

Modelaje de lo complejo

Retos para la ciencia de la computación

Chris Stephens

Coordinador de Ciencia de Datos

C3 – Centro de Ciencias de la Complejidad

Investigador Titular C

Instituto de Ciencias Nucleares

stephens@nucleares.unam.mx

Presentación en el Taller de Cómputo Científico

Posgrado en Ciencia e Ingeniería de la Computación

05/11/2020

¿Por qué se necesita el “Cómputo Científico”...

La parte científica...

- El propósito de la ciencia es **PREDECIR**
- Queremos **PREDECIR** para **DECIDIR**
- Para **PREDECIR** necesitamos **DATOS**

Prediction and science: the last 3 centuries

NO EXCEPTIONS.

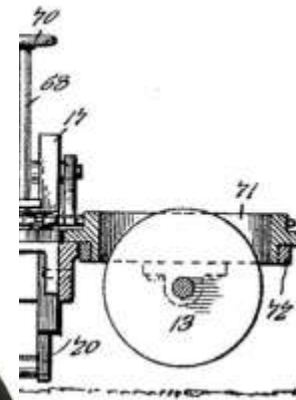
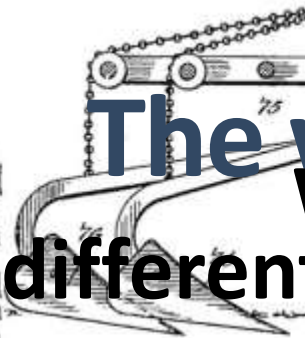
In fact...

How do you

The
With different

we are slaves to the law

Jacobs A. Moravia Invention

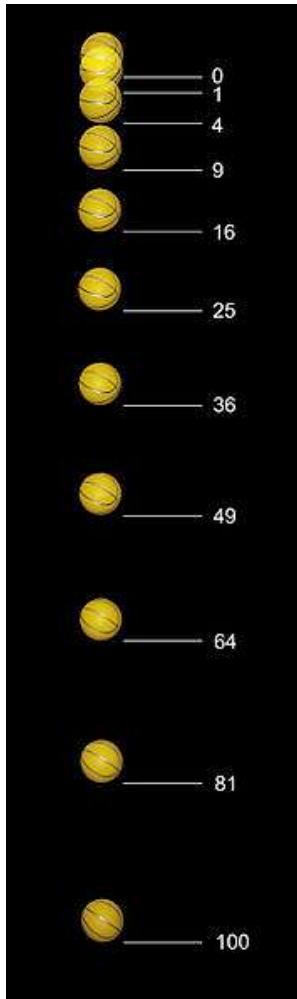


¿Por qué se necesita el “Cómputo Científico”...

La parte científica...

- El propósito de la vida es **PREDECIR**
- Queremos **PREDECIR** para **DECIDIR**
- Para **PREDECIR** necesitamos **DATOS**

Predictability in Simple Systems (Cómputo científico “tradicional”) versus Complex Systems (Cómputo científico “moderno”)



