The neglected importance of complexity in statistics and Metascience

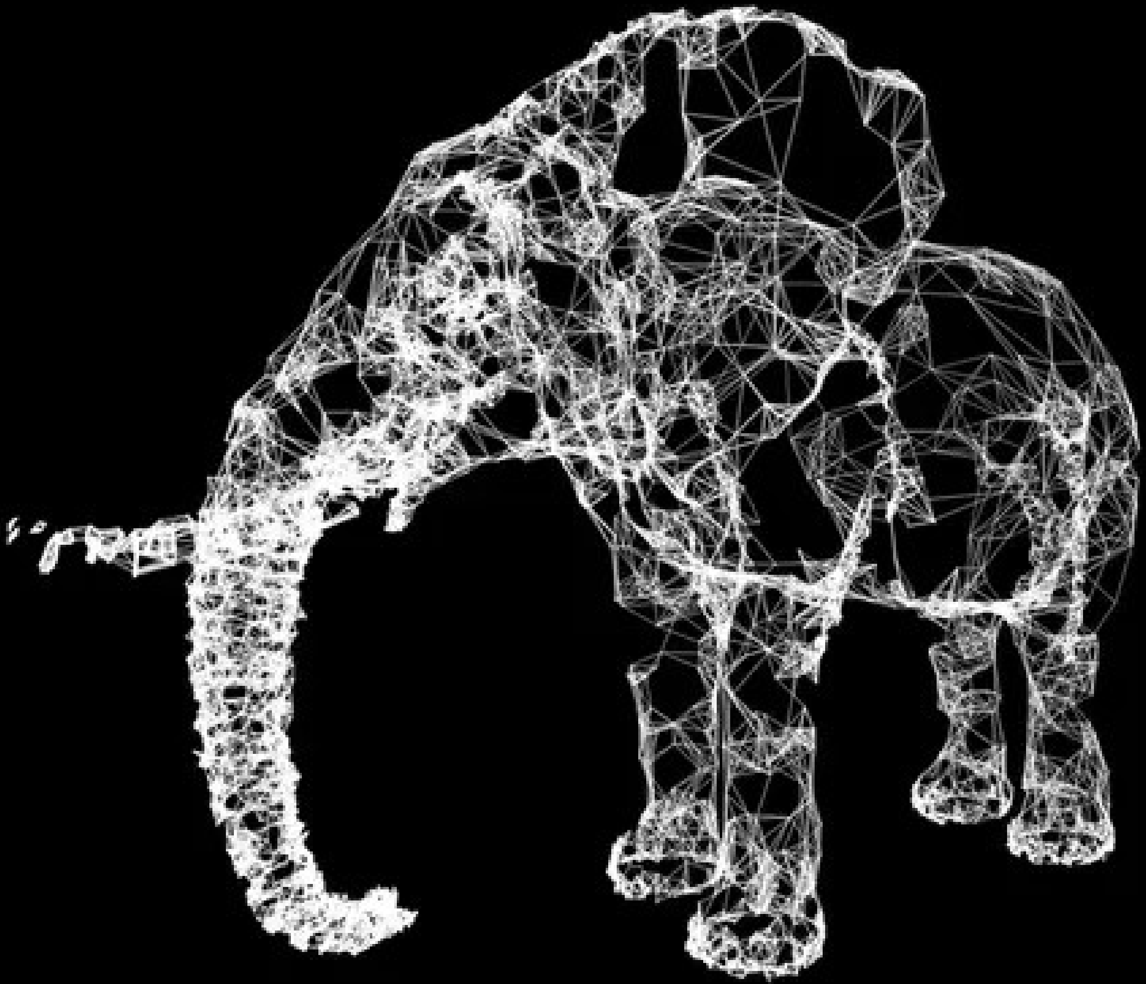
Daniele Fanelli

In this talk:

1) what’s missing in the current paradigm

2) what a new paradigm might look like

3) evidence in support of this proposal

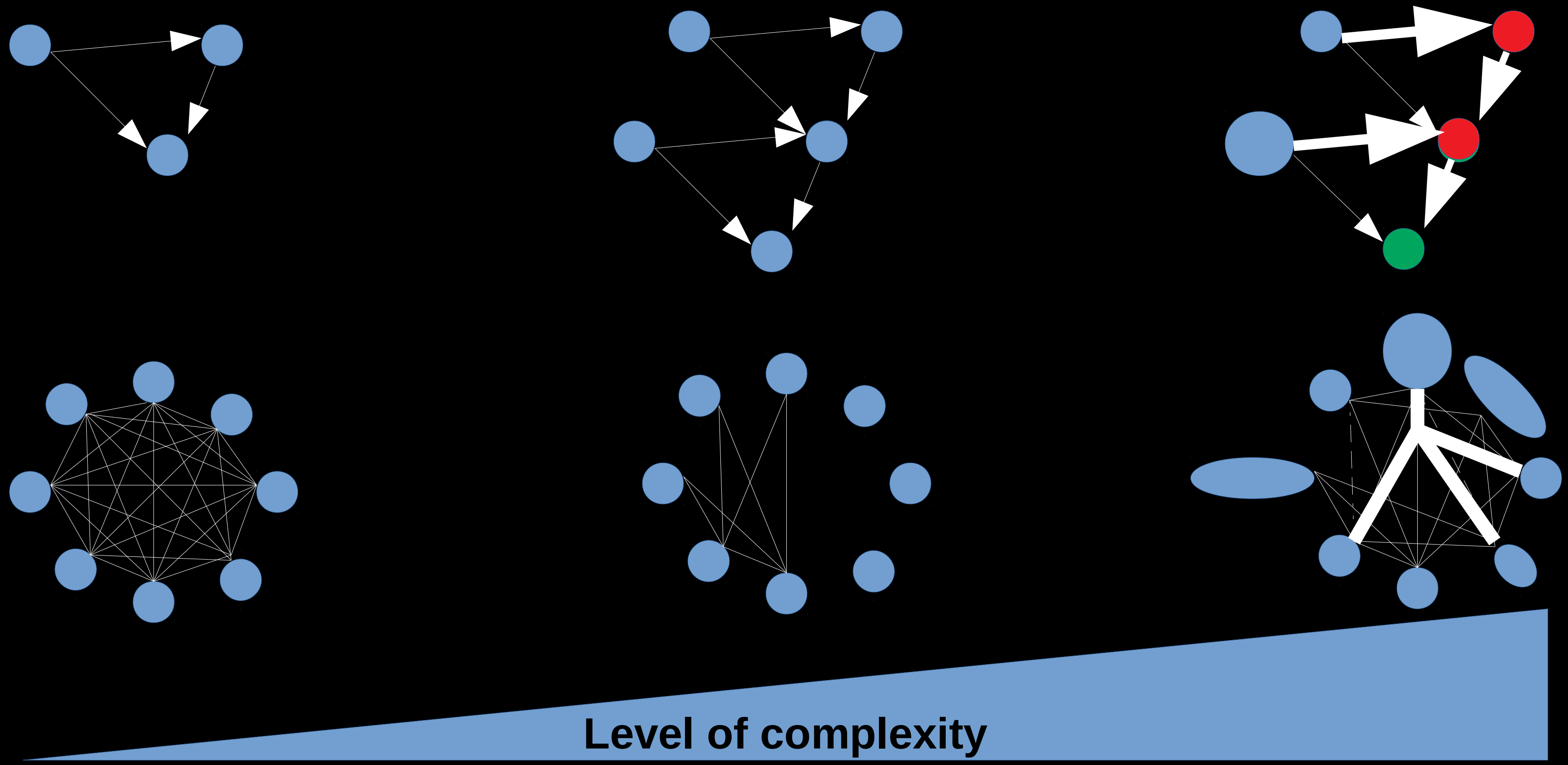


The elephant in the room of:

1) metascience
   - e.g. reproducibility
2) statistics
   - e.g. model “complexity”
   - [see SW seminar 2021, Fanelli 2019, 2022]

What is “complex”?

Many, diverse, interacting parts.
Long to describe, difficult to predict.



example 1) complexity deflates the “reproducibility crisis”.

year	project	discipline	N	result
2014	Many labs 1	**psychology, misc.**	13, 36 labs	**77%**
2016	COS	**social+cognitive psychology**	100	**36-68%**
2016	Camerer et al.	**experimental economics**	18	**61-78%**
2018	Many labs 2	**social+cognitive psychology**	28, 62 samples, 36 countries	**54%**
2018	Camerer et al.	**social studies in Nature, Science**	21	**57-67%**
2021	RPCB	**cancer biology**	188, 50 exper., 23 papers	**3-82%**

Lower reproducib:

1) complex phenomena
2) complex methods

- not just random noise
- structured, systematic diff.

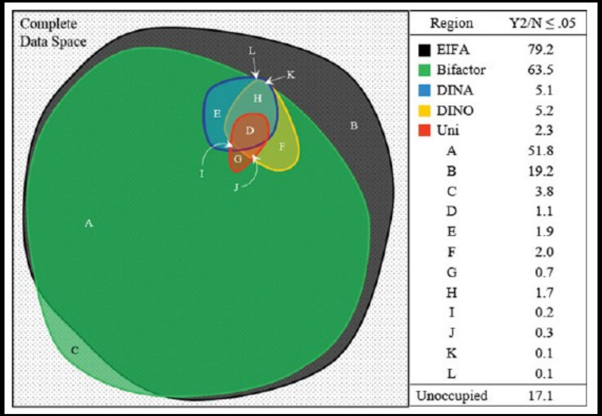
example 2) complexity confuses statistical results

$$AIC: -2\log(L) + 2k$$

models vary by "fitting propensity"

("complexity" beyond n. of parameters)

universe of possible data



(Bonifay and Cai 2017)

When is a theory <u>actually</u> supported?

How might the elephant appear?

1) integrate complexity of phenomena & methods in measuring, forecasting, correcting reproducibility
2) penalize statistical models for complexity beyond number of parameters

integrating “complexity” ~ severity of testing

Part 2: A candidate alternative

https://www.wallpaperflare.com/elephant-during-daytime-grayscale-photo-of-elephant-portrait-wallpaper-zhdat

K theory in a nutshell

$$K = \frac{H(Y) - H(Y|X,\tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$

(Fanelli 2019, Fanelli 2022)

K = consilience

$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$

explain/predict/control
more/diverse phenomena
with fewer/simpler theories/methods

what the variables represent

$$K = \frac{H(Y) - H(Y|X,\tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$

explain/predict/control

more/diverse phenomena

with fewer/simpler theories/methods

more information Y makes K GROW

$$K = \frac{H(Y) - H(Y|X,\tau)}{H(Y) + \frac{n_X H(X) + D(\tau)}{n_Y}}$$

(Fanelli 2019, Fanelli 2022)

more information Y makes K GROW

$$K = \frac{H(Y_1) + H(Y_2) + H(Y_3) - H(Y|X, \tau)}{H(Y_1) + H(Y_2) + H(Y_3) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$

(Fanelli 2019, Fanelli 2022)

K theory in a nutshell

$$K = \frac{H(Y) - H(Y|X,\tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$

