James Evans

KNOWLEDGE LAB

The Geometry of Culture & Society

Analyzing meaning & connection through embeddings of text and networks

KNOWLEDGE LAB

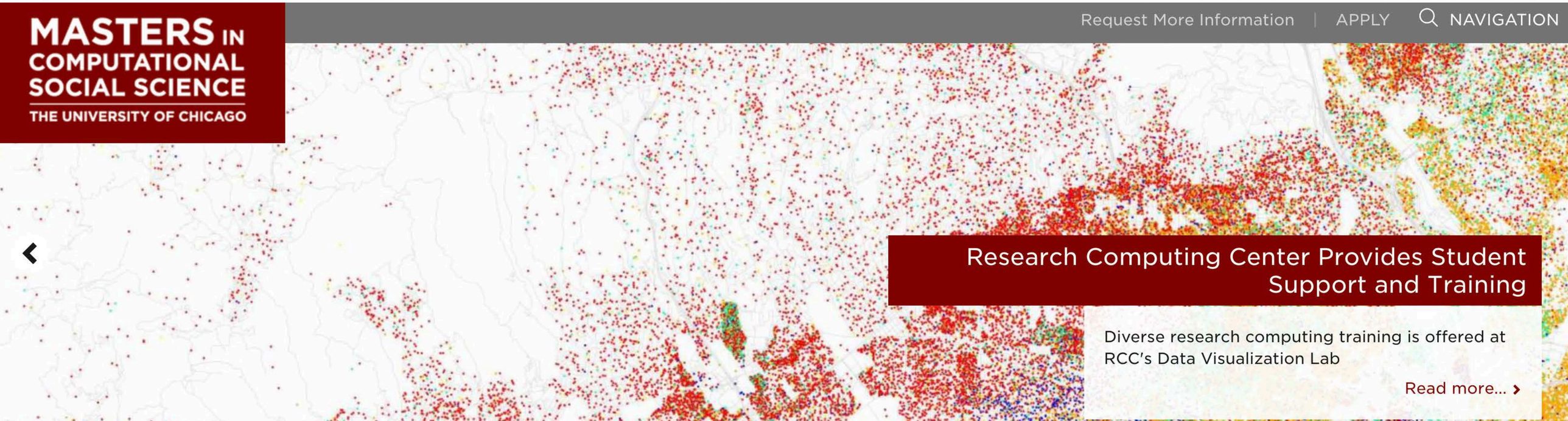
Big Data, Machine Learning and Intelligent Crowdsourcing enables us to:

1. Represent
2. Understand
3. Transform…the scientific and scholarly process

computationally enhanced

Knowledge^2

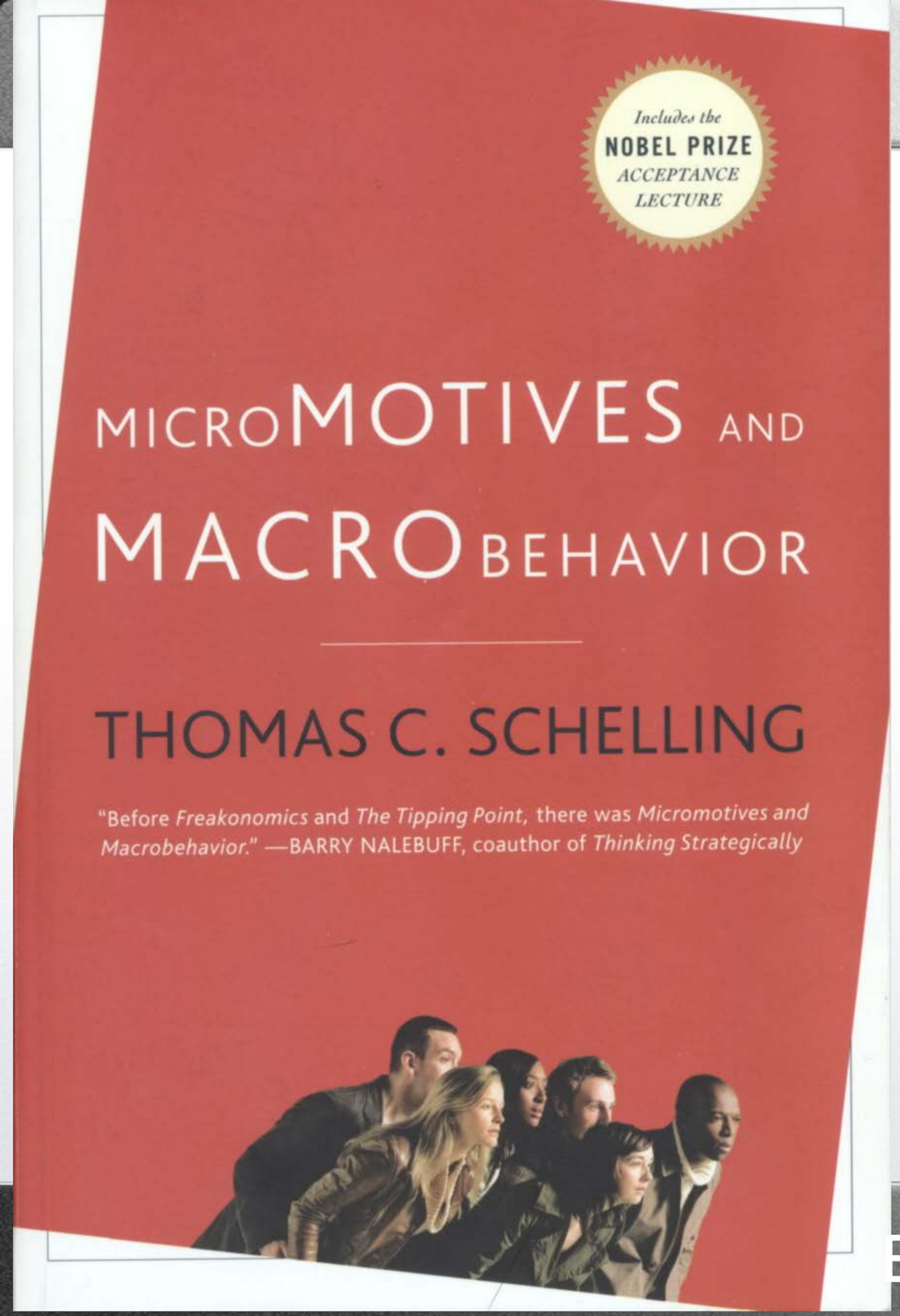
Computational Social Science



What is Computational Social Science?

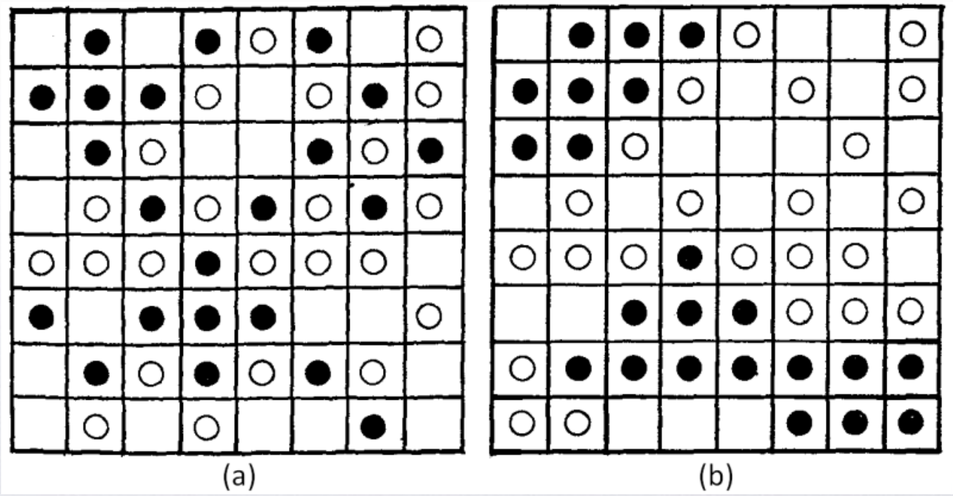
Using computers to generate data, discover patterns or generate and test explanations that you could not have without them.

…implies a shift in computationally enabled research designs, methods and theoretical standards



Before 2000

Computing consequences of theoretical assumptions…



E.g., Is rule set X sufficient to generate social world Y.

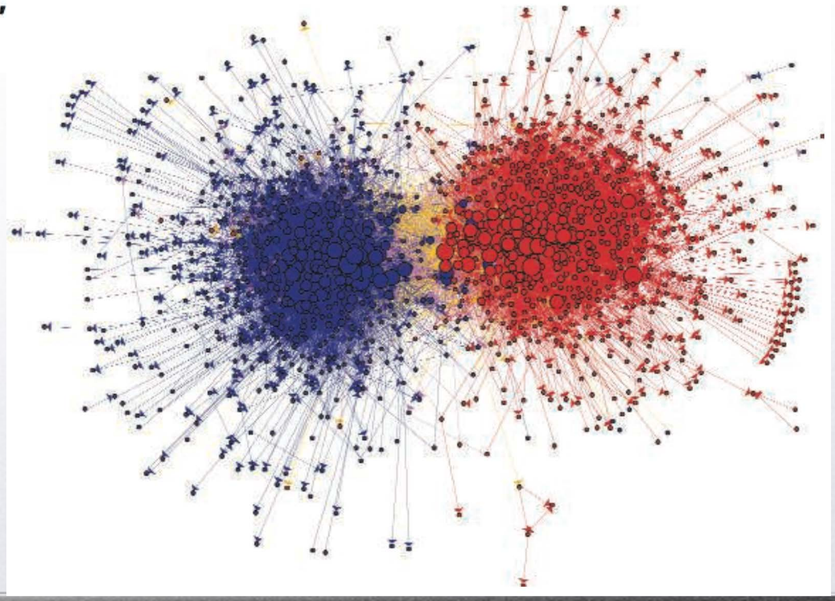
SOCIAL SCIENCE

After 2000 (relabeled 2009)

Computational Social Science

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5]
Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3]
Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

