

Our social experiences can directly shape our neural architecture.

Interpersonal experiences appear to have a direct effect on the development of explicit memory.

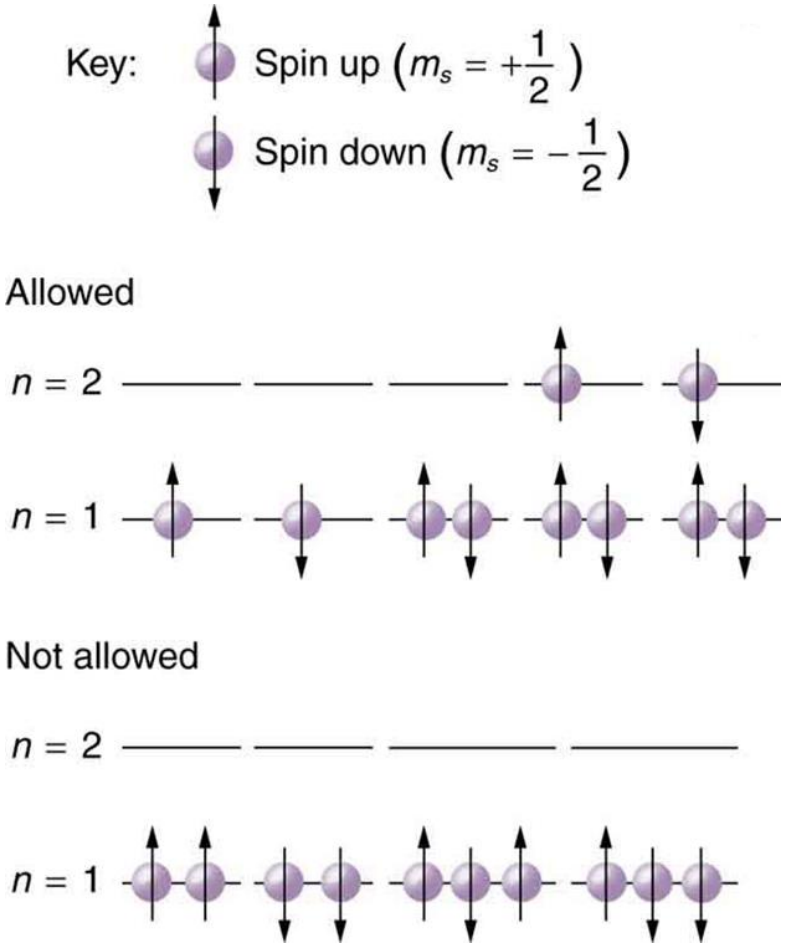
Relationships and the embodied brain are really part of one larger system.

Early experience shapes the regulation of synaptic growth and survival.

Interpersonal experiences continue to influence how our minds function throughout life.

Consciousness is the experience of being aware, the internal state of knowing that something is happening in the present moment.

The Pauli Exclusion Principle



It is impossible for two electrons of a poly-electron atom to have the same values of the four quantum numbers: n, the principal quantum number, ℓ, the angular momentum quantum number, m_ℓ, the magnetic quantum number, and m_s, the spin quantum number. For example, if two electrons reside in the same orbital, and if their n, ℓ, and m_ℓ values are the same, then their m_s must be different, and thus the electrons must have opposite half-integer spin projections of 1/2 and −1/2.

Periodic Table

Group → / Period ↓	1	2	3		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																		2 He
2	3 Li	4 Be												5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg												13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc		22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y		40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57 La	*	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	89 Ac	* *	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
				*	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	
				* *	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	

Morowitz's Three Claims

(i) PEP is a *nondynamical* principle, but it influences the dynamical behavior of electrons.

(ii) PEP has *nothing to say about the behavior of individual electrons*.

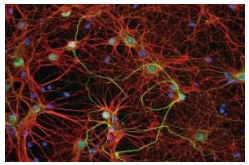
(iii) PEP is *unrelated to the other laws of physics*.

How does the [mind] work ?

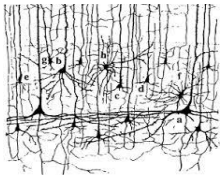
Molecules/ Genes

Neurons/Synapses

Neuronal circuits



Systems of neurons



Brain regions

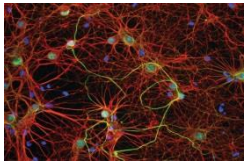
Data & state Representation + algorithms

Behavior/Cognition

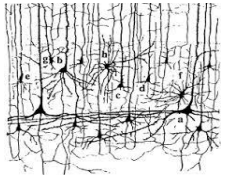
Molecules/ Genes

Neurons/Synapses

Neuronal circuits



Systems of neurons



Brain regions

Data & state Representation + algorithms

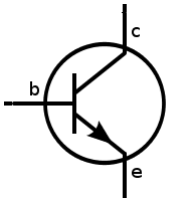
How does the [mind] work ?

"You, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact ~~no more than~~ the behavior of a vast assembly of nerve cells and their associated molecules“

-- Francis Crick (Co-discover of the structure of DNA)

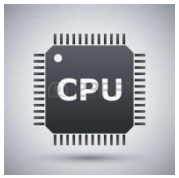
How does my favorite app work?

Transistor



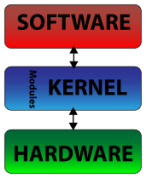
IC

CPU/GPU



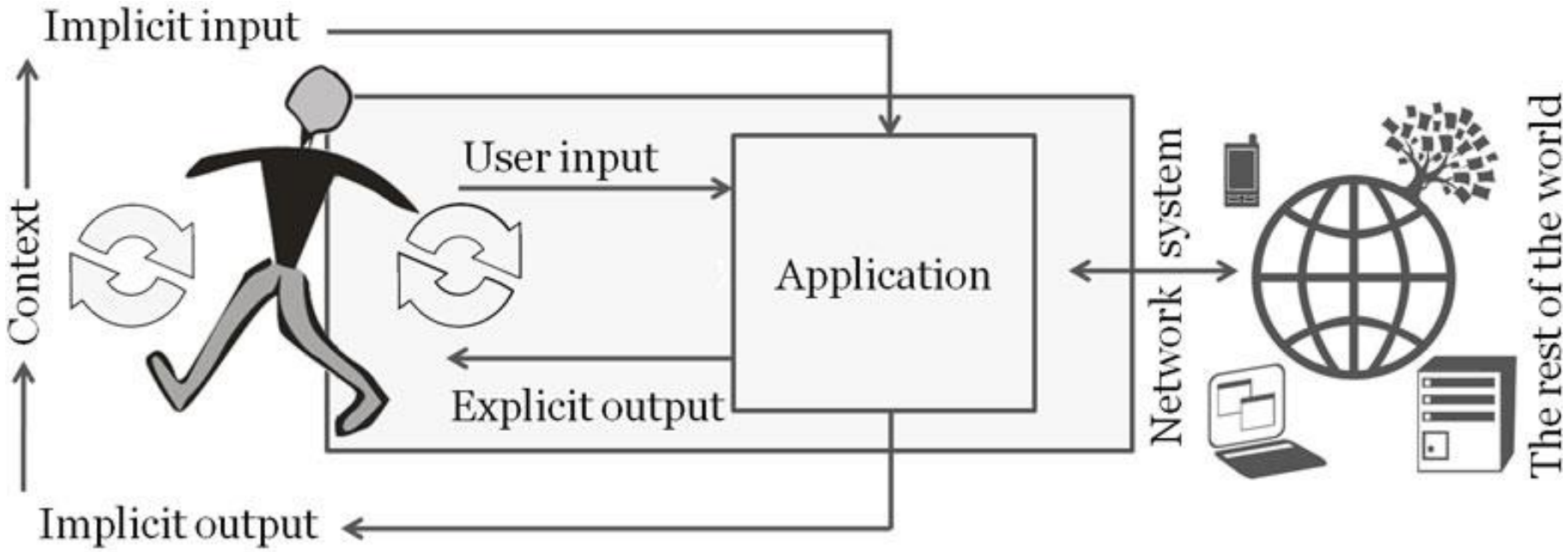
Motherboard

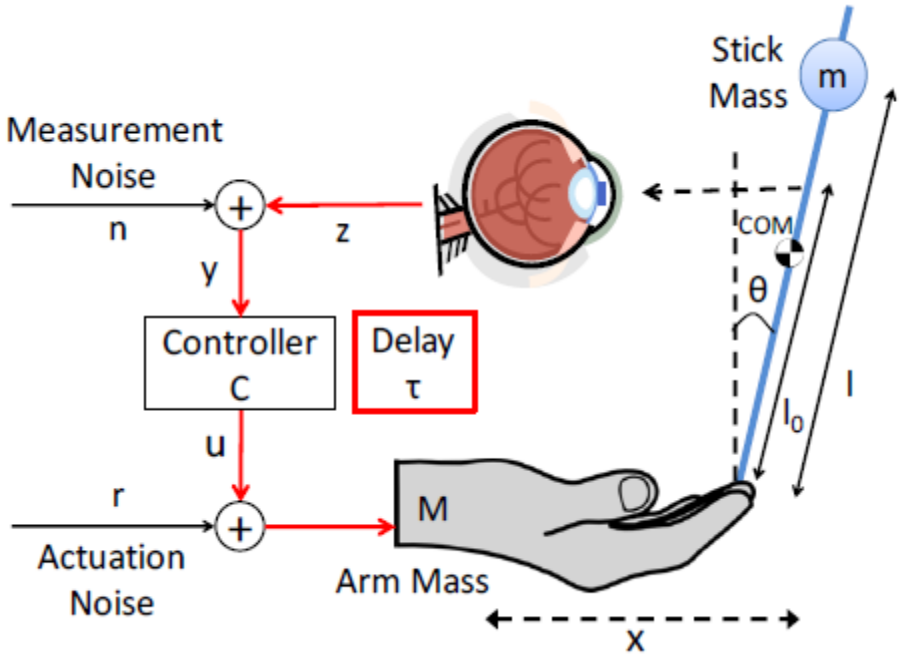
Kernel

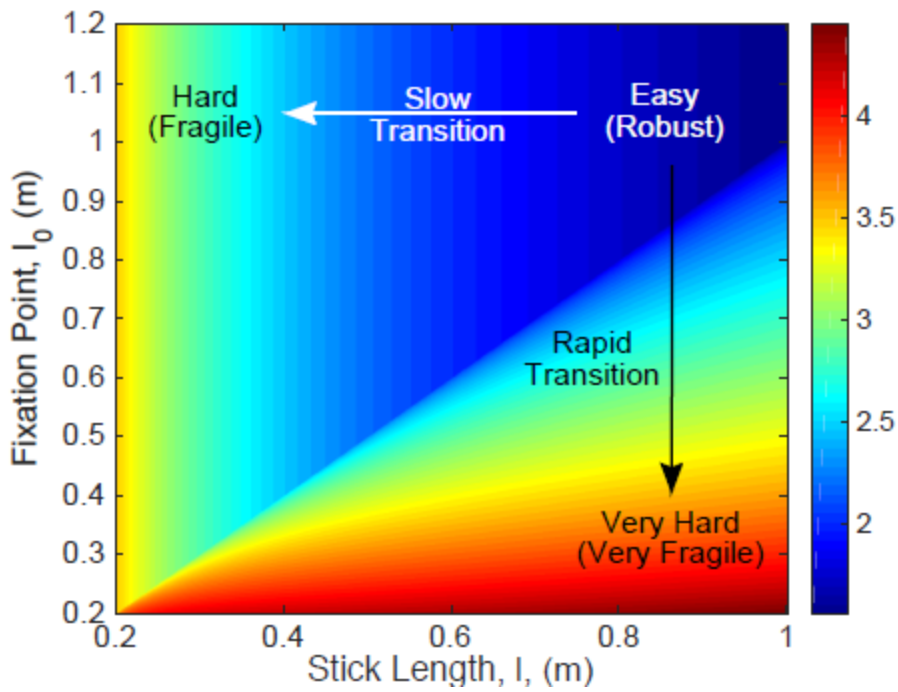


Algorithm application

User Interface


Implicit input
Context
User input
Application
Explicit output
Implicit output
Network system
The rest of the world


Stick
Mass
m
Measurement
Noise
n
z
y
COM
θ
Controller
C
Delay
τ
l
l_0
u
r
M
Actuation
Noise
Arm Mass
x


Hard
(Fragile)
Slow
Transition
Easy
(Robust)
Rapid
Transition
Very Hard
(Very Fragile)
Fixation Point, l_0 (m)
Stick Length, l, (m)
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.2
0.4
0.6
0.8
1
4
3.5
3
2.5
2

