Conductome

A Bayesian Classifier Approach to Causal Inference in Complex Adaptive Systems

Christopher R. Stephens

C3- Centro de Ciencias de la Complejidad and Instituto de Ciencias Nucleares, UNAM

Centre for Complexity Science, Imperial College, London

Centro de Estudios Mexicanos UNAM-Reino Unido, King's College, London

41st TBI Winterseminar in Bled

Some pretty important problems...

What do they have in common?

1) They’re all Complex Adaptive Systems
2) They’re all our fault
3) They’re multiobjective
4) They’re not getting (much) better

1) They're all Complex Adaptive Systems

a) "Complex" = Extremely multifactorial/multicausal
   - There are many things you can say about a CAS and each thing you say depend on many other things

b) "Adaptive" = Live in a space of configurations AND rules, strategies, laws, behaviours
   - The "laws" of physics and chemistry change from being "explainers" to being restrictions.

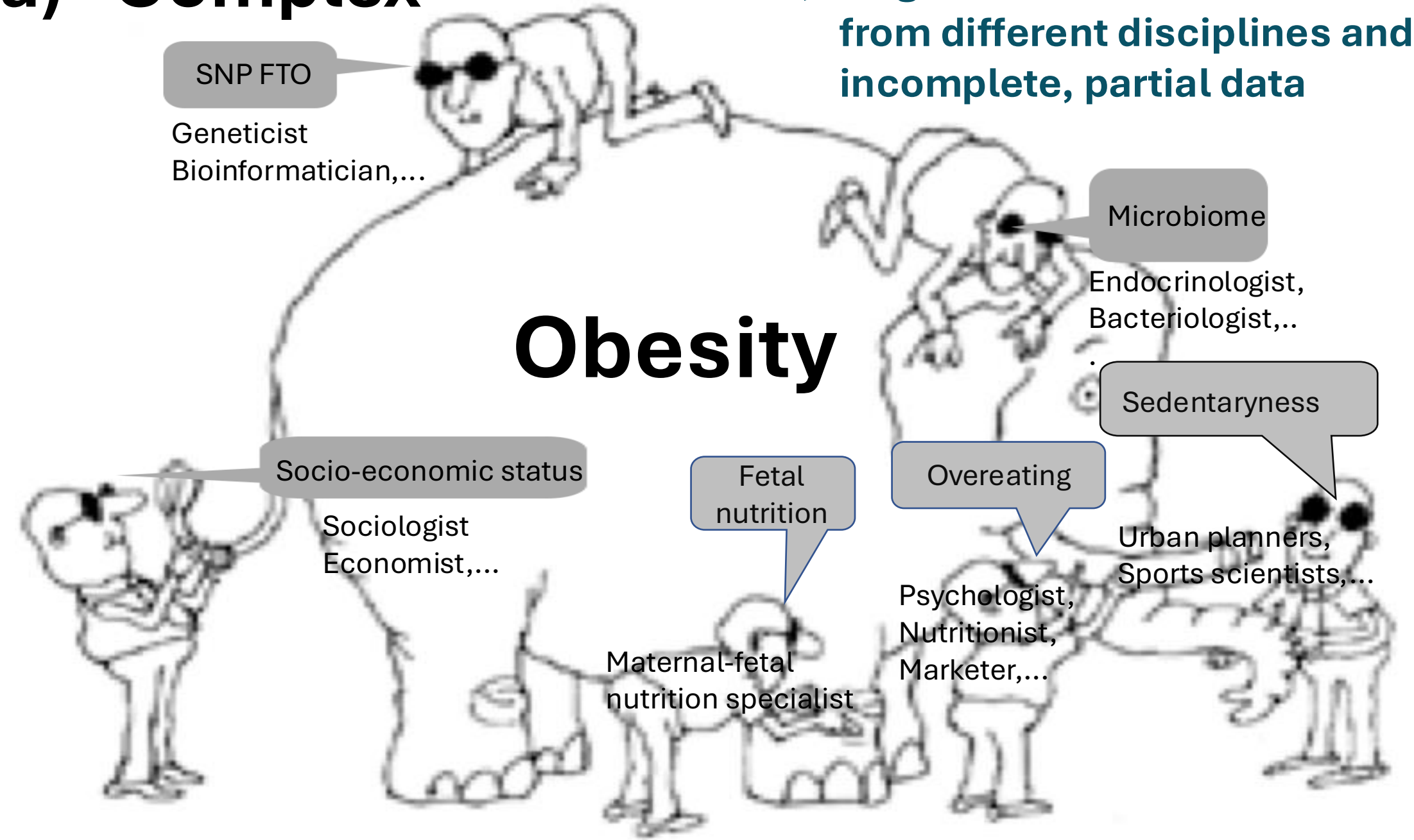
$$P(C|X(t))$$

Who, What, Where, When,…

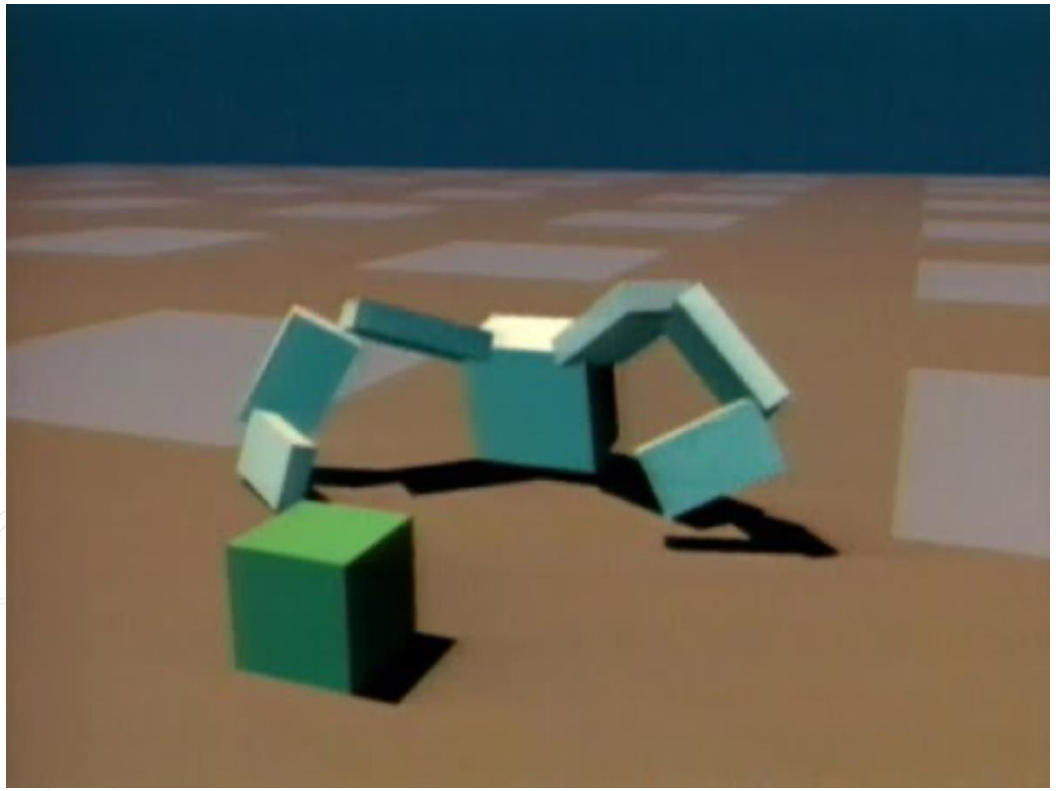
The "World" – causes, risk factors, environment,…

a) “Complex”

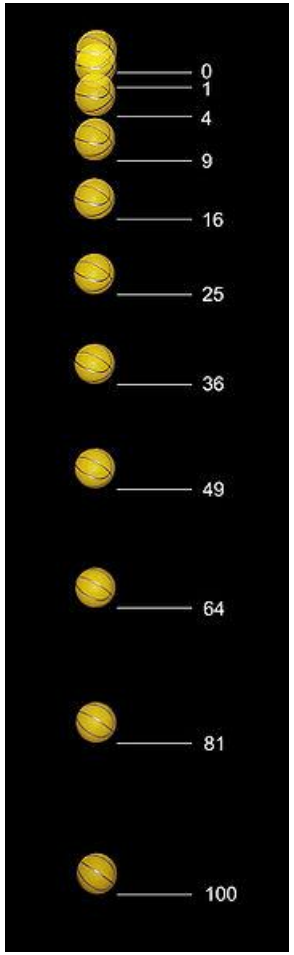
Now, imagine thousands of blind men… from different disciplines and with incomplete, partial data



b) “Adaptive”



Mechanistic



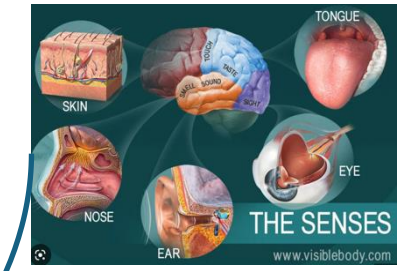
Both systems obey Newton’s 2nd Law. This is enough to describe the motion of their centers of gravity but doesn’t tell us how or why the cat lands feet first.