DATA



High Throughput Observatories

Big Data

Theoretical Entailments

Generative Standard

for social scientific epistemology:

from
Necessary but
not Sufficient
to
Sufficient but
not Necessary

Plausible
Social Science
Fiction

If you didn't grow it, you didn't explain it

NOT Text as Data

Text as Simulation

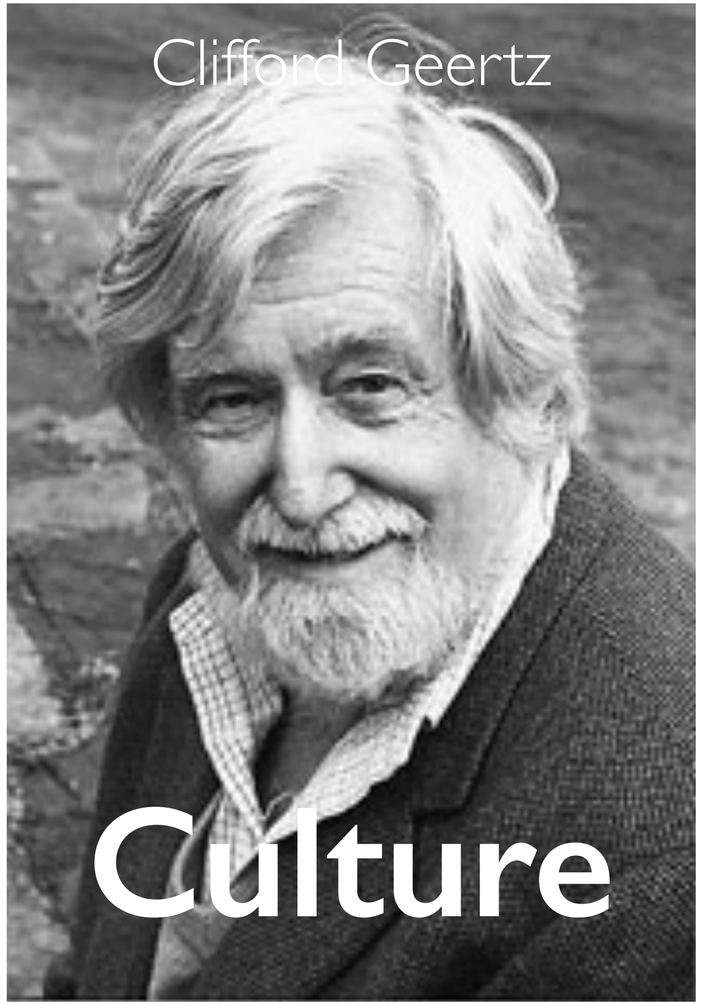
Digital Doubles

Simulated Subjects

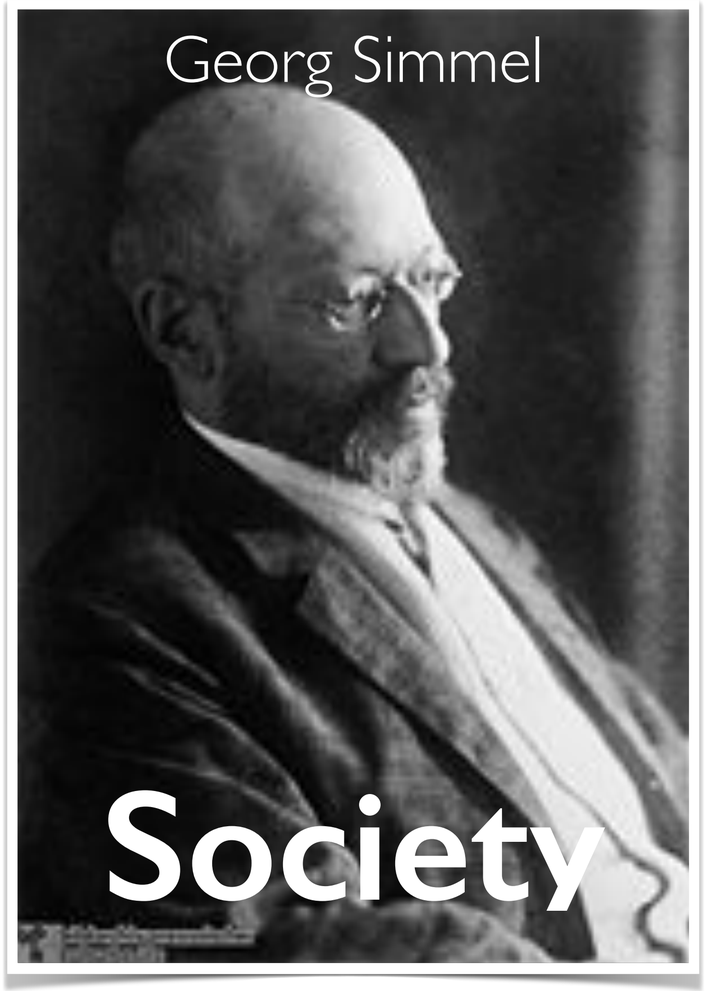
In Silico Societies

Discriminative Models

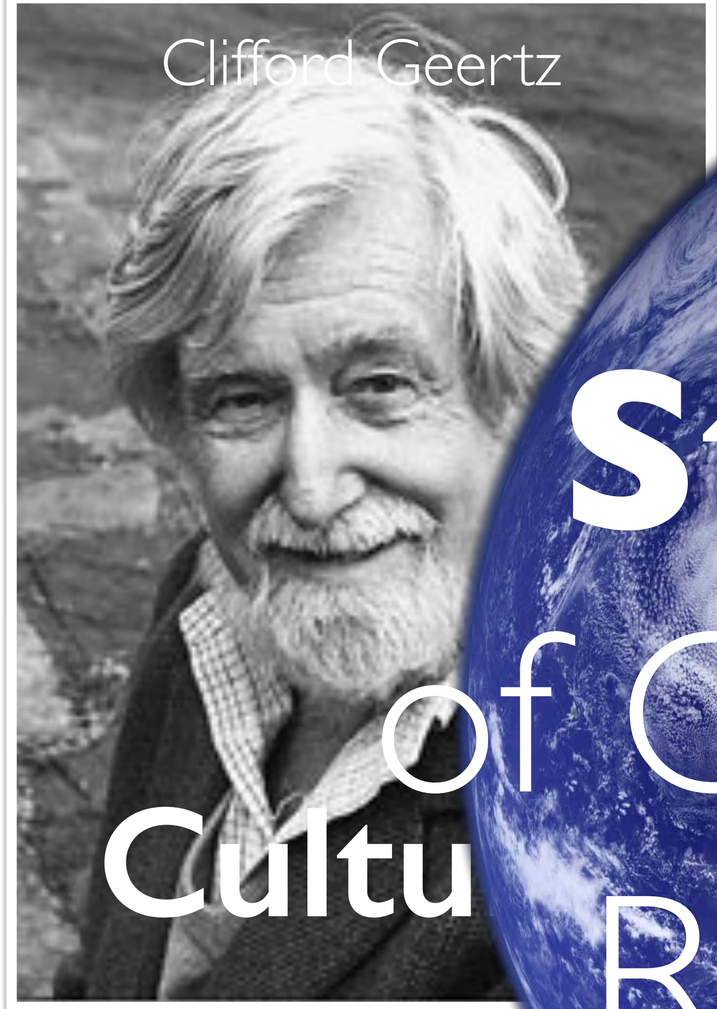
Generative Models



"All the **forms** of association by which a mere sum of separate individuals are made into a 'society'," whereby society is defined as a "higher unity," composed of individuals.

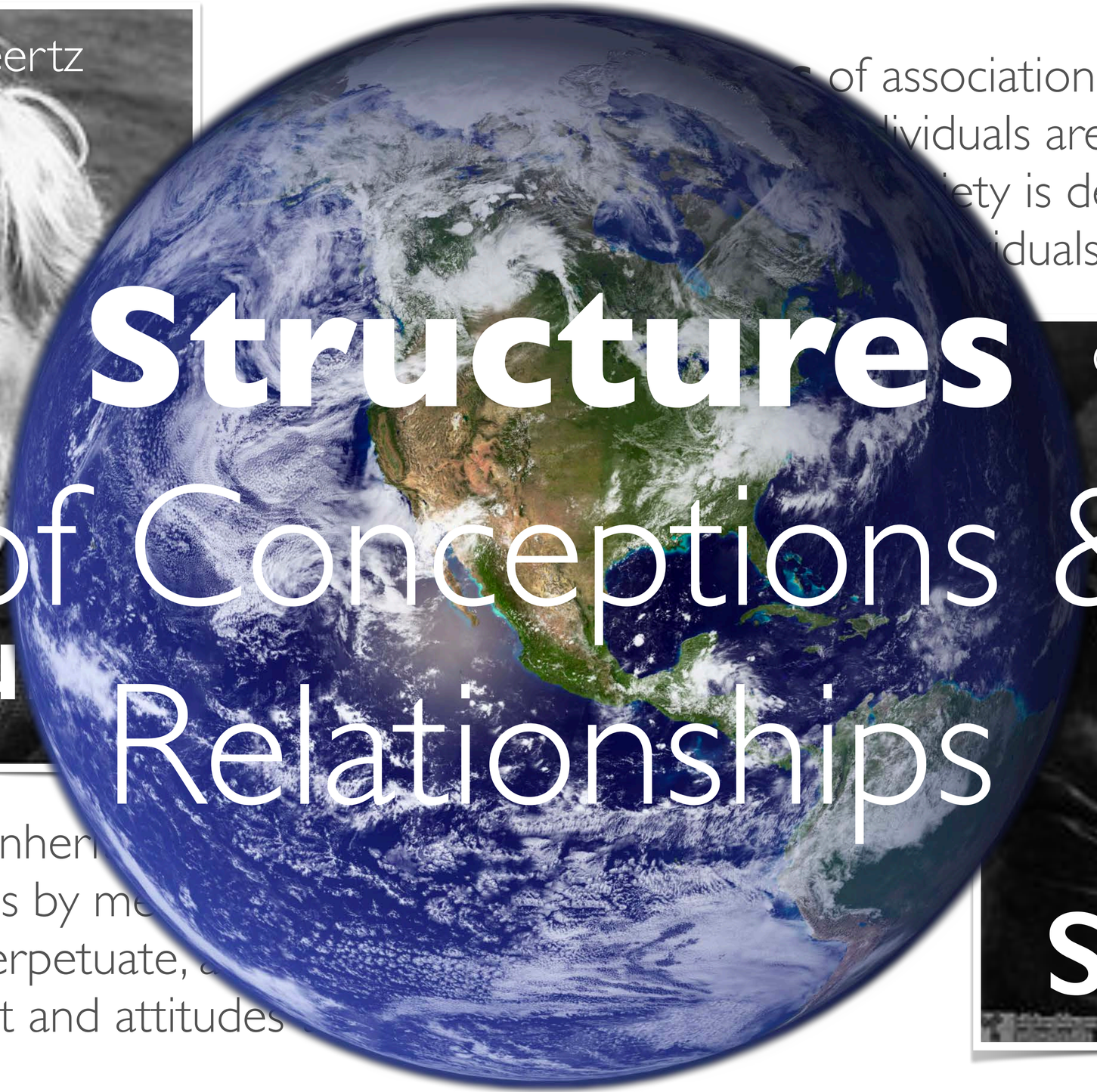


"A **system** of inherited conceptions expressed in symbolic forms by means of which men communicate, perpetuate, and develop their knowledge about and attitudes toward life."



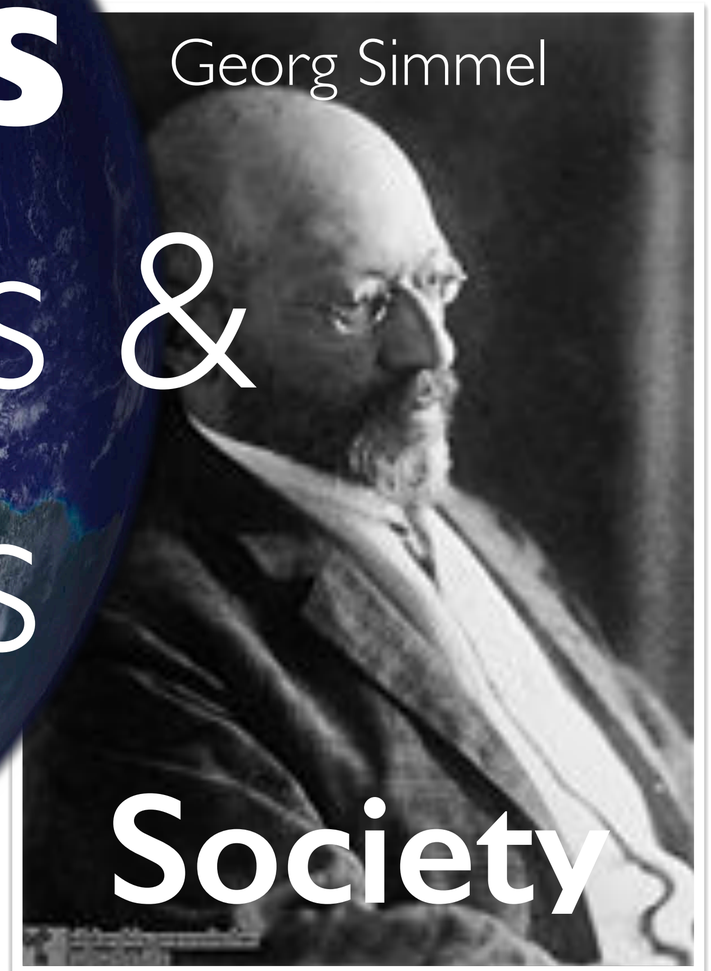
Cultu

"A **system** of inheri
in symbolic forms by me
communicate, perpetuate, a
knowledge about and attitudes



Structures of Conceptions & Relationships

s of association by which a mere
dividuals are made into a
iety is defined as a "higher
iduals.



