The Conductome: A New Paradigm for Predicting and Explaining Human Behaviour

Christopher Stephens
C3 – Centro de Ciencias de la Complejidad
Instituto de Ciencias Nucleares, UNAM
XIX Congreso Mexicano de Psicología Social

What is the Conductome?

1. A new conceptual and theoretical framework
   Because the current ones aren't sufficiently predictive or explanatory.

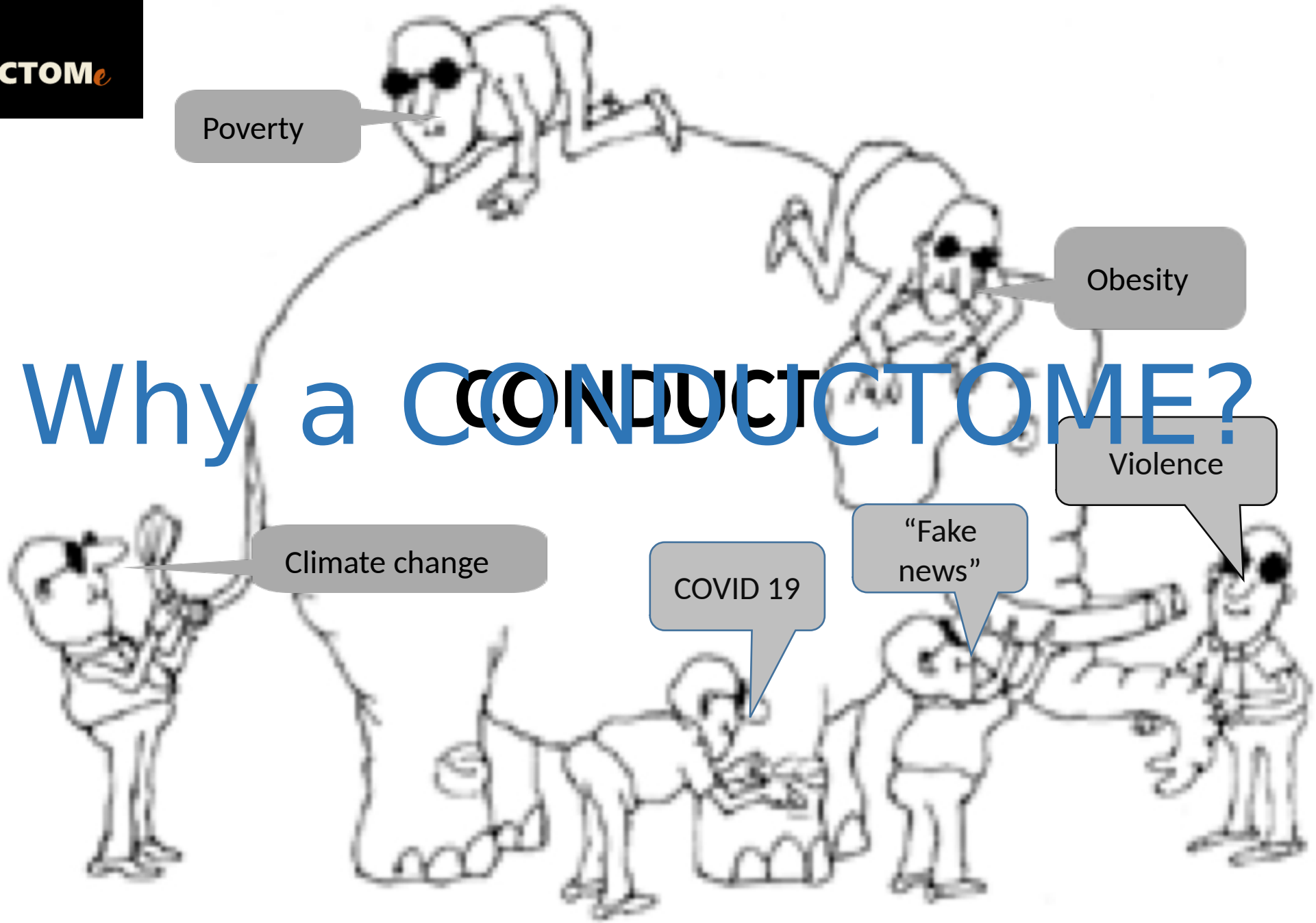
2. Based on Big, Deep data - multiscale and multi-disciplinary
   Because our conduct depends on an immense number of factors from the micro to the macro

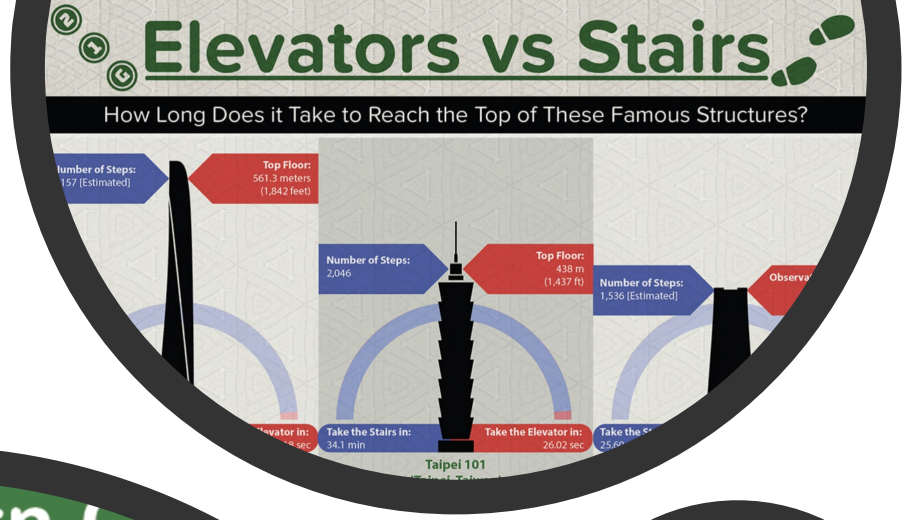
3. Based on Bayesian prediction models that are "precise", explainable and useful
   Because there is no concept of a decision (and therefore behaviour) without a corresponding prediction.