(Fanelli 2019, Fanelli 2022)

more info. Y|X, X, τ makes K small

$$K = \frac{H(Y) - H(Y|X,\tau)}{H(Y) + \frac{n_X H(X) + D(\tau)}{n_Y}}$$

(Fanelli 2019, Fanelli 2022)

more info. Y|X, X, τ makes K small

$$K = \frac{H(Y) - H(Y_1|X, \tau)+H(Y_2|X, \tau)+H(Y_3|X, \tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau_1)+D(\tau_2)+D(\tau_3)}{n_Y}}$$

(Fanelli 2019, Fanelli 2022)

K theory in a nutshell

$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}} =$$

≈1, full consilience

=0, no knowledge

<0, wrong

(Fanelli 2019, Fanelli 2022)

K theory in a nutshell

$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \frac{n_X H(X) + D(\tau)}{n_Y}}$$

(Fanelli 2019, Fanelli 2022)

K of a regression model

$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$

$$Y = \alpha + \beta X + error$$

(Fanelli 2019, Fanelli 2022)

this has been said before

$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \frac{n_X H(X) + D(\tau)}{n_Y}}$$

$$Y = \alpha + \beta X + error$$

(Fanelli 2019, Fanelli 2022)

key theoretical innovations

$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$

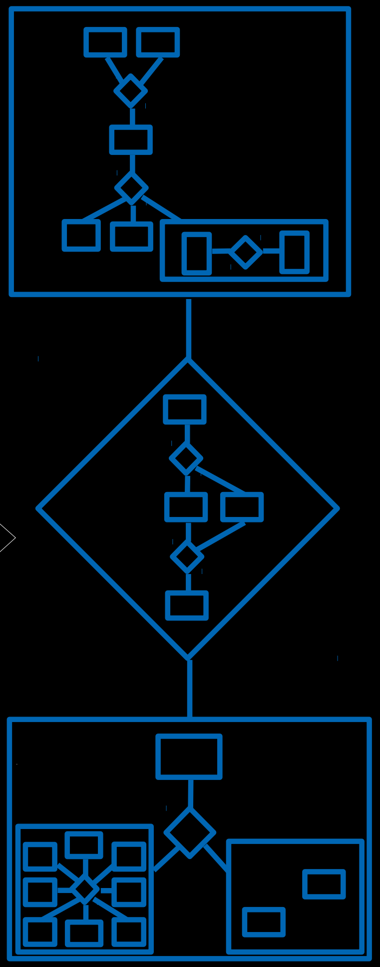
$$Y = \alpha + \beta X + error$$

(Fanelli 2019, Fanelli 2022)

key methodological innovations

$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \frac{n_X H(X) + D(\tau)}{n_Y}}$$

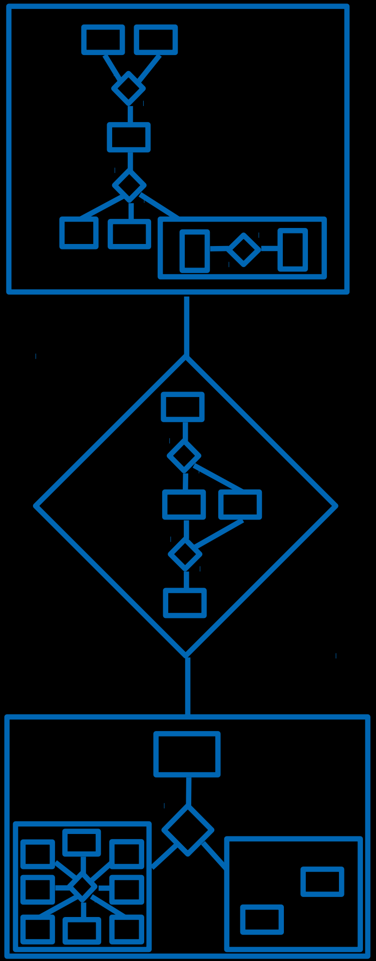
$$Y = \alpha + \beta X + \text{error}$$



(Fanelli 2019, Fanelli 2022)

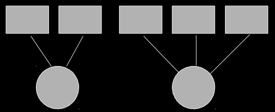
graphs are everywhere in science

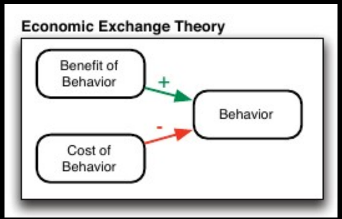
$$K = \frac{H(Y) - H(Y|X, \tau)}{H(Y) + \dfrac{n_X H(X) + D(\tau)}{n_Y}}$$



theories

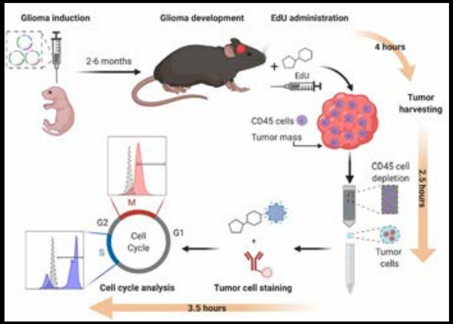
$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$





(Mueller 2015, ICSS)

methodologies



www.protocols.io/view/an-optimized-protocol-for-in-vivo-analysis-of-tumo-3byl471m2lo5/v1