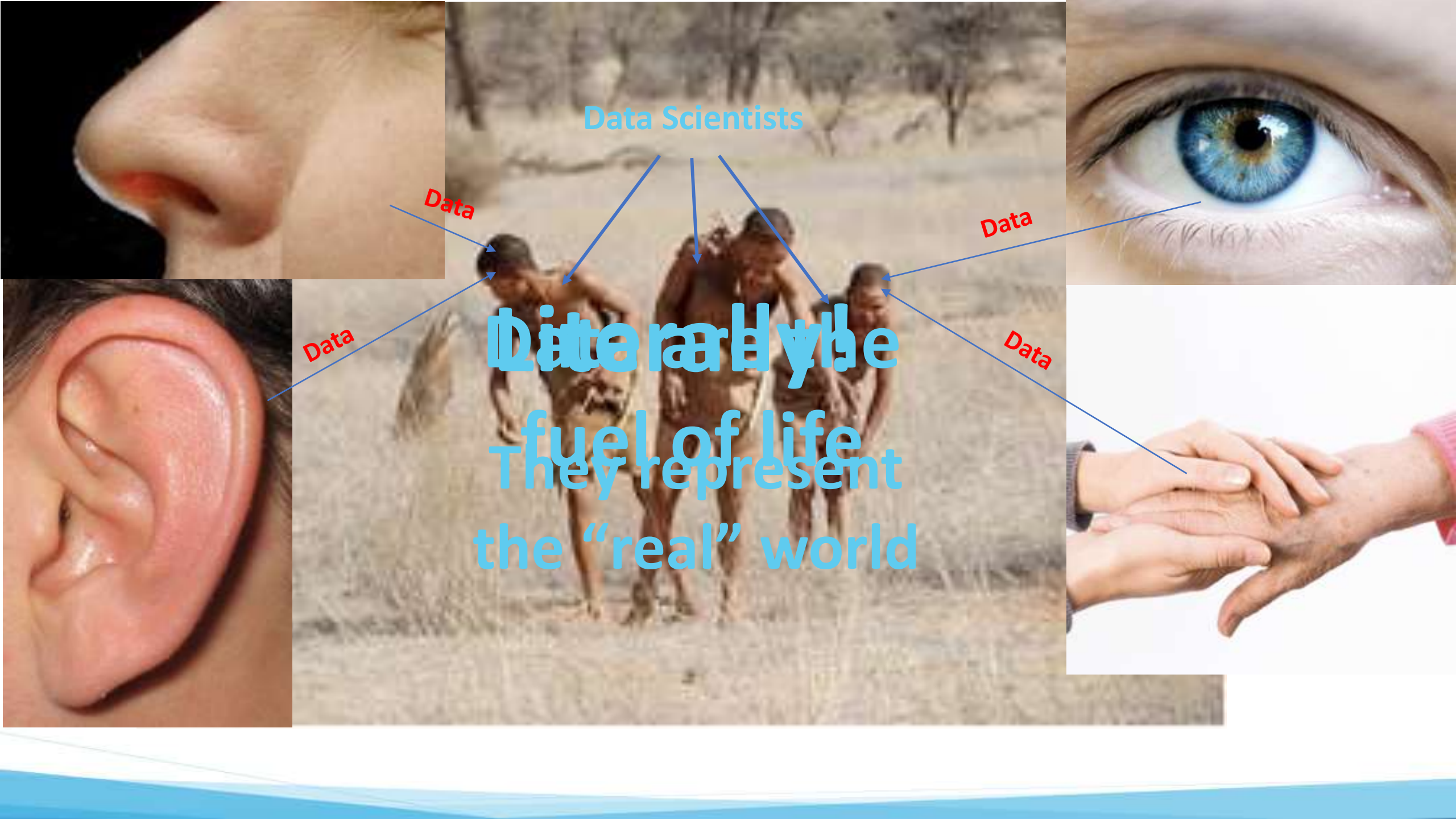
Mechanistic

Adaptive

The difference between complex and simple systems is the difference between *evolution* of function and the evolution that allows systems to escape the tyranny of the laws of physics.

Complexity is a consequence of that revolution.