The "omic" perspective is to indicate that we are trying to characterize the totality of factors that enter into the description and prediction of the behaviour

1) A New Conceptual and Theoretical Framework

Why a CONDUCTOME?





Why do we make “bad” versus “good” decisions?

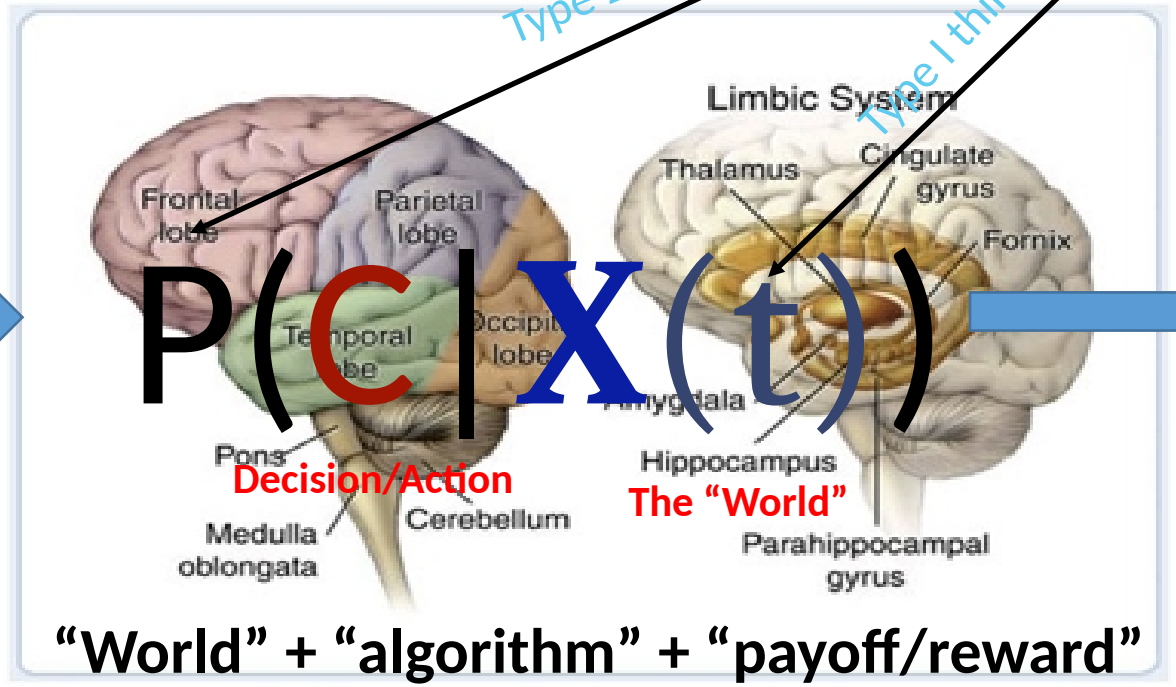
What is the Conductome?

Data + Meaning + Value
Prediction + Decision + Action

Type 2 thinking

Type I thinking

Data +
Meaning
+ Value



Prediction
+ Decision
+ Action

There are many alternative for any decision and many possible actions

"World" + "algorithm" + "payoff/reward"

Did it work?

DMV PDA is evaluated to be good or bad according to one (or more) performance criteria – payoff/reward

$P(C \mid X(t))$ can represent:

i) **Our internal model** of reality and its perception.

ii) **An external model** of reality based on observation and data

- In either case it is a **statistical inference** that is based on an **ensemble** of **"events"** (real – "**external**" ensemble or imaginary - "**internal**" ensemble)
- The "**external**" is objective and empirical – we count events and relations between effects and their potential causes and use frequencies
- The "**Internal**" is subjective and associated with our **mental model** of the world where we use Bayesian beliefs.

Now, let's go back to the beginning… What is a behaviour? – Defining C

The American Psychological Association says:

The "EFFECT"

"**behaviour** is an **organism's** activities in **response** to **external** or **internal** stimuli, including objectively observable activities, introspectively observable activities and nonconscious processes."

The "CAUSE"

1. Why is it that **only** organism's have behaviour?
2. How do we quantify EFFECTS and CAUSES?

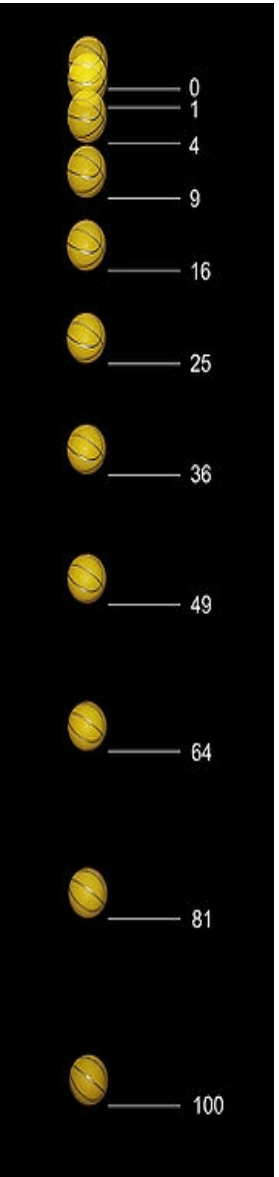
No EFFECT without a CAUSE – no behaviour without a cause

Note that many things we consider as behaviours, such as eating junk food, are considered as such without associating a specific cause