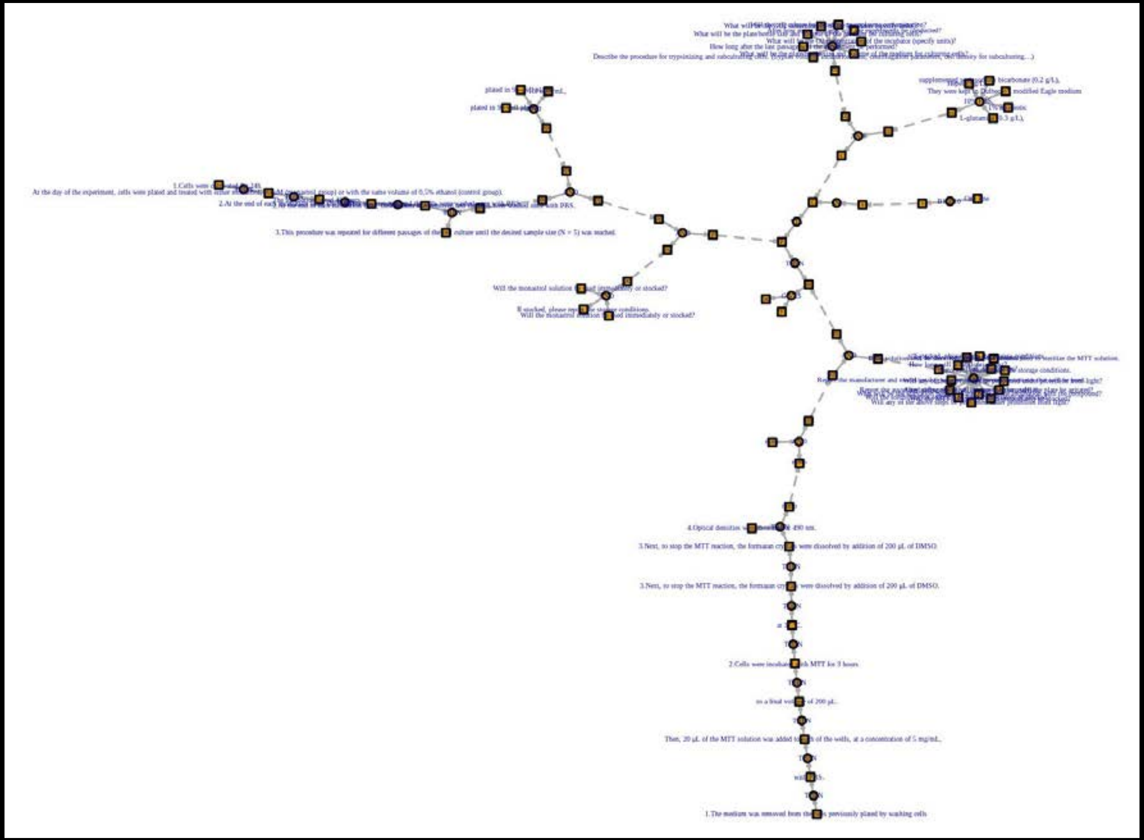
(Fanelli 2019, Fanelli 2022)

Part 3: supporting evidence

https://www.wallpaperflare.com/elephant-in-black-and-white-elephant-photo-animal-grey-wallpaper-tkqgk

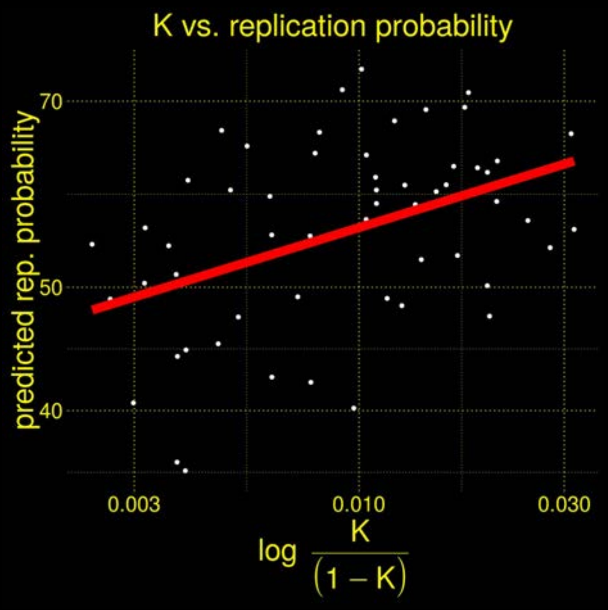
1) K predicts perceived
and actual reproducibility

“tau” of biological experiments

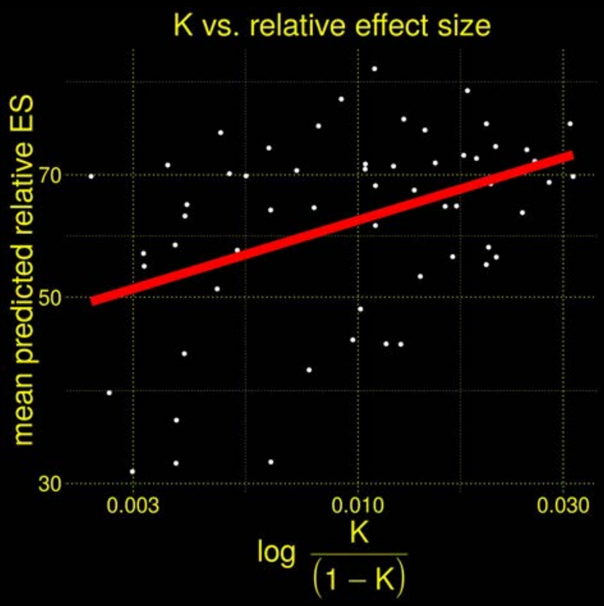


(Fanelli, Tan, Amaral & Neves, 2022, MetaArxiv)