In the first case Newton’s laws explain the phenomenon but in the second case they become a restriction to be obeyed.

Dogs, cats and balls all have different rules when falling from a tree. For CAS we can call them behaviours

Adaptive

2) They're all our fault



- Limited symbolic processing power ~ 30 bytes/sec
- We don't speak "database"

- Computers don't "understand" anything
- Produces "machines" - specialists

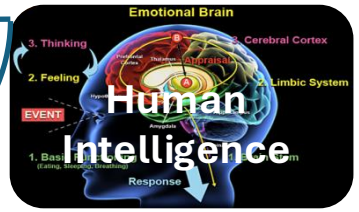
Consequences

Objetivos y metas estratégicos

López-Gatell, AMLO say no to enforcing use of coronavirus face masks

MND Staff October 28, 2020

The effectivness of face masks is 'overstated,' says Mexico's coronavirus czar.



Hybrid Intelligence

$P(C|X)$

Where? Who? What? When?

Cond

IBERO CIUDAD DE MÉXICO

UNIVERSIDAD DE GUANAJUATO

IMSS

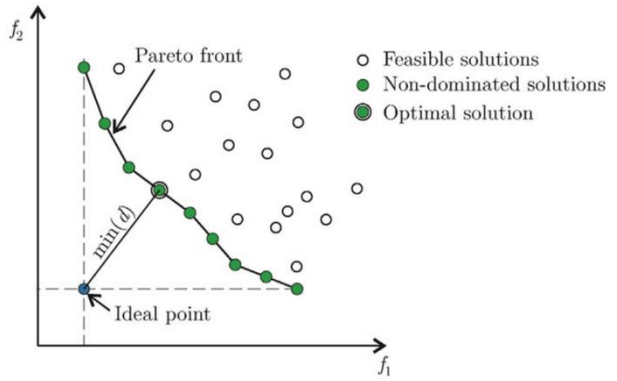
3) They're all multiobjective

When is a decision "good" versus "bad"?

We need a metric to measure it.
"Rational" decision making implies that there is a unique utility function that we optimize

What is the utility function here?

DT = the time difference between taking the stairs and taking the elevator?
DE = the difference in physical energy expended in one route versus another?
DH = the relative perceived health benefit of one versus the other (think how this can change pre-COVID versus post-COVID!)?
DS = the socialization benefit in the case you would take one versus the other with coworkers?
Others?



How do we balance these tradeoffs and make our decision?

What are the ingredients in creating a general framework for modelling these problems?

1. An appropriate conceptual, theoretical and algorithmic framework
2. “Deep”, commensurable data sets
3. A different way to how we do science now

The Bayesian Models

$P(C|\mathbf{X}) = P(C|X_1, X_2, \ldots, X_N) \qquad \mathbf{X} = (X_1, X_2, \ldots, X_N)$

As a number this can tell us the probability of the "effect" given the "causes" **X**. It won't tell us the "whys".

For the Decision Cycle we need to know the "whys" - which X_i are:

- Important
- Believable (statistically)
- Causal (direct, indirect, very indirect,…)
- Actionable (is it a variable we can change, e.g., climate or age = no, government program or diet = yes)

Use a Bayesian approach: P(C|**X**) = P(**X**|C)P(C)/P(**X**). And then P(C|**X**,**X**') = P(**X**´|C,**X**)P(C|**X**)/P(**X**´|**X**). And then…

$P(\mathbf{X}|C) \rightarrow \prod_{i=1}^{M} P(\xi_i|C)$ Assume a factorization of the likelihood (weight of evidence)

$$S(C|\boldsymbol{X}) = \ln\left(\frac{P(C|\boldsymbol{X})}{P(\bar{C}|\boldsymbol{X})}\right) = \sum_{i=1}^{M} s_i + s_0. \qquad s_i = \ln(P(\xi_i|C)/P(\xi_i|\bar{C})). \qquad s_0 = \ln\left(\frac{P(C)}{P(\bar{C})}\right).$$

This is a generative not a discriminitive classifier – depends on the distribution of X

Importance/effect size. The "Weight of Evidence" in favour of C versus $\bar{C}$

Bayesian prior

The Bayesian Models

$$P(\xi_i|C) = \frac{N(C\xi_i)}{N(C)};\ P(C) = N(C)/N; P(\bar{C}|X) = (N(\xi_i) - N(C\xi_i))/N(\xi_i);\ P(\\ (N - N(C))/N; P(\xi_i) = \frac{N(\xi_i)}{N}.$$

Are the Building Blocks of everything . Everything is just counting!