Society

Structures

Clusters

Spaces

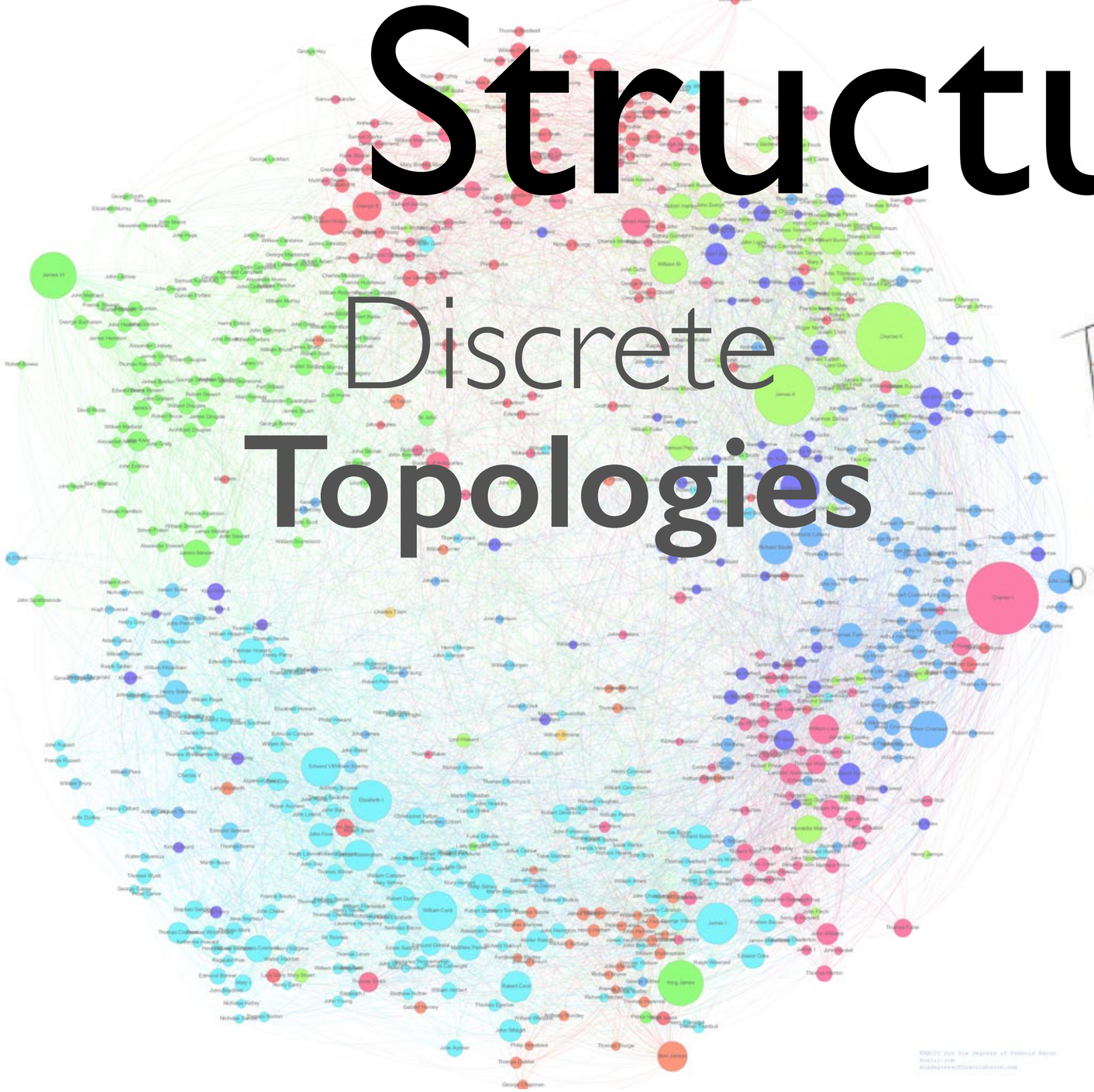
Orders

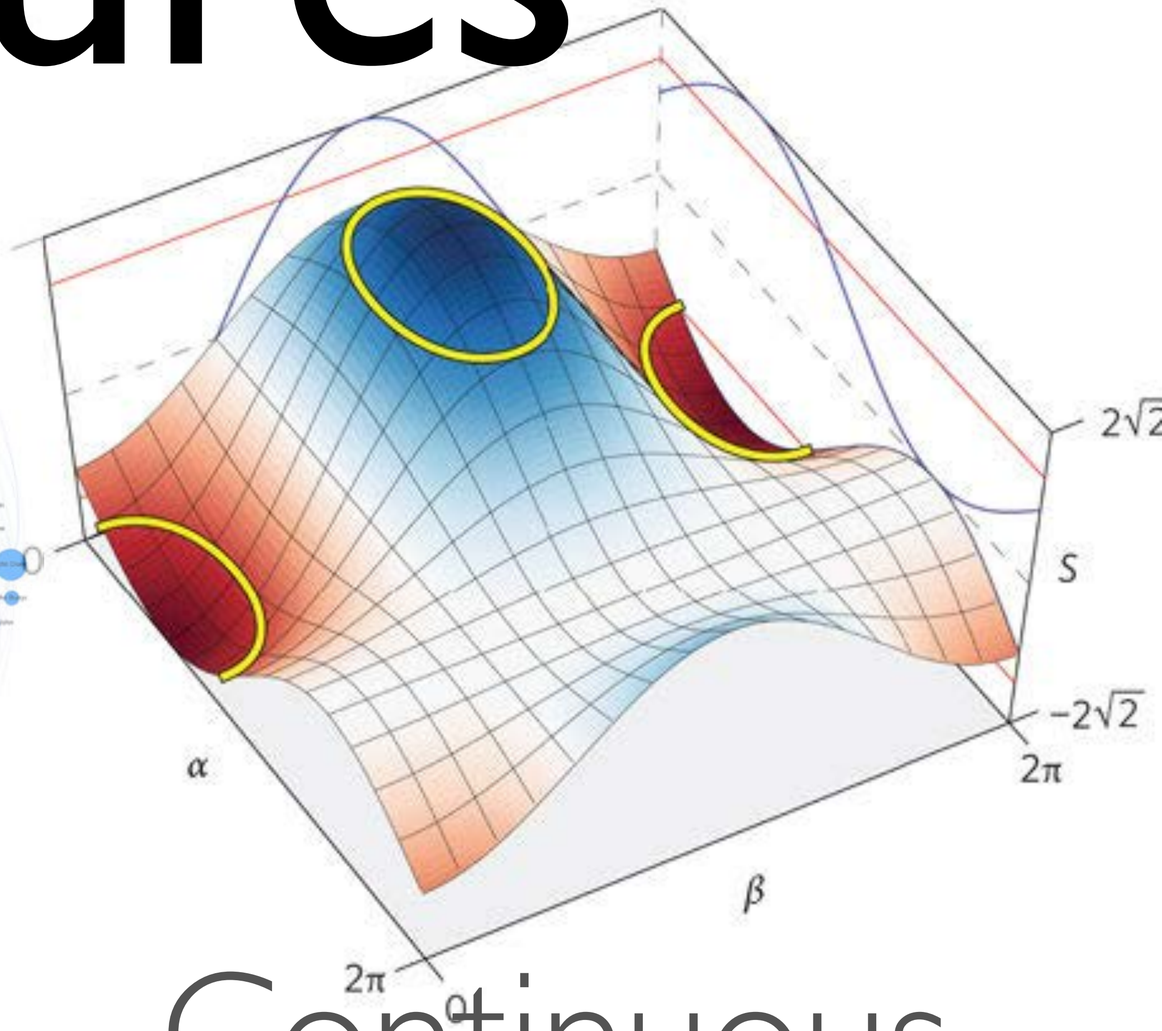
(Hyper)graphs

Hierarchies

Structures

Discrete
Topologies

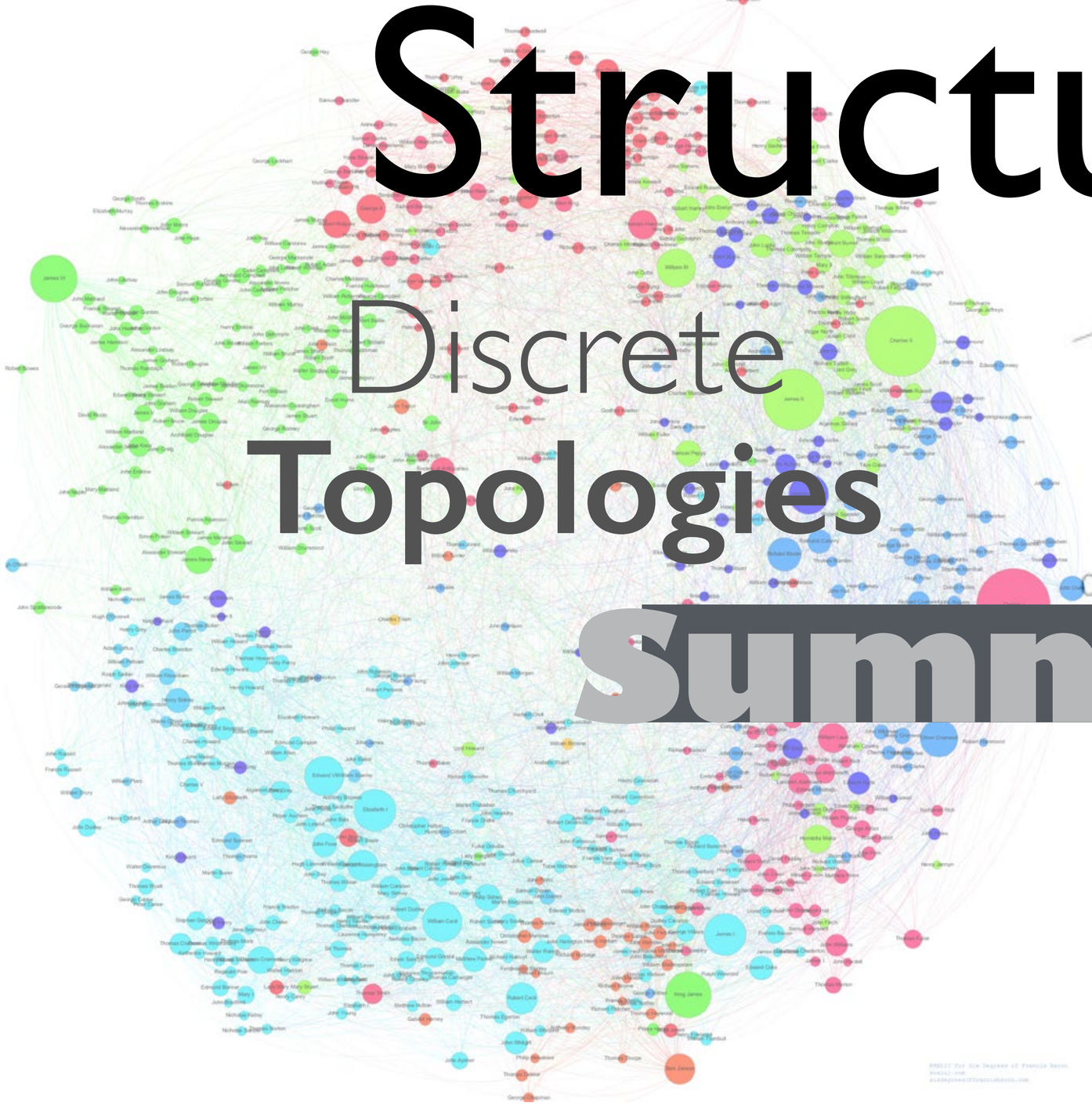




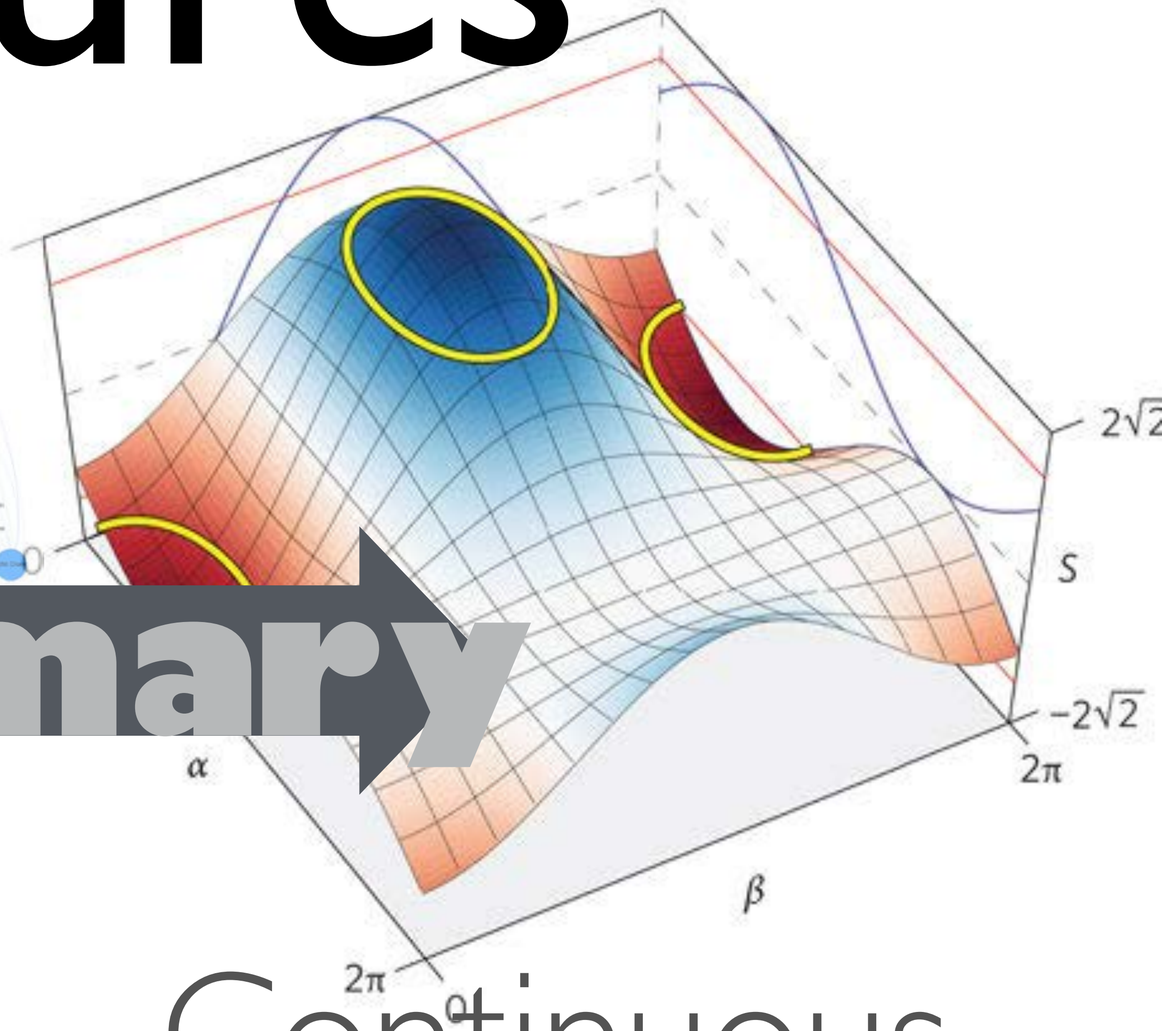
Continuous
Geometries

Structures

Discrete
Topologies

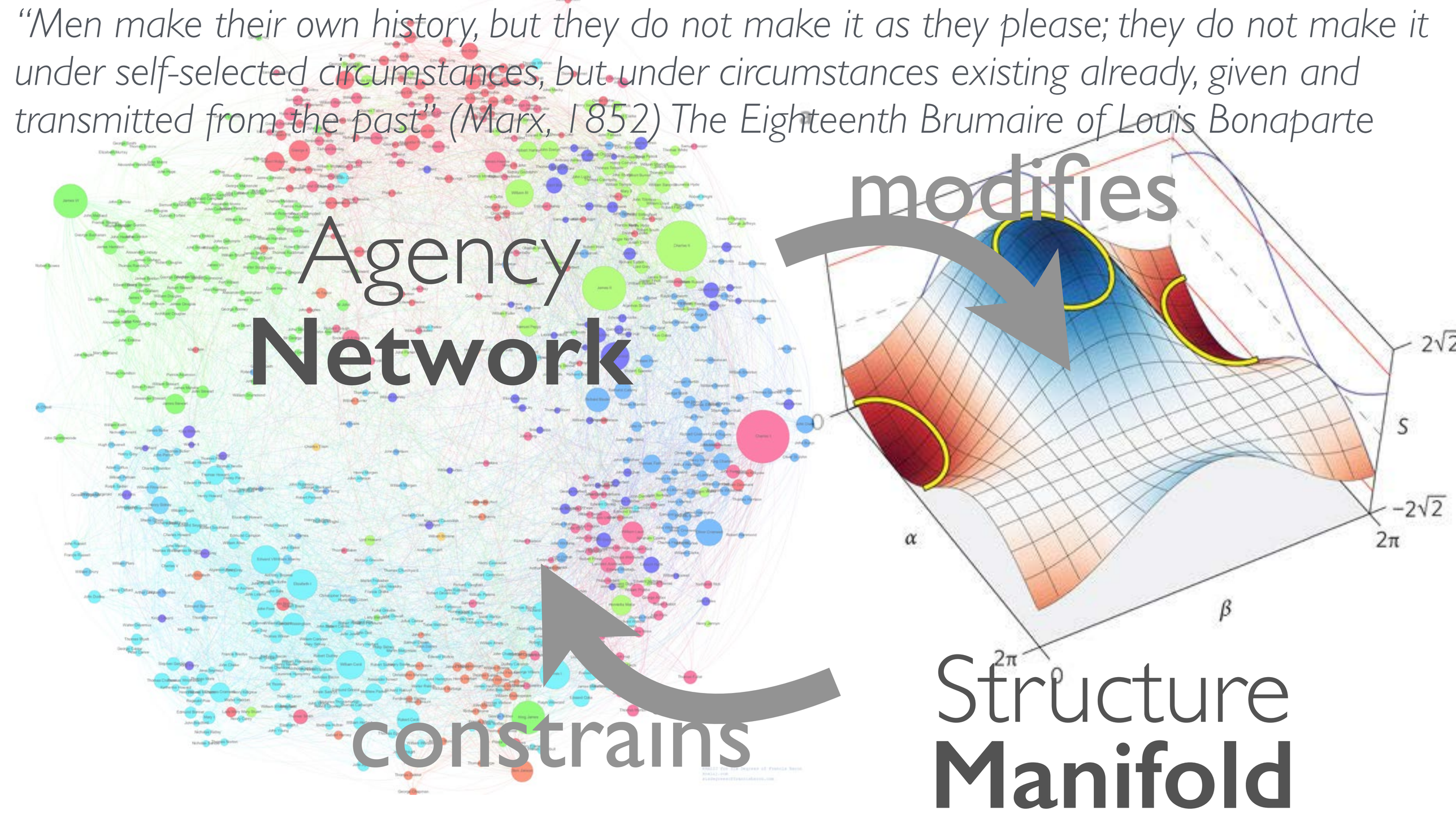


Summary



Continuous
