Using Deep Data and AI to Address Emerging Challenges:

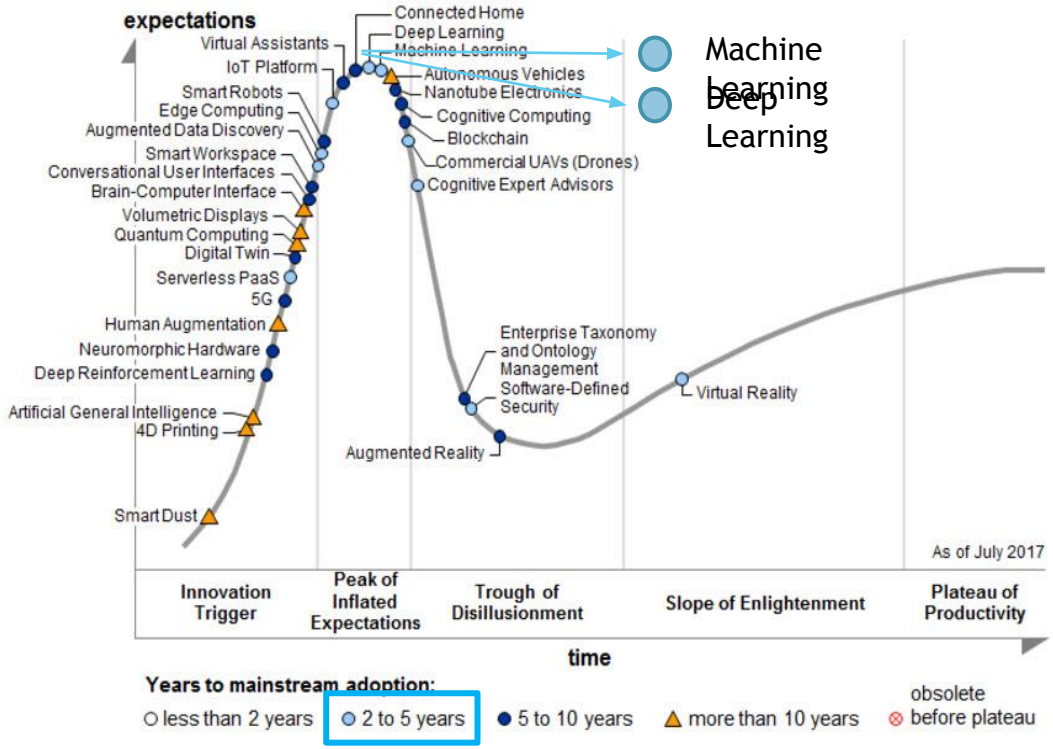
How we live and die through data

Christopher R. Stephens
Director of Data Science, C3 – Center for Complexity Science, UNAM
Full Professor, Institute for Nuclear Sciences, UNAM
Founder, Presage Research
Seminar - Thunderbird, ASU 7th February, 2020

PRESAGE RESEARCH
science to solutions

Gartner's 2017 Hype Cycle for Emerging Technologies

Hype Cycle for Emerging Technologies, 2017

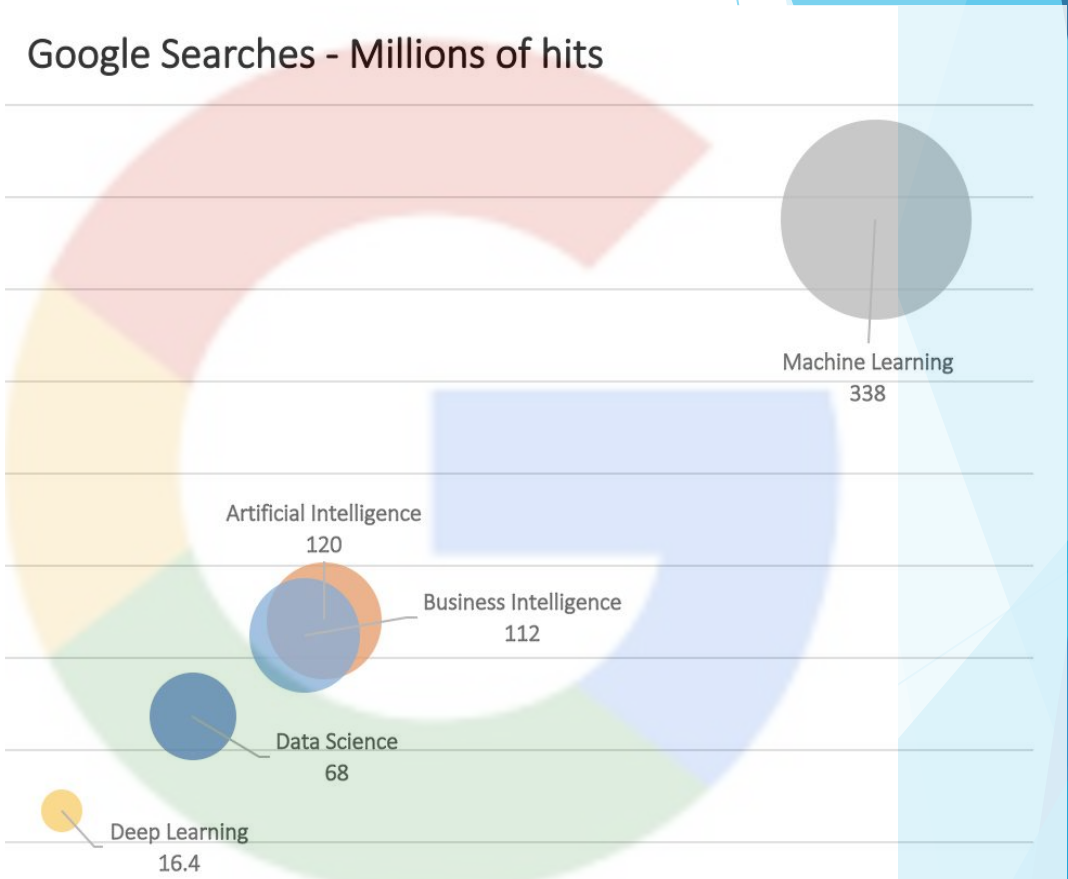


Machine learning

Deep Learning

Note: PaaS = platform as a service; UAVs = unmanned aerial vehicles

Source: Gartner (July 2017)



This is an Emerging Challenge

00100010111111100010001000111011111100100
10010100110000011111000110010010010000010
Энэ бол шинэ сорилт юм
This is an Emerging Challenge
Дэлхий даяар коронавирусээр халдварласан
Virus Infected Worldwide
Infected with Corona Virus
Вирусээр халдварласан дэлхий
Дэлхий даяар коронавирусээр халдварласан

Data

NO

0010001011111110001000100011101111100100100101001100000111110001100100100100001001110011011001100100011011001101110111110001001110011011111100100010111111100011001001001000100011000111100100011010111011111000110010001011111110010010000101000100001111011101111100011001001111110011010001

NO Энэ өрөөнд байгаа хэн нэгэн нь корона вирусын халдвар авсан байж магадгүй юм

NO Dh ’fhaodadh gum bi bhìoras corona air cuideigin san t-seòmar seo

YES **Someone in this room may be infected with corona virus**

Meaning

Meaning

NONE	Spanish flu killed 20-50 million people worldwide in 1918. Its mortality rate was 10-20%
LOW	Coronavirus has killed 391 people in 2020, none in the US. Its mortality rate is about 2%
HIGH	Someone in this room may be infected with coronavirus
VERY HIGH	The person sitting next to me is infected with coronavirus and wants to shake my hand

Valu

Valu

Spanish flu killed 20-50 million people worldwide in 1918. Its mortality rate was 10-20%

e

Prediction I can't be infected with Spanish flu

Coronavirus has infected over 17000 people in China and killed 391. There are 11 cases in the US.

Prediction The risk is low that anyone I've had contact with has coronavirus

Someone in this room may be infected with corona virus

Prediction I am at risk of being infected, potentially high risk if the infected person is someone I've had contact with

The person sitting next to me is infected with coronavirus and wants to shake my hand

Prediction I am at high risk and even more so if I shake hands

Prediction

Prediction

I can't be infected with Spanish flu

Decision Don't do anything

The risk is low that anyone I've had contact with has coronavirus

Decision Avoid travel to high risk areas or with high risk individuals

I am at risk of being infected, potentially high risk if the infected person is someone I've had contact with

Decision Report to my doctor for a test and try to determine who the infected person is

I am at high risk and even more so if I shake hands

Decision Decide not to shake hands and report to my doctor

Decision

Decision

Don’t do anything

Action Nothing

Avoid travel to high risk areas or with high risk individuals

Action Cancel business trip to Wuhan

Report to my doctor for a test and try to determine who the infected person is

Action Visit doctor, explain situation, consult with ASU health authorities

Decide not to shake hands and report to my doctor

Action Don’t shake hands, apologize to person, visit doctor, explain situation

Action

DMV ▫ PDA

Data + Meaning + Value ▫ Prediction + Decision + Action

Example: DMV ▫ PDA

Data = person sitting next to me is infected with coronavirus
and wants to shake my hand
Meaning (key features) = “physical proximity”, “contact”,
“transmission”, “infected”

Value = High – value on potential very negative health consequences
and on reliability of the data

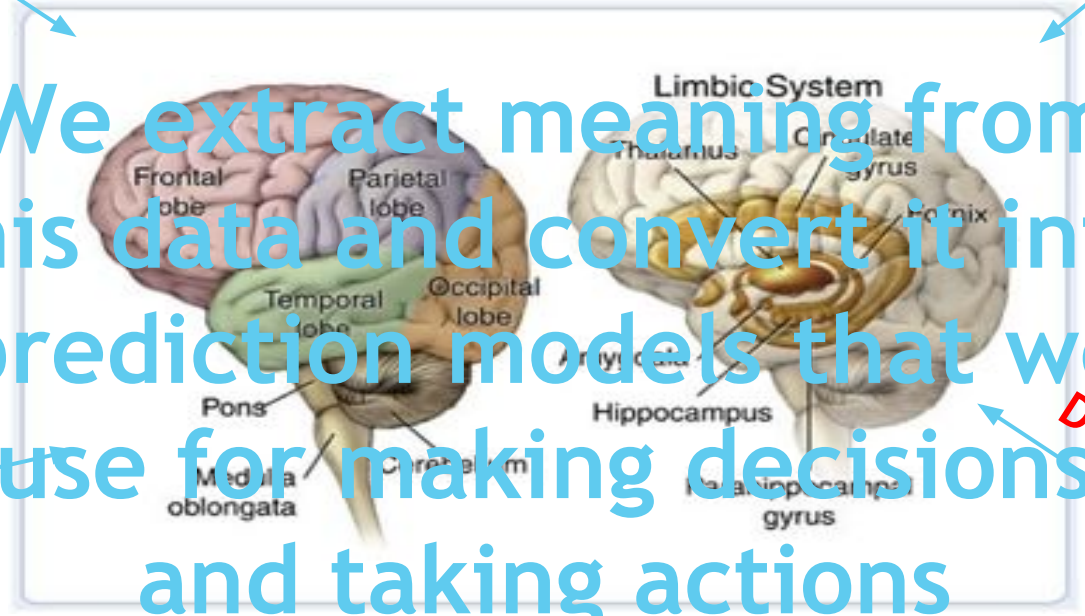
Prediction = I may be infected with coronavirus if I shake hands
Decision = To politely refuse to shake hands
Action = Don’t shake hands