Sigmund Freud

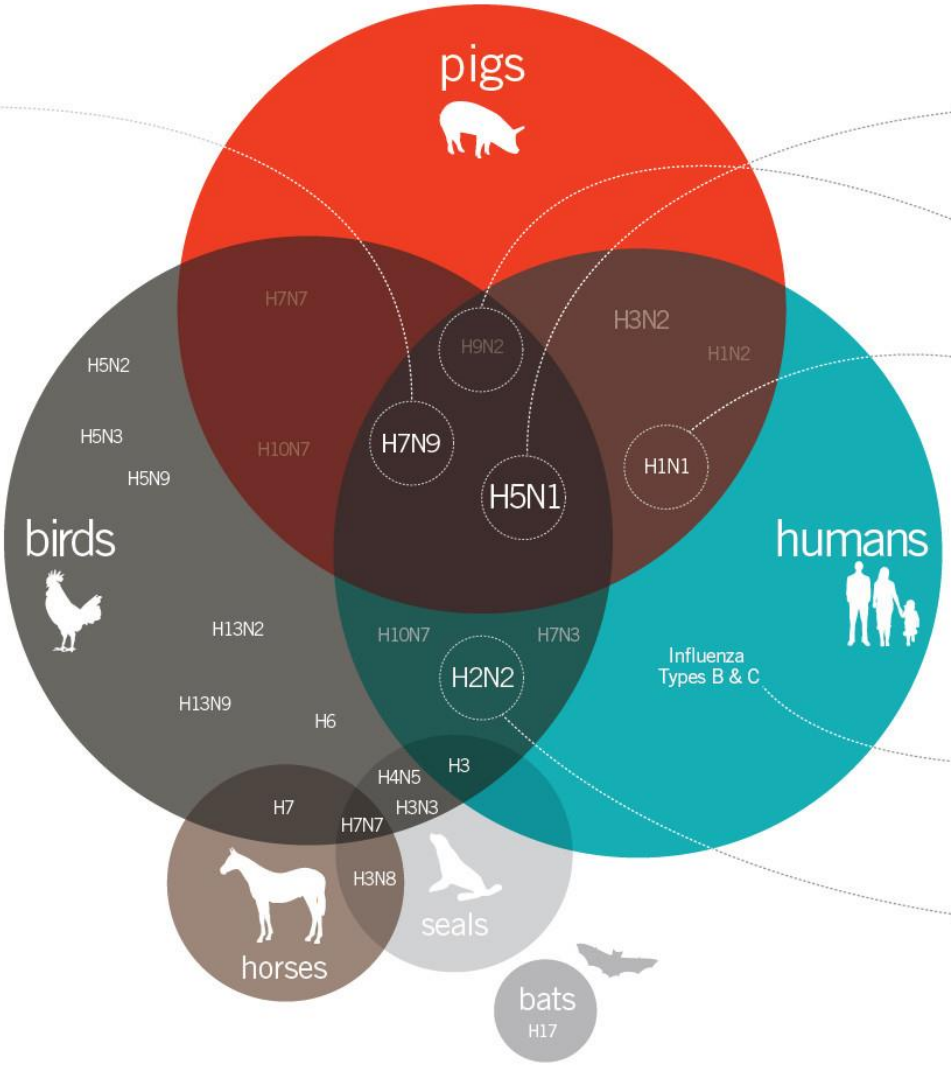
Influ-Venn-Za

Who can catch which flu?

Pigs often a source of flu pandemics as they can be infected by bird, human & swine flus.

Worse-case, they act as a bridge for newly evolved virus strains to cross from birds to humans.

April 22nd 2013 - suspected mutation of an avian virus emerged outside Shanghai, China. Human fatality rate is unknown but out of 130 people infected, 40 have died (30%).



The "Bird Flu" most mentioned in the media. Kills 60% of humans it infects. But direct human-to-human transmission has not been reported.

Lesser known "bird flu" endemic in poultry in Eurasia. Rarely seen in humans.

The most common variant of "Swine Flu". As "Spanish Flu" it killed 50-100 million people in 1918. The 2009-10 pandemic killed 15,000 worldwide.

Influenza Type A is divided into H & N strains (i.e. H1N1) referring to different combinations of:

H = hemagglutinin (binds to cells)

N = neuraminidase (surface enzyme)

"Common flu" only found in humans. Less harmful than type A. Does not cause pandemics.

Caused "Asian Flu" pandemic in 1957 then disappeared from human population. Still circulates in birds.

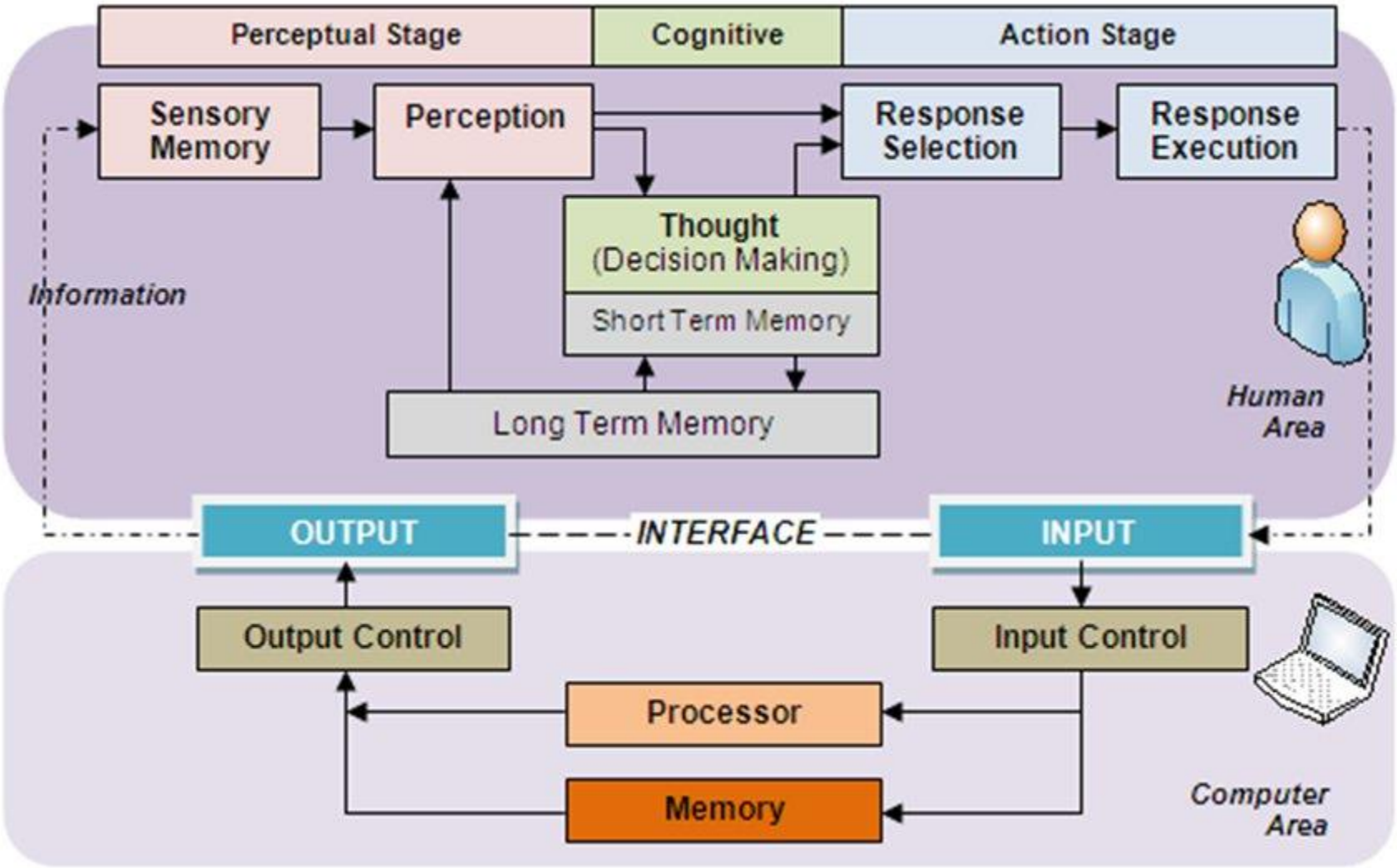
text SIZE = human fatality rate

LIGHT TEXT = rarely infects humans

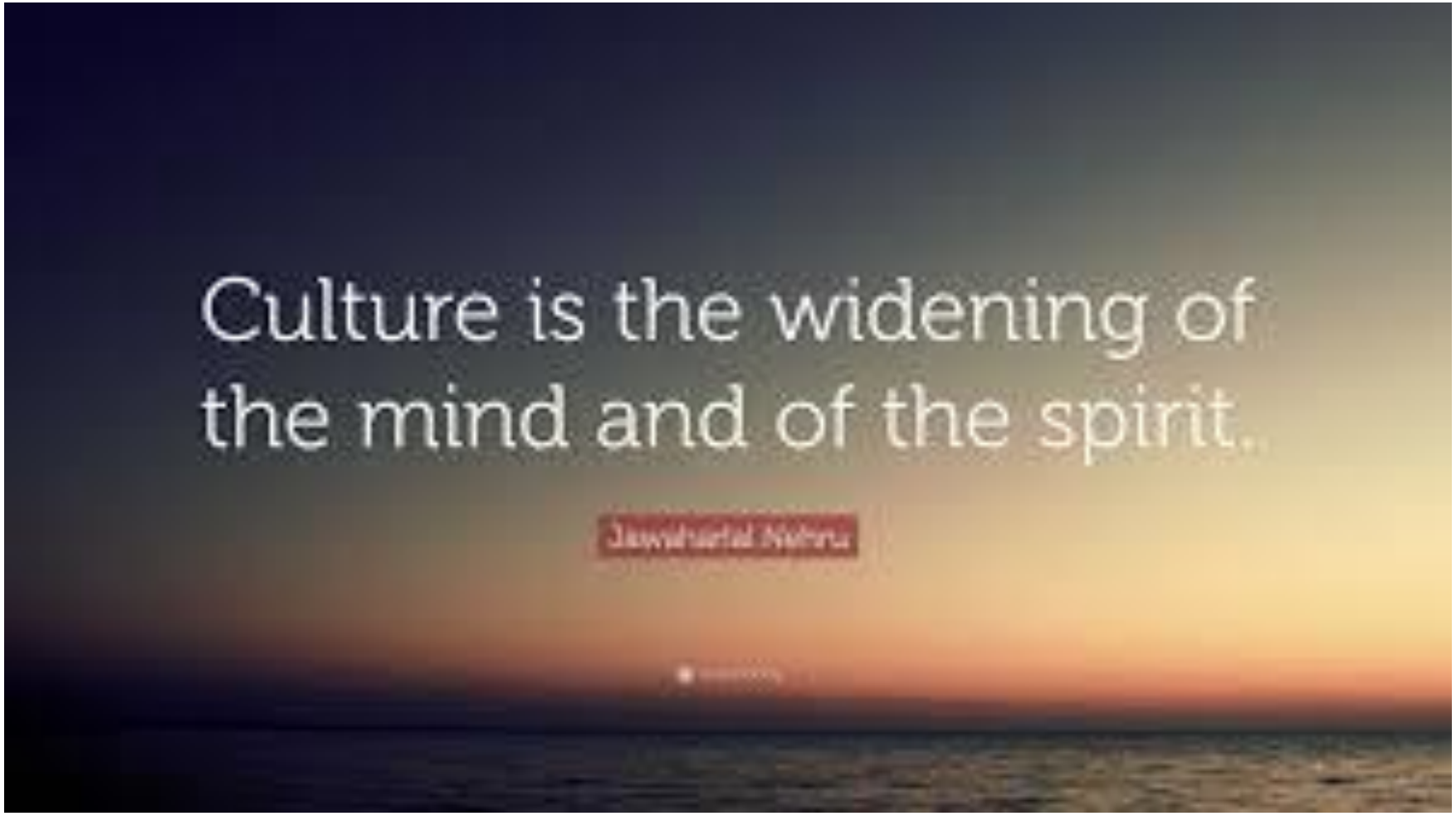
Concept & Design: David McCandless
Research: Ella Hollowood // Additional design:Phillipa Thomas
Version 1.22 / August 2013

informationisbeautiful.net

Sources: Centres for Disease Control, WHO
data: http://bit.ly/KIB_influenza

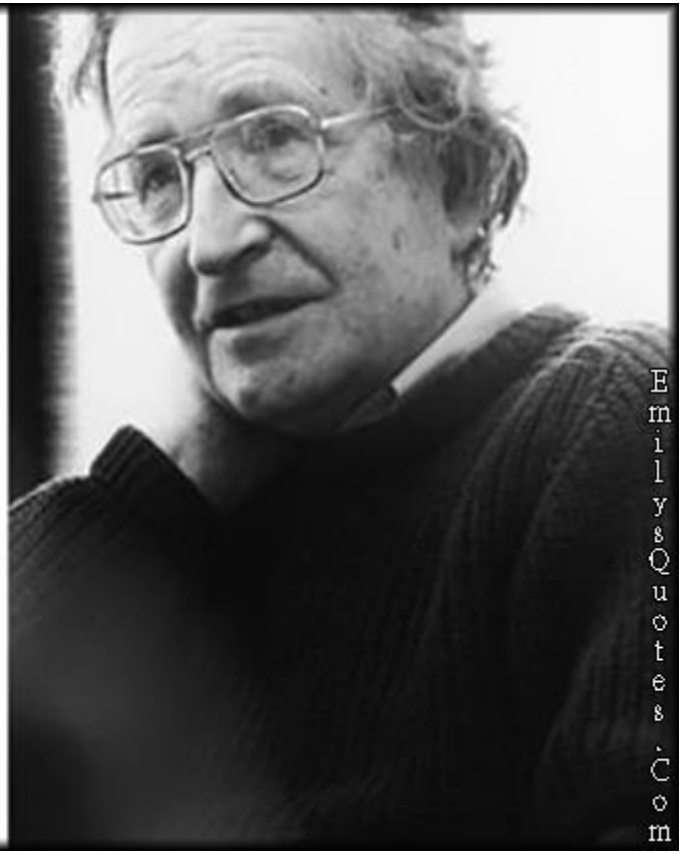

Perceptual Stage
Cognitive
Action Stage
Sensory Memory
Perception
Response Selection
Response Execution
Thought (Decision Making)
Short Term Memory
Long Term Memory
Information
Human Area
OUTPUT
INTERFACE
INPUT
Output Control
Input Control
Processor
Memory
Computer Area


social
information
research
web
users
user
people
design
applications
collaboration
mobile
tools
experience
technologies
new
systems
software
technology
human
speech
accessibility
computer
ibm
project
also
using
interaction
email
use
content
collaborative
work
development
understanding
improve
problem
management
computing
feeds
working
capture
network
application
devices
group
personal
methods
business
allows
input
university
including
science
program
gesture
many
data
develop
pages
interfaces
create
feed
unified
psychology
chi
infrastructure
exploring
hci
text
center
system
support
various
complexity
activities
multiple
networks
services
presence
individual
usability
used


Culture is the widening of
the mind and of the spirit.
Jawaharlal Nehru

Hundreds of billions of dollars are spent every year to control the public mind.