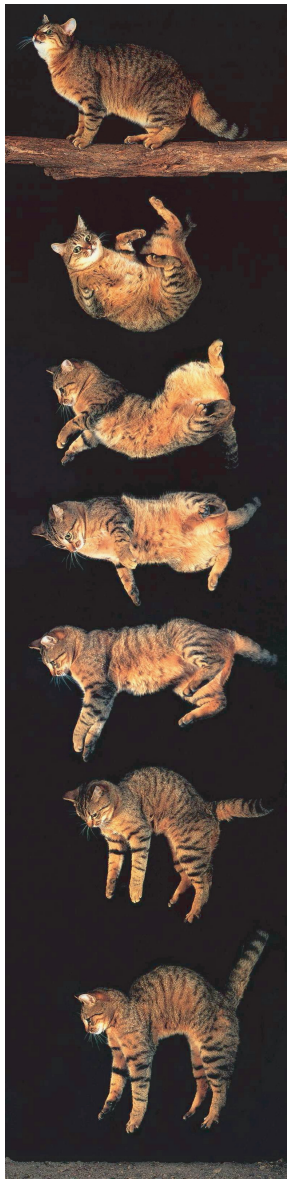
CONDUCTOMe

Defining a CONDUCT…

A CONDUCT, considered as an EFFECT on a subject S with an underlying cause c is characterized by four properties



1. A change of state of S: $\mathbf{X}(\mathbf{S},t)$ --> $\mathbf{X}(\mathbf{S}, t+1)$
   - where state is defined by a vector of state variables $\mathbf{X}(\mathbf{S},t) = (X_1(\mathbf{S},t), X_2(\mathbf{S},t)_{,...,} X_N(\mathbf{S},t))$
   - these state variables may be external – position, velocity etc. - or internal – happy/sad, hungry/satiated etc.
2. Multiple update rules F_c for that state: $\mathbf{X}(S, t+1) = F_c(\mathbf{X}(S, t))$
3. A statistical ensemble of events $\mathbf{X}(\mathbf{S},t)$ --> $\mathbf{X}(\mathbf{S}, t+1)$ and/or $\mathbf{X}(S, t+1) = F_c(\mathbf{X}(S, t))$
   - There is no such thing as a behaviour associated with a unique event
   - This ensemble can be internal (Bayesian) or external (frequentist) or a combination
   - Our internal models construct ensembles ubiquitously
4. A CONDUCT, as distinct to a behaviour, should provide a hypothesis as to the WHY? behind the EFFECT, i.e. the potential causes and also the potential payoffs/goals.



Defining the predictors **X**…

The factors that can affect a given conduct, C, i.e., to make one posible action more or less likely than another can be considered as those associated with the subject, S

$$\mathbf{X}(\mathbf{S},t) = (X_1(\mathbf{S},t), X_2(\mathbf{S},t)_{,\ldots,} X_N(\mathbf{S},t))$$

And those associated with the environment, **E**

$$\mathbf{X}(\mathbf{E},t) = (X_1(\mathbf{E},t), X_2(\mathbf{E},t)_{,\ldots,} X_N(\mathbf{E},t))$$

Among the subject variables are external states, e.g., having an altered FTO gene or being overweight; internal, perceived, states, e.g., tired or hungry; and perceptions of the environment, e.g., healthy food is expensive.

For most conducts of interest, the number of potentially important predictors is enormous and covers an enormous range of disciplines – from genetics to sociology
It is an enormous challenge to collect data that is representative of this huge spectrum of factors

Putting the behaviour C and the predictors X together – the external ensemble

- P(C |**X**) – the probability of the behaviour C given the predictors **X**

- Bayes theorem $$P(C|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|C)P(C)}{P(\boldsymbol{X})}$$

-)

  where is the score/weight of the predictor/"risk" factor

If the set of predictors **X** of the Conductome is such that > 0 then that combination of factors indicates a higher probability to be in the class C, and vice versa for < 0.

We define the behaviour C and, in frequentist terms, count from an ensemble of events N, N_C , the number of times the event occurred in the ensemble, is the number of times C did not occur. The ensemble may be transverse – a population considered without a time element – longitudinal – a person being followed in time – or both, a population over time. P(C) = N_C/N is the probability of observing the behaviour. is the number of events in which the predictor occurred and the number of times C and co-occurred. is the likelihood of the evidence in the events C. Note that we can't just count when X is highly multifactorial

CONDUCTOMe

2) Big, Deep data in Project 42

Project 42 and the Conductome

(CONACyT Fronteras, CONACyT Redes, PAPIIT, SECTEI and Microsoft Academic Relations)

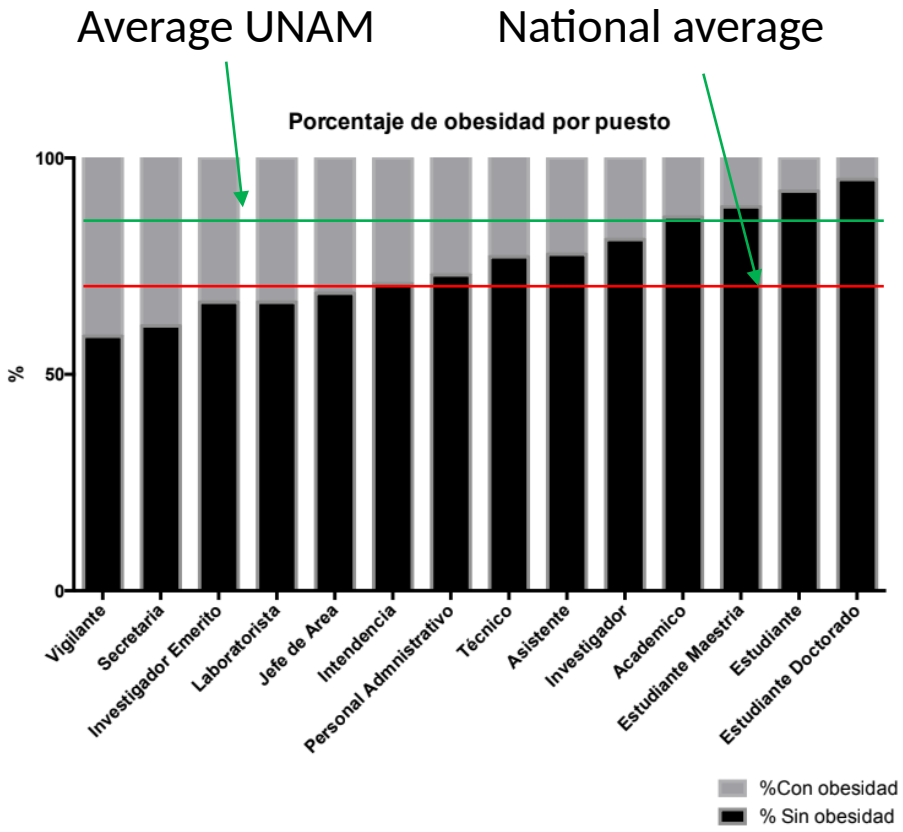
Phase I: (03-05/2014) 1,076 academics and non-academics from the UNAM (ICN, IFC, FC, IB, II, IG, IF, IM, IIMAS)
2,524 variables - Genetic, epidemiological, physiological,...
Epidemiological: Personal (81), Personal history (130), Family History (548), Self-health evaluation (226), Nutrition (220), Lifestyle (390), Health knowledge (293); **Genetic** (772); **Anthropometric and physiological** (49).