New data → New counts: $N \to N + 1; N(C) \to N(C)\ or\ N(C) + 1;\ N(\xi_i) \to 1; N(C\xi_i) \to N(C\xi_i)\ or\ N(C\xi_i) + 1;$

Dynamics/**Adaptation**: $C\ or\ \xi_i$ can be states or rules.

But is the correlation between C and ξ_i statistically reliable? Use:

$$\varepsilon(C|\xi_i) = N(\xi_i)(\mathrm{P}(\mathrm{C}|\xi_i) - P(C))/\sqrt{N(\xi_i)(P(C)(1 - P(C))}$$

We are comparing the relation between C and ξ_i with the null hypothesis P(C)

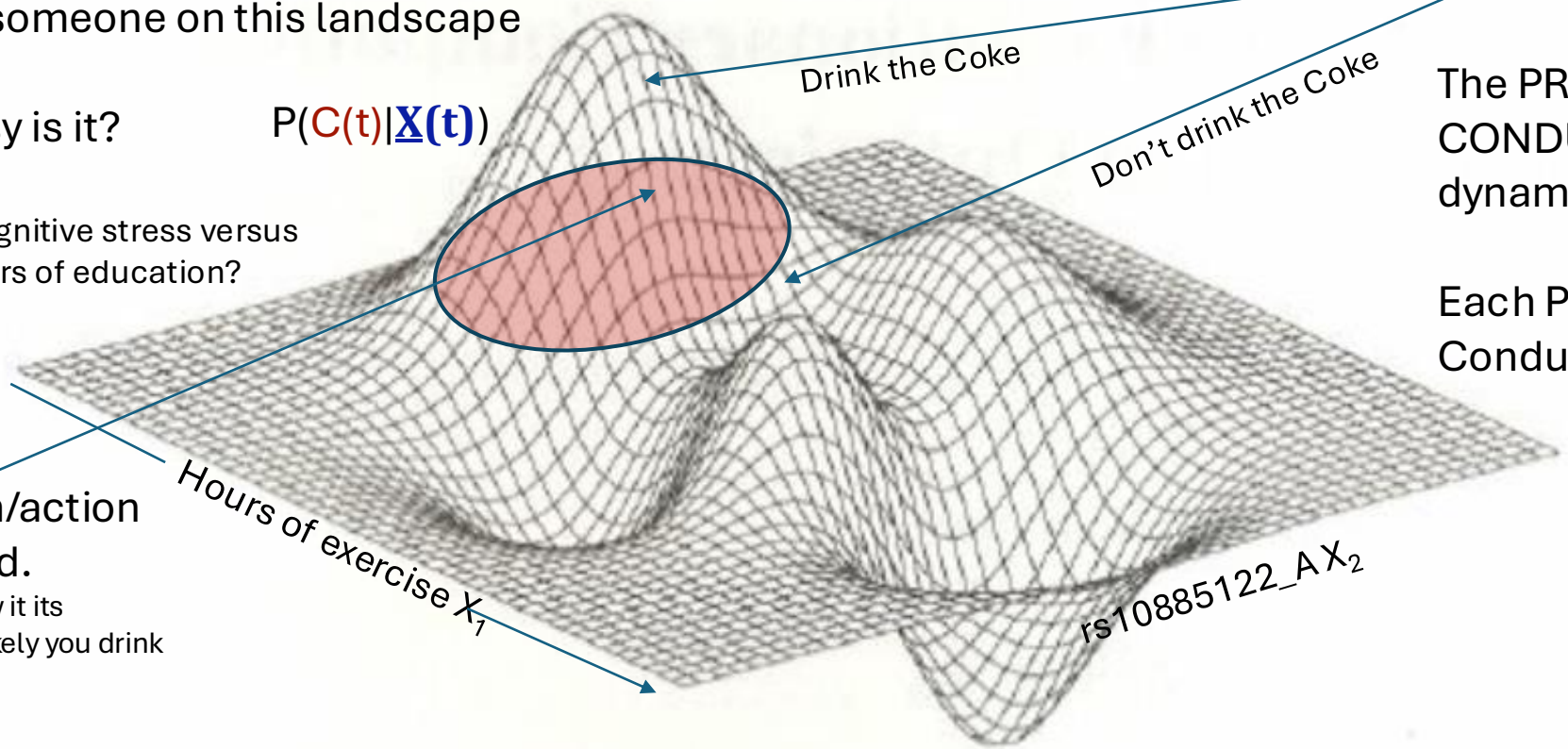
Concepts – The PREDICTABILITY Landscape
(CONDUCTOME in the case that C is a behaviour/conduct)