K vs. perceived reproducibility



R^2	R[adj]^2
0.187	0.144



R^2	R[adj]^2
0.279	0.241

(Fanelli, Tan, Amaral & Neves, 2022, MetaArxiv)

K vs. actual reproducibility

$$K_r = K_o 2^{-\lambda \cdot d}$$

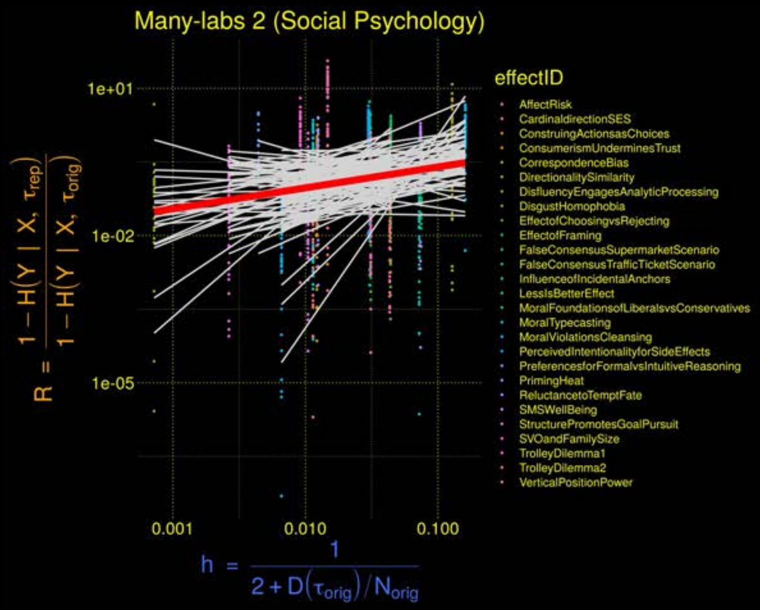
$$k_r h_r = k_o h_o 2^{-\lambda \cdot d}$$

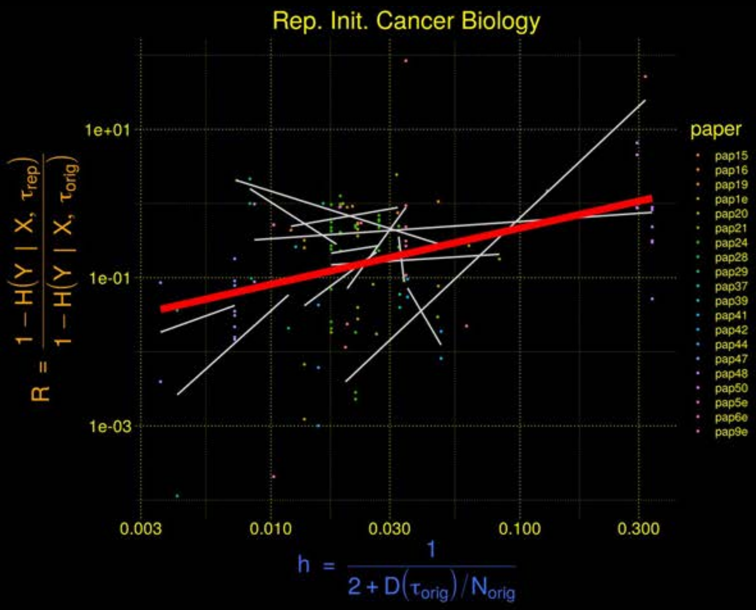
$$\log \frac{k_r}{k_o} = \log \frac{h_o}{h_r} - \lambda \cdot d$$

$$R \equiv \log \frac{H(Y) - H(Y|X, \tau_r)}{H(Y) - H(Y|X, \tau_o)} = \alpha + \beta \log \frac{1}{D(\tau_r)/N}$$

1) K predicts actual reproducibility

(part of collaboration with Brazilian Reproducibility Initiative)





(Fanelli, Tan, Amaral & Neves, in prep)

independent new predictor

multiple regression, Y=reproducibility

	social psychology	cancer biol.
median (ir)reproducibility	-4.76 [-7.01; -2.5] ***	-2.65 [-3.21; -2.1] ***
$-\log\left(2+D\left(\tau_{orig}\right)\right)$	0.47 [0.25; 0.69] ***	1.27 [0.68; 1.86] ***
$\log\left(P_{orig}\right)$	-0.13 [-0.15; -0.1] ***	0 [0; 0.01]
$\sqrt{N_{orig}}$	0.01 [-0.11; 0.13]	0.19 [0.08; 0.3] ***

(Fanelli, Tan, Amaral & Neves, in prep)

very easy to measure, automatize

multiple regression, Y=reproducibility

	social psychology	cancer biol.
median (ir)reproducibility	-4.76 [-7.01; -2.5] ***	-2.65 [-3.21; -2.1] ***
$-\log\left(2+D\left(\tau_{orig}\right)\right)$	0.47 [0.25; 0.69] ***	1.27 [0.68; 1.86] ***
$\log\left(P_{orig}\right)$	-0.13 [-0.15; -0.1] ***	0 [0; 0.01]
$\sqrt{N_{orig}}$	0.01 [-0.11; 0.13]	0.19 [0.08; 0.3] ***

(here $D(\tau)$ based on sentences in replication protocol!)