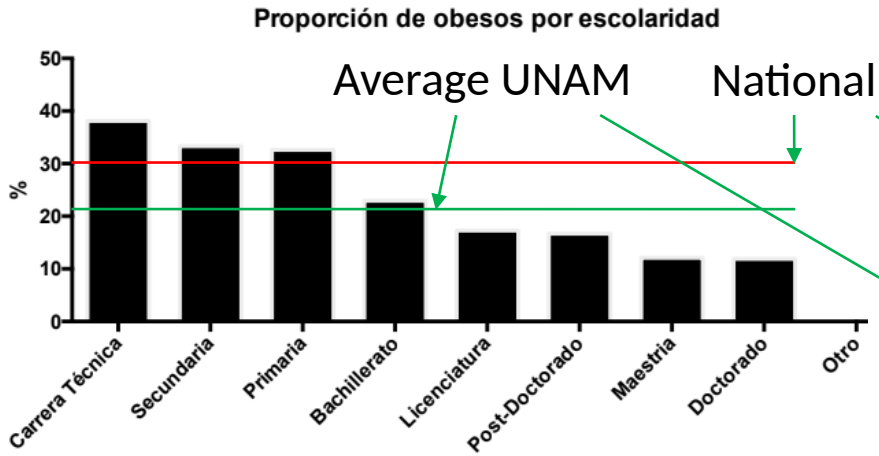
Phase II: (2017-2018) 282 undergraduate students from the FM; (06/17) 400 workers and teachers from the FM. Psychological variables were added.

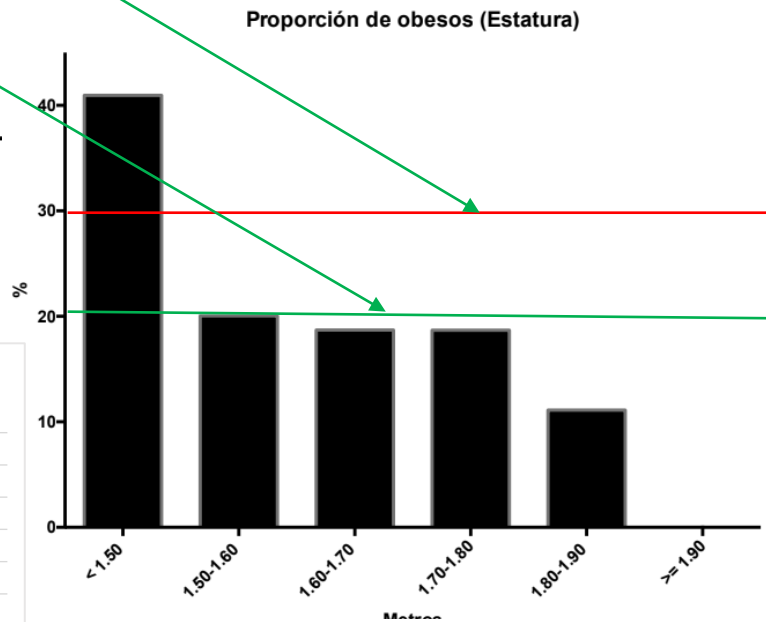
Phase III: (2019) Follow up of Phase I and new participants. Follow up blood analysis, pshychological test, detailed ”a day in your life” data, more physiological variables.

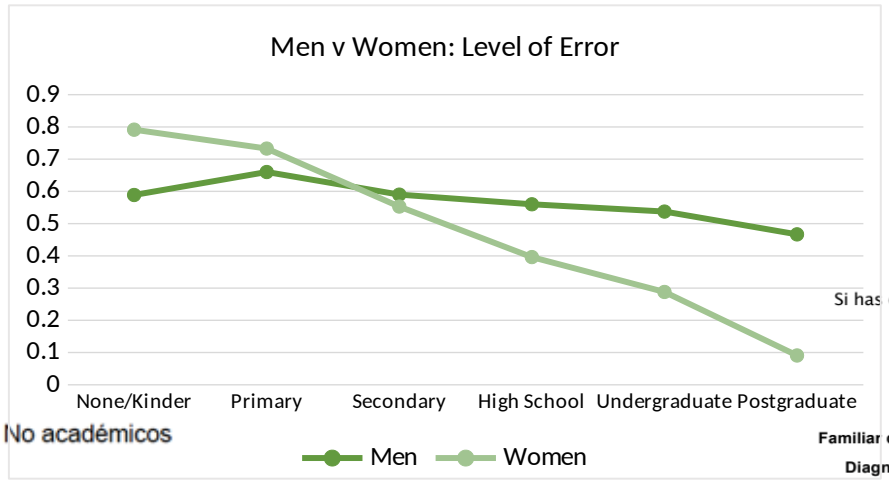
Phase IV: (2020-21) 700 undergraduate students from different institutions - FM, Fac. Psic., Ibero, FES Zaragoza
Multiple psychological instruments used. Impact of COVID 19. Link to the program Salud en tu Vida of the Mexico City Government.