Data Scientists

Data

Data

Data

Data

~~Data~~ is the
fuel of life
They represent
the "real" world



Datos

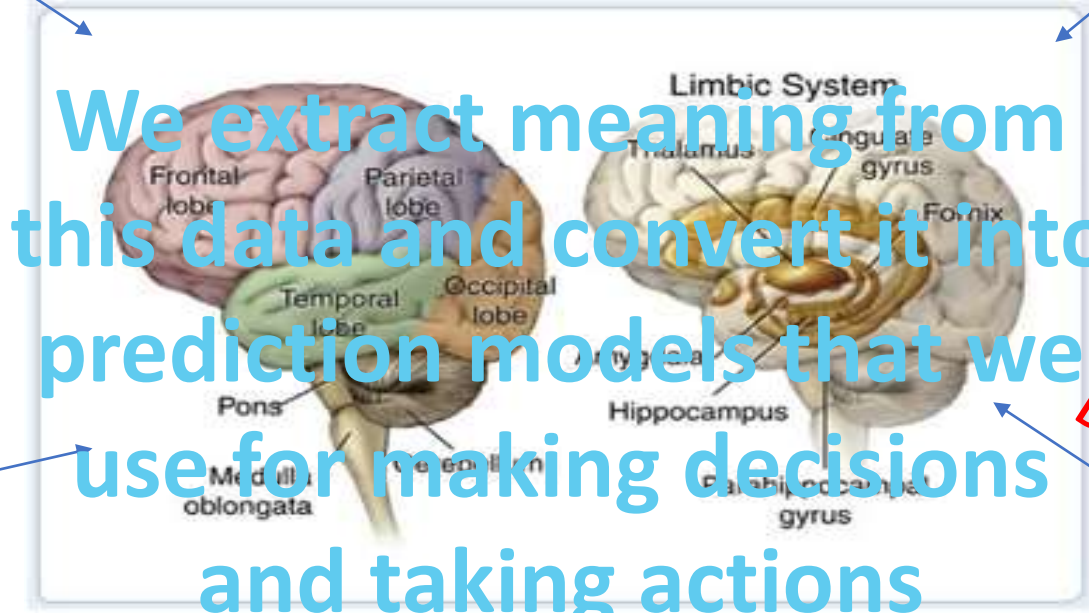


Datos

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

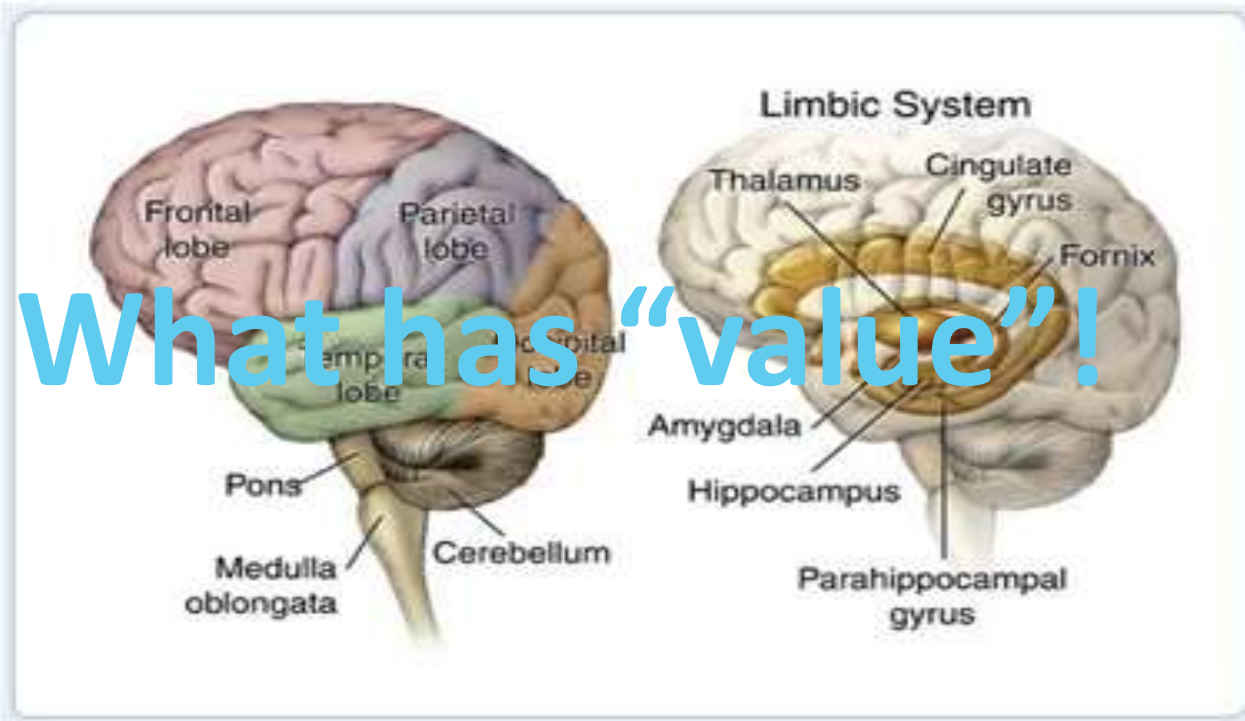


Datos

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



The human brain can store about 100,000,000,000,000,000 bytes
“hard disc” would be full in 3 months
What does it “choose” to store?





01/11/1980

5538151775

SECR590417



	2010	2009
Notes	\$	\$
EXPENSES		
Employee benefits	1,344,800	210,750
Supplies expenses	70,550	8,100
Depreciation and amortization	82,000	4,000
Financial costs	3,162	3,734
TOTAL EXPENSES	2,207,805	2,085,070
LESS:		
OWN-SOURCE INCOME		
Own-source revenue		
Rental income	4A 51,843	51,843
Other revenue	4B 7,200	7,200
Total own-source revenue	1,800	60,043
Gains		
Other	4C 18,000	18,000
Total gains	18,000	18,000
Total own-source income	19,800	77,043
Net cost of service	2,188,005	2,008,027
Revenue from Government	4D 2,179,000	2,167,000
Surplus	41,038	158,989