Conduct/behavior change means moving someone on this landscape

How do we move someone from here to here?

How easy is it? $P(C(t)|\underline{X(t)})$

Reduce cognitive stress versus 5 more years of education?



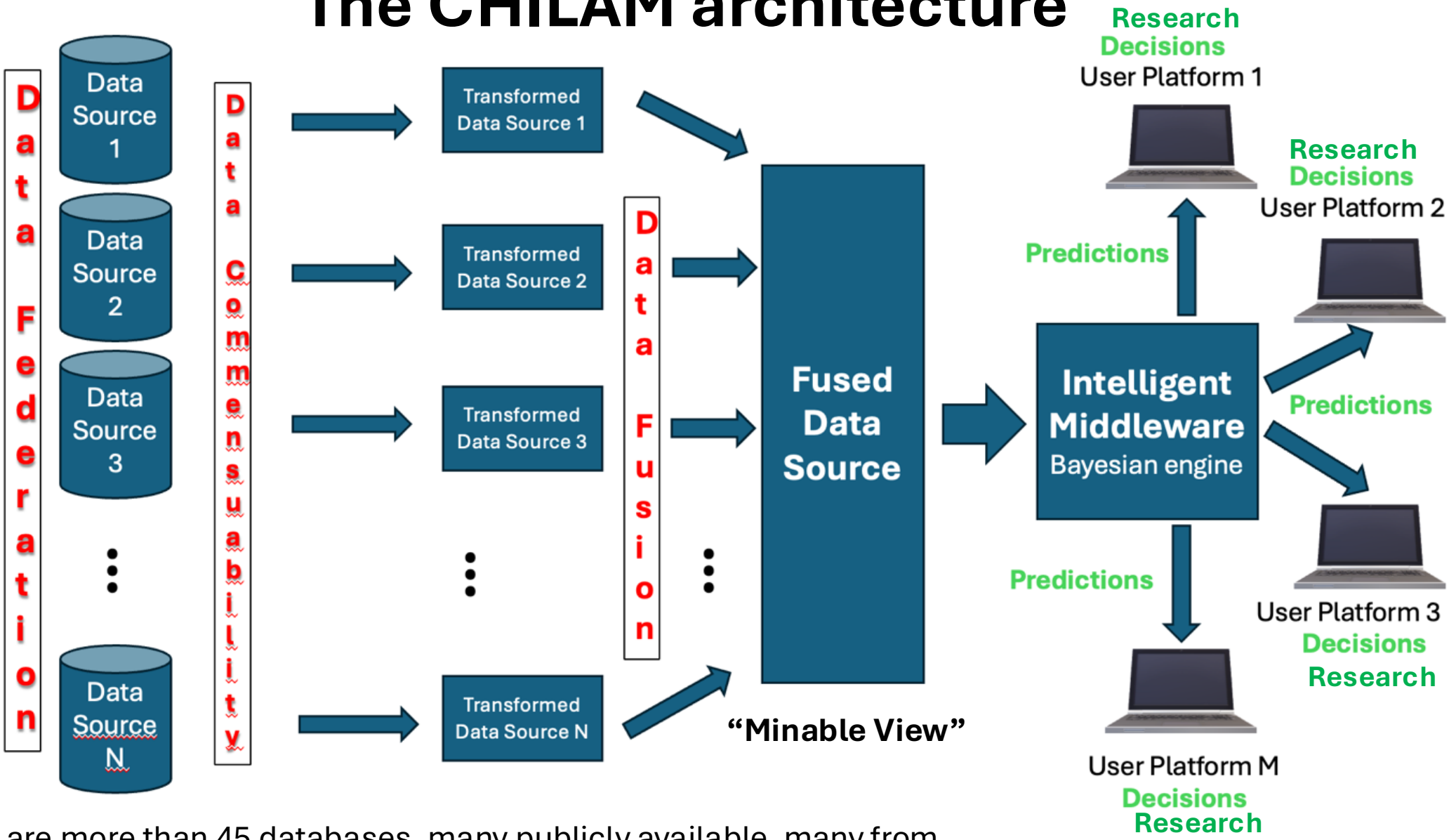
Decision/action threshold.
Above/below it its more/less likely you drink the Coke