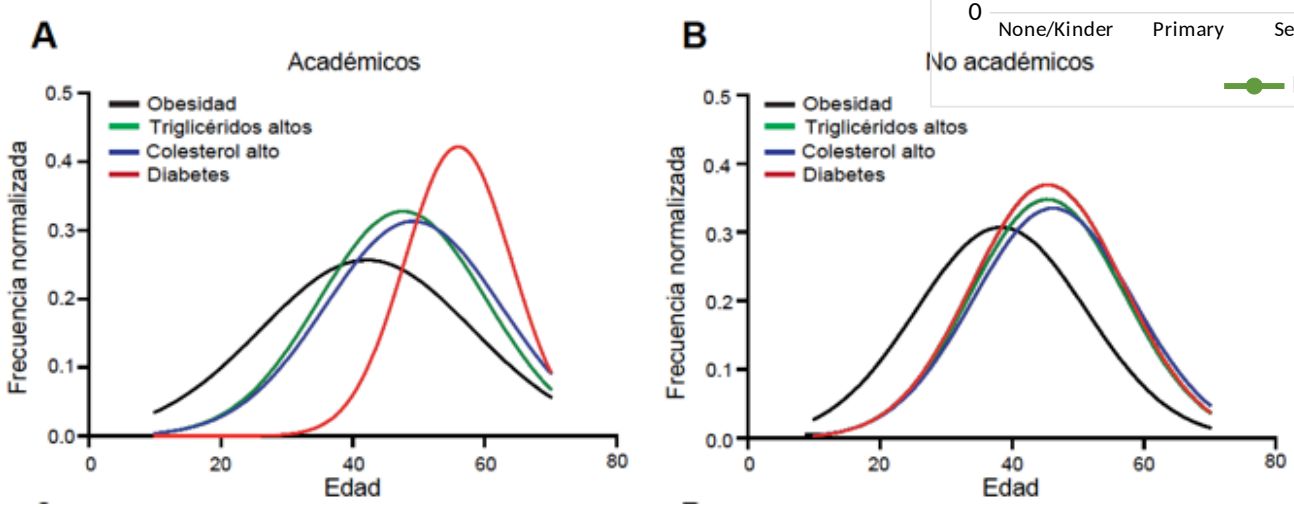
In total we now have ”deep” data on more than 3000 subjects and ”shallow” data on more than 3000 others.



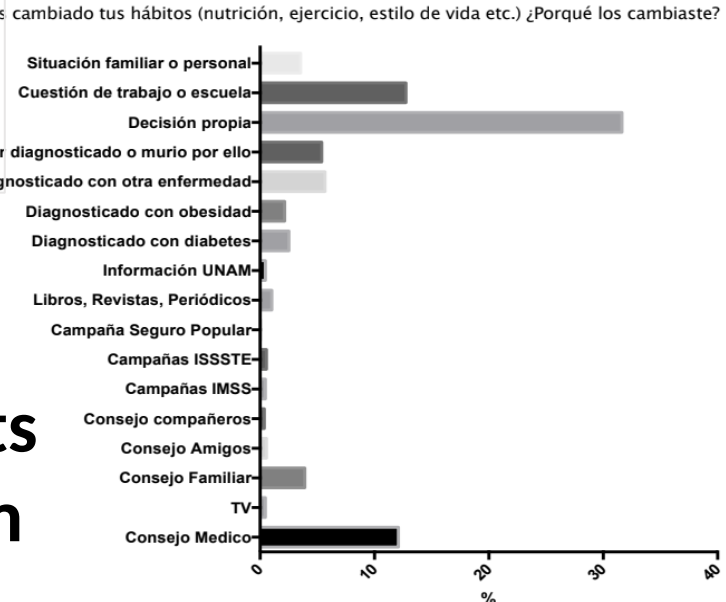








Behind these results are the conducts that cause them



3) Machine learning-based Bayesian prediction models

An explicit Conductome: “¿Haces ejercicio entre semana?” ”Do you exercise during the week?”

- P(C = Haces ejercicio entre semana = NO| external and internal factors **X**)
-)
- When is a variable X_i considered to be important in the model – through its score, but its score does not take into account the statistical significance of the relation between C and X_i hence we use a statistical diagnostic, such as a binomial test, to determine if the relation between C and X_i is significantly different to a null hypothesis
-
- If |epsilon| > 1.96 then observations are not consistent with the null hypothesis at the 95% confidence level

For our Phase 3 population - The most important “ris factors in the Conductome for the conduct – No haces ejercicio entre semana