Data Scientists
Data
Data
Data
Data
literally
fuel of life
They represent
the "real" world

Datos

Each person is processing about
11 Mb/s from our senses
(3 photos in terms of pixels)

Datos



We extract meaning from this data and convert it into prediction models that we use for making decisions and taking actions

Datos

Datos

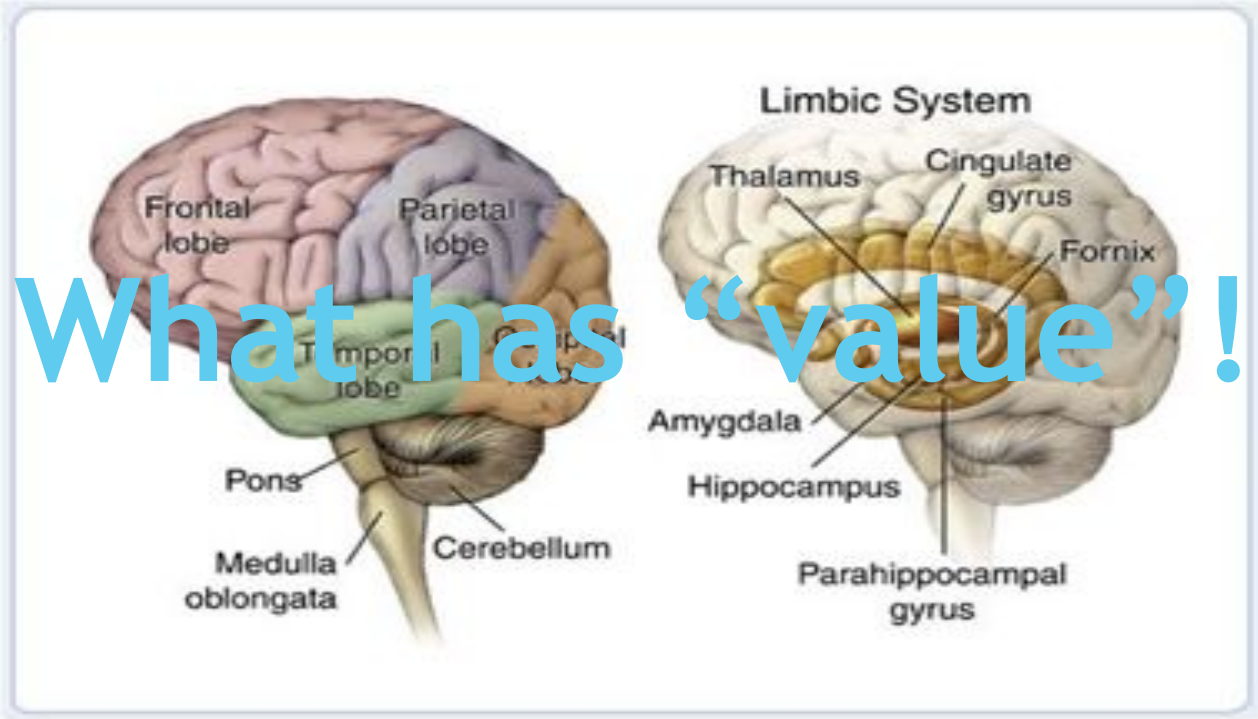
The human race is processing about
1 yottabyte (10^{24} b/year) from our senses

The human brain can store about 10-100Tb

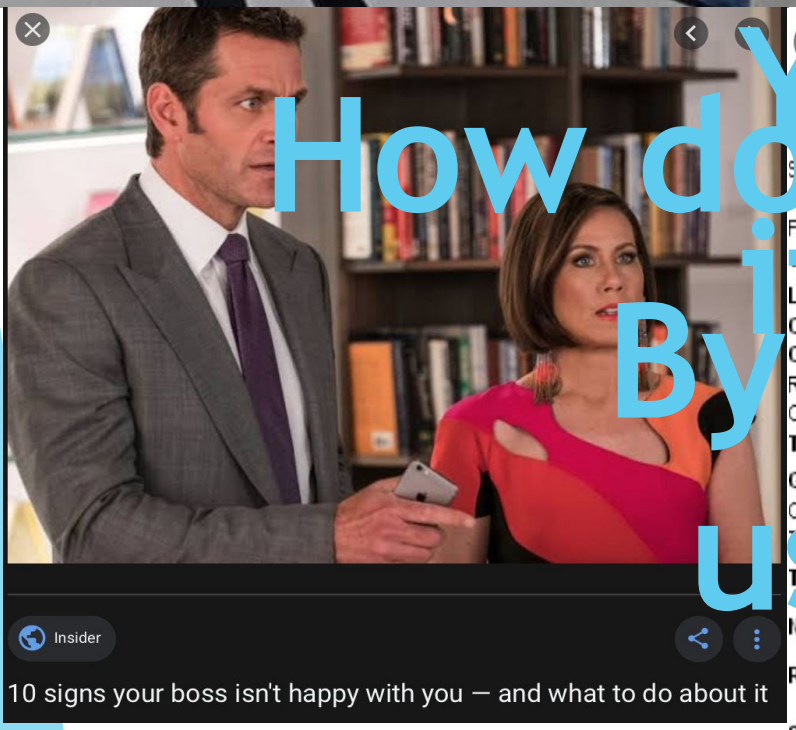
If it stored all the 1Mb/s we'd

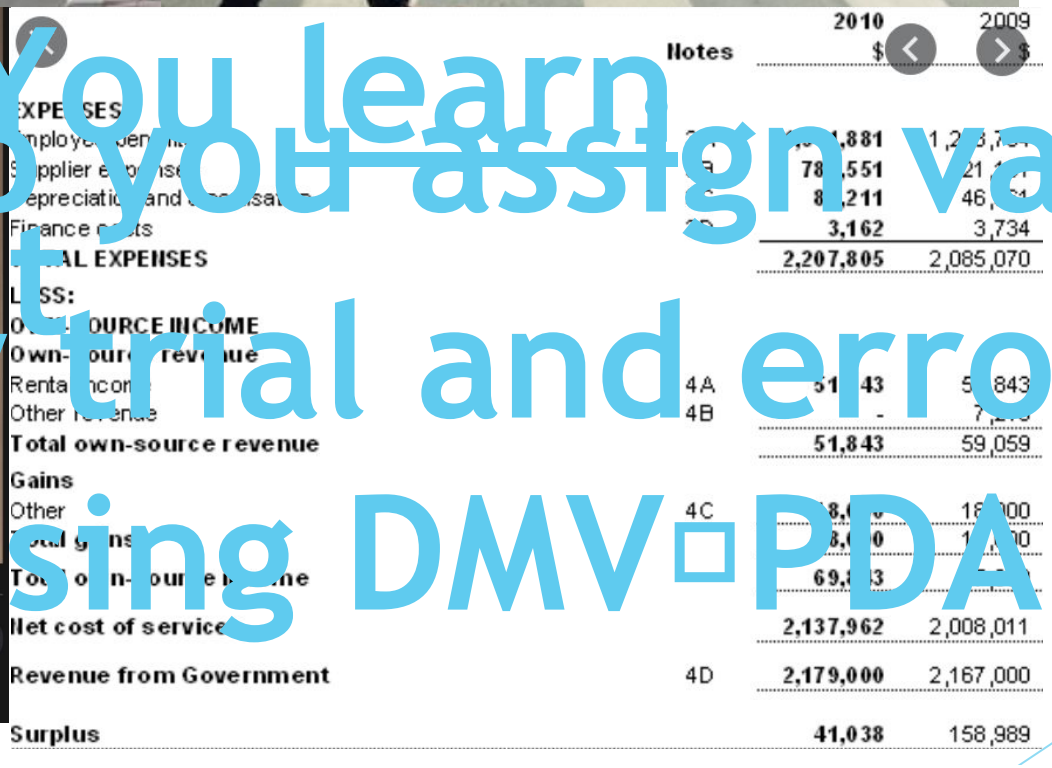
“hard disc” would be full in 3 months

What does it “choose” to store?



What has “value”!





	Notes	2010 $	2009 $
EXPENSES			
Employee benefits		[illegible],881	1,2[illegible]
Supplier expenses		78[illegible],551	21[illegible]
Depreciation and [illegible]		8[illegible],211	46,[illegible]
Finance costs		3,162	3,734
TOTAL EXPENSES		2,207,805	2,085,070
LESS:			
OWN-SOURCE INCOME			
Own-source revenue			
Rental income	4A	51[illegible]43	5[illegible]843
Other revenue	4B	-	7,[illegible]
Total own-source revenue		51,843	59,059
Gains			
Other	4C	[illegible]	18[illegible]00
Total gains		[illegible]	[illegible]
Total own-source income		69,[illegible]	[illegible]
Net cost of service		2,137,962	2,008,011
Revenue from Government	4D	2,179,000	2,167,000
Surplus		41,038	158,989

You learn
How do you assign value?
it
By trial and error
using DMV-PDA

01/11/1980

5538151775

SECR590417

Reyas

Susan Cordts

F = ma

s(x) = ln(P(x|C)/P(x|C'))

P(C | X(t)) represents **our model** of reality and perception

Heuristic: we don't know what it is in humans.

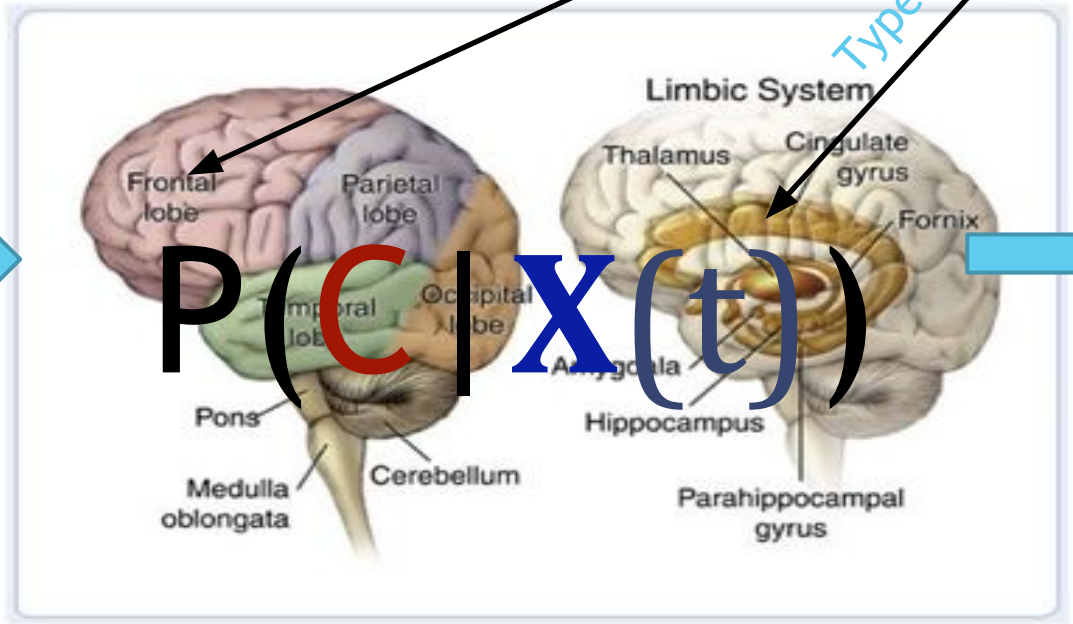
It is a model of the world.