Geometries

"Men make their own history, but they do not make it as they please; they do not make it under self-selected circumstances, but under circumstances existing already, given and transmitted from the past" (Marx, 1852) The Eighteenth Brumaire of Louis Bonaparte



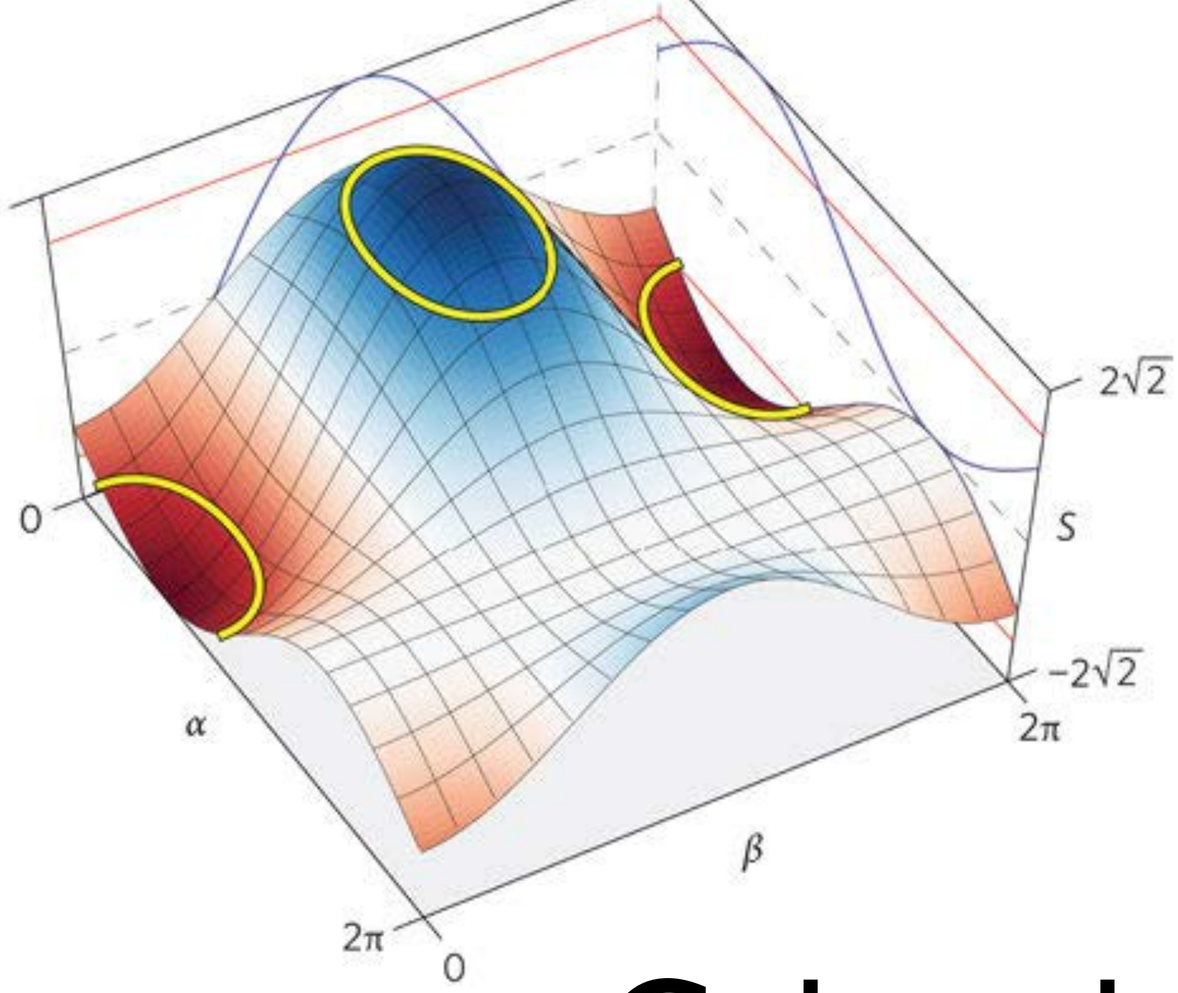
How do we explore the complex relationship between **culture** & **social structures**?