Reyas

Susan Cordts

F = ma

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

using DMV → PDA

Data+Meaning+Value-->Prediction+Decision+Action

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

DMV → PDA

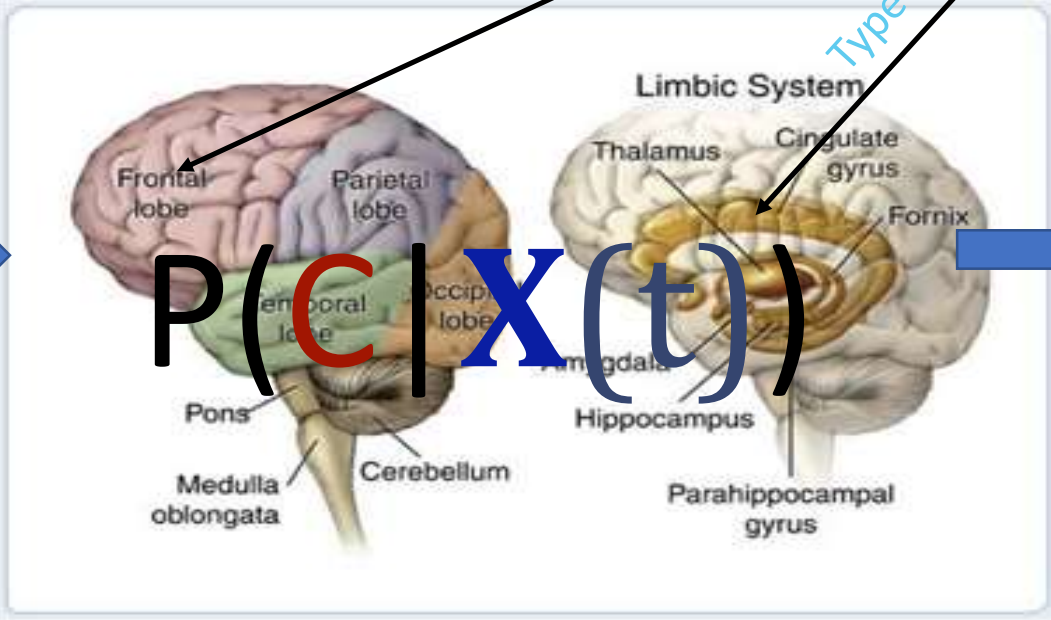
Heuristic: we don't know what it is in humans.

It is a DMV world.

Adaptability

Type 2 thinking

Type 1 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

¿Por qué se necesita el “Cómputo Científico” ...

La parte de “cómputo” ...

- El propósito de la ciencia es **PREDECIR**
- Queremos **PREDECIR** para **DECIDIR**
- Para **PREDECIR** necesitamos **DATOS**
- ¿Cuáles datos?



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

```
D:\C++\calculates.cpp *
#include <stdio.h>
void main(void)
{
    int mangoes;
    int persons;
    int perPerson;
    int left;
    printf("\nEnter the No. of mangoes:");
    scanf("%d",&mangoes);
    printf("\nEnter the No. of persons:");
    scanf("%d",&persons);
    perPerson=mangoes/persons;
    left=mangoes%persons;
    printf("\nEach person gets %d mangoes",perPerson);
    printf("\nWe have %d mangoes leftover",left);
}
```

“Recently we have learned how to represent the world in a new way or the formation of a new friend, so that was not local in space and time.”

CAMPAIGN SUMMARY

FACEBOOK COST: \$2,205.74

FACEBOOK CLICKS: 389

ADWORDS COST: \$15,332.60

ADWORDS CLICKS: 3,690

ADWORDS CPA: \$38.43

ADWORDS CONVERSIONS: 399

Marketing Intern
ABC Agency
05/2015 - 10/2015

- 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.