DMV ▫ PDA

Adaptability

Type II thinking

Type I thinking



PDA

There are many alternatives to any decision and many possible actions

Human Intelligence

Did it work?

DMV ▫ PDA is judged to be good or bad according to a performance criterion

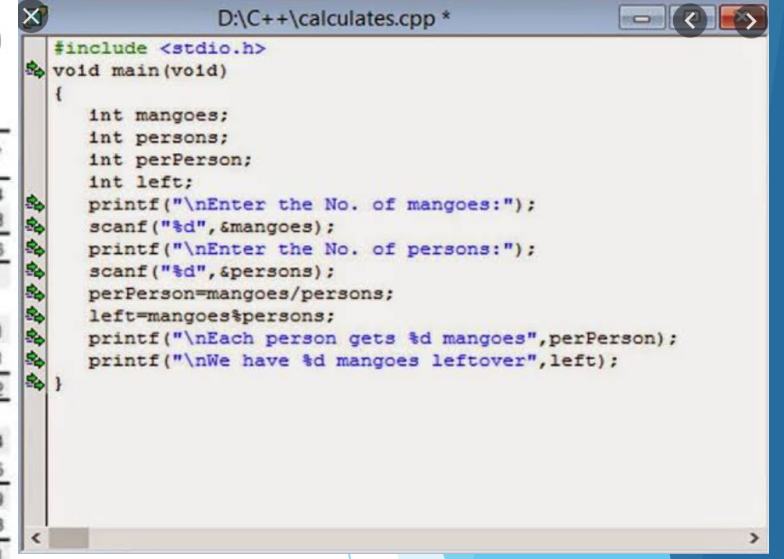
This represents the culmination of about 4.5 billion years of evolution and represents how, until recently, humanity did everything

There’s just one problem…

CONDENSED CONSOLIDATED STATEMENTS OF OPERATIONS (Unaudited)

(In millions, except number of shares which are reflected in thousands and per share amounts)

	Three Months Ended		Twelve Months Ended	
	September 29, 2018	September 30, 2017	September 29, 2018	September 30, 2017
Net sales	$ 62,900	$ 52,579	$ 265,595	$ 229,234
Cost of sales [1]	38,816	32,648	163,756	141,048
Gross margin	24,084	19,931	101,839	88,186
Operating expenses:				
Research and development [1]	3,750	2,997	14,236	11,581
Selling, general and administrative [1]	4,216	3,814	16,705	15,261
Total operating expenses	7,966	6,811	30,941	26,842
Operating income	16,118	13,120	70,898	61,344
Other income/(expense), net	303	797	2,005	2,745
Income before provision for income taxes	16,421	13,917	72,903	64,089
Provision for income taxes	2,296	3,203	13,372	15,738
Net income	$ 14,125	$ 10,714	$ 59,531	$ 48,351



- Increased lead generation of the agency by 20% within the first 6 months by creating 4 "How-to" guides on the blog.
- Contributed to the increase by 35% of new clients in 2017 compared to 2016.

Marketing Intern
ABC Agency
05/2015 – 12/2015
Achievements
- Increased social media presence of the company by 23% in the last 3 months.
- Drafted a new on-boarding email process that increased the email open rate from 31% to 36%.
- Created 2 marketing-related articles per month which had on average 150 social media shares each increasing brand awareness.

VOLUNTEER EXPERIENCE

Project/Campaign Management

Public Speaking

CERTIFICATES

Hootsuite Social Marketing Certificatic (03/2019)

Google Ads Certification (08/2018)

LANGUAGES

English
Native

Spanish
Full Professional Proficiency

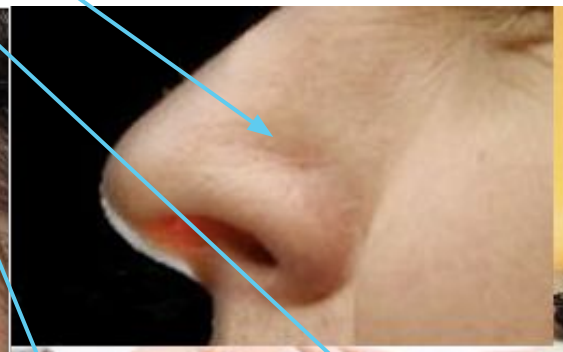
French
Limited Working Proficiency

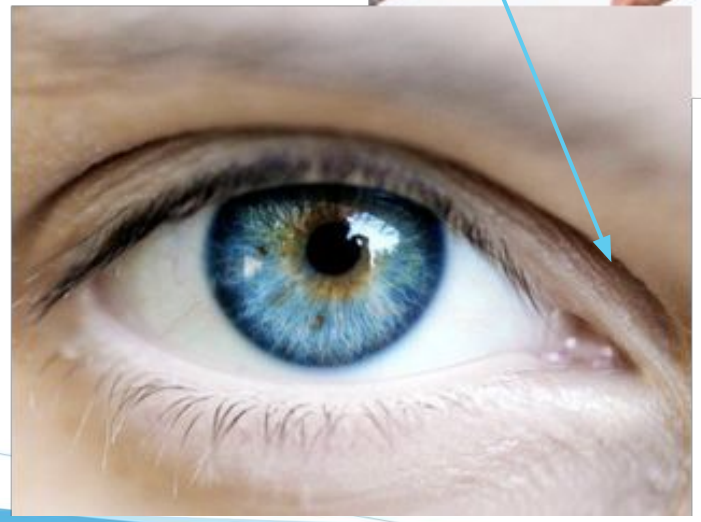
Italian

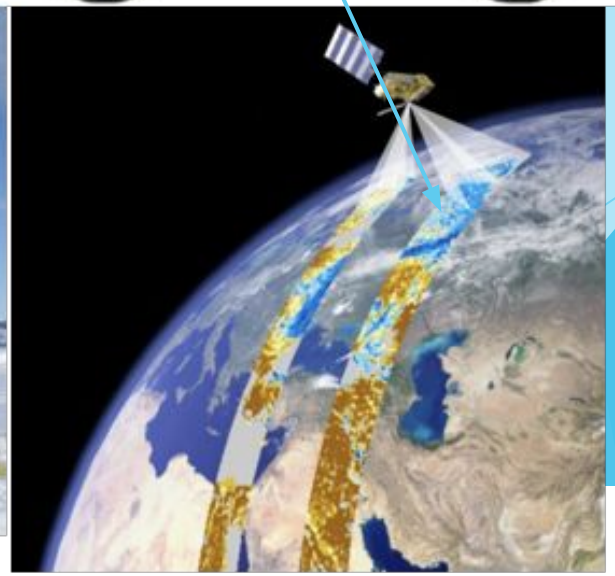
"Recently I've learned how to
Everything I've learned how to
Unfortunately I've learned how to
Every time I've learned how to
Every time I've learned how to
Every time