~Noam Chomsky



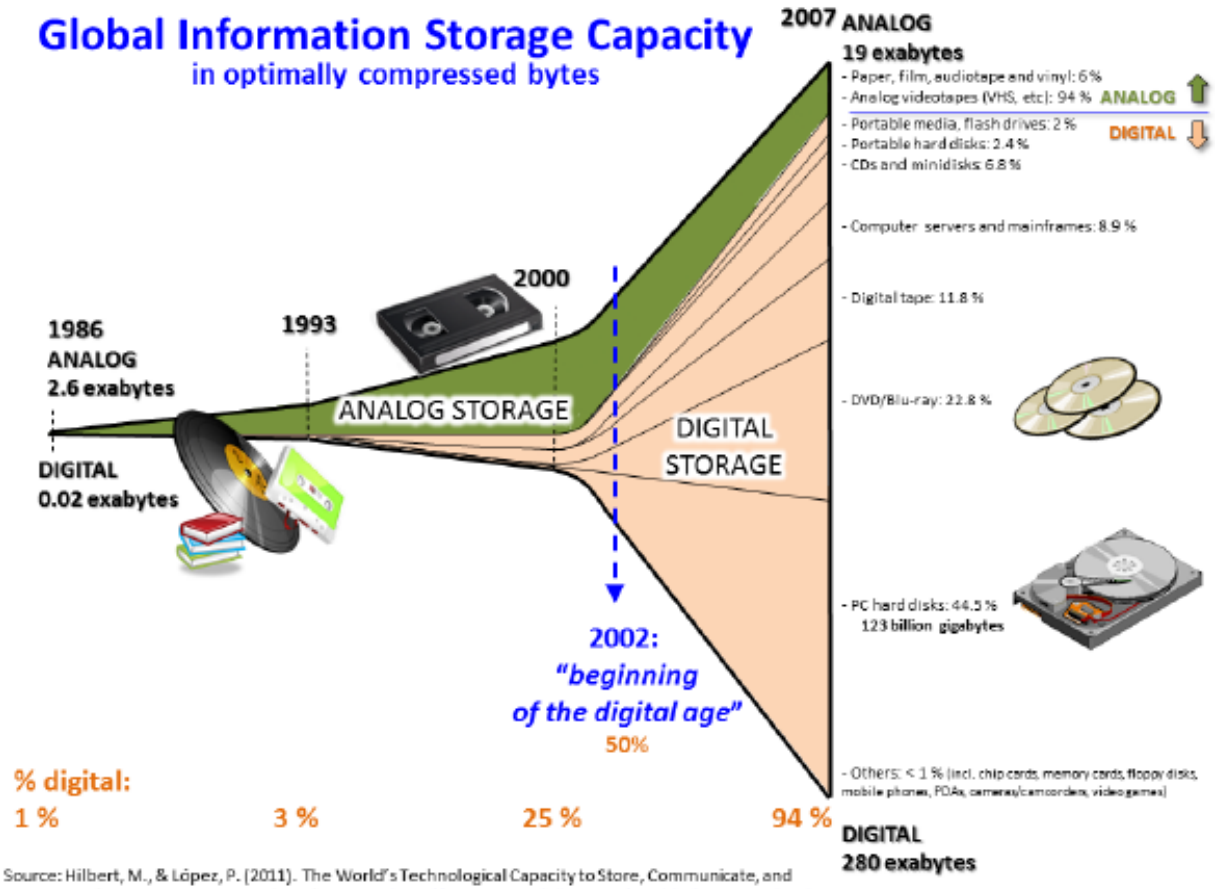
Data Science

The (Long) History of Data Processing

- (…) Manual Data Processing
- (1832) Punch cards
- (1936) Turing Machine
- (1944) First IBM Computers (SSEC)
- (1945) Von Neumann Architecture
- (1950s) Sort & Search Algorithms
- (1950s) Heuristics Methods
- (1936) Pattern Recognition
- (1951) The First Neural Network Machine
- (1955) Concerns on data explosion by Fremont Rider, Wesleyan University Librarian
- (1960s) DBMS
- (1960s) Data Analysis Methods (Bayesian, Time Series, Stochastic, …)
- (1968) Knuth – The Art of Computer Programming
- (1970s) Relational DBMS
- (1974) SQL
- (1975) First PC (MITS Altair 8800)
- (1990s) Data Mining / KDD
- (1990s) Complex-Event Processing
- (1990s) Data Stream Processing
- (1990s) Social Network Analysis
- (1998) The Term "Big Data" was first coined by John Mashey
- (1999) Internet of Things
- (2001) Volume, Velocity, Variety by Doug Laney
- (2001) "Data Science" by William Cleveland
- (2004) MapReduce
- (2009) No-SQL
- (2011) Global Information Storage Capacity grows at 25% annually by Martin Hilbert, Priscila López

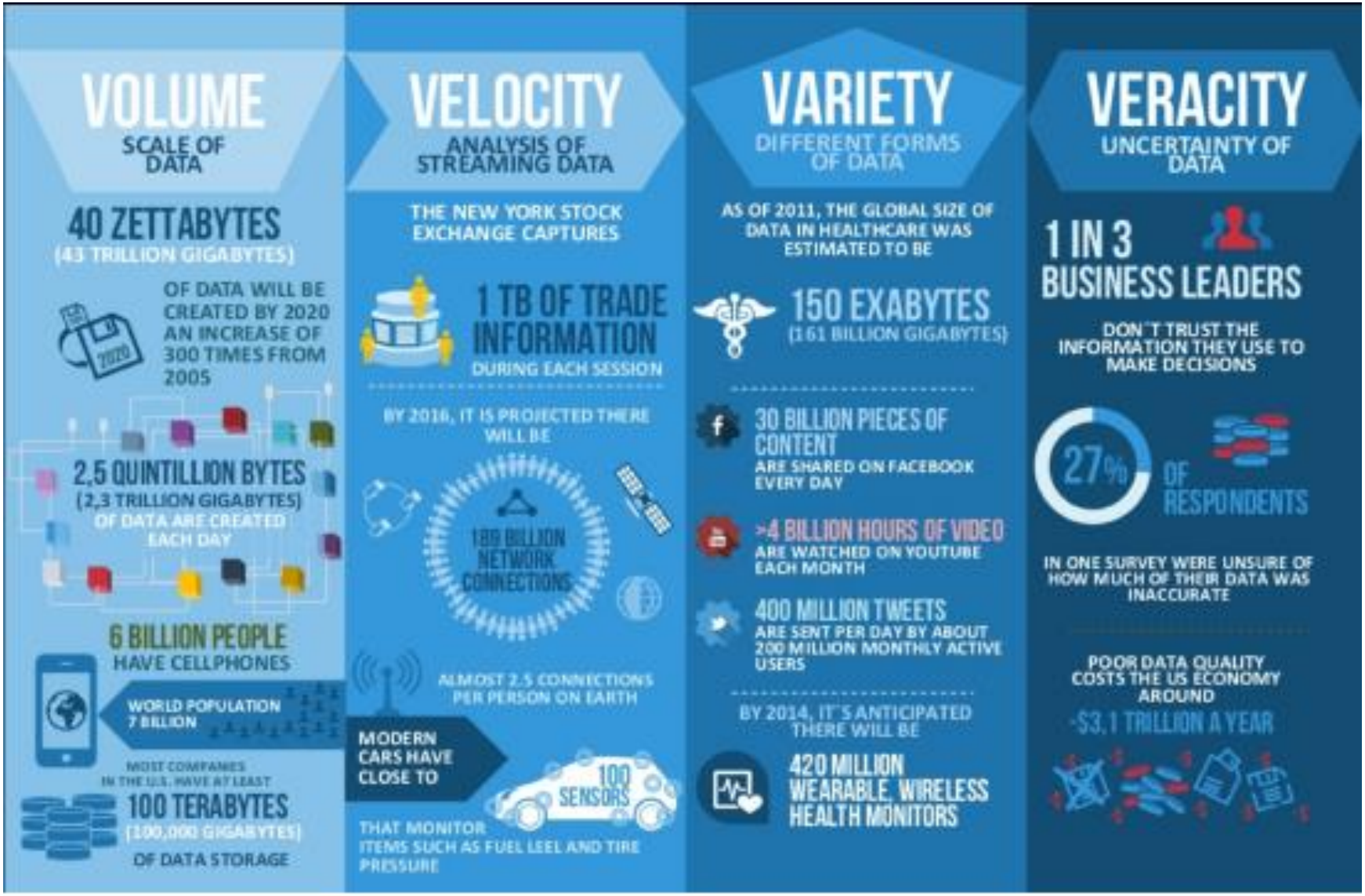
Pre-1900 | 1900-1950 | 1960-1970 | 1980 | 1990 | 2000 | 2010 | …

How Big is Big?



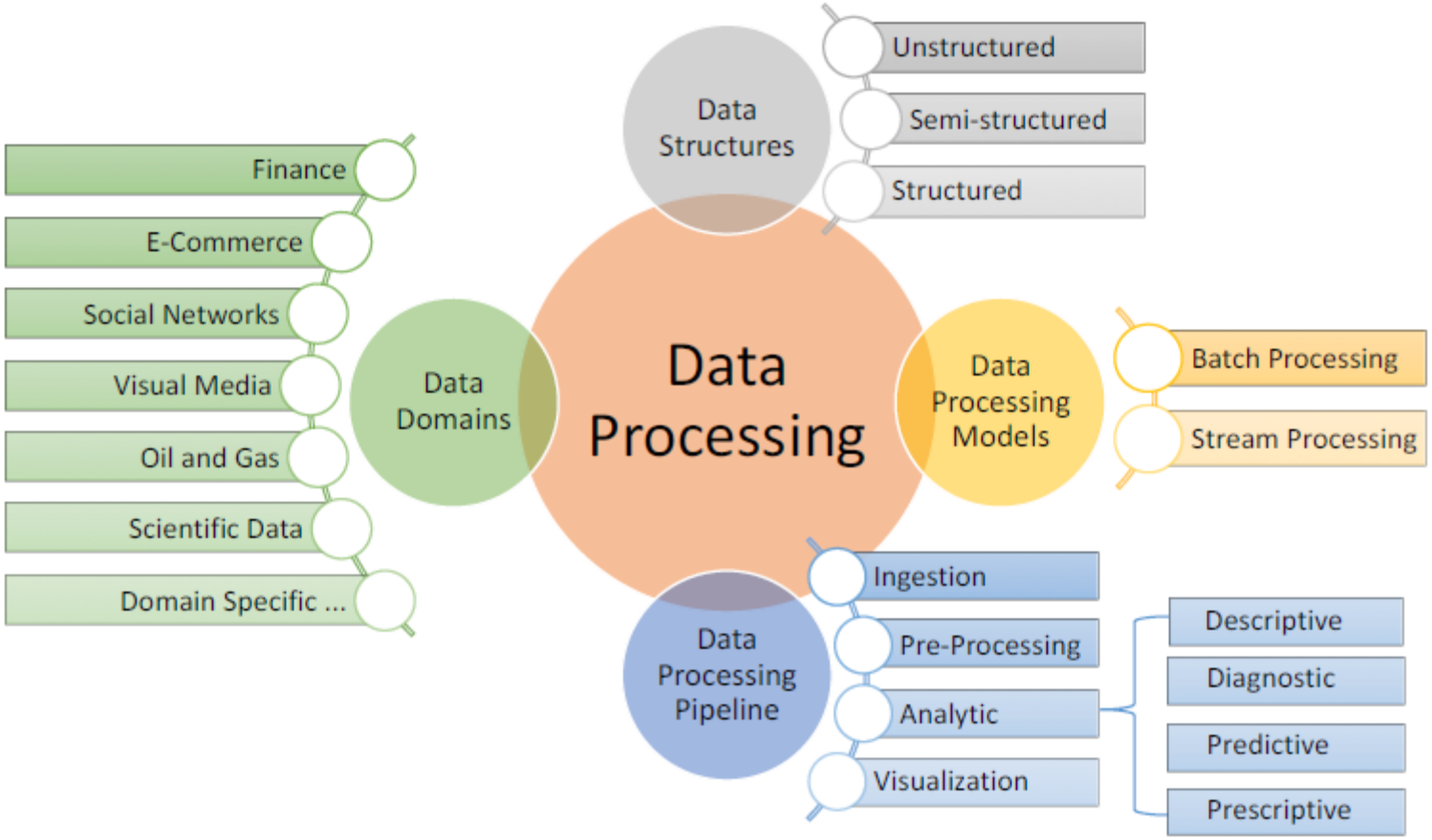
Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