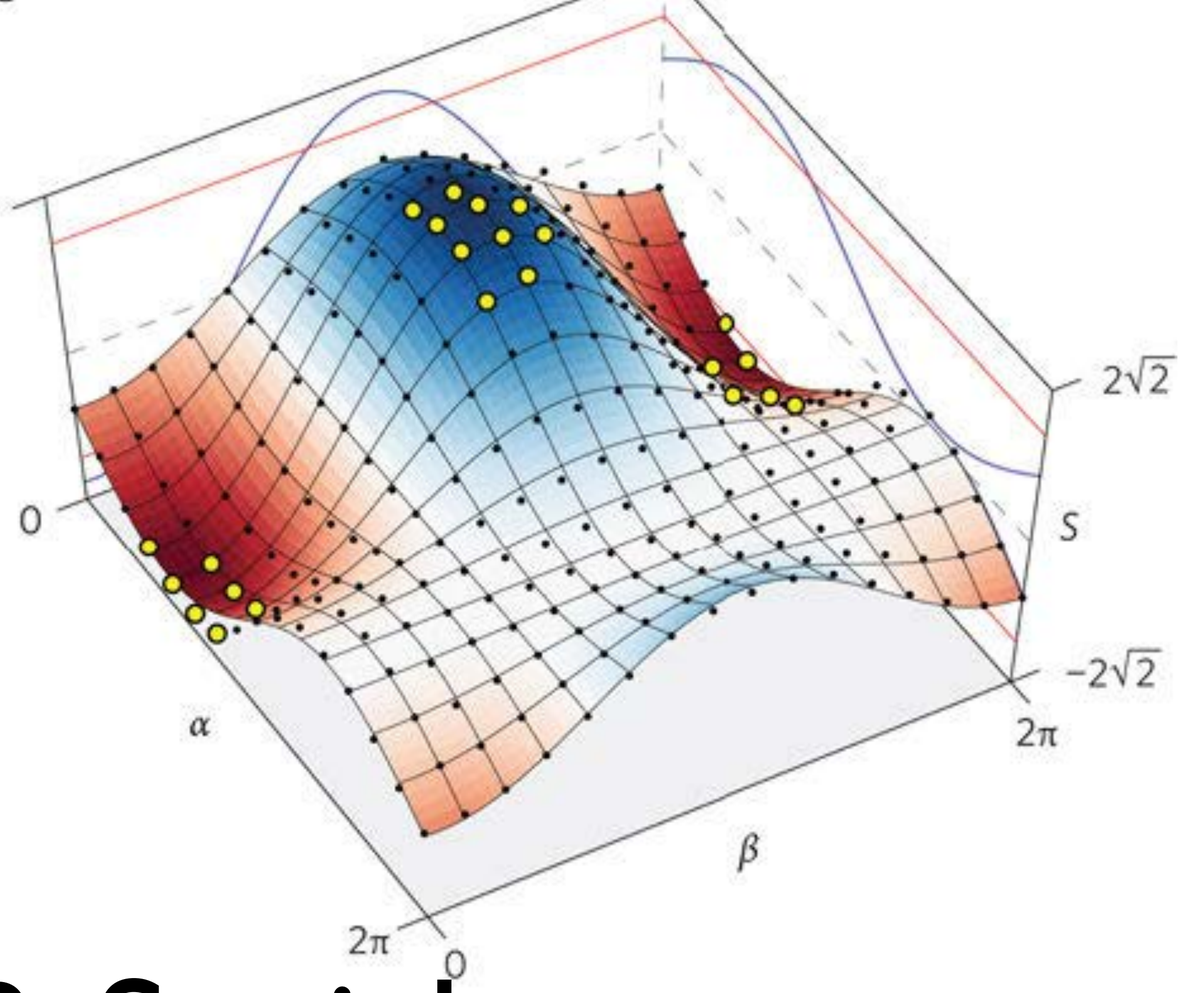
same register

Either:

Cultural & Social Clusters / Networks

Discrete, simple **structures**…that reduce to the lowest resolution





Cultural & Social Fields

Continuous, complex **geometries**…that elevate to the highest resolution

Continuous Meaning Spaces

J. R. Firth:

"You shall know a word by the company it keeps"

Wittgenstein:

"For a large class of cases of the employment of the word 'meaning'—though not for all—this way can be explained in this way: the meaning of a word is its use in the language" (PI 43)

Overlapping contexts create quasi-continuous metric

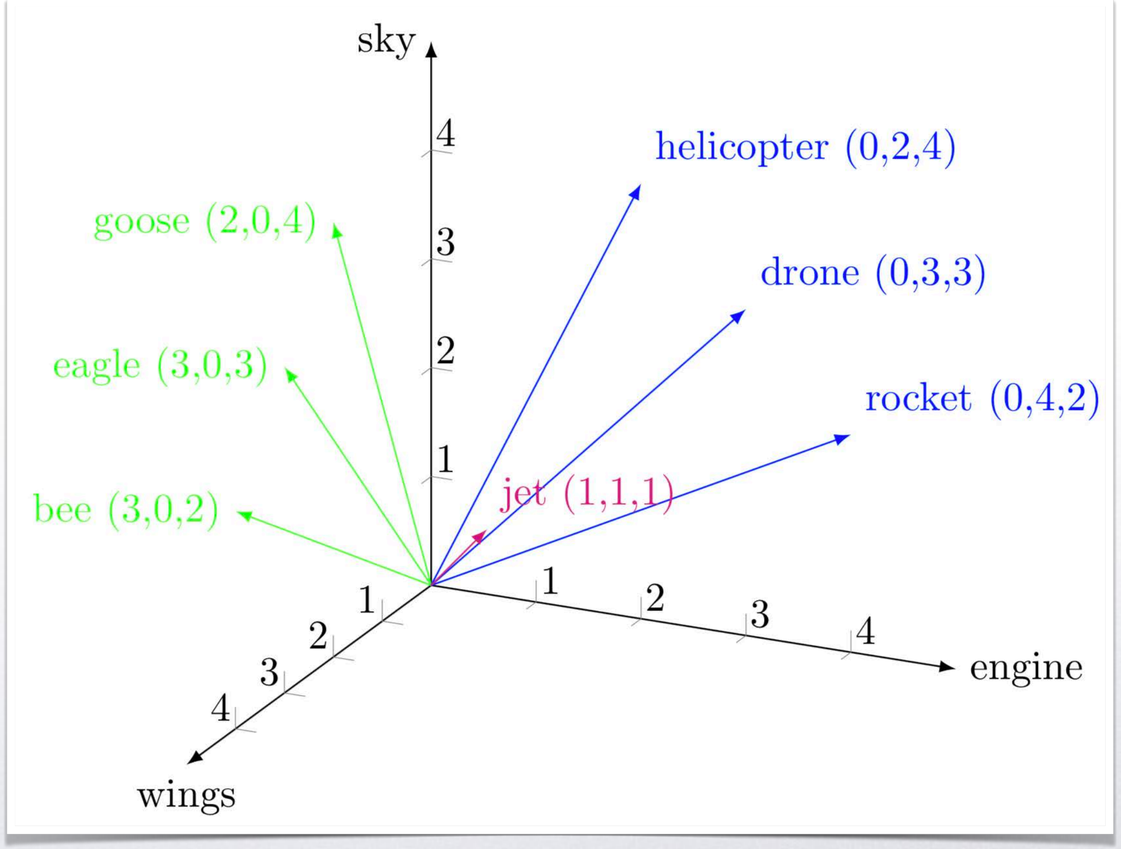
Continuous Meaning Spaces

A bottle of tesgu¨ino is on the table.
Everybody likes tesgu¨ino.
Tesgu¨ino makes you drunk.
We make tesgu¨ino out of corn.

What can be inferred?