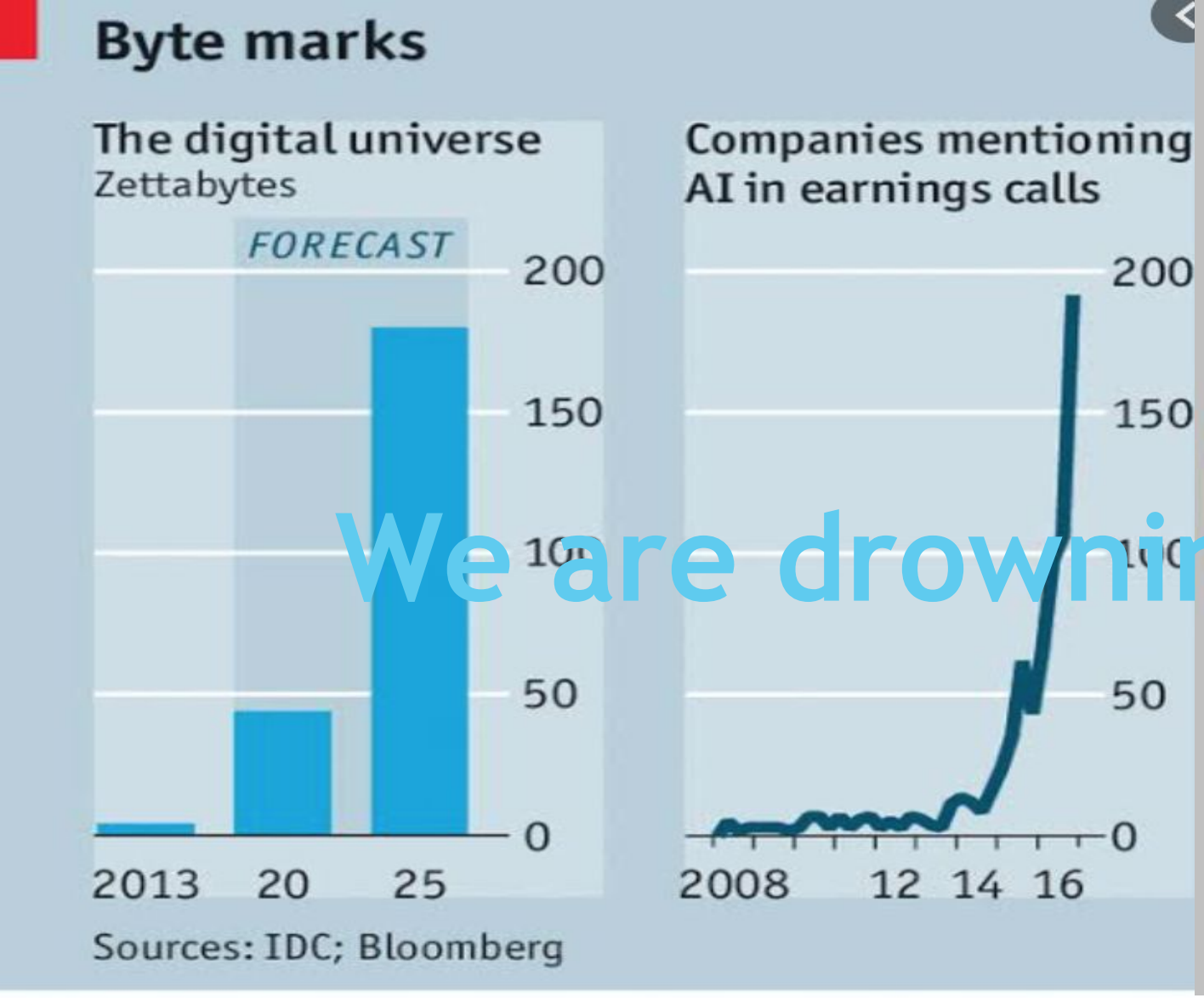
However, we've gone from this world…

To this





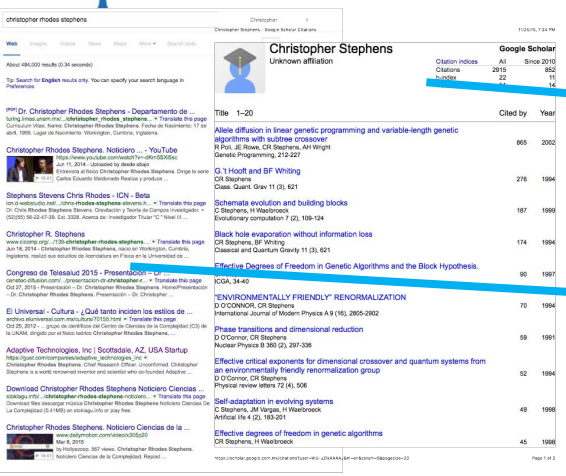


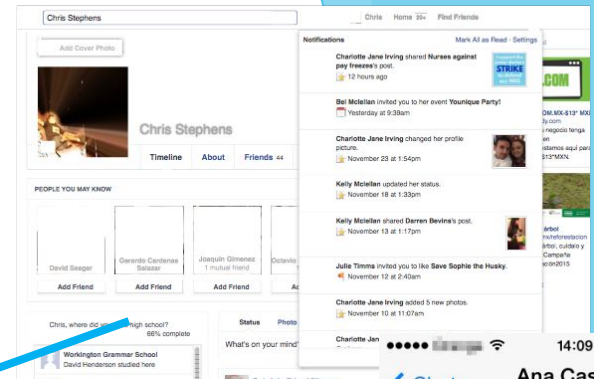


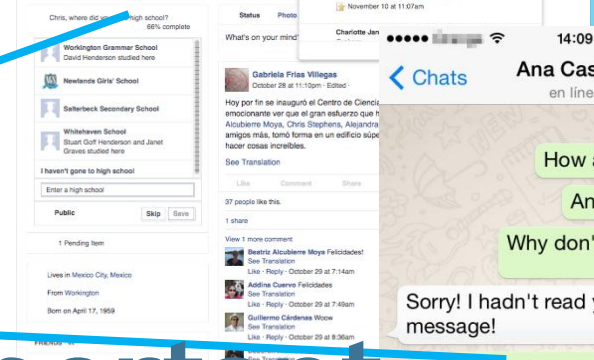
We are drowning in data!

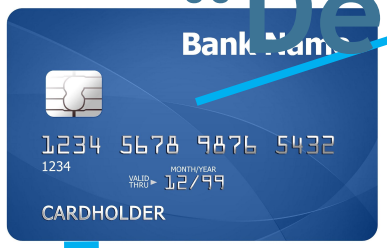
The Data Revolution

The Data Revolution

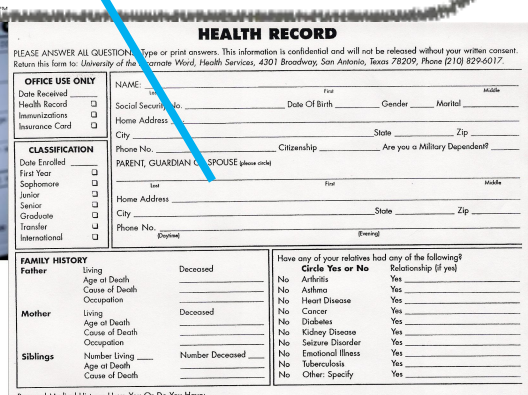








"Deep" data is more important than "Big" data



P(C | X(t)) represents the **algorithm's model** of reality and perception

DMV ▫ PDA

Type 2 thinking

Heuristic: we know exactly what it is. It's a mathematical model.

DMV → P(C | X(t)) → PDA

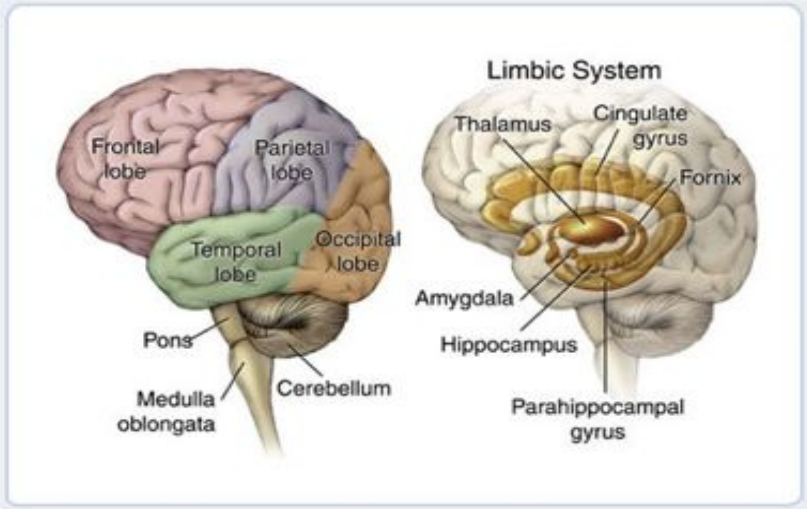
There are many alternatives to any decision and many possible actions

Artificial "Intelligence"

Did it work?

DMV ▫ PDA is judged to be good or bad according to a performance criterion

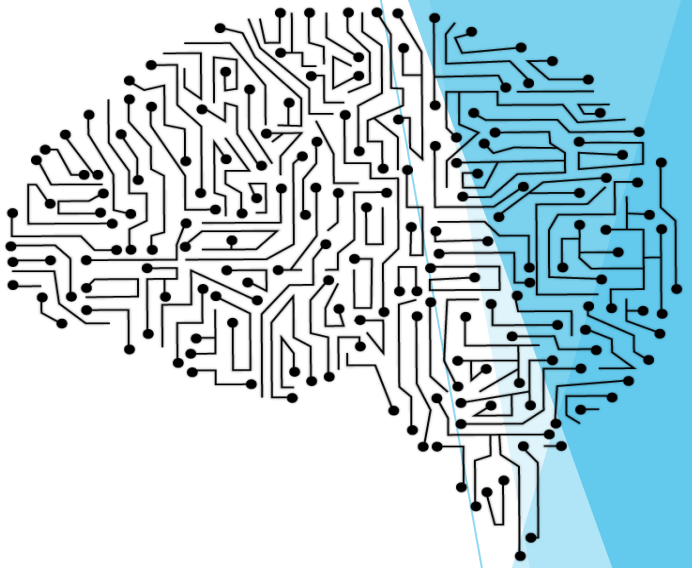
Human versus Artificial Intelligence



$$P(C \mid X(t))$$

A representation of DMV▫PDA

It's a Machine



Sensorial/physical/analog –
Always hugely multi-factorial
Always has meaning

$X(t)$

Electronic/digital/binary - very biased
And not very multi-factorial
No meaning

There are a huge number of goals, each with an associated value

C

There is just one goal that we have to pre-specify (with bias) as well as specify the value (with bias)

We don't know the heuristic but do know its VERY biased

$P(\ \mid\ )$

We know exactly the heuristic and its (weak) bias and variance

What is a machine?

Machines are created to do one particular task better than a human

These are physical machines

Machines are specialists

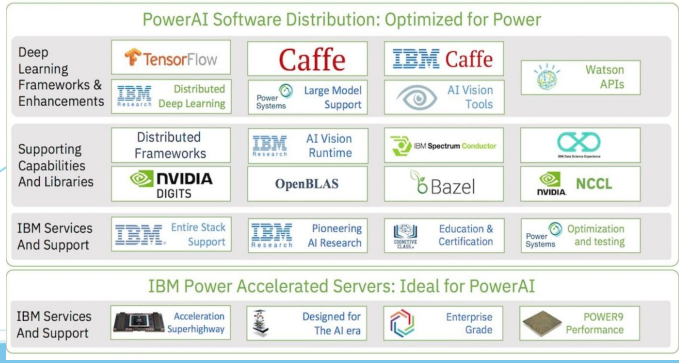
- You can't drive a hammer. You can't prepare coffee with a watch. A human being can prepare coffee, tell the time and decide to drive to work – if he has the appropriate tools. Humans are generalists with Generalized Intelligence. We are meta-machines that use other machines.

Machine Learning is associated with the development of algorithms that perform like physical machines, i.e. they do one thing well. Sometimes very well. Sometimes better than a human. Just like a hammer!

$$P(C \mid \mathbf{X}(t))$$

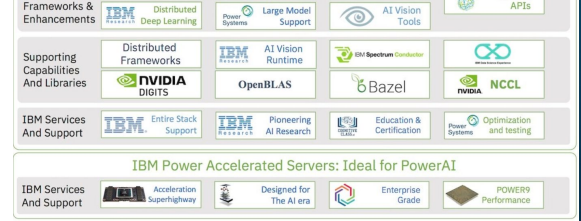
It's a machine that takes DV as an input and gives a P and potentially a DA as an output.