As of 2012, every day 2.5 exabytes (2.5×10^{18}) of data are generated
[IBM, "What is Big Data"]



https://www.slideshare.net/PresentationLoad/big-data-51890676

The Taxonomy of Data Processing



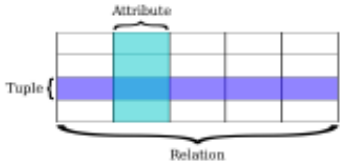
Data Structures

Unstructured Data

Semi-structured

Structured Data

Relational

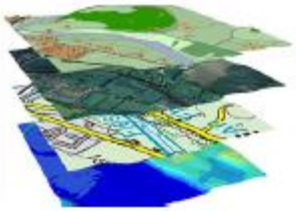


Around 80-90% of all potentially usable business information may originate in unstructured form [Merill Lynch, 1998]

Graph DB → Network Model DB e.g., CODASYL (1969)

Document DB → Hierarchical Model DB e.g., IBM IMS (1969)

Other Data Structures



Spatial / Geospatial Data

Spatio-temporal Data

e.g., Moving Objects

Biological Data

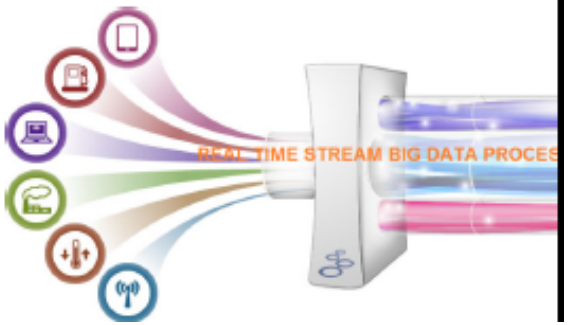
1000 Genomes Project → >200 Terabytes
https://aws.amazon.com/1000genomes/

Million Human Genomes project → ???

Data Stream

- … is an ordered sequence of instances that in many applications can be read only once or a small number of times using limited computing and storage capabilities*

- Data Stream Processing Applications:
  - IoT applications
  - Live datamart
  - Pattern mining on live data
  - …

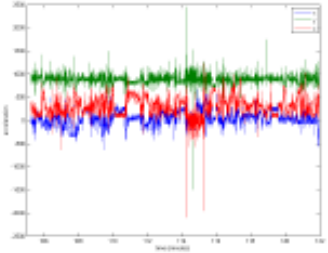


Data Stream Characteristics

- Continuous flow of data
- Infinite length
  - not just BIG but "UNLIMITED"
  - impractical to store all data
- Examples:

Call Detail Records (CDR)

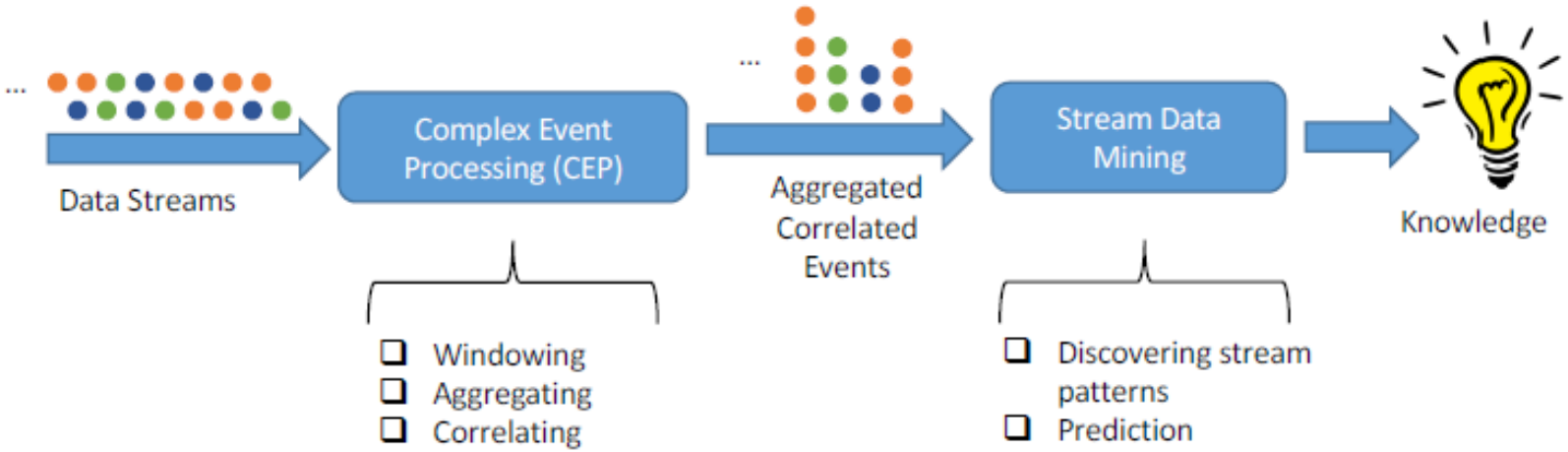
Tweets



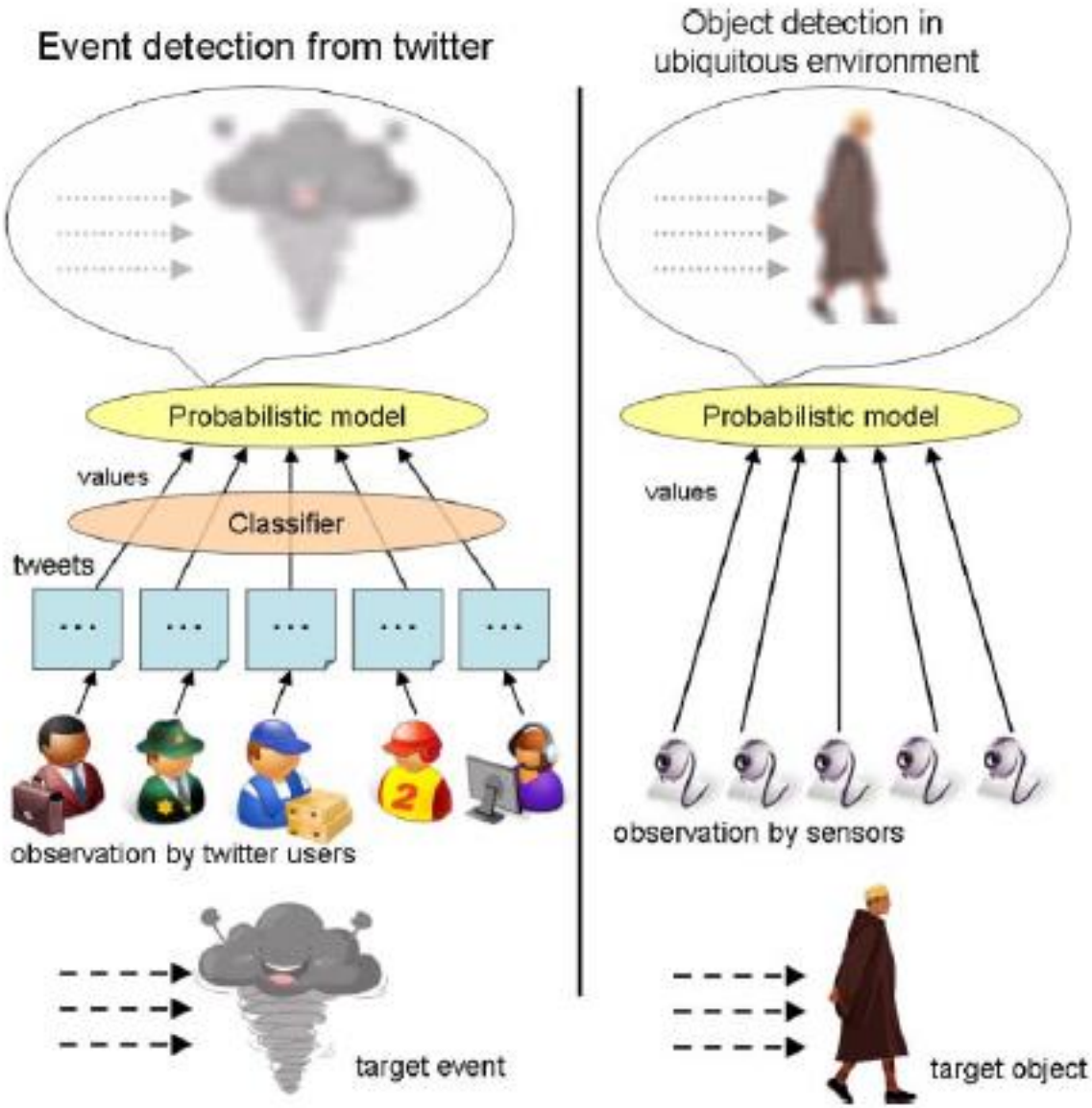
Sensor Data

Data Stream Processing

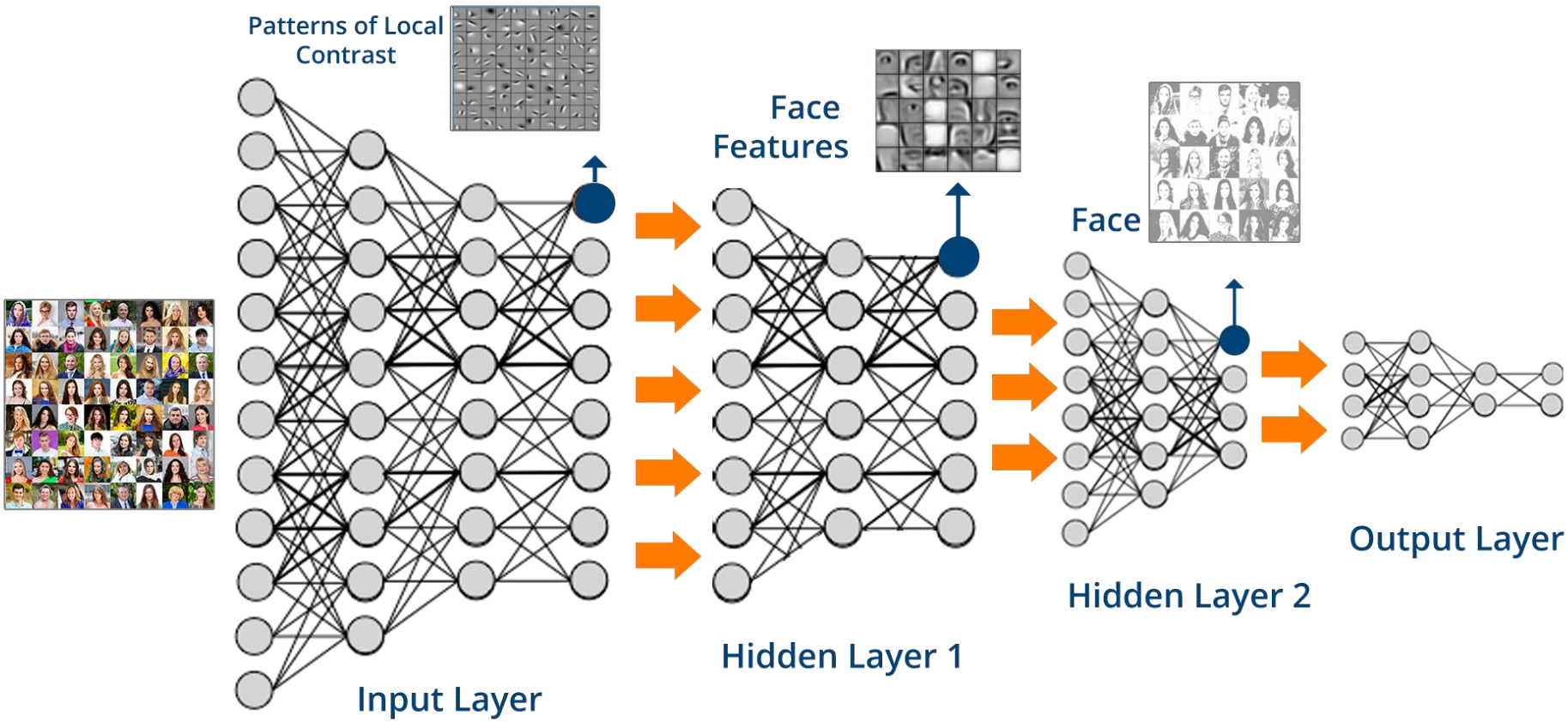
An example of a typical data stream processing flow



Digital vs Analog World



Deep Learning



Big Data
Correlation vs Causation

Is correlation more important than causation?