CERTIFICATES

- Hootsuite Social Marketing Certificate (03/2016)
- Google Ads Certification (08/2016)

LANGUAGES

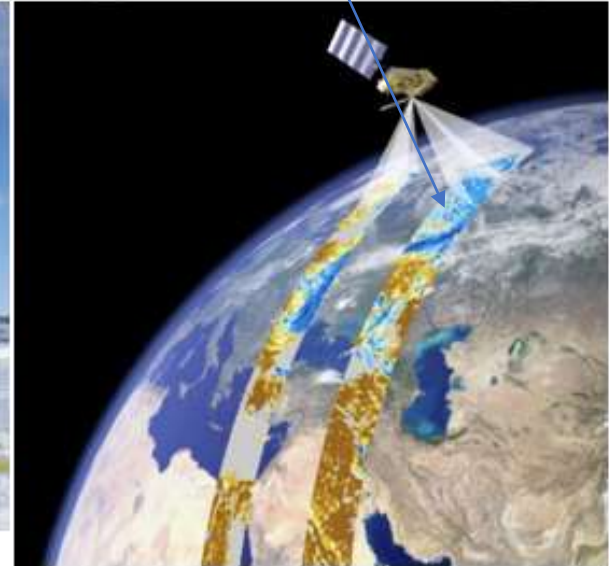
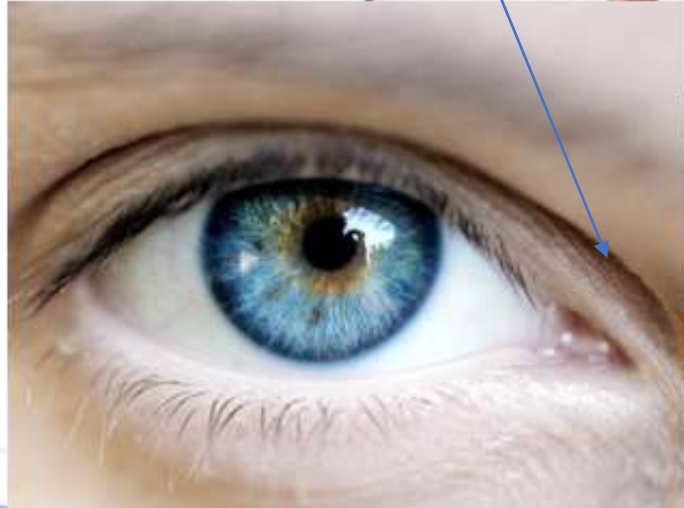
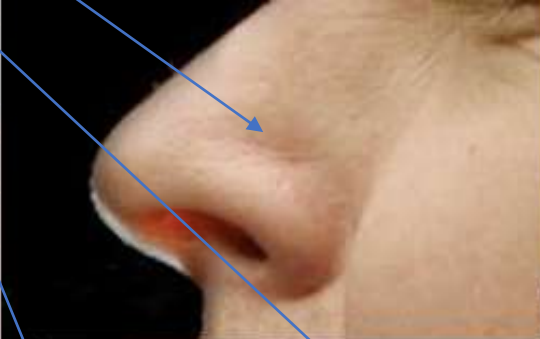
- English
- Spanish
- French
- Italian





However, we've gone
from this world...

To this



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

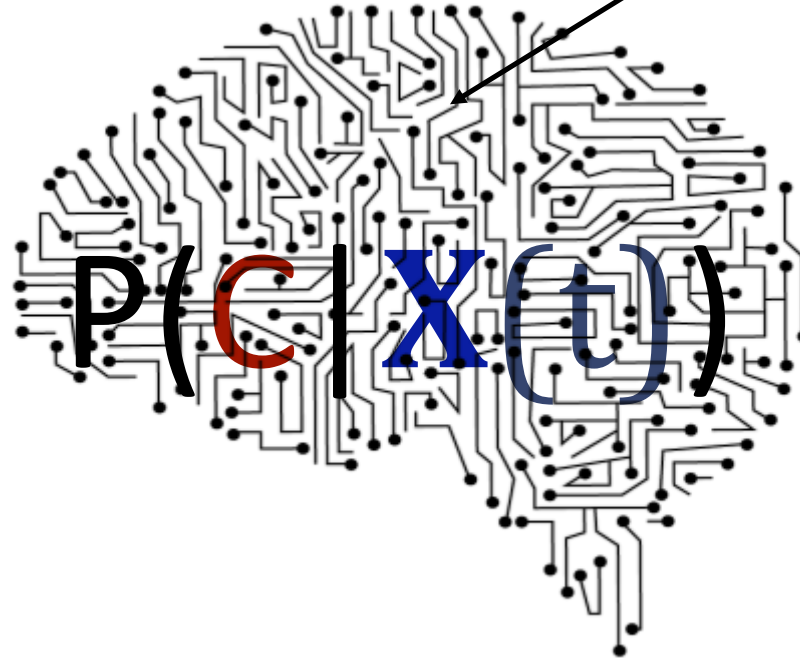
Heuristic: we know exactly what it is. It's a mathematical model.