(Fanelli, Tan, Amaral & Neves, in prep)

better than just P-values and N

multiple regression, Y=reproducibility R

	social psychology	cancer biol.
median (ir)reproducibility	-4.76 [-7.01; -2.5] ***	-2.65 [-3.21; -2.1] ***
$-\log(2 + D(\tau_{orig}))$	0.47 [0.25; 0.69] ***	1.27 [0.68; 1.86] ***
$\log(P_{orig})$	-0.13 [-0.15; -0.1] ***	0 [0; 0.01]
$\sqrt{N_{orig}}$	0.01 [-0.11; 0.13]	0.19 [0.08; 0.3] ***

(Fanelli, Tan, Amaral & Neves, in prep)

leads to progress in metascience

multiple regression, Y=reproducibility

	social psychology	cancer biol.
median (ir)reproducibility	-4.76 [-7.01; -2.5] ***	-2.65 [-3.21; -2.1] ***
$-\log(2 + D(\tau_{orig}))$	0.47 [0.25; 0.69] ***	1.27 [0.68; 1.86] ***
$\log(P_{orig})$	-0.13 [-0.15; -0.1] ***	0 [0; 0.01]
$\sqrt{N_{orig}}$	0.01 [-0.11; 0.13]	0.19 [0.08; 0.3] ***

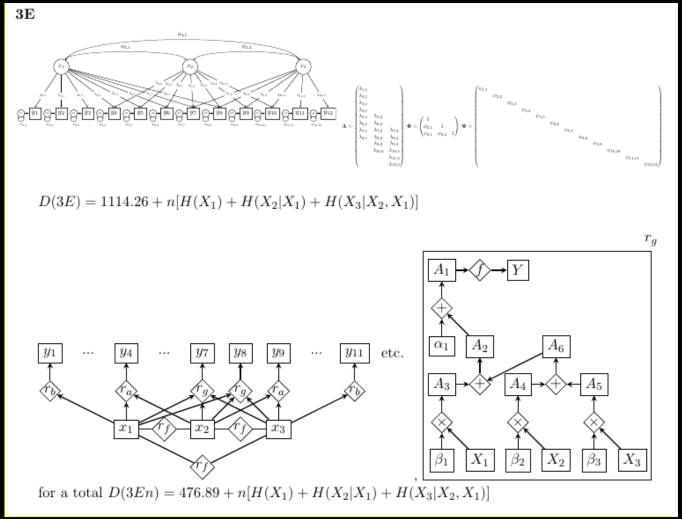
(Fanelli, Tan, Amaral & Neves, in prep)

2) D(τ) might
predict fitting propensity

preregistered test

(Fanelli & Bonifay, in prep)

preregistered test



1) generated N=**20** models, all with 36 parameters

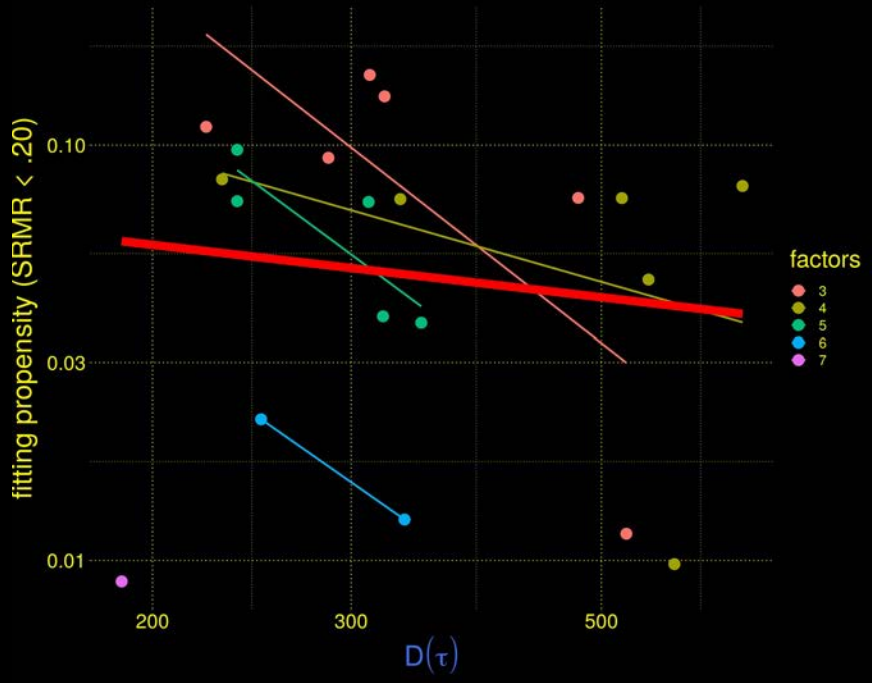
2) derived D(τ), predicted their **fitting propensity**

3) tested them on 20,000 random covariance matrices

(Fanelli & Bonifay, in prep)

D(τ) uniquely reflects model complexity

(pre-registered test)



(Fanelli & Bonifay, in progress)

NO alternative theory explains this!

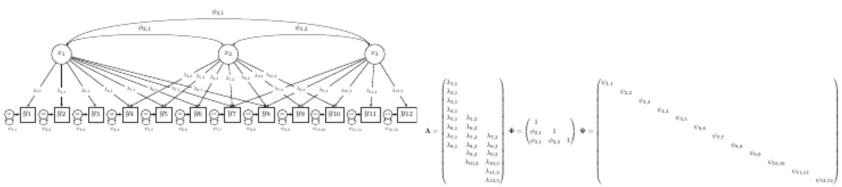
(pre-registered test)

multiple regression: Y=fitting propensity

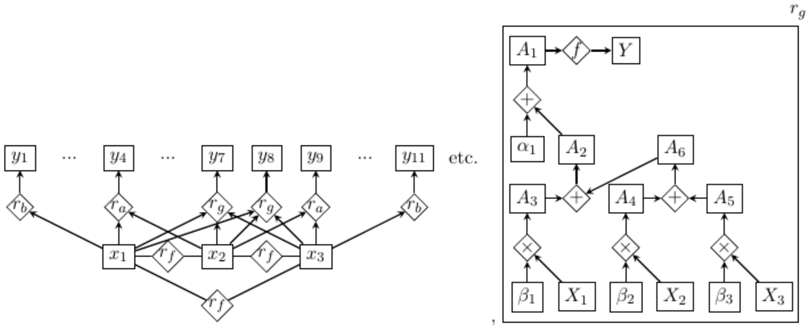
intercept	0.23 [0.16; 0.31] ***
$D(\tau)$	-1.46 [-2.43; -0.48] **
n. factors	-2.71 [-3.86; -1.55] ***