Correlations play an important role as heuristic devices [but] have to be further analyzed [. . .] to assign them a meaning”

The correlations may not tell us precisely why something is happening, but they alert us that it is happening.

Big Data
Correlation vs Causation

Thomas Kuhn:

Anomalies, by definition, For such discoveries to occur, establishing that there is something that does not match our expectations is not enough. We have also to find out what it is. This process does not arise directly from data or numbers, but rather from a change in how we look at them, and it involves a reassessment of our beliefs and methodologies.

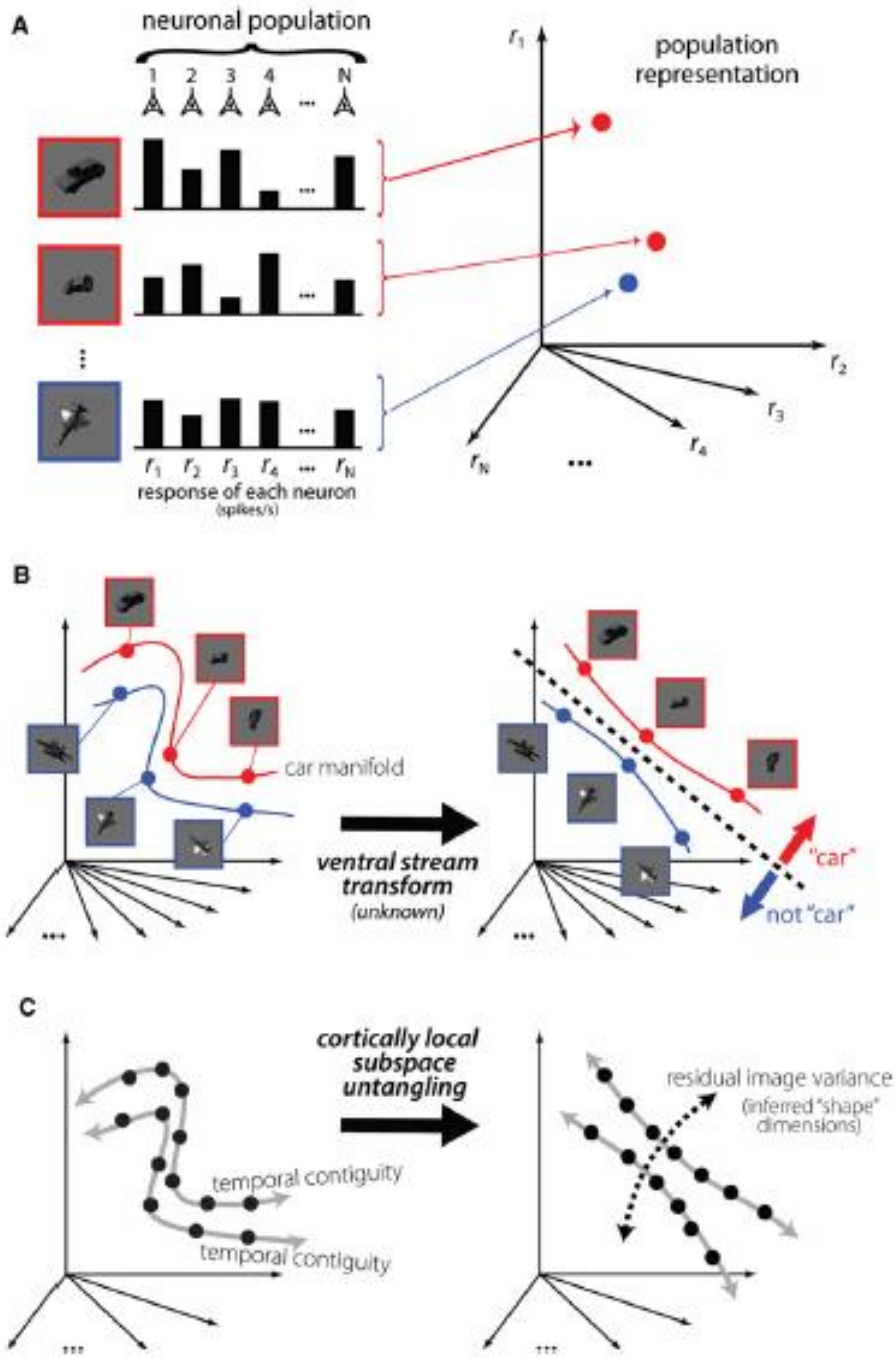
Big Data
Correlation vs Causation

"Big Data, distributed computing and sophisticated data analysis all played a crucial role in the discovery of the Higgs boson [. . .] But the discovery of the Higgs boson was not data-driven."

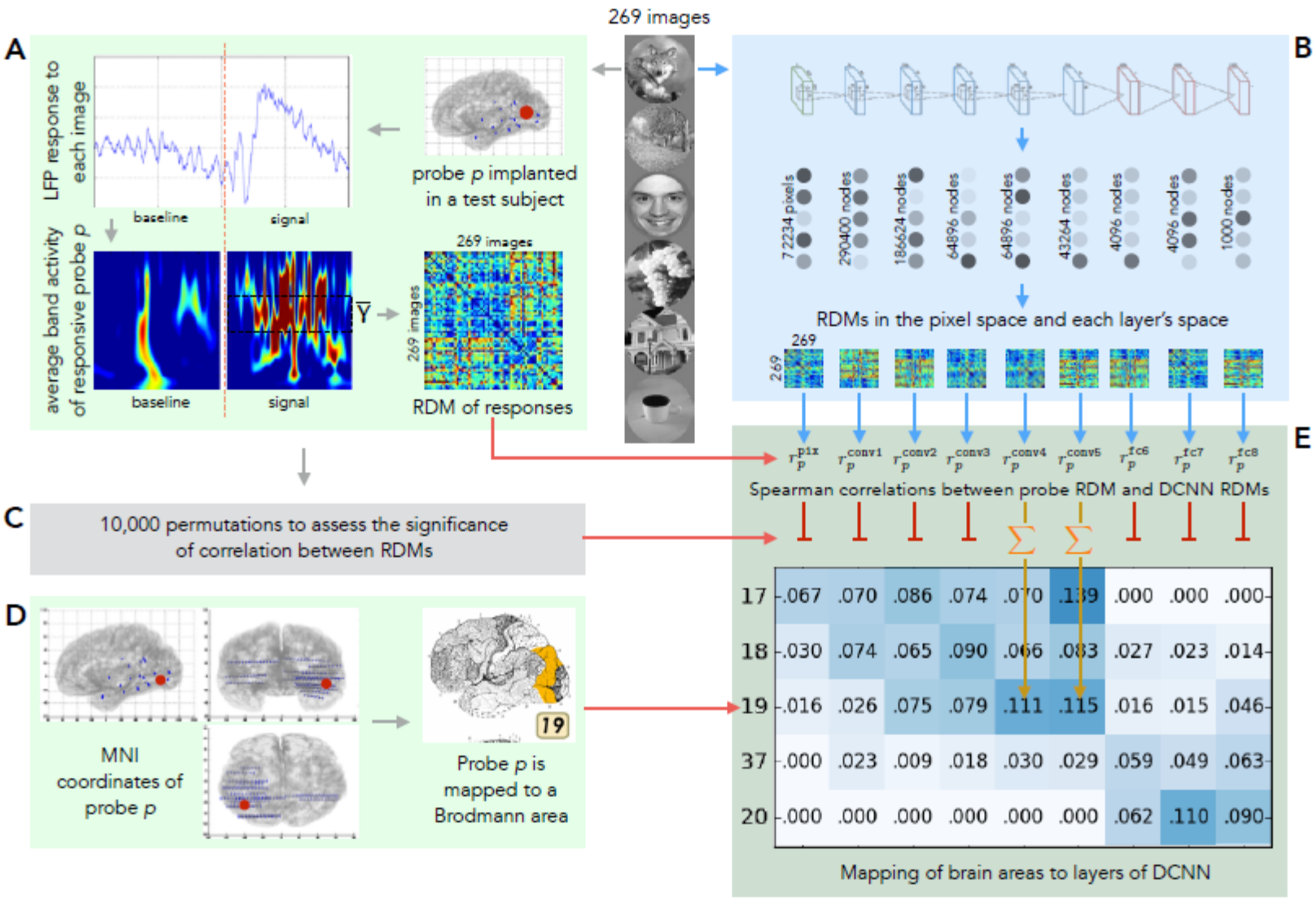
"The data-driven approach constitutes a novel tool for scientific research. Yet this does not imply that it will supersede cognitive and methodological procedures. . ."

Some Applications

How does the Brain solve visual objective recognition

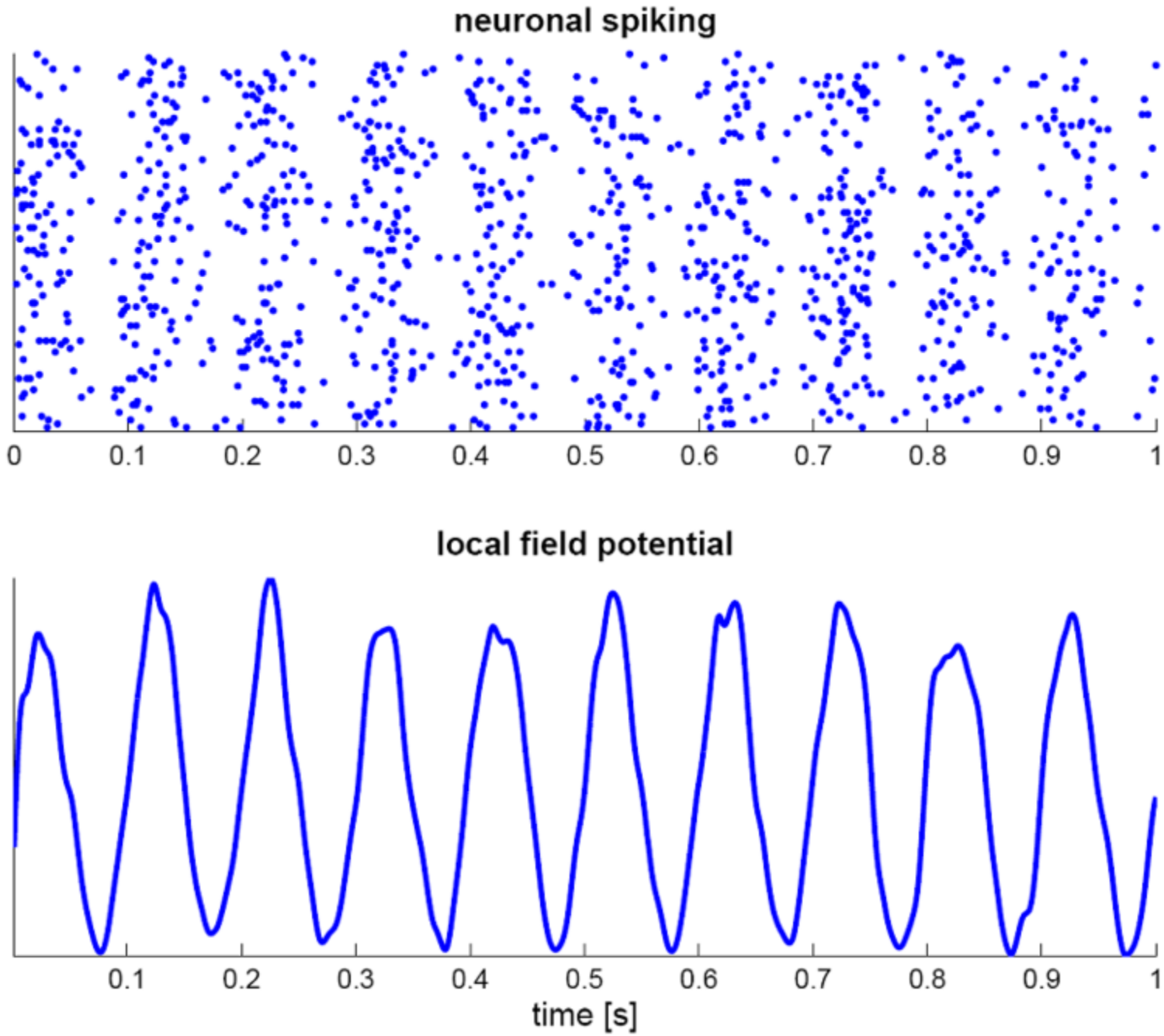

A
neuronal population
1 2 3 4 ... N
r_1 r_2 r_3 r_4 ... r_N
response of each neuron
(spikes/s)
population representation
r_1
r_2
r_3
r_4
r_N
B
car manifold
ventral stream transform
(unknown)
"car"
not "car"
C
cortically local subspace untangling
temporal contiguity
temporal contiguity
residual image variance
(inferred "shape" dimensions)