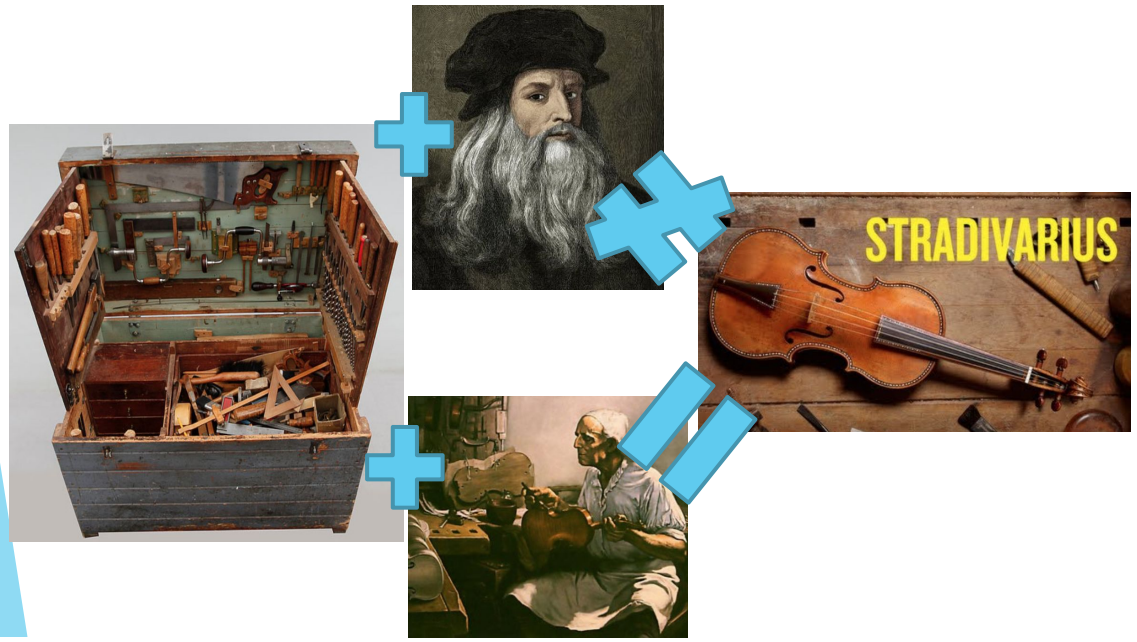
IBM PowerAI Platform

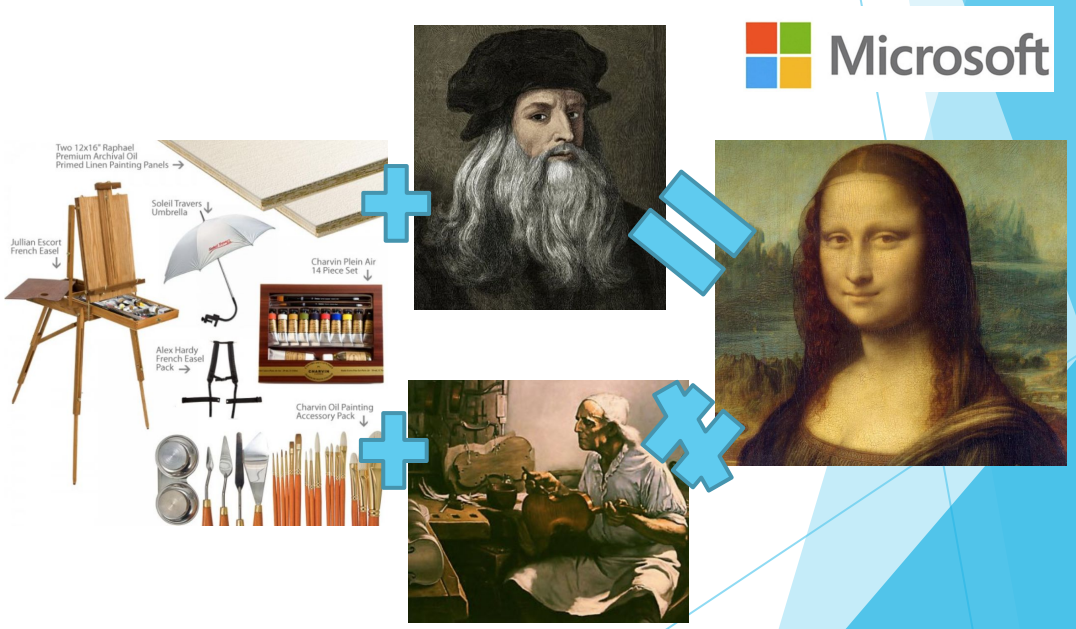


Selecting a P(|) is like selecting a tool

These are some possible tool sets, but…







Obesity and Metabolic disease

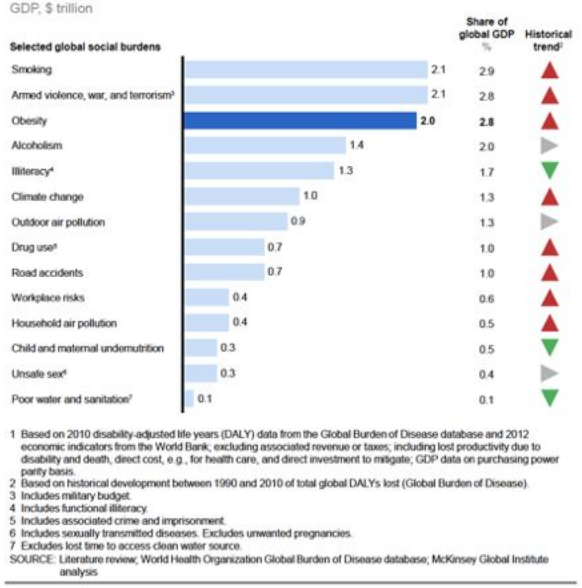
Imagine a problem that cost …

and you invested …

Obesity is one of the top three global social burdens generated by human beings

Estimated annual global direct economic impact and investment to mitigate selected global burdens, 2012[1]

GDP, $ trillion



1 Based on 2010 disability-adjusted life years (DALY) data from the Global Burden of Disease database and 2012 economic indicators from the World Bank; excluding associated revenue or taxes; including lost productivity due to disability and death, direct cost, e.g., for health care, and direct investment to mitigate; GDP data on purchasing power parity basis.
2 Based on historical development between 1990 and 2010 of total global DALYs lost (Global Burden of Disease).
3 Includes military budget.
4 Includes functional illiteracy.
5 Includes associated crime and imprisonment.
6 Includes sexually transmitted diseases. Excludes unwanted pregnancies.
7 Excludes lost time to access clean water source.
SOURCE: Literature review; World Health Organization Global Burden of Disease database; McKinsey Global Institute analysis

Estimates of Funding for Various Research, Condition, and Disease Categories NIH

Research/Disease Areas	FY 2016Actual (Dollars in millions)	2015 US Mortality	2015 US Prevalence (Standard deviation)
Cancer	5589.00	652,672	8.7% (0.20%)
Cardiovascular	2108.00	1,464,485	-
Chronic Obstructive Pulmonary Disease	97.00	292,471	6.2% (0.18%)
Diabetes 4/	1084.00	252,806	9.7% (0.22%)
Digestive Diseases	1745.00	-	-
Heart Disease	1289.00	1,202,319	11.7% (0.26%)
Heart Disease - Coronary Heart Disease	419.00	536,339	6.1% (0.17%)
Hypertension	224.00	427,631	27.0% (0.33%)
Inflammatory Bowel Disease	126.00	2,966	-
Obesity	965.00	39,590	30.0% (0.38%)
Stroke	308.00	234,867	-
	13,954	5,106,146	
% of total	17%	73%	
Physical Activity	392.00	-	-
Prevention	7566.00	-	-
Tobacco	299.00	-	-
Nutrition	1615.00	-	-
Basic Behavioral and Social Science	1804.00	-	-
Behavioral and Social Science	4137.00	-	-
	15,813		

nd it got v

eople" make "

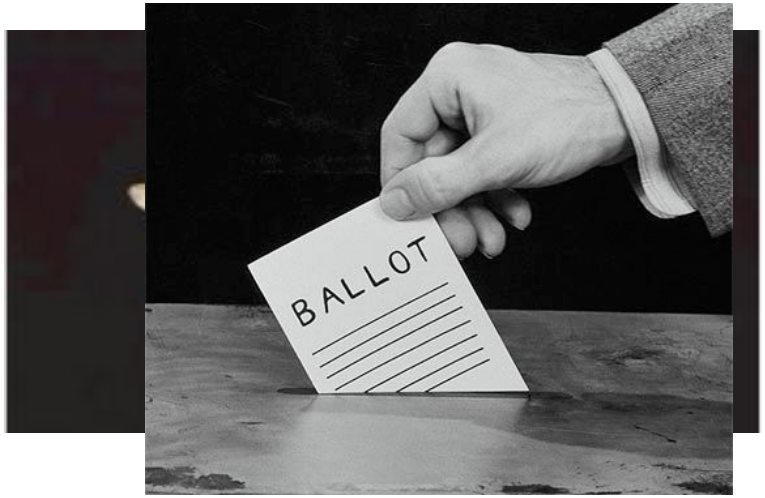
le are stupid, they deserve what they get?

Rapacious capitalism?

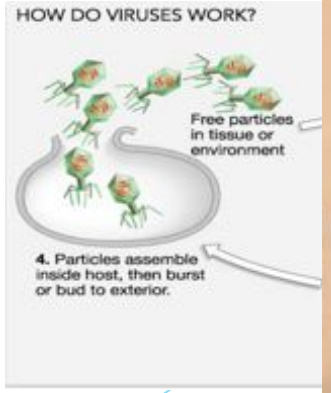
Evolutionary determinism?

Or is it all of these and thousands of other factors that also influence it?

To understand better we're going to go backwards… ADP▫VMD. First, AD…



RNA



And now P and V!

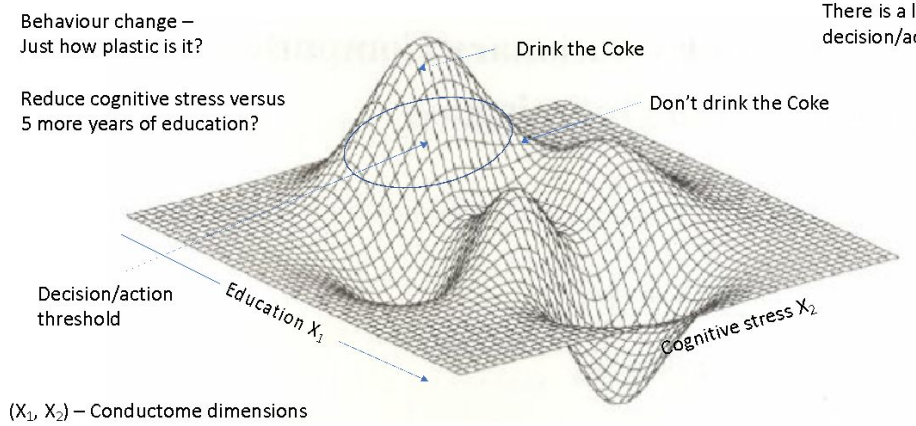
1. **Prediction** – If I eat this, I will enjoy it
Value – High - Physiologically and psychologically set – Type I thinking

2. **Prediction** – If I eat this, I will be diabetic and hypertensive in 20 years
Value – Low - Cognitively set – Type 2 thinking

The **Mental Model** P(C|**X**(t)) necessary to make 2. take precedence over 1. is enormously difficult to obtain. The evolutionary cards are stacked against us.

The Conductome Landscape

The Conductome landscape is dynamic and adaptive. There is a landscape for every decision/action/behaviour

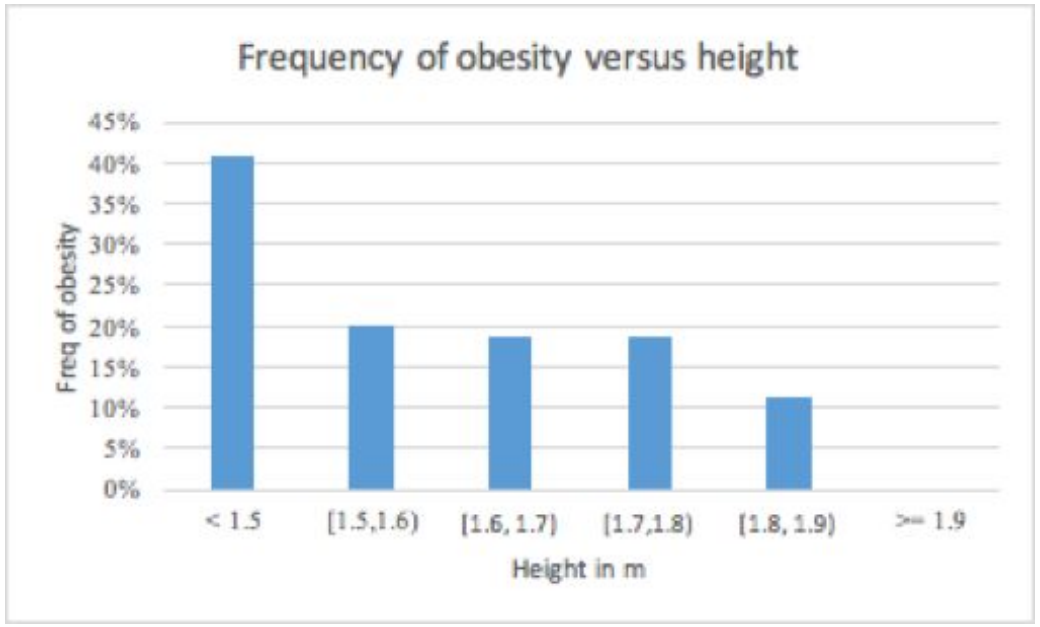


And now M and D! Here's a D.