The PREDICTABILITY-CONDUCTOME Landscape is dynamic and adaptive.

Each Problem-Decision-Action-Conduct has its own Landscape

(X_1, X_2) – Two Conductome dimensions (among many) – represents the state of the person ("who" and "what" - old/young, happy/sad,...) and/or environment ("where" and "when" - at home/work, in Mexico/France,...)

The CHILAM architecture



There are more than 45 databases, many publicly available, many from government and academic institutions - INEGI, SNIB, CONEVAL, DGE, Sec. Health., WorldClim,.... **chilam.c3.unam.mx**

Clave	Nombre	Año	Población (e= estudiantes, t= trabajadores)	n	Estatus limpieza
14UNI1_CAQ	Cuestionario + Antropometría + Química Sanguínea	2014	UNAM, e, t	1075	100%
14UNI1_G	Genética	2014	UNAM, e, t	1068	100%
17MED1_CA	Med1: Cuestionario + Antropometría	2017	UNAM, medicina e	329	90%
17MED1_Q	Med1: Química Sanguinea	2017	UNAM, medicina e	291	90%
17MED2_CA	Med2: Cuestionario + Antropometría	2017	UNAM, medicina e	131	90%
17MED2_Q	Med2: Química Sanguinea	2017	UNAM, medicina e	119	90%
19MED4_CA	Med4: Cuestionario + Antropometría	2019	UNAM, medicina e	280	90%
19MED4_Q	Med4: Química Sanguinea	2019	UNAM, medicina e	201	90%
19UNI2_Cs	Cuestionario Grande seguimiento 2019	2019	UNAM, e, t	292	0%
19UNI1_Cn	Cuestionario Grande 2014 nuevo	2019	UNAM, e, t	1170	0%
19UNI2_GC	Entrega Genética y Cuestionario	2019	UNAM, e, t	182	0%
19UNI1_M	Menú	2019	UNAM, e, t	927	0%
19UNI1_F	Foto porción - Seguimiento Cuestionario 2014	2019	UNAM, e, t	834	0%
19UNI1_A	Antropometría	2019	UNAM, e, t	1070	0%
19UNI1_Q	Química Sanguinea	2019	UNAM, e, t	1242	0%
19UNI1_Cp	Cuestionario Psicológico por correo	2019	UNAM, e, t	523	0%
19PDM1_C	Cuestionario Diabetes Mellitus	2019	pacientes con diabetes	99	80%
21UNI1_C1	Cuestionario antecedentes personales	2021	UNAM, IBERO, otras universidades, e	721	100%
21UNI1_C2	Cuestionario ancestros	2021	UNAM, IBERO, otras universidades, e	704	40%
21UNI1_C3	Cuestionario hermanos	2021	UNAM, IBERO, otras universidades, e	717	0%
21UNI1_C4	Cuestionario pareja y amigos	2021	UNAM, IBERO, otras universidades, e	700	0%
21UNI1_Cp1	Cuestionario Psicológico 1	2021	UNAM, IBERO, otras universidades, e	384	100%
21UNI1_Cp2	Cuestionario Psiclológico 2	2021	UNAM, IBERO, otras universidades, e	265	100%
21UNI1_Cp2.2	Cuestionario Psicológico 2.2	2021	UNAM, IBERO, otras universidades, e	265	100%
22ISAa_C	Cuestionario Salud en tu Vida a	2022	Universidad de la Salud aspirantes	3186	10%
22ISAi_C	Cuestionario Salud en tu Vida i	2022	Universidad de la Salud ingresados	334	0%
22IRC1_C	Cuestionario Salud en tu Vida IRC	2022	Instituto Rosario Castellano aspirantes, ingresados	253	0%

An Example and its Challenges

- 1076 workers, academics and students of the UNAM collected in 2014
- Prediction model for C = BMI > 29.9 = obesity