Natural images vs artificial vision system

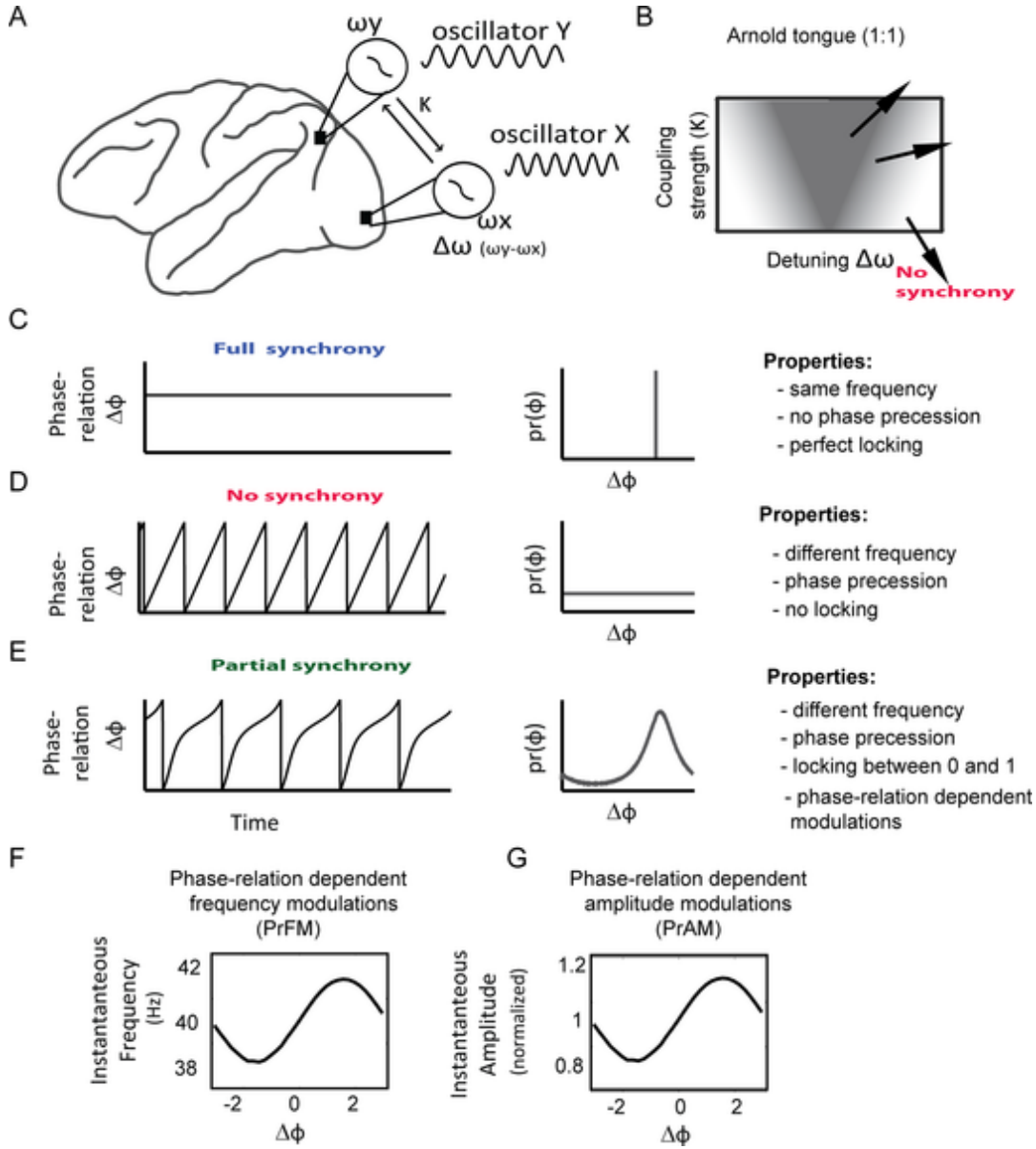


The phenomenon of synchronization

- Circadian rhythms
- Electrical generators
- Heart, intestinal muscles
- Menstrual cycles
- Fireflies
- Applause (esp in Eastern Europe)


neuronal spiking
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
local field potential
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
time [s]

The dynamic emergence of coherent physiological activity, such as phase-locked high-frequency electromagnetic oscillations,



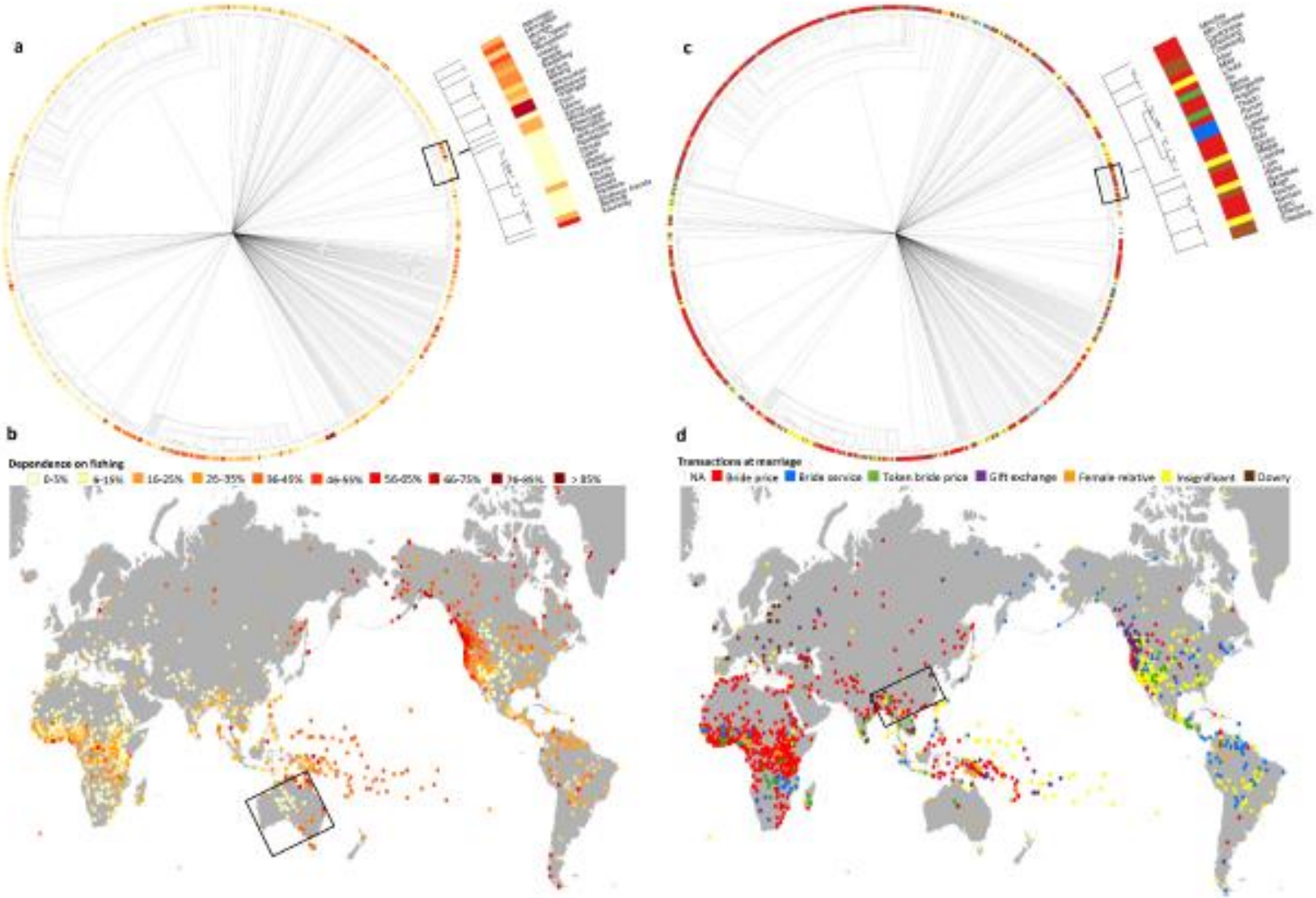
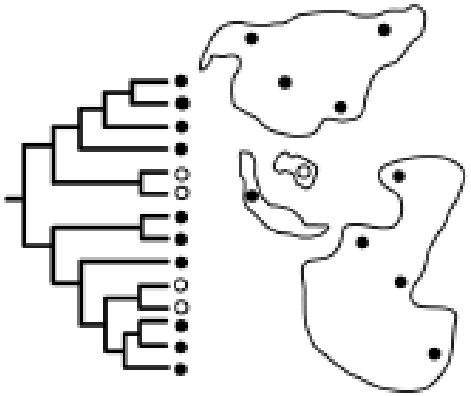


Fig 1. D-PLACE links cultural information to language classifications and phylogenies (a, c) and to geographic locations and environmental features (b, d). This allows users to consider the relative influence of cultural ancestry, spatial proximity, and environment on diverse cultural practices. For example, panels a and b illustrate variation among societies in their dependence on fishing relative to other subsistence activities, based on data from the Ethnographic Atlas (EA) [11–15] and the Binford Hunter-Gatherer dataset [16,17]. Panels c and d highlight diversity in the most common economic transaction at marriage, based on data from the EA. In addition to providing global results, D-PLACE allows users to focus a search on a particular geographic region or linguistic family. Here, results for societies speaking Pama-Nyungan languages (a, b) or Sino-Tibetan languages (c, d) are magnified and outlined in black boxes on the global tree and map.

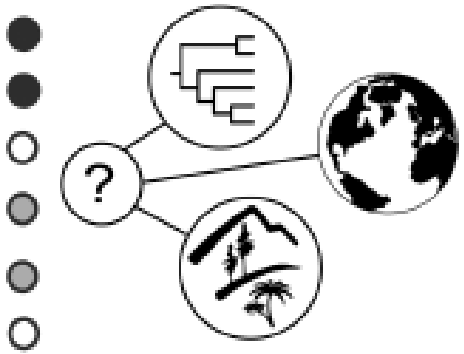
a. Exploratory

How are features distributed across societies?



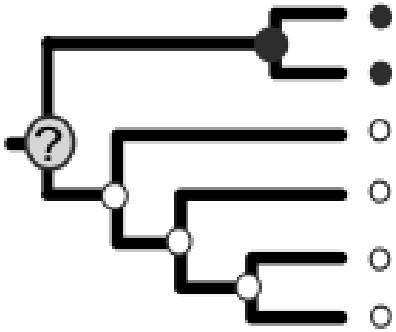
b. Regression Analysis

What predicts patterns of cultural diversity?



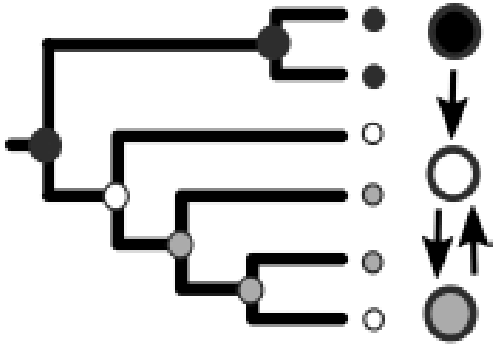
c. Ancestral States

What was the earlier form of a feature?



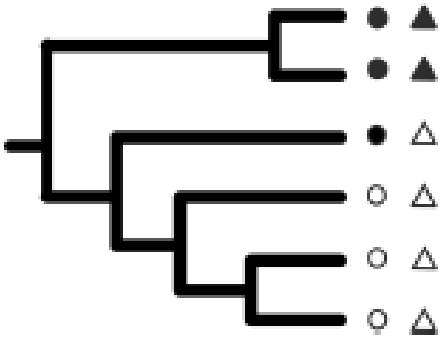
d. Transformation

How do cultural features change form?



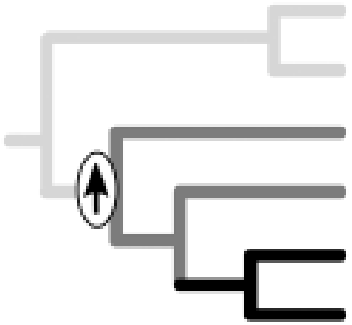
e. Correlated Evolution

Do features change together?