Project 42 – the goal – to build the "deepest" database on the planet for obesity and metabolic disease. Over 3000 university academics, workers and students and over 2000 variables.

Variable	Valor	Epsilon	Nx	Nxc	N	Nc	Pc	Pxc	Descripción
Abba	1	-2.62561	910	160	1074	227	0.2114	0.1758	HBA (valor para diagnosticar la diabetes) <= 5.6 : 1 = Normal o r
Aestatura	1	4.801461	91	38	1076	228	0.2119	0.4176	Estatura que estima tener el encuestado < 1.5 : 1
Aestatura	2	-0.92449	399	77	1076	228	0.2119	0.193	Estatura que estima tener el encuestado [1.5, 1.6) : 2
Aestatura	3	-1.09413	366	69	1076	228	0.2119	0.1885	Estatura que estima tener el encuestado [1.6, 1.7) : 3
Aestatura	4	0.143796	185	40	1076	228	0.2119	0.2162	Estatura que estima tener el encuestado [1.7, 1.8) : 4
Aestatura	5	-1.63546	32	3	1076	228	0.2119	0.0938	Estatura que estima tener el encuestado [1.8, 1.9) : 5
Aestatura	6	-0.7333	2	0	1076	228	0.2119	0	Estatura que estima tener el encuestado [1.9, 2.0) : 6
Aestatura	7	1.928548	1	1	1076	228	0.2119	1	Estatura que estima tener el encuestado > 2.0) : 7
Apeso	1	-3.77209	62	1	1076	228	0.2119	0.0161	Peso que estima tener el encuestado <= 50 : 1
Apeso	2	-4.05811	79	2	1076	228	0.2119	0.0253	Peso que estima tener el encuestado (50, 55) : 2
Apeso	3	-5.74441	132	1	1076	228	0.2119	0.0076	Peso que estima tener el encuestado [55, 60) : 3
Apeso	4	-5.1211	172	9	1076	228	0.2119	0.0523	Peso que estima tener el encuestado [60, 65) : 4
Apeso	5	-1.86651	142	21	1076	228	0.2119	0.1479	Peso que estima tener el encuestado [65, 70) : 5
Apeso	6	-2.34173	138	18	1076	228	0.2119	0.1304	Peso que estima tener el encuestado [70, 75) : 6
Apeso	7	0.84116	106	26	1076	228	0.2119	0.2453	Peso que estima tener el encuestado [75, 80) : 7
Apeso	8	8.123762	143	70	1076	228	0.2119	0.4895	Peso que estima tener el encuestado [80, 90) : 8
Apeso	9	14.14686	102	80	1076	228	0.2119	0.7843	Peso que estima tener el encuestado >= 90 : 9
condi_act	1	5.045429	44	23	1076	228	0.2119	0.5227	¿Cómo consideras tu condición física actual? 1 : Muy mala
condi_act	2	5.865344	189	73	1076	228	0.2119	0.3862	¿Cómo consideras tu condición física actual? 2 : Mala
condi_act	3	-0.57931	429	86	1076	228	0.2119	0.2005	¿Cómo consideras tu condición física actual? 3 : Regular
condi_act	4	-4.18504	355	43	1076	228	0.2119	0.1211	¿Cómo consideras tu condición física actual? 4 : Buena
condi_act	5	-2.94241	57	3	1076	228	0.2119	0.0526	¿Cómo consideras tu condición física actual? 5 : Muy buena
condi_act	8	-0.7333	2	0	1076	228	0.2119	0	¿Cómo consideras tu condición física actual? 8 : No quiero res
condi1	1	3.176688	41	17	1076	228	0.2119	0.4146	¿Cómo consideras tu condición física hace un año? 1 : Muy ma
condi1	2	4.71648	180	64	1076	228	0.2119	0.3556	¿Cómo consideras tu condición física hace un año? 2 : Mala
condi1	3	0.133941	396	85	1076	228	0.2119	0.2146	¿Cómo consideras tu condición física hace un año? 3 : Regular
condi1	4	-2.65254	367	57	1076	228	0.2119	0.1553	¿Cómo consideras tu condición física hace un año? 4 : Buena

And the M?



Why do shorter people tend to be more obese?

Big Mac meal
for a large person

Big Mac meal
for a short person

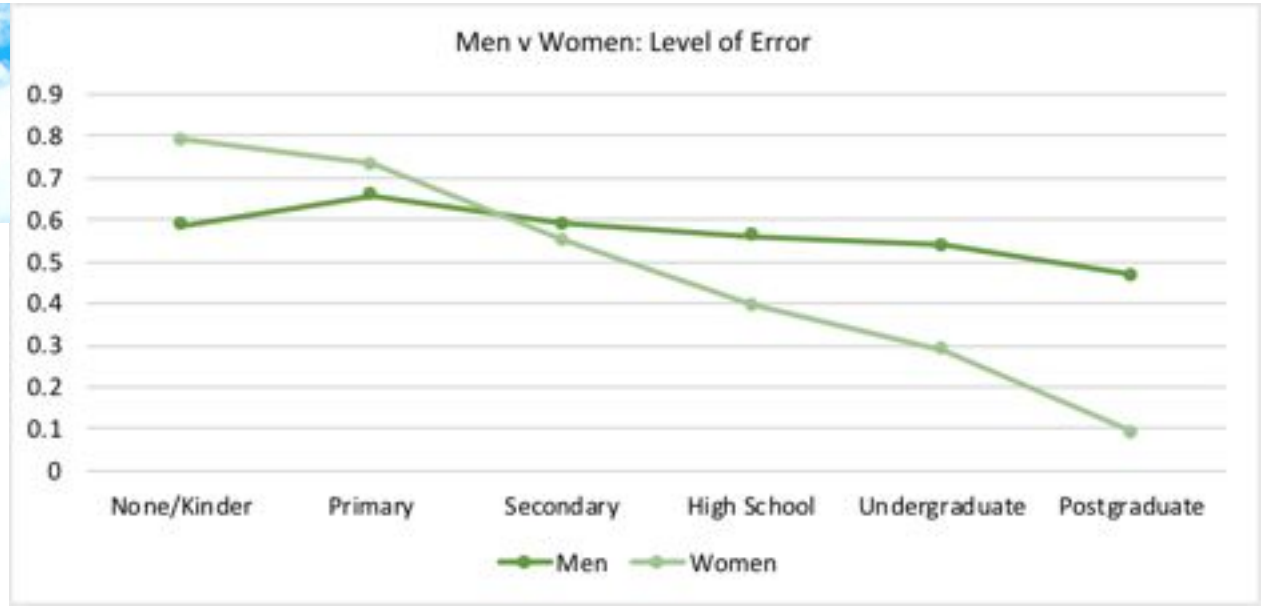
Torta Cubana
for a large person

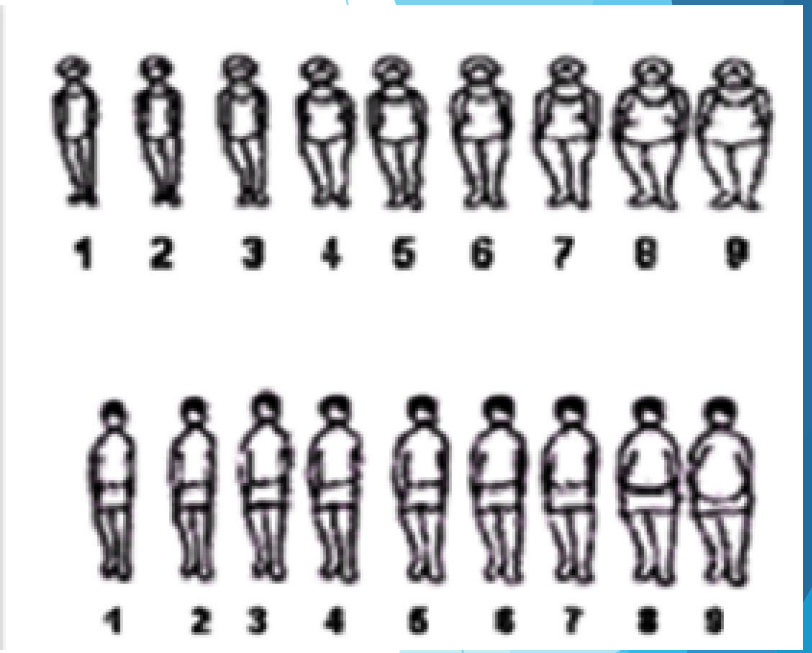
Torta Cubana
for a short person

Have you ever seen portions that were sized to your size?

And the M?

Difference in perceived BMI versus real BMI





We all live in an alternative reality.

For some its really different.

We can quantify the difference and ask why some people have bigger differences than others.

There is a reality

We just don't have the data to reconstruct it

Yet

…

The Data Revolution is making it feasible

But Human Intelligence is so biased it can just ignore the data

Oil Pipeline leaks DMV

The problem: Predict the risk of an oil pipeline leak while minimizing false positives and still capturing the leaks

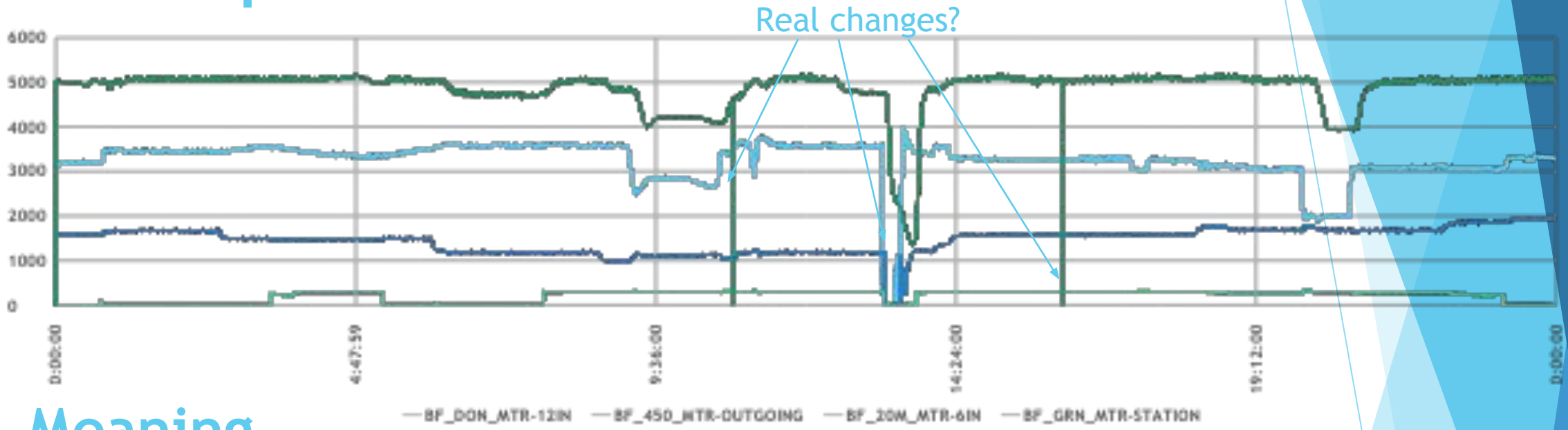
Data: Time series of flowrates, pressures, densities, pump states, drag reducing agents, valve positions, weather, time of year, pipeline topography, pipeline topology, exogeneous/endogenous changes,…