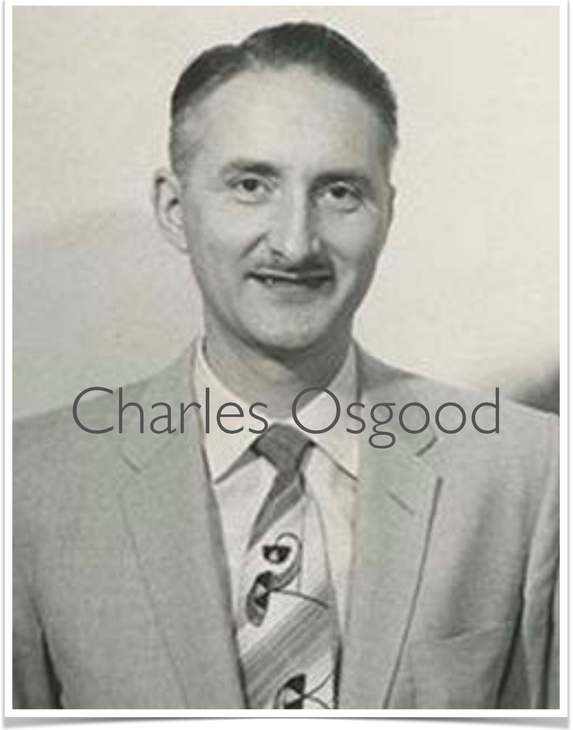
Distributional Hypothesis

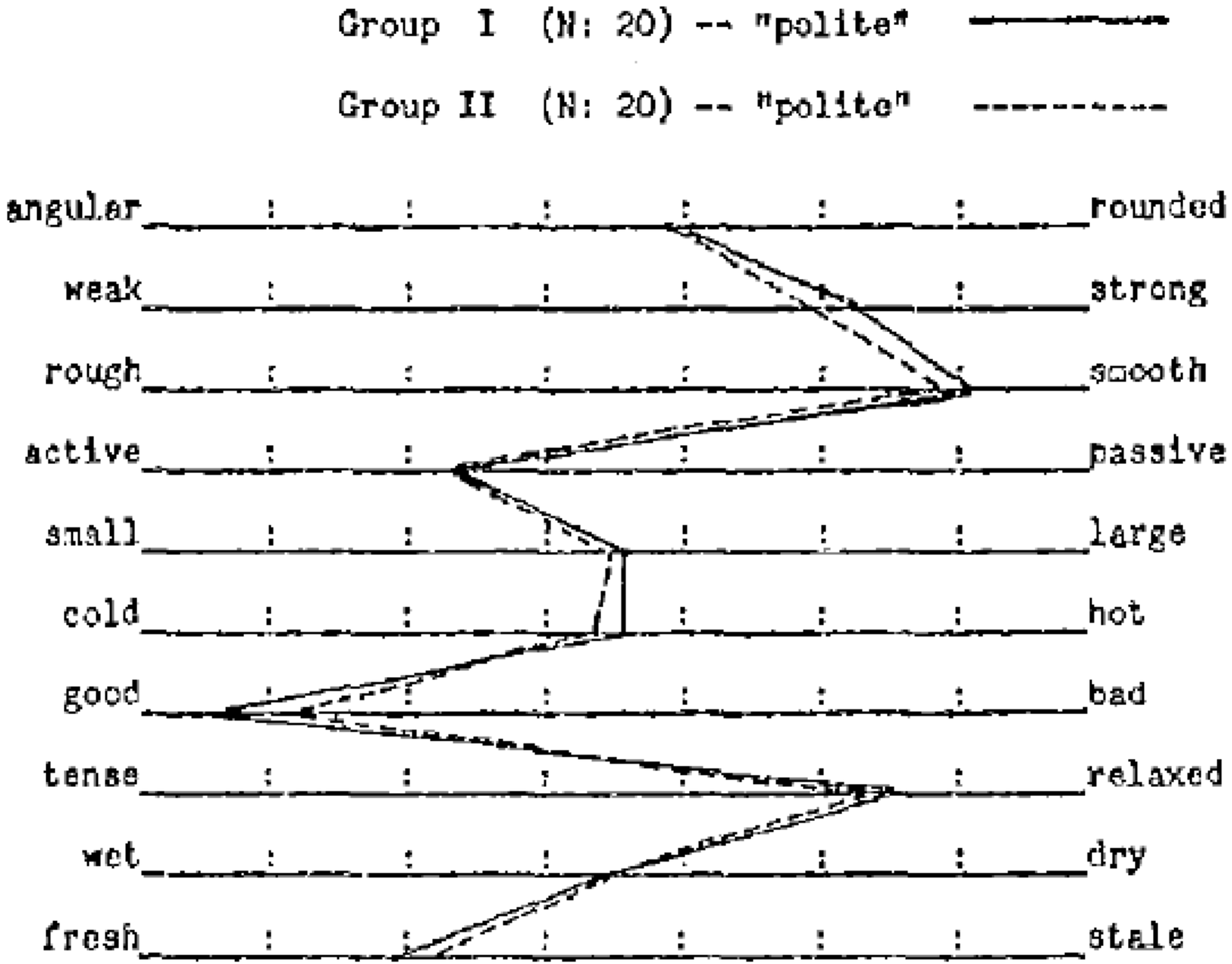
Similarity in how words are distributed suggests similarity in what they mean…with the amount of meaning difference between two words "corresponding roughly to the amount of difference in their environments" (Zellig Harris, 1954).



Osgood's Semantic Measurement

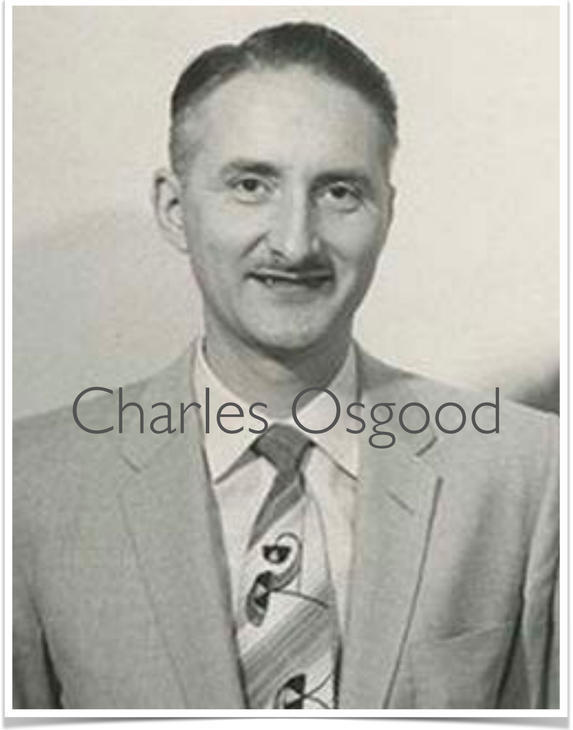
Behavioral theory of meaning as tracing the similarity of evoked response

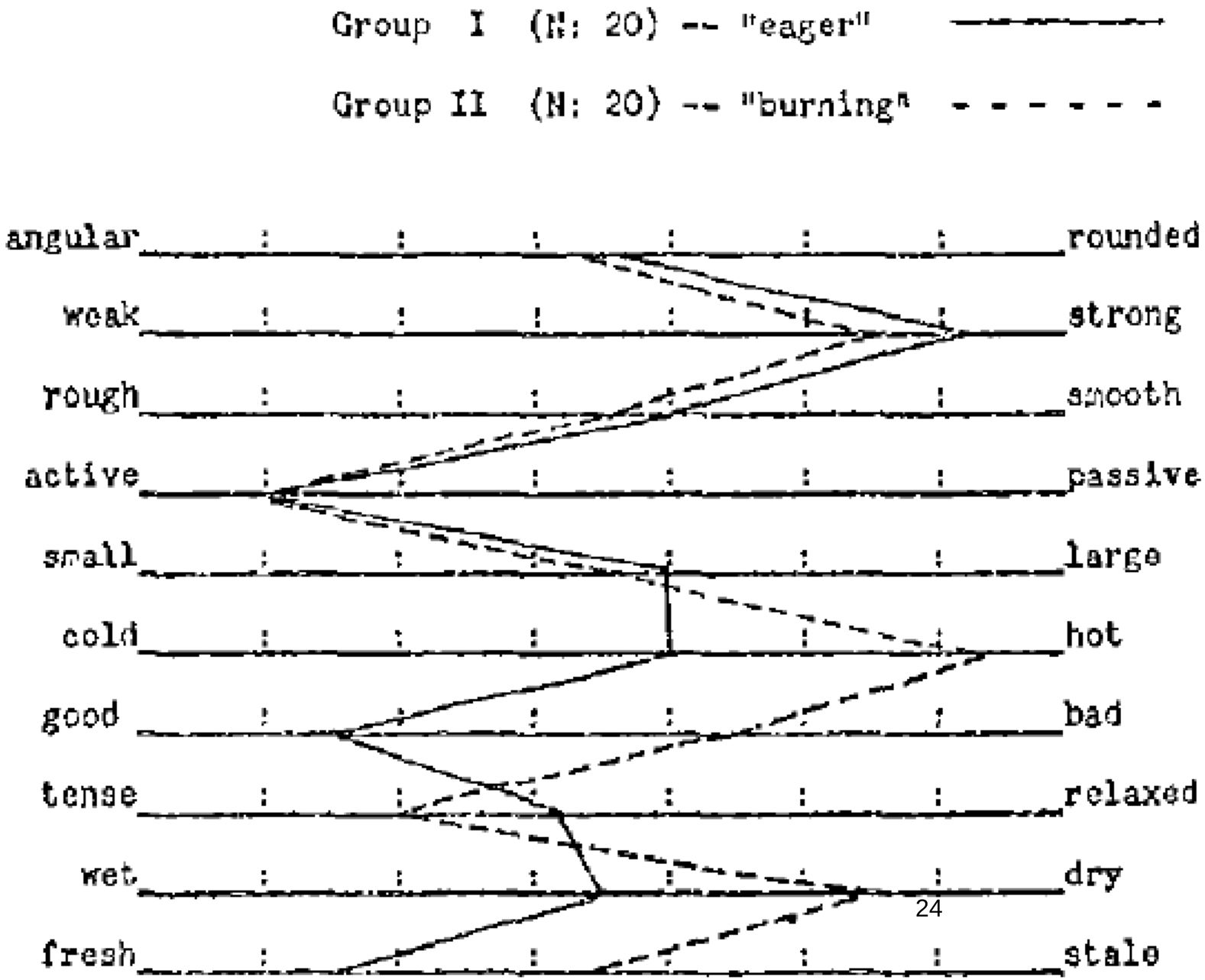




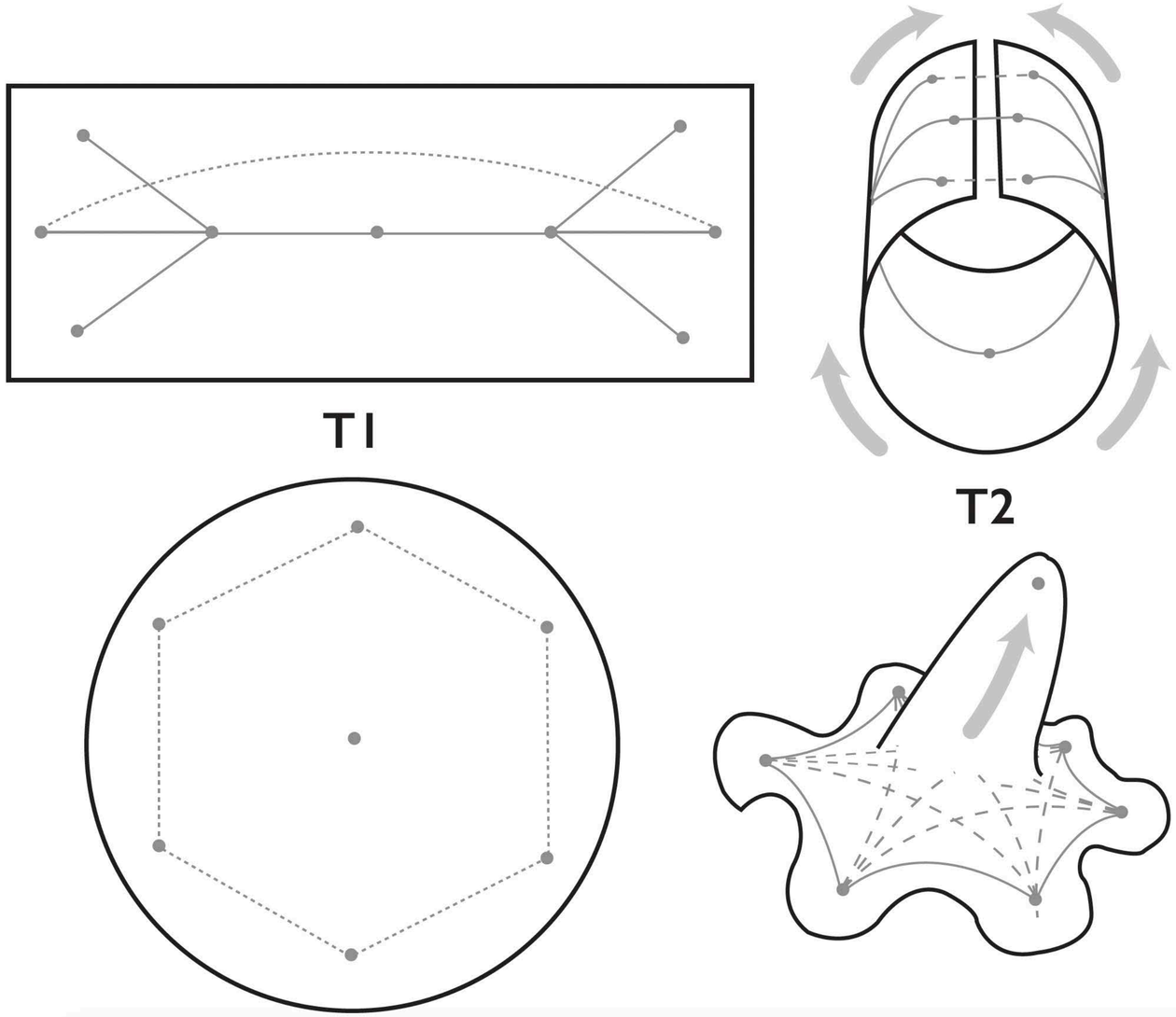
Osgood's Semantic Measurement

Behavioral theory of meaning as tracing the similarity of evoked response





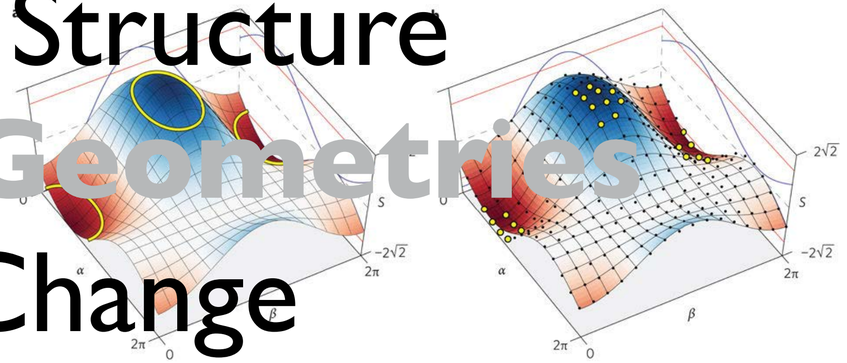
Continuous Social Spaces Predict Future Ties