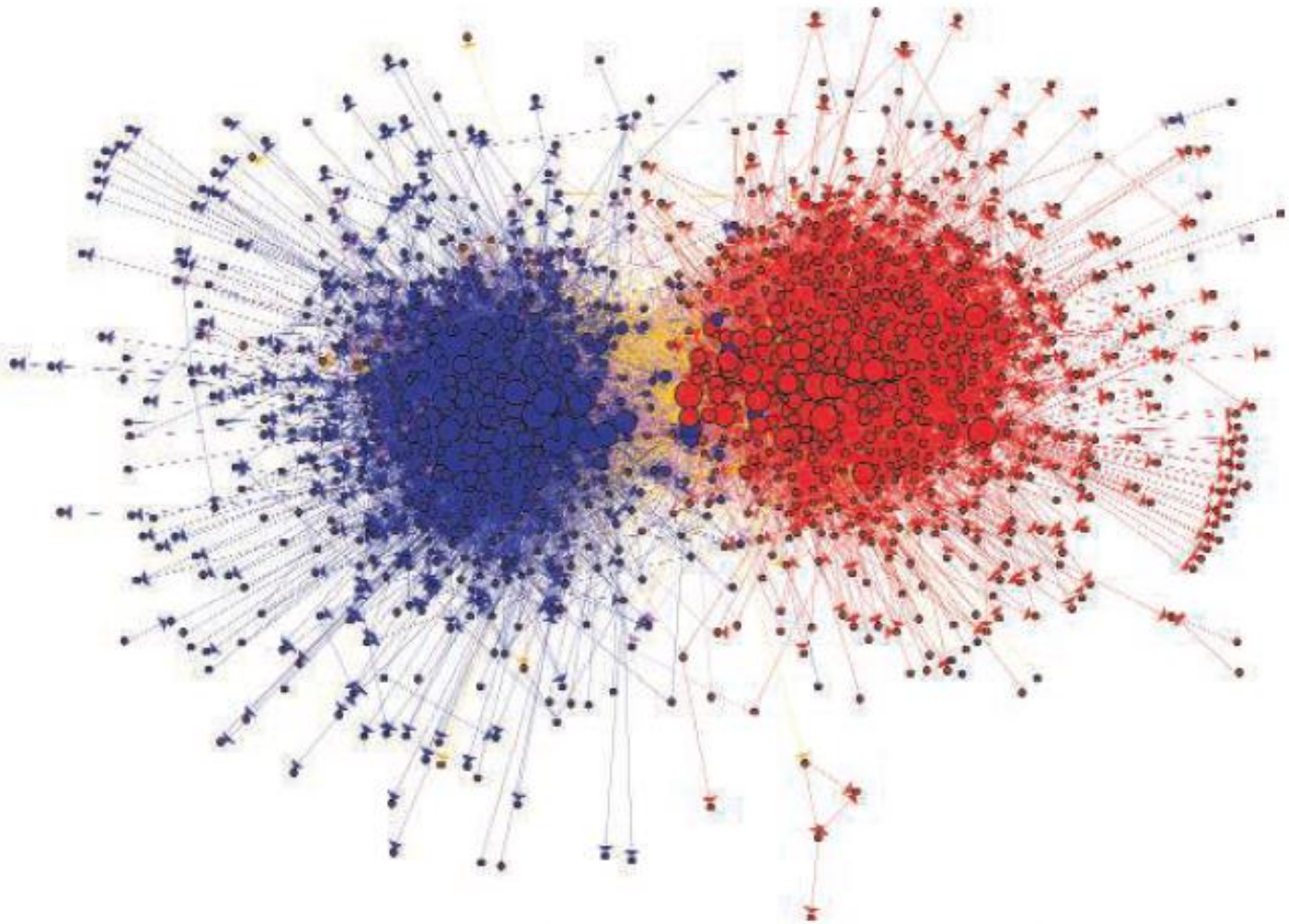


f. Mode and Tempo

How and when do features diversify?

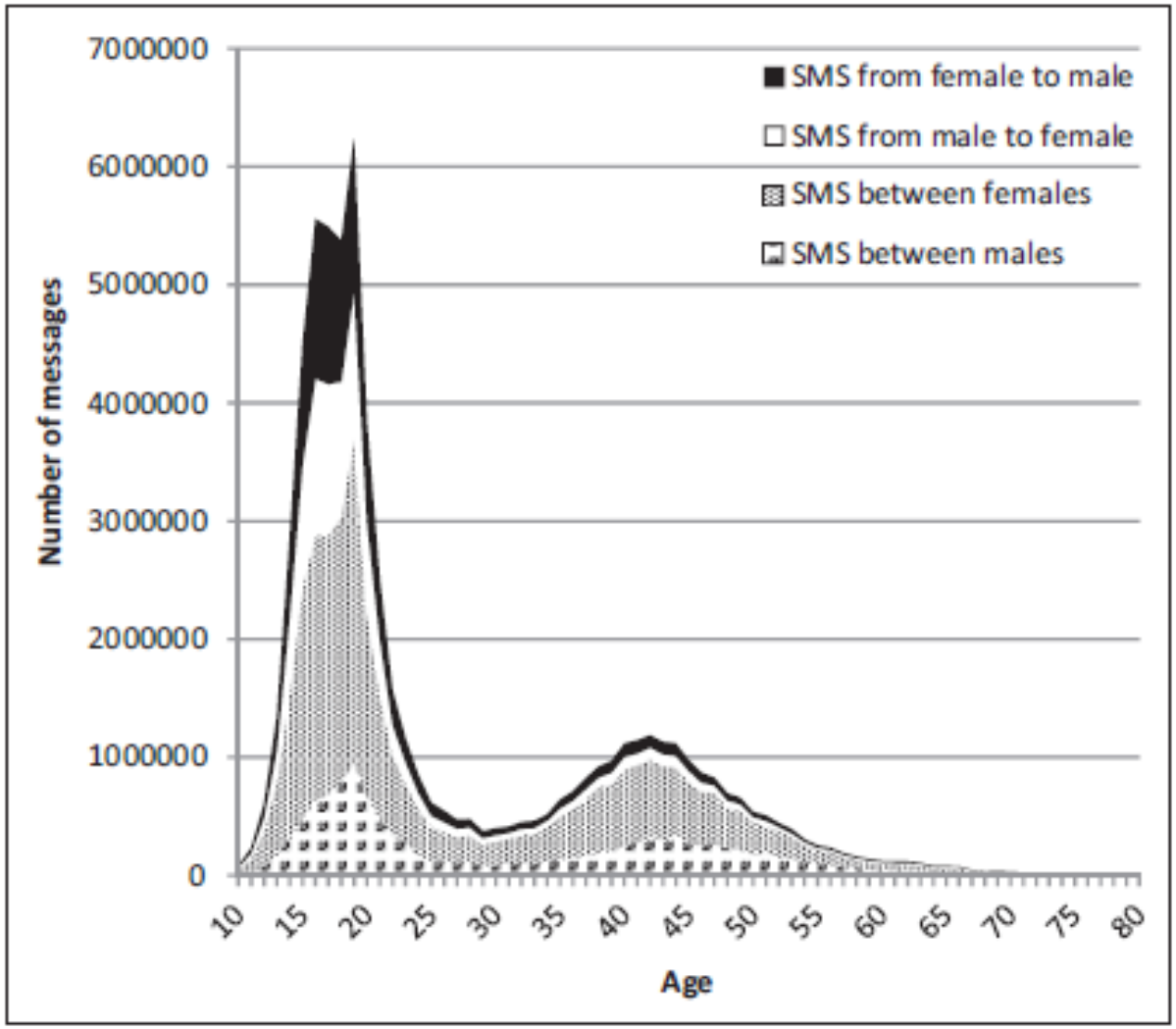


Political Blogs

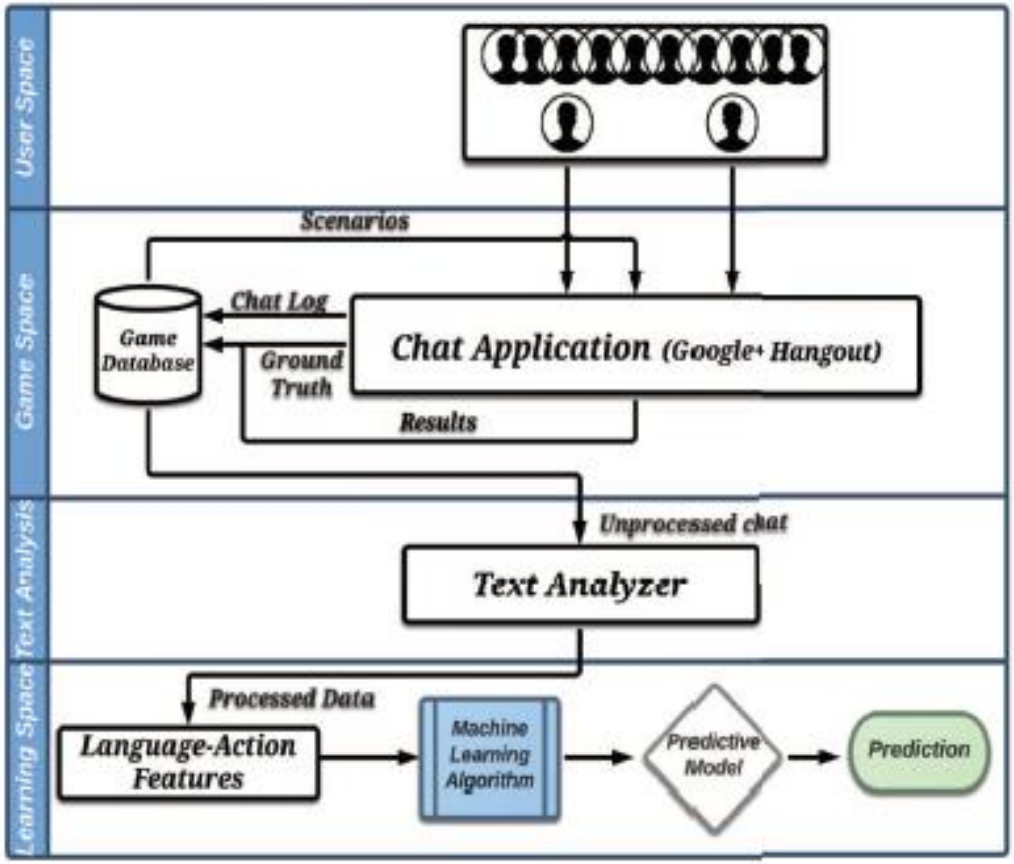


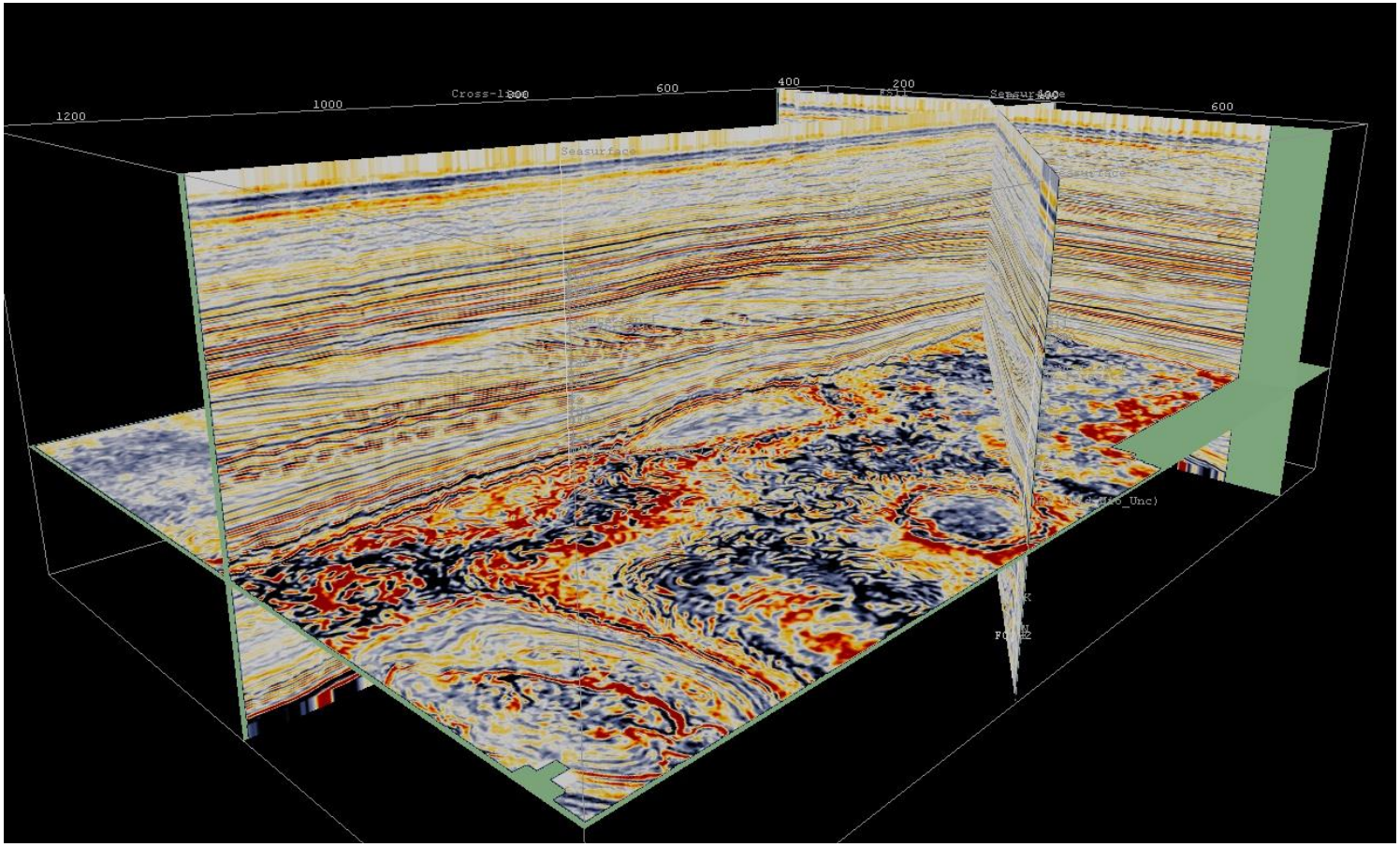
Data from the blogosphere. Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it. [Reproduced from (*8*) with permission from the Association for Computing Machinery]