Pregunta	Respuesta	Número de personas con X	Número de personas que no hacen ejercicio y X	N	Nc	% que no hacen ejercicio	% que no hacen ejercicio y X	Predictive model weight (score)	Statistical reliability (Epsilon)	Causa o consecuencia
¿Qué quehaceres realiza?: Cuidado de niños	Sí	41	29	292	120	41.10%	70.73%	0.38	3.86	Causa
¿Qué tan regular es su horario para ir a dormir?	1 - 2 hrs	17	14	292	120	41.10%	82.35%	0.67	3.46	Ambos
¿Realiza ejercicio en fin de semana?	No	182	97	292	120	41.10%	53.30%	0.06	3.35	Ambos
¿A qué hora se transporta a su casa?	15:00	32	22	292	120	41.10%	68.75%	0.34	3.18	Causa
¿Cuántas horas duerme entre semana?	4-5 horas	65	39	292	120	41.10%	60.00%	0.18	3.1	Consecuencia
¿Aproximadamente cuantas horas libres tiene al día entre semana?: No sé	Sí	24	17	292	120	41.10%	70.83%	0.39	2.96	Causa
¿Dónde come entre semana?: Posición 2	En puestos de la calle	9	8	292	120	41.10%	88.89%	0.9	2.91	Ambos
¿Qué quehaceres realiza?: Lavar el baño	Sí	172	89	292	120	41.10%	51.74%	0.03	2.84	Causa
¿Qué quehaceres realiza?: Sacudir	Sí	158	82	292	120	41.10%	51.90%	0.03	2.76	Causa
¿En qué tipo de vehículo se transporta de su casa al trabajo? y ¿Cuánto dura cada uno aproximadamente EN MINUTOS?: Metro: Valor	60 min	11	9	292	120	41.10%	81.82%	0.65	2.75	Causa
¿Cómo consigue sus colaciones?: La compro en un puesto	Sí	50	30	292	120	41.10%	60.00%	0.18	2.72	Ambos
¿Qué quehaceres realiza el fin de semana?: Cuidado de niños	Sí	60	35	292	120	41.10%	58.33%	0.15	2.71	Causa
¿Dónde desayuna? Seleccione por orden de frecuencia.: Posición 1	En la cocina del trabajo	27	18	292	120	41.10%	66.67%	0.3	2.7	Ambos

The most important “risk” factors in the Conductome for the conduct – Haces ejercicio entre semana (Do you exercise midw

Pregunta	Valor	Respuesta	Número de personas con X	Número de personas que no hacen ejercicio y X	% que no hacen ejercicio	% que no hacen ejercicio y X	Predictive model weight (score)	Statistical reliability (Epsilon)
¿En qué tipo de vehículo se transporta de su casa al trabajo? y ¿Cuánto dura cada uno aproximadamente EN MINUTOS?: Auto propio	Y	Sí	164	55	41.10%	33.54%	-0.3	-1.97
Cintura	(8.199, 76.28]		30	7	41.10%	23.33%	-0.52	-1.98
¿Aproximadamente cuantas horas libres tiene al día entre semana?: Tarde (En Horas)	Y	Sí	111	35	41.10%	31.53%	-0.34	-2.05
¿Aproximadamente cuantas horas libres tiene al día en fin de semana?: Noche (En Horas): Valor	2		40	10	41.10%	25.00%	-0.48	-2.07
¿Dónde realiza la mayoría de su ejercicio?	A2	Calle	18	3	41.10%	16.67%	-0.7	-2.11
¿Cómo considera que es su comida?	A4	Ligero	25	5	41.10%	20.00%	-0.6	-2.14
¿Cómo realiza su jornada laboral? y ¿Cuánto tiempo (en HORAS) aproximadamente? En movimiento: Valor	6		16	2	41.10%	12.50%	-0.85	-2.32
¿Aproximadamente cuantas horas libres tiene al día entre semana?: Mañana (En Horas): Valor	1		16	2	41.10%	12.50%	-0.85	-2.32
¿Aproximadamente cuantas horas libres tiene al día entre semana?: Tarde (En Horas): Valor	2		30	6	41.10%	20.00%	-0.6	-2.35
¿Aproximadamente cuantas horas libres tiene al día entre semana?: Mañana (En Horas)	Y	Sí	40	8	41.10%	20.00%	-0.6	-2.71
¿En qué tipo de vehículo se transporta de su casa al trabajo? y ¿Cuánto dura cada uno aproximadamente EN MINUTOS?: Auto propio: Valor	30		27	4	41.10%	14.81%	-0.76	-2.78
Grado de estudios	Doctorado		47	9	41.10%	19.15%	-0.63	-3.06
Puesto	Investigador		24	2	41.10%	8.33%	-1.04	-3.26
¿Realiza ejercicio en fin de semana?	Y	Sí	107	23	41.10%	21.50%	-0.56	-4.12

CONDUCTOMe

For our Phase 4 population of undergraduate students these are only the top 30 of 275 statistically significant predictors at the 95% and above confidence level from a total of 1800

This is multi-factoriality!