(Fanelli & Bonifay, in progress)

3E

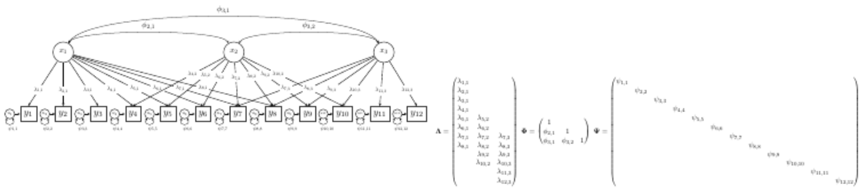


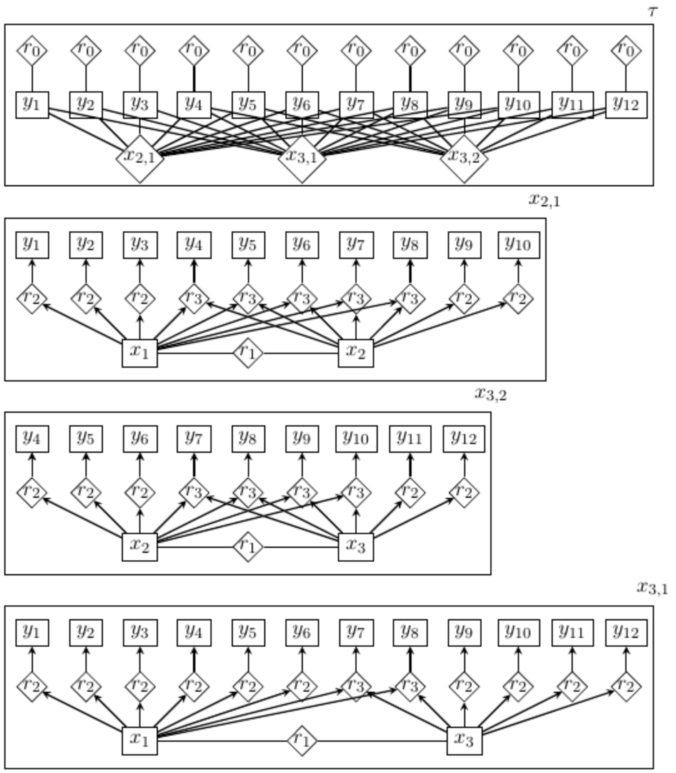
$$D(3E) = 1114.26 + n[H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1)]$$

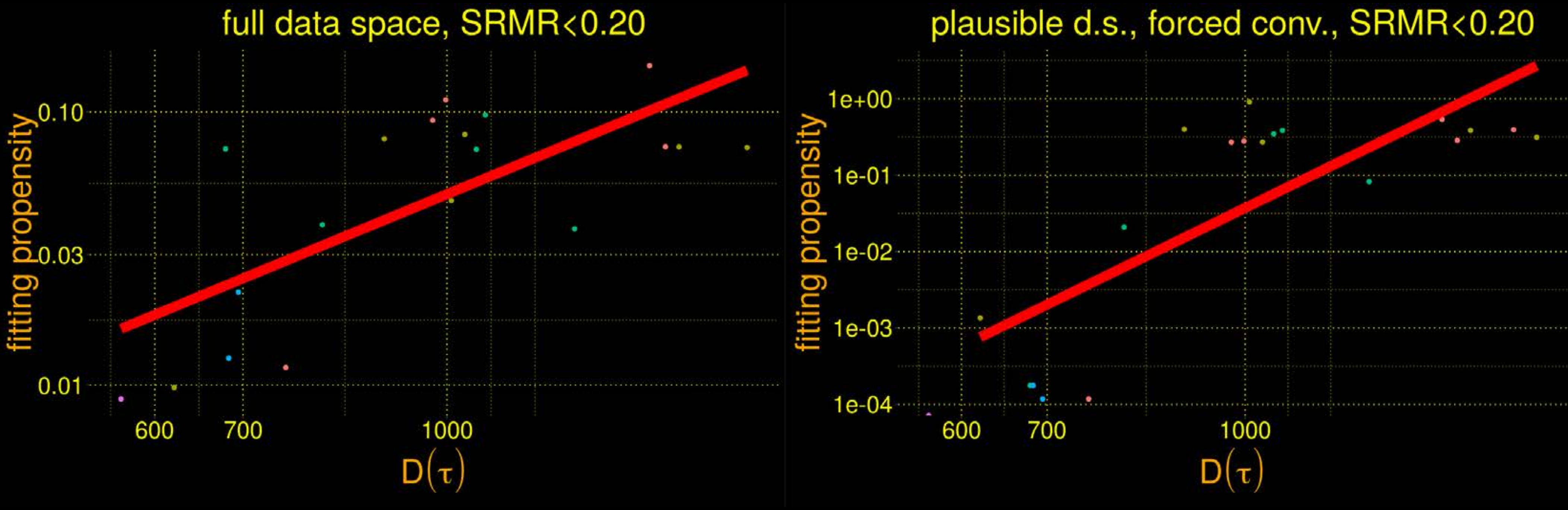


for a total $D(3En) = 476.89 + n[H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1)]$