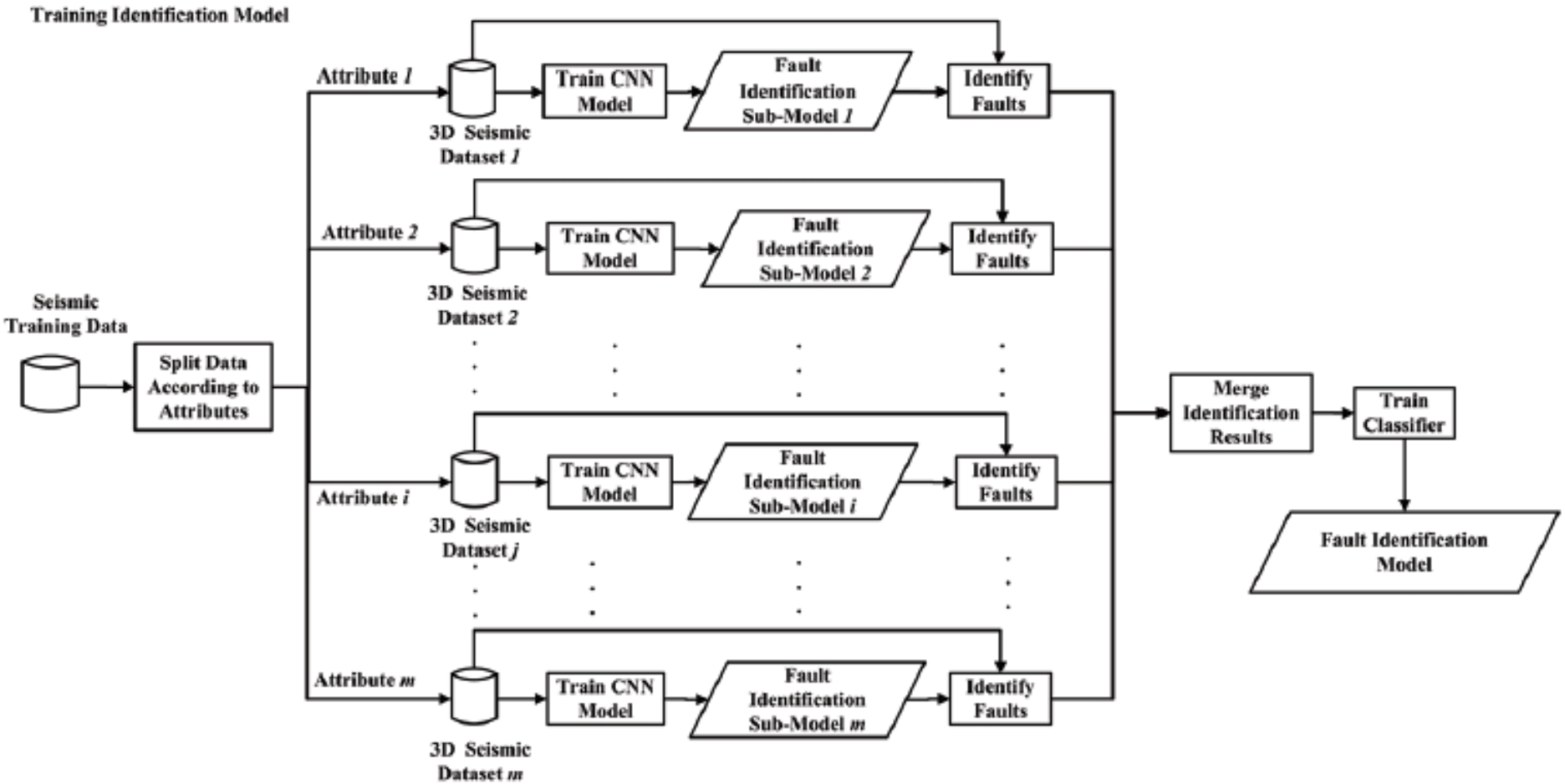
Texting

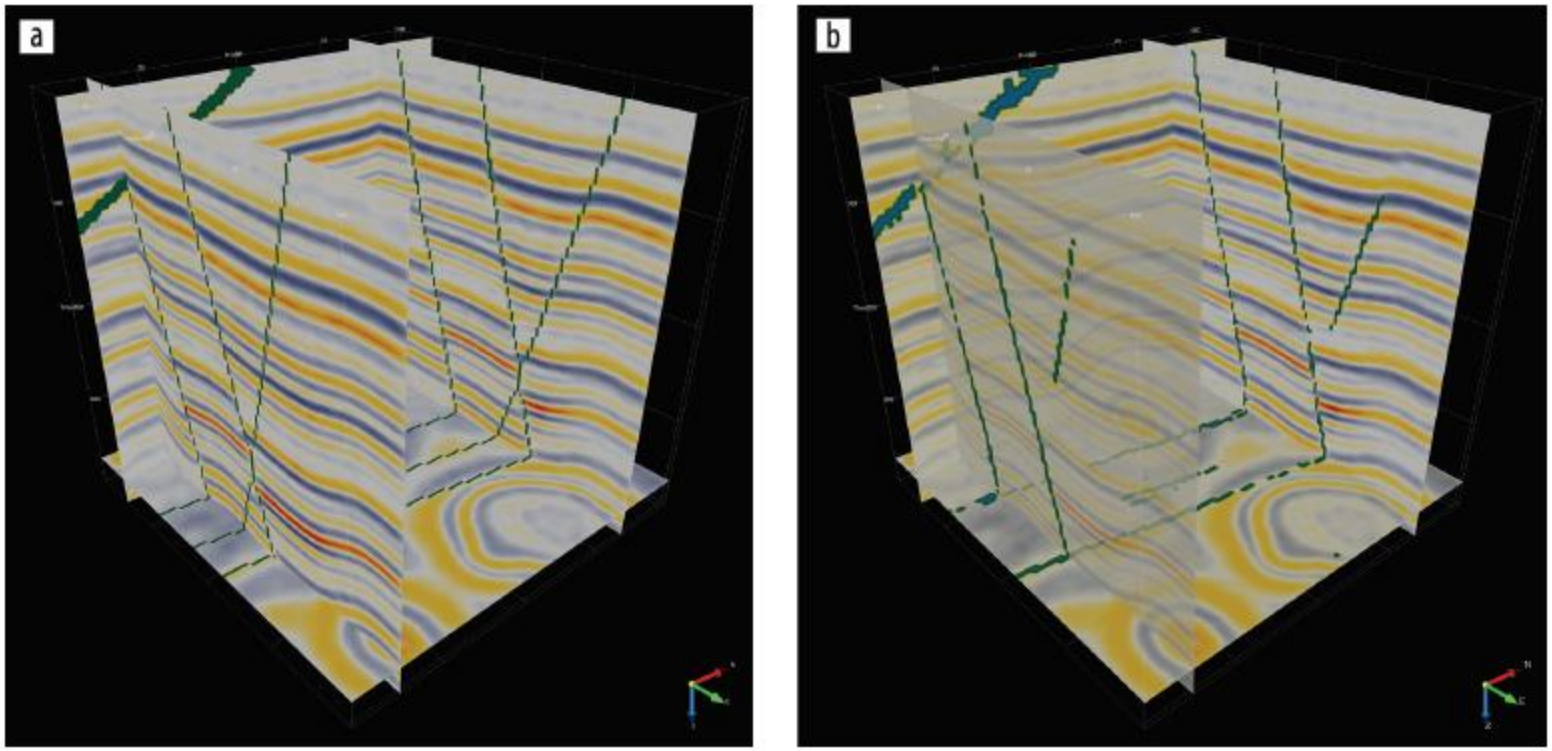

SMS from female to male
SMS from male to female
SMS between females
SMS between males
Number of messages
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
Age

Automated Detection of Deceptive Language-Action Cues




1200
1000
Cross-line
800
600
400
200
Seasurface
600


Training Identification Model
Seismic Training Data
Split Data According to Attributes
Attribute 1
3D Seismic Dataset 1
Train CNN Model
Fault Identification Sub-Model 1
Identify Faults
Attribute 2
3D Seismic Dataset 2
Train CNN Model
Fault Identification Sub-Model 2
Identify Faults
Attribute i
3D Seismic Dataset j
Train CNN Model
Fault Identification Sub-Model i
Identify Faults
Attribute m
3D Seismic Dataset m
Train CNN Model
Fault Identification Sub-Model m
Identify Faults
Merge Identification Results
Train Classifier
Fault Identification Model

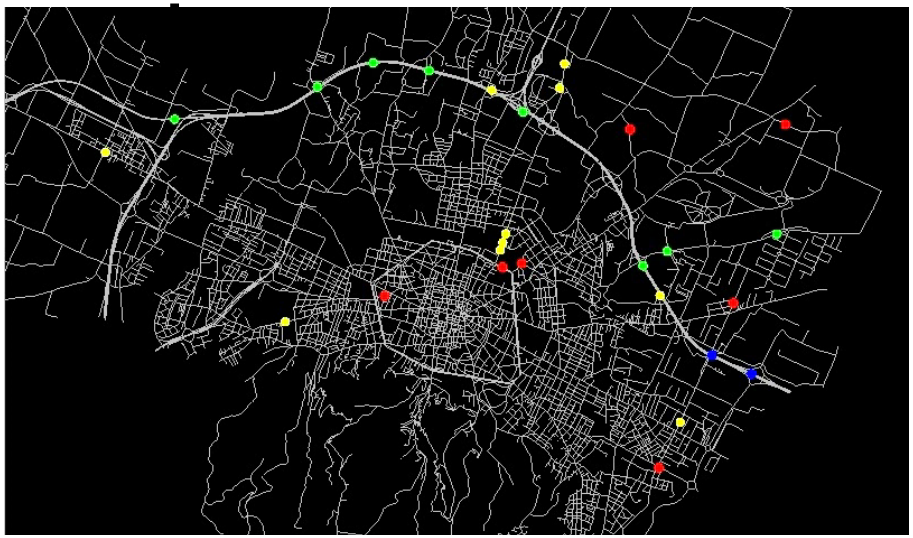

a
b

URBAN TRAFFIC

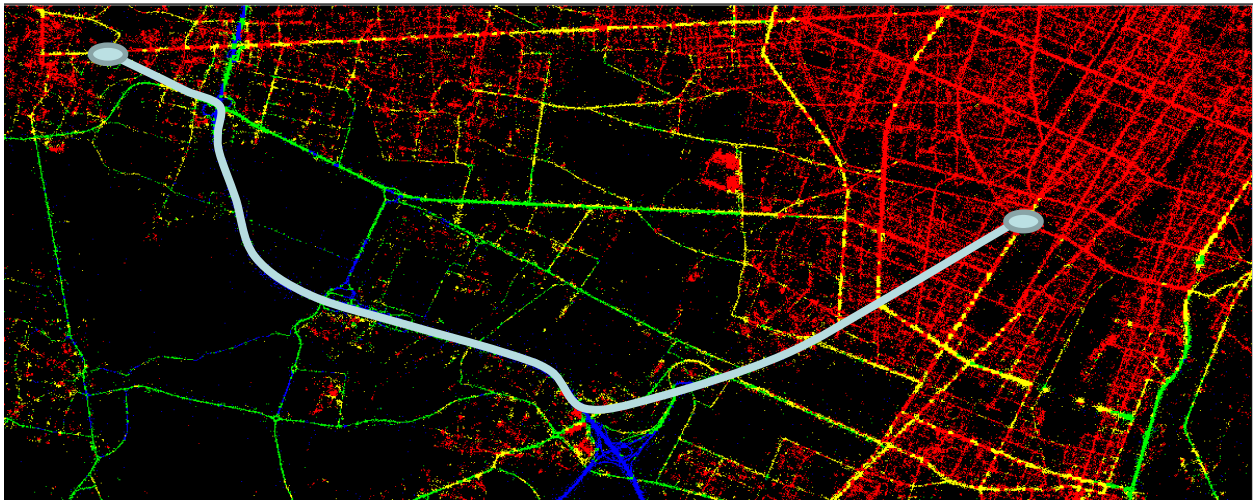
Heterogeneity

- Spatial and Temporal
- Congestion Level
- Topology
- Modes of transport
- Sensing equipment

Sparse multi-sensor



Develop travel time field and route choice information



Pictures provided by Armando Bazzani (Un. of Bologna)