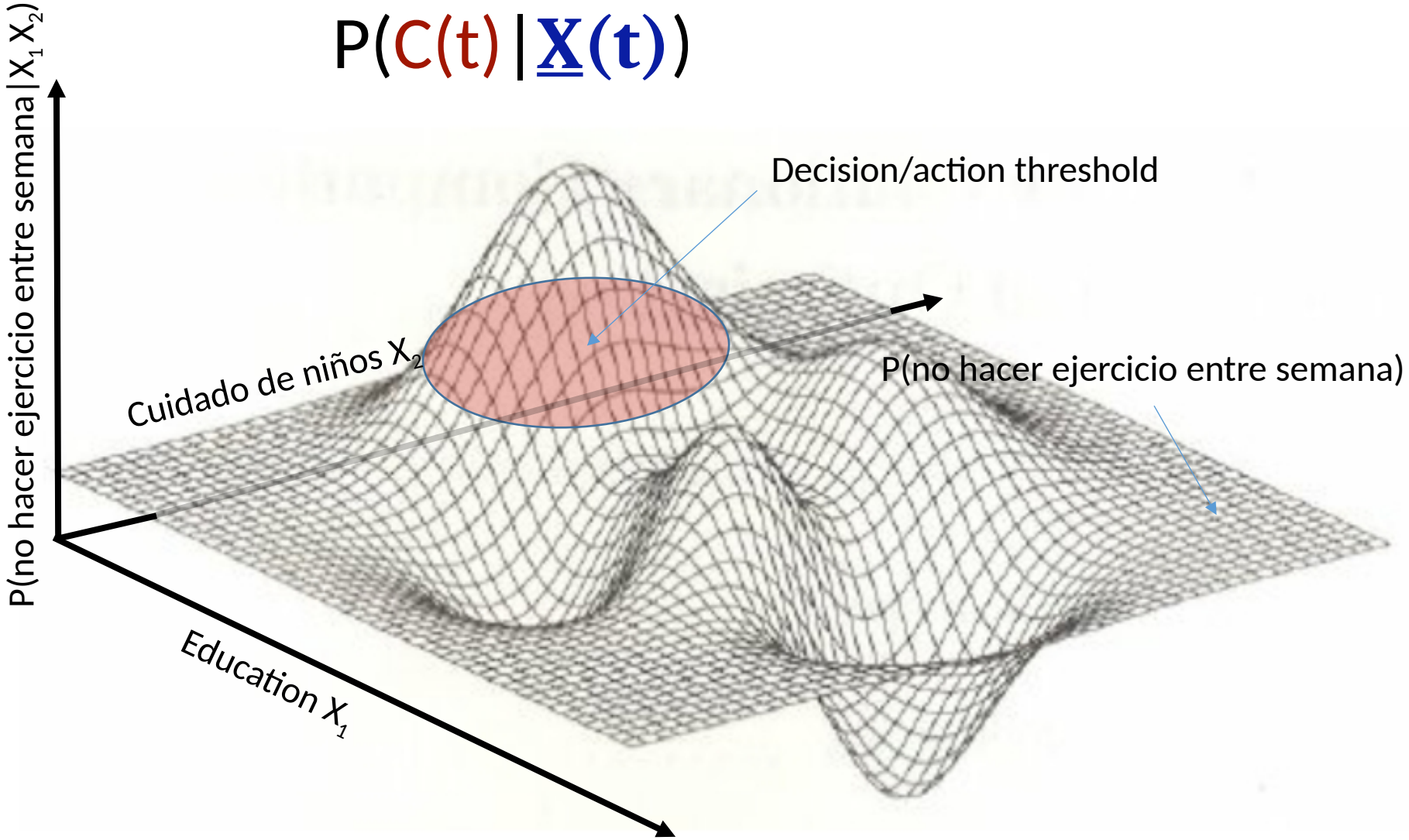
This is complexity!

We need help!

item content	answer given	answer option1	answer option2	answer option3	answer option 4	answer option5	Ncx/Nx	Epsilon	Score si
¿Cómo consideras que es tu condición física durante la carrera (con pandemia)?	A5	A2 = Muy mala	A3 = Mala_2	A4 = Regular	A5 = Buena_4	A6 = Muy buena_5	0.82	7.80	1.42
¿Haces /has hecho ejercicio habitualmente?	A2	A2 = Sí_1	A3 = No_2				0.71	6.94	0.76
¿Cómo consideras que es tu salud durante la carrera (con pandemia)?	A6	A2 = Muy mala	A3 = Mala_2	A4 = Regular	A5 = Buena_4	A6 = Muy buena_5	0.83	5.24	1.46
¿Cómo consideras que es tu condición física durante la carrera (con pandemia)?	A6	A2 = Muy mala	A3 = Mala_2	A4 = Regular	A5 = Buena_4	A6 = Muy buena_5	0.88	5.20	1.94
donde 1 significa "no restringirme en comer" (comer lo que quiera, donde quiera) y 8 significa "restringirme en comer" (limitarse constantemente y nunca comerlo): ¿Qué número podría describirte mejor?	A5	A2 = 1-2	A3 = 3-4	A4 = 5-6	A5 = 7-8		0.94	4.95	2.71
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Mi falta de iniciativa me impide hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.81	4.84	1.54
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Mi falta de voluntad me impide hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.81	4.52	1.52
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Aunque me lo propongo, nunca logro hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.72	4.45	1.04
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Después de acabar mis actividades no me queda tiempo para hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.85	4.36	1.84
¿Cuánto consideras que comías durante la carrera (con pandemia)?	A4	A2 = Mucho Me	A3 = Menos	A4 = Lo recom	A5 = Más de	A6 = Mucho más de	0.67	4.29	0.57
una rutina de ejercicio. ¿Qué tan segura(o) estas de poder ajustarte a la rutina de manera regular o consistente (más de tres veces a la semana) a pesar de estas situaciones? Sin el apoyo de familia o amigos.	A6	A2 = 0-1	A3 = 2-3	A4 = 4-6	A5 = 7-8	A6 = 9-10	0.73	4.18	1.07
Retraso_de_gratificación_respuestas_de_frecuencia: Selecciona la respuesta que más te caracteriza. Siempre he tratado de comer saludable porque es una buena decisión a futuro	A6	A2 = Nunca_1	A3 = Casi nur	A4 = Algunas	A5 = Casi sie	A6 = Siempre_5	0.77	4.09	1.18
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Mi falta de organización me impide hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.79	4.09	1.41
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Mi inconstancia me impide hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.78	4.01	1.34
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Mis deberes me impiden hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.84	4.01	1.73
Tu o algún integrante de este hogar tiene: ¿Personas de limpieza?	Y	Y = Sí_1	N = No_2				0.68	3.89	0.64
¿Comes saludable durante la carrera (con pandemia)?	Y	Y = Sí_1	N = No_2				0.63	3.79	0.40
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Mis responsabilidades me impiden realizar ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.81	3.78	1.55
Autoeficacia_ejercicio: Se describen situaciones que podrían reflejar la dificultad que implica apegarse a una rutina de ejercicio. ¿Qué tan segura(o) estas de poder ajustarte a la rutina de manera regular o consistente (más de tres veces a la semana) a pesar de estas situaciones? Después de recuperarte de una enfermedad que te impidió seguir con el ejercicio.	A6	A2 = 0-1	A3 = 2-3	A4 = 4-6	A5 = 7-8	A6 = 9-10	0.75	3.76	1.18
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Quisiera hacer ejercicio pero no sé cómo hacerlo	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.67	3.74	0.81
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Lo único que me resulta posible respecto al ejercicio es caminar	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.67	3.68	0.81
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Distraerme en muchas cosas me impide hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.76	3.67	1.25
Retraso_de_gratificación_respuestas_de_nivel (orignal): Selecciona la respuesta que más te caracteriza. Siempre he tratado de comer saludable porque es una buena decisión a futuro	A2	A6 = Totalment	A5 = Desacue	A4 = Ni en ac	A3 = De acue	A2 = Totalmente de	0.73	3.65	0.97
Locus_de_control_alimento: Selecciona el grado en que te describe la afirmacion: Como porque tengo que comer	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.89	3.48	2.16
Locus_de_control_ejercicio: Selecciona el grado en que te describe la afirmacion: Me resulta imposible hacer ejercicio	A6	A6 = No me des	A5 = Casi no	A4 = Me desc	A3 = Me desc	A2 = Me describe to	0.64	3.47	0.67
¿Cómo consideras que es tu peso durante la carrera (con pandemia)?	A4	A2 = Muy bajo	A3 = Bajo_2	A4 = Normal	A5 = Con sob	A6 = Obeso_5	0.61	3.40	0.37
G_uni_campus	A3	A1 = UNAM CIU	A2 = UNAM	A3 = IBEROAMERICANA SANTA FE_5			0.64	3.38	0.51
Autoeficacia_ejercicio: Se describen situaciones que podrían reflejar la dificultad que implica apegarse a una rutina de ejercicio. ¿Qué tan segura(o) estas de poder ajustarte a la rutina de manera regular o consistente (más de tres veces a la semana) a pesar de estas situaciones? Después de vivir problemas familiares.	A6	A2 = 0-1	A3 = 2-3	A4 = 4-6	A5 = 7-8	A6 = 9-10	0.72	3.35	1.05