"IoT"

Meaning: Conservation of mass, Newton's laws, confounding factors – need an engineer's or physicist's intuition

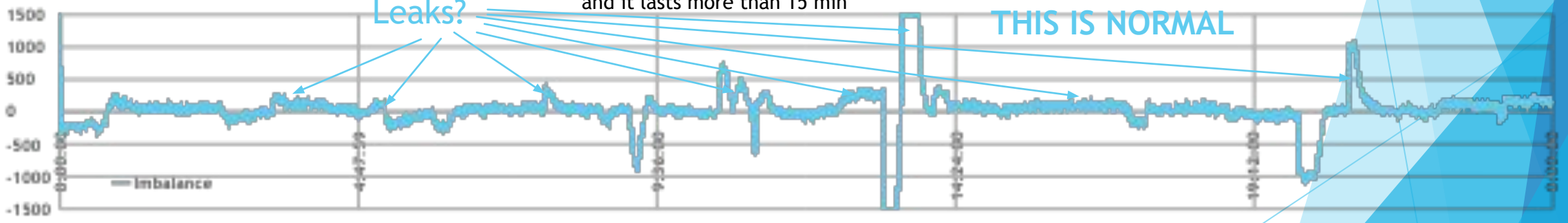
Value: Avoid environmental damage, brand loss, litigation, operator burnout, expensive pipeline revisions,…

Oil Pipeline leaks



Meaning

Looks like a leak of over a 1000 bph
and it lasts more than 15 min



Oil Pipeline leaks PDA

Prediction: Probability that there is a "leak-like" event

$$P(C \mid \mathbf{X}(t))$$

This is meta-model – a model whose inputs X_1, X_2,… are also models

Decision: IF ($P(C \mid \mathbf{X}(t)) > \text{threshold}$) THEN (leak)

Action: Shut down the pumps

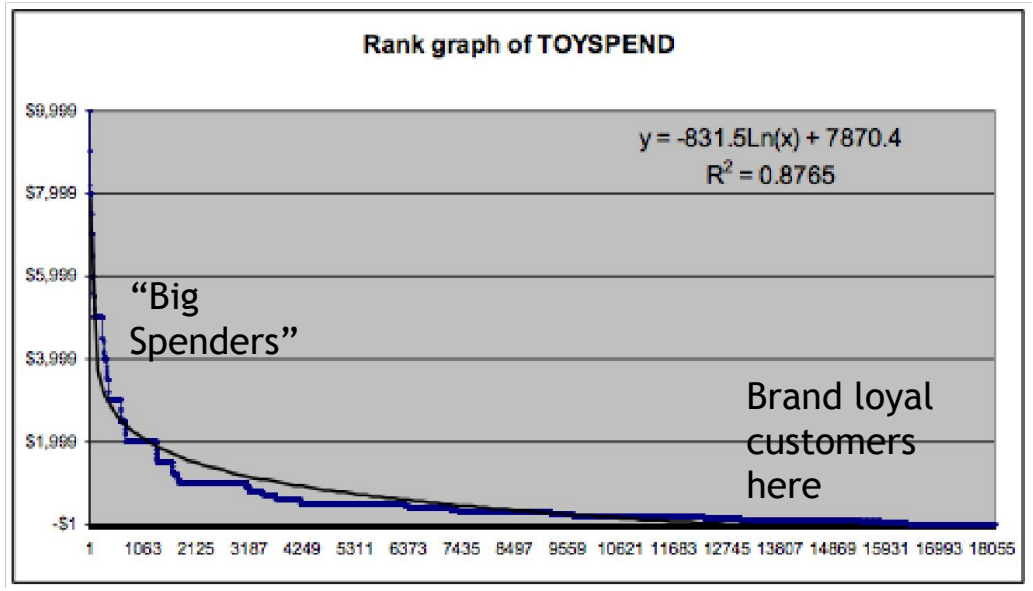
In DMV▫PDA, which part is Human Intelligence or Artificial Intelligence or both?
D = "machine"
M = human
V = human
P = "machine"
D = "machine"
A = human

Marketing

Human Intelligence

- **V:** Company A was losing market share to competitor
- **D+M ▫ P** It was losing brand loyal consumers (measured in > 90% of wallet spend) to competitor
- **DA** to be taken by client afterwards

D ▫



Biggest spender

Lowest spender

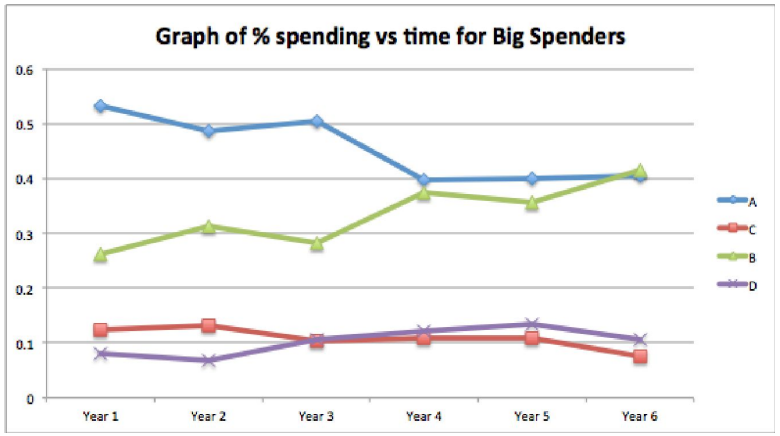
M - loyals aren't very profitable

20% of $1000 is better than 100% of $100!

Marketing

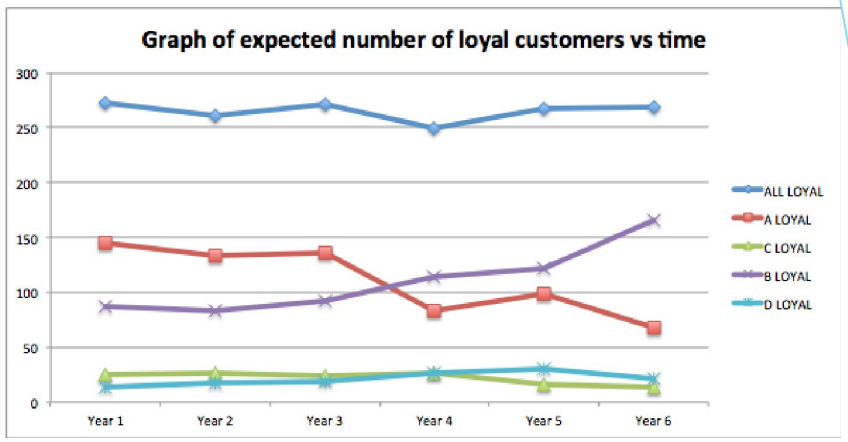
"Artificial" Intelligence –
just less biased human intelligence really

D ▫



M – Retailer A has lost all its competitive advantage to competitor B in the segment with the highest ROI

D ▫



M – Retailer A's most brand "loyal" customers have been defecting to competitor B.

P – Given that there was an economic downturn between year 1 and year 4 Brand "Loyal" customers as defined by wallet spend are more price sensitive. These Loyalists are actually Switchers!

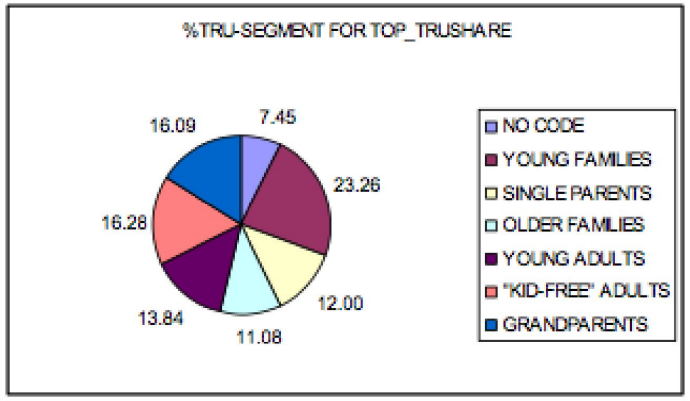
Marketing

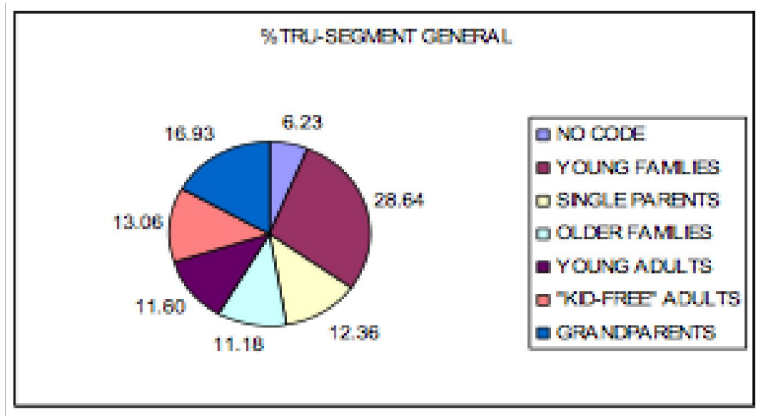
D ▫

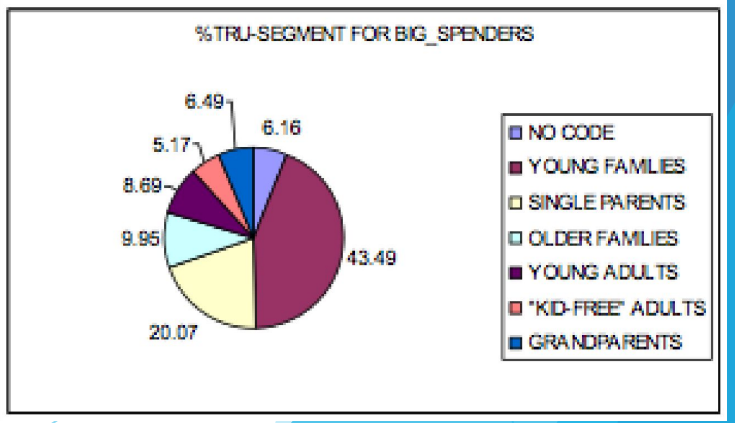
	Profile variable	Value	Statistical signficance (Std. devs)
	Children under 18?	Yes	13.1
	Second child's age	4-6	9.97
	First child's age	4-6	9.2
	Third child's age	7-9	8.87
	Your age?	28-32	8.28
	Persons in household	5	7.86
	Total family income	> $100,000	8.31
	Gender	Male	7.31

P - Big Spenders are: "Young, high earning males with several young children"

P - Loyals are: "Low earners who live close to retailer A's stores and in their home territory". It's easy to be loyal when its convenient







Loyals look like John Doe

Marketing

D + M

Year 1 attitudes Big Spenders	Statistical significance (std. devs)
Retailer A...their sales and specials on toys?	7.97
Retailer A...their selection of toys and games?	7.87
Retailer A...their return and exchange policy...problem?	7.83
Retailer A...having the newest toys and games being advertised?	7.58
Retailer A...feeling comfortable in the store?	7.54
Retailer A...knowledgeability of their employees...question?	7.4
Retailer A...being in stock on the items you want to buy?	7.09
Retailer A...the friendliness of their employees?	6.67
Retailer A...assembling a bike or other purchase...?	6.43
Retailer A...the prices they charge for toys and games?	6.42
Retailer A...finding someone in the store if you need help?	5.84
Retailer A...speed and accuracy of checkout.	5.63

P - $P(C \mid X(t))$

Predicting who will be a Big Spender – profile changes in Time

A good example of how to lose brand identity quickly.