Overlapping connections create a continuous space

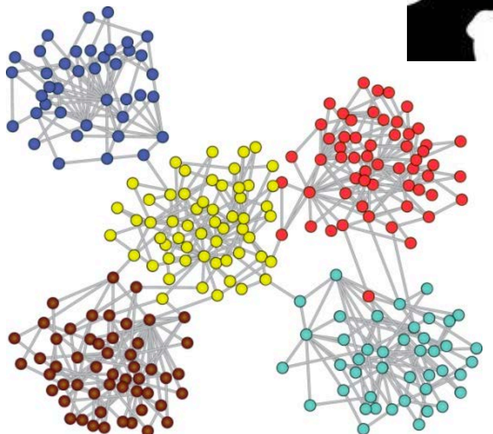
Chapters

1. Neural Networks Auto-Encode Structure
   ...into High Dimensional Geometries



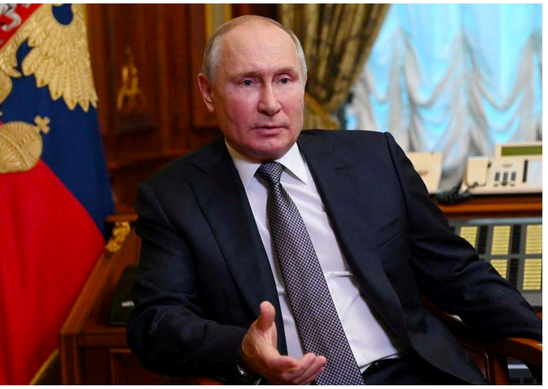
2. Cultural Structure, Variation & Change
   Conception and Communication

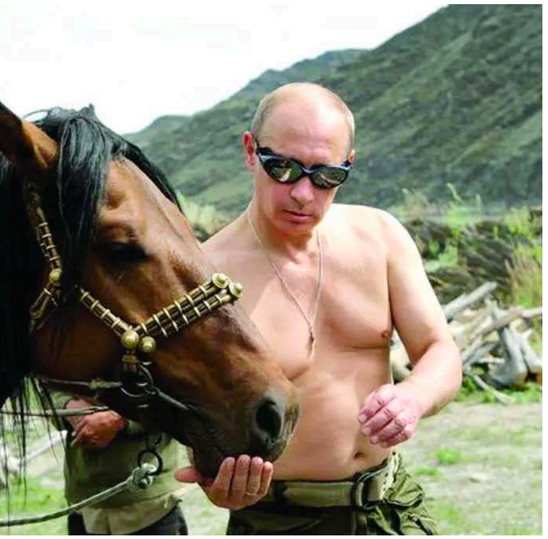
3. Social Structure, Change & Variation
   Network Evolution