3E







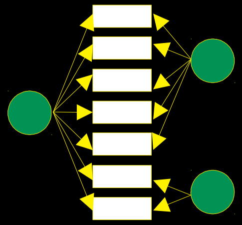
Spearman's ρ = 0.64, P<0.002

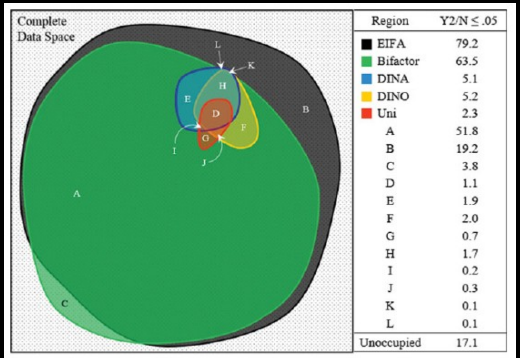
Spearman's ρ = 0.73, P<0.001

how K improves theory testing

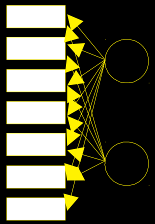
example of application

Bifactor
confirmatory
(theoretical)
20 parameters





(Bonifay and Cai 2017)

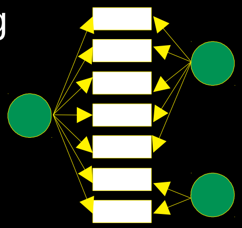


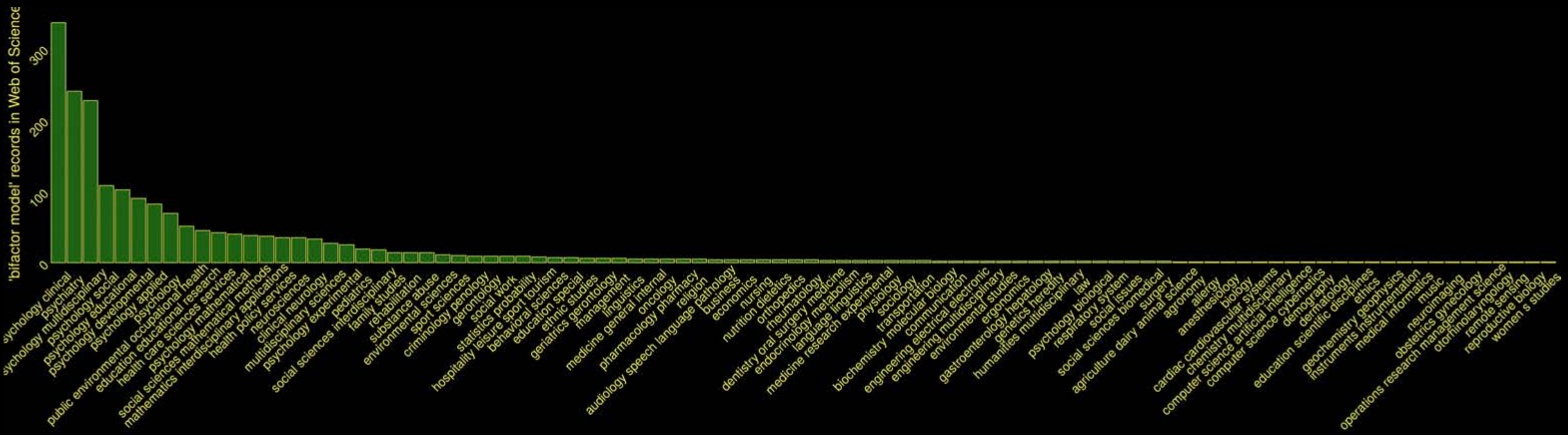
EIFA:
exploratory
(a-theoretical)
20 parameters

(Fanelli & Bonifay, in progress)

Bifactor widely used to “test” theories

theories invoking general factor (e.g. IQ, stress, psychosis)

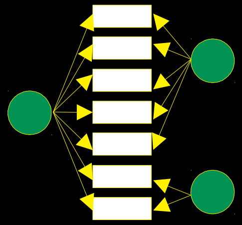


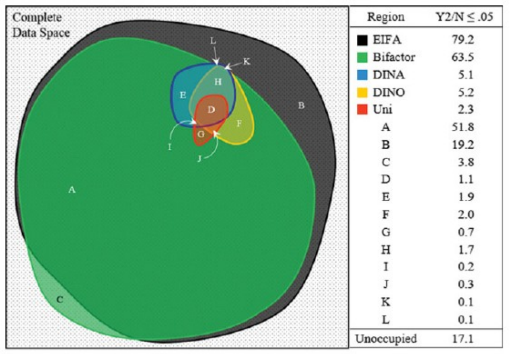


when is a theory *actually* supported?

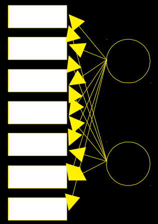
example of application

Bifactor
confirmatory
(theoretical)
20 parameters





(Bonifay and Cai 2017)



EIFA:
exploratory
(a-theoretical)
20 parameters

~~P<0.05~~

K(Bifactor) ≥ K(EIFA)

"The [bifactor-encoded] theory
is specifically supported
by the data"

(Fanelli & Bonifay, in progress)

Summary of this talk:

1) what's missing in the current paradigm?
   - we pretend complexity is irrelevant
2) what might a new paradigm look like?
   - measuring D(τ), integrating/penalizing with K
3) evidence in support of this proposal?
   - increasingly promising
   - any alternative, better suggestions?