DMV → PDA

Type 2 thinking



DMV →



→ PDA

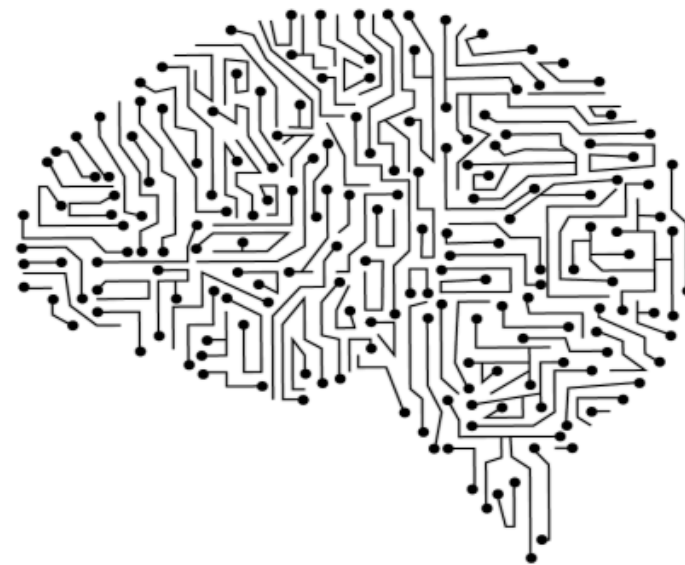
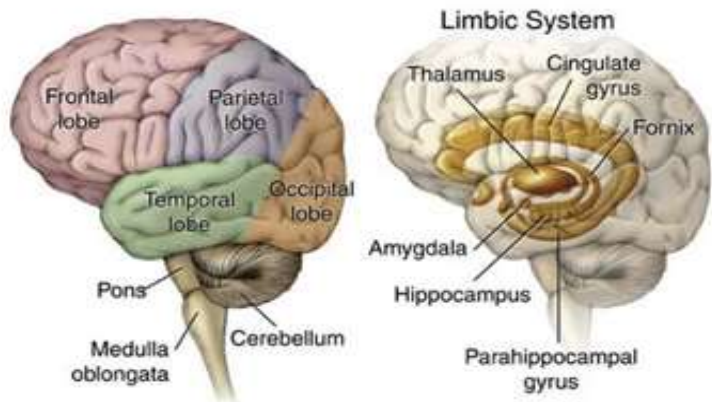
There are many alternatives to any decision and many possible actions

Artificial "Intelligence"
Did it work?

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



Human versus Artificial Intelligence



$$P(C | X(t))$$

A representation of DMV → PDA
It's a Machine

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

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

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

$$X(t)$$

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

$$C$$

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

$$P(|)$$

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

- Mi trabajo es para usar la computadora (dispositivos digitales) para hacer lo que hace bien la computadora (dispositivos digitales) y el cerebro humano para hacer lo que el hace bien

INTELIGENCIA HIBRIDA

¿Qué hace la computadora (dispositivos digitales) bien?

- Colecta de datos digitales
- Almacenamiento de datos digitales
- Procesamiento de datos digitales
- ¿Qué hace el ser humano bien?
 - Entender el **significado** de los datos

Proyectos: Modelaje de Sistemas Complejos

stephens@nucleares.unam.mx

- Proyecto 42
 - Para crear modelos predictivos para la obesidad y enfermedades metabólicas
 - Para crear la base de datos más profunda en el mundo para estas enfermedades
 - Datos de más de 3000 universitarios con más de 3000 variables asociados
 - Desarrollo de una Plataforma (project42.c3.unam.mx) disponible al público para crear modelos de ML
 - Oportunidades para hacer contribuciones en: Desarrollo de software, modelaje de datos, desarrollo de algoritmos de ML
- Proyecto Epi-PUMA
 - Para crear modelos predictivos para COVID19 y otras enfermedades emergentes
 - Desarrollo de una Plataforma (covid19.c3.unam.mx) disponible al público para crear modelos de ML
 - Oportunidades para hacer contribuciones en: Desarrollo de software, modelaje de datos, desarrollo de algoritmos de ML
- Proyectos “teóricos” – causalidad, No free lunch theorem – cuál algoritmo es mejor para cuál problema