The Conductome Landscape

$P(C(t)|\underline{\mathbf{X}(t)})$



This construction es purely phenomenological – we put all the (**X**) dentro in the machine P(C|X) and see what comes. This is the AI part. The role of Human Intelligence is to interpret what it means. A vital part is the search for causality.

The landscape has a multitude of dimensions – genetic, epigenetic, physiological, psychological, sociological, environmental, economic, political,…

1) A New Conceptual and Theoretical Framework – Reprise: The Internal Ensemble – How do make predictions to make decisions?

What does an organism have to do with its internal models? It has to satisfy multiple obectives. However, …

- According to “Rational Choice Theory” we are optimizing agents that optimize a unique utility function – there is only one objective
- For example, consider an experiment where we have a choice between taking the stairs or the elevator. There are multiple payoffs/goals:
  - DT = the time difference between one and the other;
  - DE = the difference in expended energy;
  - DH = the perceived health benefit;
  - DS = the perceived social benefit;
  - And potentially many more.
- The simultaneous optimization of these goals is not possible. The system is ”frustrated”. E.g., the difference in time versus energy.
  - The relative value of one versus the other is a function of the states of the subject and their environment
    - E.g., if one is tired (subject effect) taking thte elevator will be more likely. If the elevator is very small or the stairs slippery, this may affect the decision (environment effect).

How is a decision made?
What's your decision worth?

There are N_v value functions/utilities: $V = (v_1, v_2, \ldots, v_{Nv})$ – energy expenditure, time etc.

An action A_k causes changes to these functions $DV(A_k) = (Dv_1(A_k), Dv_2(A_k), \ldots, Dv_{Nv}(A_k))$.

$DV(A_k)$ may take different values depending on whether it is measured before or after the action

Post-action, $DV(A_k)$ represents actual or perceived payoff - results. They are usually subjective but can sometimes be compared to reality.

Pre-action, $<Dv_1(A_k)>$ represents a prediction of the gain due to the action. <…> does not necessarily mean that it is an expected value associated with an external set, although it could be, but rather, it means that one has an internal prediction model to estimate the change.

It is hypothesized that the probability for action A_k is

$P(A_k \mid <DV(A_K)>) = P(A_k \mid <Dv_1(A_k)>, <Dv_2(A_k)>, \ldots, <Dv_{Nv}(A_k)>)$

Thus, the probability for the A_k action is conditioned on the predicted gains of the action associated with the payoff functions.

- A decision is "good" or "bad" with respect to a value function v_i if $Dv_i(A_k) > 0$ ("good" decision) versus $Dv_i(A_k) < 0$ ("bad" decision). Thus, decisions can be frustrated – good for some value functions and bad for others.

How do you evaluate a utility?

If you assume that a decision and a corresponding action are taken according to the predicted changes in a set of utilities, you have to ask, how do you make the predictions?

P(Dvi(Ak) | X(Subject), X(Object), X(Environment))

<Dvi(Ak)> = F(P(Dv_i(Ak) | X(Subject), X(Object), X(Environment))

i.e., within our mental model of the world there is a prediction model that estimates the probability of a certain payoff given that action is implemented in a given state of the world (subject, object, environment) and an expected payoff.

E.g., if the subject is in the "very tired" state, the probability for a perceived large increase in effort, SD, from taking the stairs would be higher than in the "non-tired" state with the consequence that P(stairs | <DE(stairs)>1) < P(stairs | <OF(stairs)>0), where <DE(stairs)>1 is the predicted effort to take the stairs since the subject is in state 1 = "very tired" and <DE(stairs)>0 is the predicted effort to take the stairs since the subject is in state 0 = "not tired"

•

Conclusions

The Conductome Project and the construction of a specific Conductome is feasible – P(C|X) can be calculated from Project 42 data using Bayesian Machine Learning algorithms in a way that exhibits predictability, explainability and usefulness.

There are many behaviors to model.

- There are a myriad of factors that contribute to a particular Conductome – how do you overcome the disciplinarity associated with concentrating on one subset and ignoring others?
- How do you deal with such a high degree of multifactoriality?
- How do we move from a phenomenological, external ensemble approach P(C|X) to a more "mechanical" and causal understanding of our mental models? P(C|X) is an indirect but powerful window. How do we go further than that?