	Total features	# p < 0.05	% significant
Personal data	80	16	20
Anthropometry	49	26	53
Family history	549	73	13
Personal history.	131	23	18
Laboratory results	44	11	25
Health self-evaluation	227	70	31
Health knowledge	294	30	10
Nutrition	220	23	10
Lifestyle	391	54	14
Genetics	1077	34	3
Total	3062	360	12

Conclusions

1. All the world's major problems are CAS – multifactorial-multiscale-multidisciplinary-multiobjective and…
2. Our fault - caused by Decision Cycles based on "bad" predictions (there's no such thing as global warming) or "good" predictions based on "bad" objectives (short-term pleasure versus long-term health)
3. Standard frameworks for CAS are descriptive rather than predictive
4. Bayesian classifiers P(C|X) offer a "universal" phenomenologically-based (data-based) approach to modelling CAS
   a) We have lots of Tycho Brahes, lots of would-be Newtons and not enough Keplers
   b) Features can be "configuration-type" (obese, cancerous, age, SNP, external locus of control, high ghrelin,…) or "behaviour-type" (exercises, smokes,…)
5. To apply the formalism sucessfully, we need
   1. "Deep", commensurable data sets
   2. Explainable AI to analyse them
   3. Interdisciplinary groups using Human Intelligence to interpret (causality-actionability) and implement them
6. We also need to change the way we create and apply science –

Don't hold your breath!

THANK YOU! COME VISIT!

Conductome | Centro de Ciencias de la Complejidad | Facultad de Psicología | UNAM | IBERO CIUDAD DE MÉXICO | UNIVERSIDAD DE GUANAJUATO | IMSS

CHILAM: Laboratory for the Simulation of Complex Adaptive Systems

Mission

- To provide virtual data-based laboratories that enable research at the forefront of human knowledge on a range of problems of high social impact
- To provide open access predictive modelling engines available as Platform-as-a-Service to a wide variety of stakeholders and decision makers as well as the general public
- To promote evidence-based decision making by utilising the best of human intelligence and artificial intelligence

chilam.c3.unam.mx

Vision

- To be a center of innovation and excellence in the use of big, deep data and the application of Artificial Intelligence to global problems in health, social welfare and sustainability.

Centro de Ciencias de la Complejidad (C3 - UNAM)

Comisión Nacional para el conocimiento y uso de la Biodiversidad (CONABIO)

Instituto de Geografia (IGg - UNAM)

Centro de Información Geográfica y Ambiental (CIGA - UNAM)

Centro de Investigaciones en Políticas, Poblaciones y Salud (CIPPS - UNAM)

Instituto de Ciencias de la Atmósfera y Cambio Climático (CCAyCC - UNAM)

Facultad de Medicina (FM - UNAM)

Instituto de Ciencias Nucleares (ICN - UNAM)

Facultad de Psicologia (Fac. Psic - UNAM)

Instituto de Biologia (IB - UNAM)

Facultad de Medicina Veterinaria y Zootecnia (FMVZ - UNAM)

Secretaría de Salud (SALUD)

Centro Nacional de Prevención y Control de Enfermedades

Facultad de Estudios Superiores, Zaragoza (FES-Z - UNAM)

Hospital General de México (HGM)

Hospital Juárez de México (HJM)

Instituto de Neurobiología (INB - UNAM)

Instituto Nacional De Medicina Genómica (INMEGEN)

Universidad de Guadalajara (UdG)

Universidad de Guanajuato (UGto)

Universidad Iberoamericana (Ibero)

Universidad Veracruzana

Interdisciplinary teams

Consejo Nacional de Ciencia y Tecnología (CONACyT)

Microsoft Academic Relations

Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT, DGAPA - UNAM)

Fundación Carlos Slim

Secretaría de Educación, Ciencia, Tecnología e Innovación de la Ciudad de México (SECTEI - CDMX)

Funding

chilam.c3.unam.mx

SPECIES (Niche and community ecological analysis) https://species.conabio.gob.mx/
EpI-SPECIES (10 emerging/neglected diseases) https://epispecies.c3.unam.mx/
EpI-PUMA (COVID19) https://covid19.c3.unam.mx/ and https://epipuma20.c3.unam.mx/
Project42 (obesity and metabolic disease) https://project42.c3.unam.mx/
Dyana (General purpose platform) https://dyana.c3.unam.mx