AI P is better than HI P

Less biased
More multi-factorial

D + M

Year 4 attitudes Big Spenders	Statistical significance (std. devs)
Competitor B...Providing good value for the money?	3.23
Retailer A...assembling a bike or other purchase...?	3.18
Competitor B...the prices they charge for toys and games?	3.14
Competitor B...their selection of toys and games?	3.06
Competitor B...store where kids like to shop?	3.01
Competitor B...is a store that you would recommend to a friend for toys and games?	2.93
Competitor B...their sales and specials on toys?	2.87
Competitor B...assembling a bike or other purchase...?	2.68
Retailer A...their return and exchange policy...problem?	2.58
Competitor B...being the place that is fun for the whole family?	2.57
Competitor B...being in stock on the items you want to buy?	2.55
Competitor B...Makes you feel like a valued customer?	2.55

Wrap up…

Human Intelligence

- Everything you as an individual or as an organization do or will do in your life is governed by DMV▫PDA, with Data as the Fuel of Life
- "Darwinian" evolution has selected those who do it well
- All data goes through our senses – in two forms – symbolic and non-symbolic
  Non-symbolic data we process very fast; symbolic data we process very slow
- Meaning is an emergent property from data processing and is a result of millions of years of evolution
- Our data processing is VERY biased and a source of most of the world's most intractable problems – obesity, war, poverty, addictions,…

Wrap up…

Artificial Intelligence

- The Data Revolution produces only symbolic information - a huge amount of it – we're drowning in it. If every person on the planet were a "Google" there'd still be too much data to process
- We don't speak "data base"
- Only Machine learning/AI algorithms can speak data base and process the data fast enough
- Unfortunately, they don't understand what any of it means

Wrap up…

Hybrid Intelligence

- The data that we are generating now is radically changing how we think about the world and act in it – for good and for bad
  - There is information there to cure and prevent diseases, reduce crime, reduce poverty, promote democracy, make money,…
  - There is information there to to acerbate diseases, to increase crime, to increase poverty, violate human rights, lose money,…
  - Only humans choose which data to look at, what value it has, what it means and what decisions and actions to take
- AI on its own won't solve anything complex – all we have are just electronic hammers – just very sophisticated ones
- However, AI can do things that no human can
- So, we need **Hybrid Intelligence** – Human + Machine
  - Render unto Caesar…
- We are all challenged to seek the optimal way to DMV ▫ PDA – both in our jobs and our personal lives, embracing AI where it can help, where it can offset our cognitive biases, and understanding where it can't

Good luck with the challenge!

Thank you

Really Big Data At Walmart: Real-Time Insights From Their 40+ Petabyte Data Cloud

Bernard Marr Contributor ⓘ
Enterprise & Cloud

https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/#677426616c10

Walmart – the world's biggest retailer with over 20,000 stores in 28 countries, is in the process of building the world' biggest private cloud, to process 2.5 petabytes of data every hour. Data

To make sense of all of this information, and put it to work solving problems, the company has created what it calls its Data Café – a state-of-the-art analytics hub located within its Bentonville, Arkansas headquarters.

Here, over 200 streams of internal and external data, including Data 40 petabytes of recent transactional data, Data can be modelled, manipulated and visualized. Teams from any part of the business are invited to bring their problems to the analytics experts and then see a solution appear before their eyes on the nerve centre's touch screen "smart boards".

He said “If you can’t get insights until you’ve analyzed your sales for a week or a month, then you’ve lost sales within that time.

“If you can cut down that time from two or three weeks to 20 or 30 minutes, then that saves a lot of money for Walmart and stopped us losing sales. That’s the real value of what we have built with the data café.”

Value

For example Naveen told me about a grocery team who could not understand why sales had suddenly declined in a particular product category. The team came to the café to find out why, and by drilling into the data were quickly able to see that pricing miscalculations had been made, leading to the products being listed at a higher price than they should have been, in some regions.

Meaning
Value
Prediction
Human Intelligence

In another example, during Halloween, sales analysts were able to see in real-time that although a particular novelty cookie was very popular in most stores, there were two stores where it wasn't selling at all. The alert allowed the situation to be quickly investigated, and it was found that a simple stocking oversight had led to the cookies not being put on the shelves. The company was able to then rectify the situation immediately, avoiding further lost sales.

Meaning
Value
Prediction
Action
Human Intelligence

As well as 200 billion rows of transactional data (representing only the past few weeks!), the Café pulls in information from 200 sources including meteorological data, economic data, Nielsen data, telecom data, social media data, gas prices, and local events databases.

Anything within these vast and varied datasets could hold the key to the solution to a particular problem, and Walmart's algorithms are designed to blaze through them in microseconds to come up with real-time solutions.

Thanks to 4.5 billon years of evolution we can…

- Extract 11 Mb/s of sensory data that offer a “useful” representation of our environment
- Create predictive models by processing these data in our brains
- These predictive models enable us to make decisions and then act on the environment

Value comes from meaning

These actions have consequences…

- They imply a “value” for the data, its extraction, the predictive model, the decision and the action
- Without adequate data there are no effective decisions or actions

Sounds cool doesn’t it?

Walmart – the world's biggest retailer with over 20,000 stores in 28 countries, is in the process of building the world' biggest private cloud, to process 2.5 petabytes of data every hour.

To make sense of all of this information, and put it to work solving problems, the company has created what it calls its Data Café – a state-of-the-art analytics hub located within its Bentonville, Arkansas headquarters.

Here, over 200 streams of internal and external data, including 40 petabytes of recent transactional data, can be modelled, manipulated and visualized. Teams from any part of the business are invited to bring their problems to the analytics experts and then see a solution appear before their eyes on the nerve centre's touch screen "smart boards".

Algorithms can't make "sense" of data

Are all these streams of equal value?

What are the relevant variables?

How do you stimulate people to "see" the experts?

How many data scientists do you need?
Where will you find them?

After generating a new insight…
How quickly can you organize a response/intervention?

He said “If you can’t get insights until you’ve analyzed your sales for a week or a month, then you’ve lost sales within that time.

How ready are your business people to accept and implement a result?

“If you can cut down that time from two or three weeks to 20 or 30 minutes, then that saves a lot of money for Walmart and stopped us losing sales. That’s the real value of what we have built with the data café.”

What is the “value” of the insight?

For example Naveen told me about a grocery team who could not understand why sales had suddenly declined in a particular product category. The team came to the café to find out why, and by drilling into the data were quickly able to see that pricing miscalculations had been made, leading to the products being listed at a higher price than they should have been, in some regions.

In another example, during Halloween, sales analysts were able to see in real-time that although a particular novelty cookie was very popular in most stores, there were two stores where it wasn't selling at all. The alert allowed the situation to be quickly investigated, and it was found that a simple stocking oversight had led to the cookies not being put on the shelves. The company was able to then rectify the situation immediately, avoiding further lost sales.

How many possible insights do you think there are in the whole organization?

What's its value?

Who assigns value?
The analysts?
The teams?
Both?

How do you anticipate problems?

How do we predict?

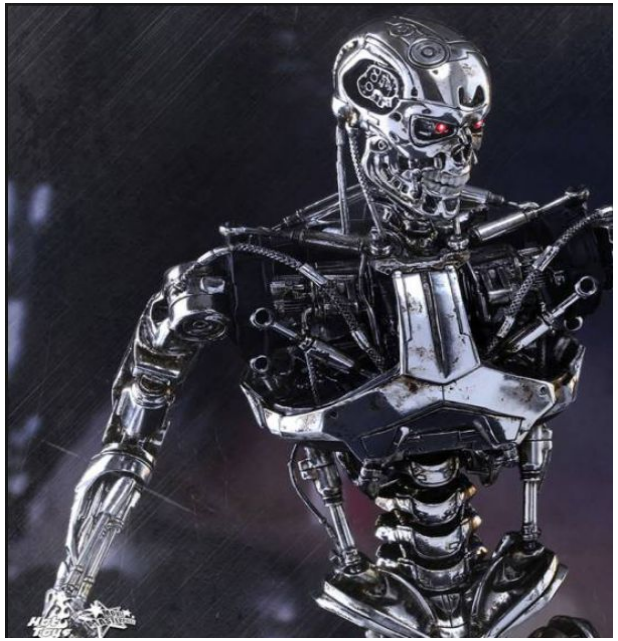
Is this the real reason?
Correlation vs causation.

As well as 200 billion rows of transactional data (representing only the past few weeks!), the Café pulls in information from 200 sources including meteorological data, economic data, Nielsen data, telecom data, social media data, gas prices, and local events databases.

The problem!

Anything within these vast and varied datasets could hold the key to the solution to a particular problem, and Walmart's algorithms are designed to blaze through them in microseconds to come up with real-time solutions.

Human “Intelligence” or Artificial “Intelligence”



Man and Machine
or
Man versus Machine

P(C | X(t)) are prediction models

You can answer all your questions and solve all your problems!

Where do I get one?

First, choose your "black box" P(|)

Where do a I get a "black box"? Make your choice!

SAP

IBM Watson

ORACLE

Microsoft

Wolfram Mathematica

sas

IBM SPSS

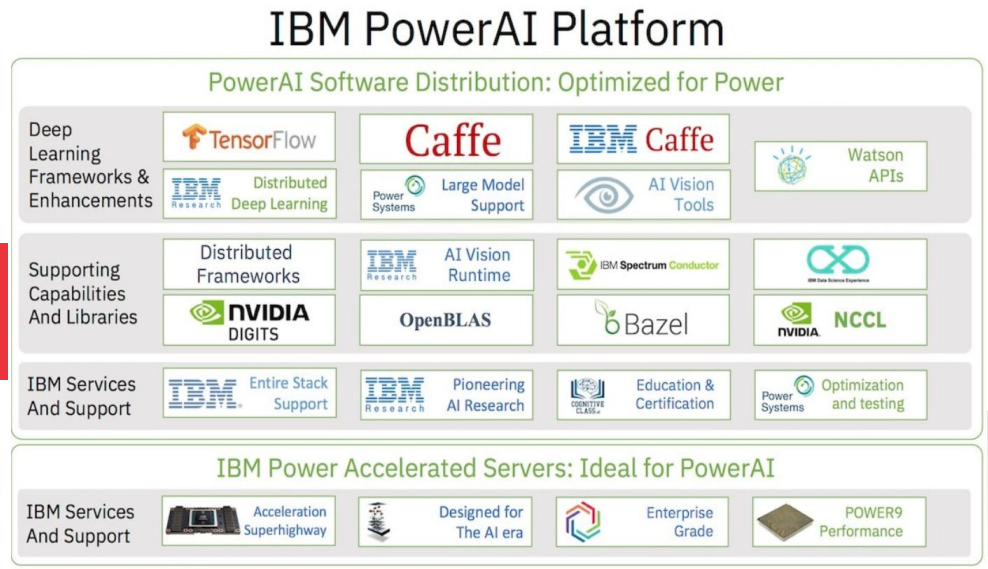



tableau SOFTWARE