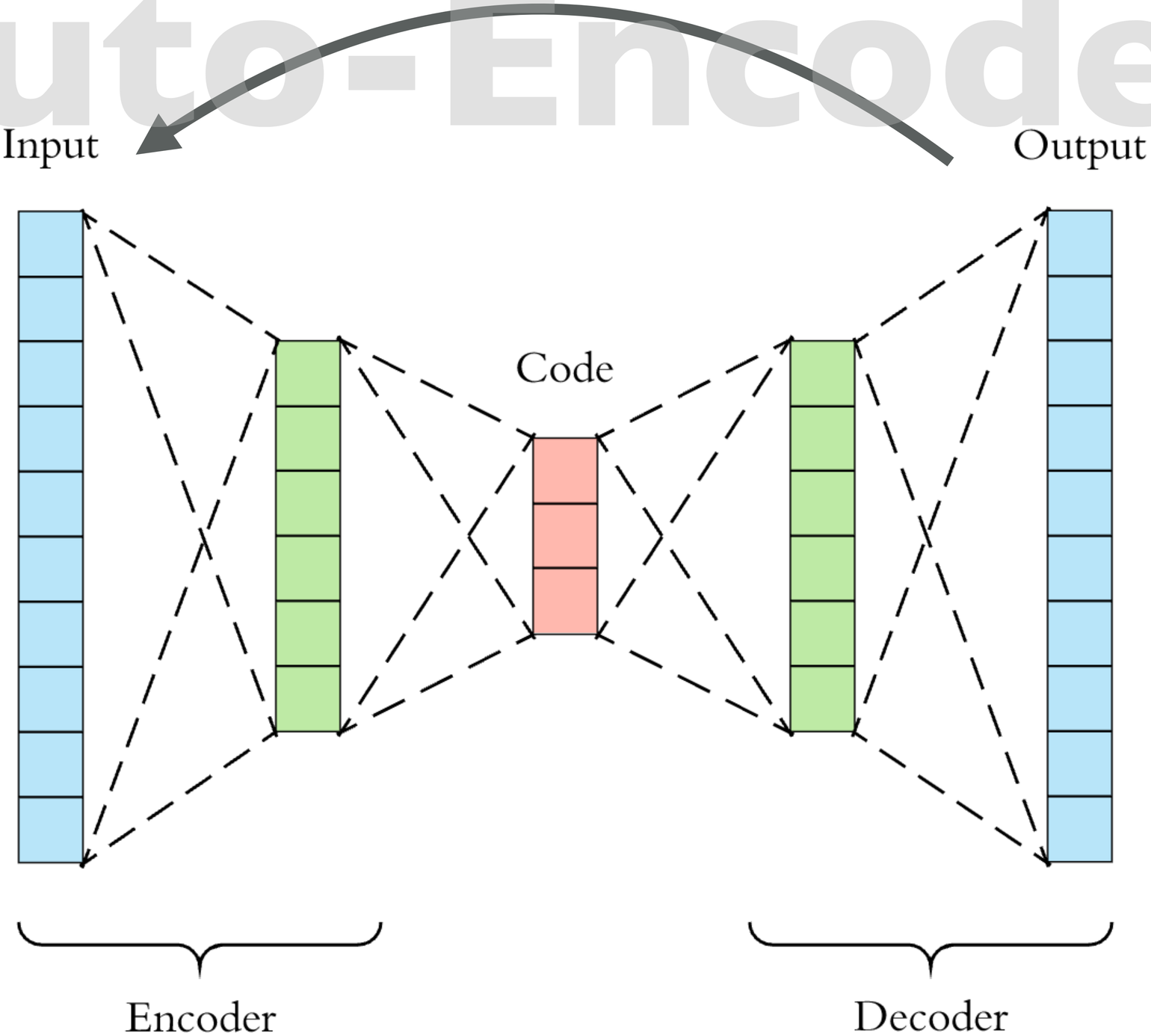


4. Culture & Society **Together**

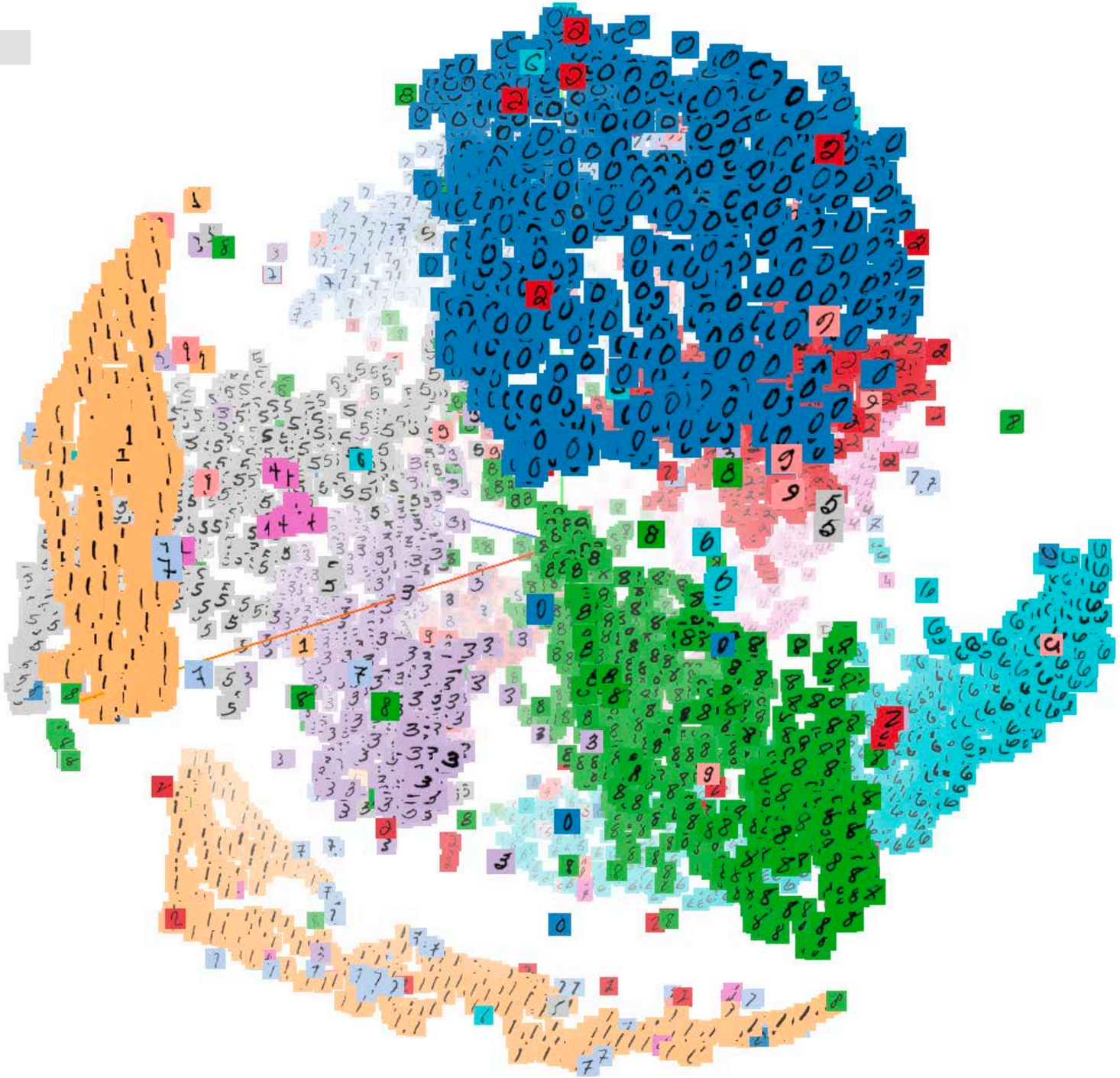
Auto-Encoder





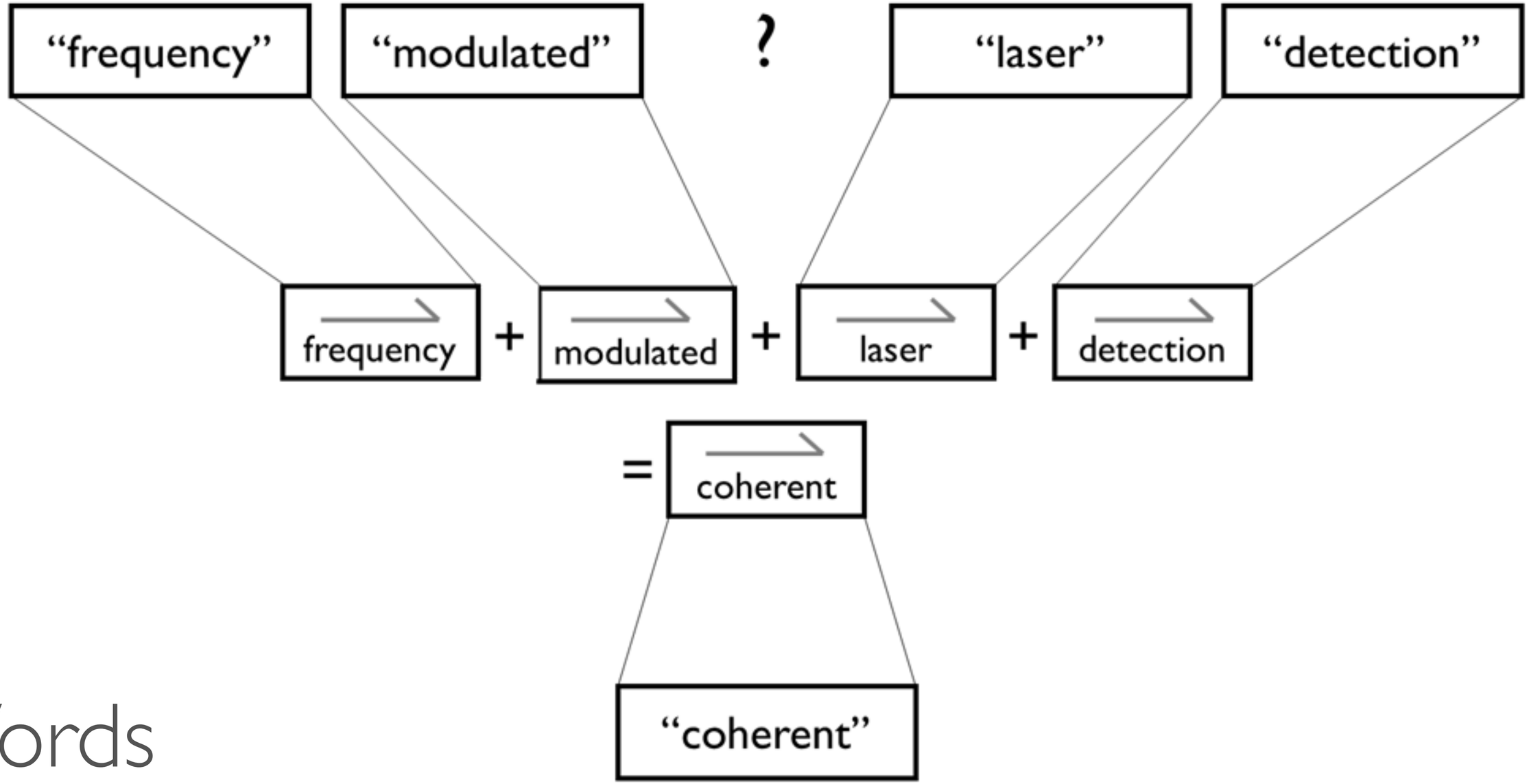


MNIST



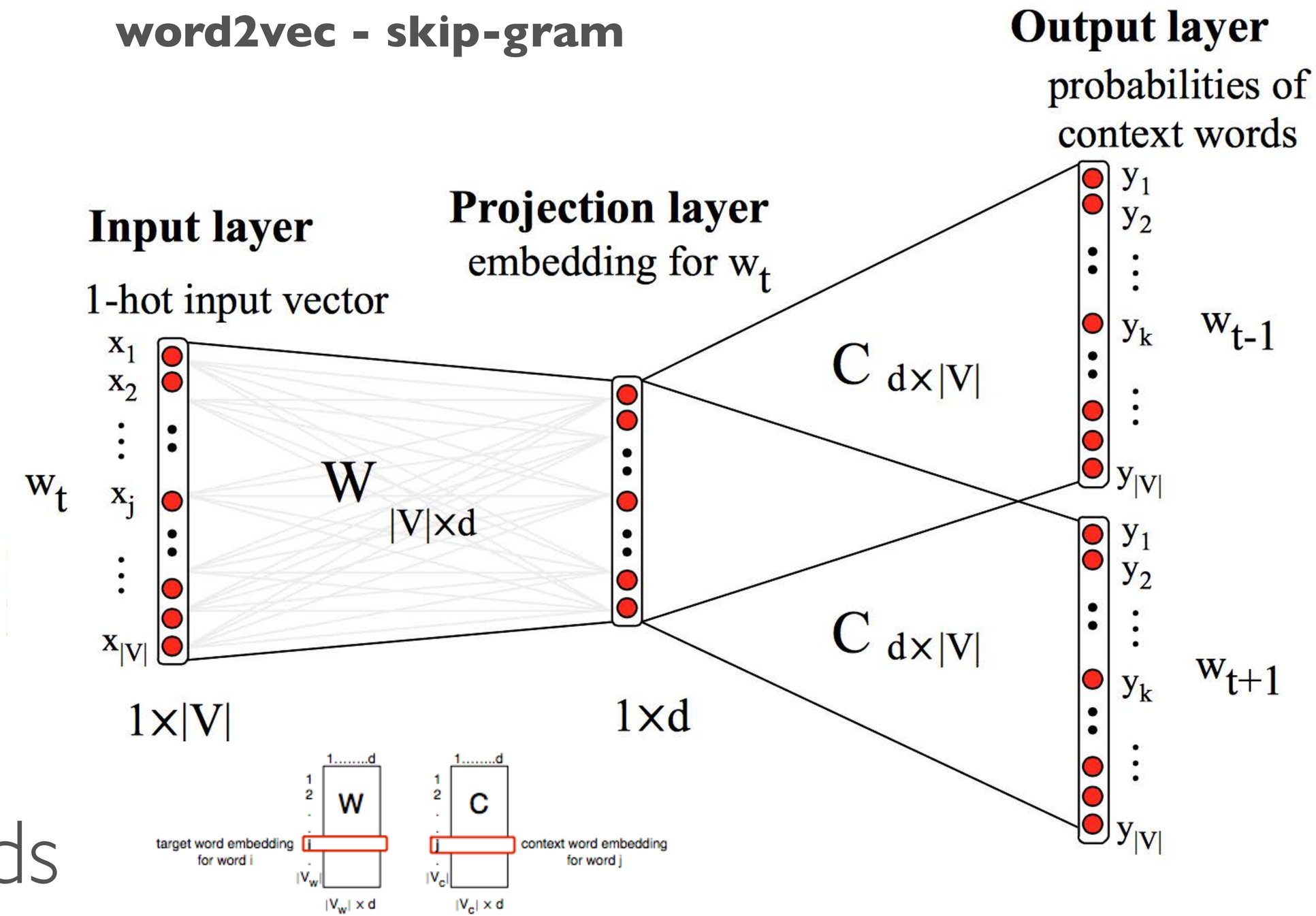
Pixels

"A **frequency modulated coherent laser detection** and ranging system…"



Words

word2vec - skip-gram



Words

Anonymized user id | Track identifier | Identifier of algorithmically guided listening

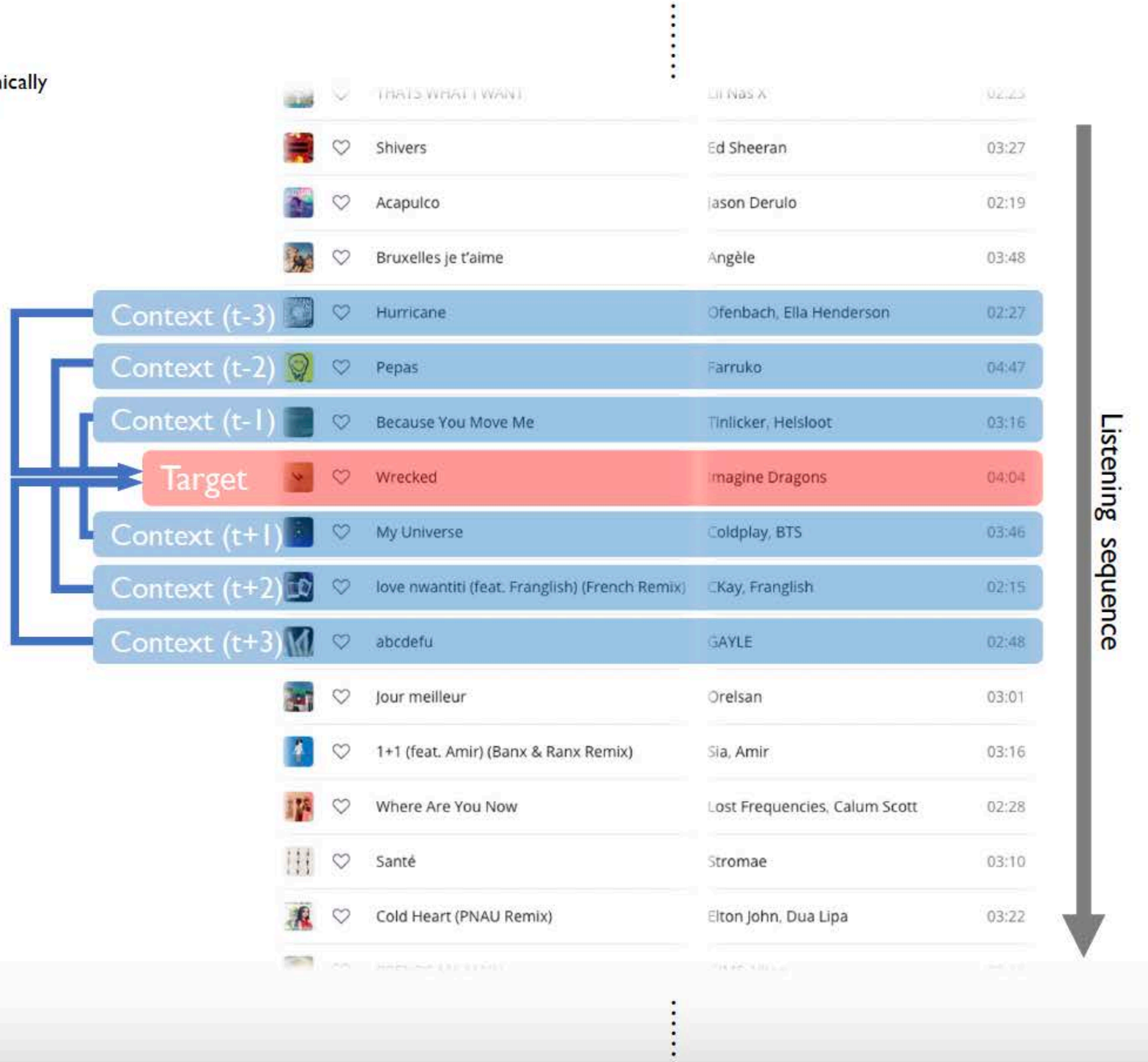
519d4bd750a4aeb2f4d9ed607d1a2d6908a1d1e7	67539452	guided_auto
519d4bd750a4aeb2f4d9ed607d1a2d6908a1d1e7	936899	guided_auto
519d4bd750a4aeb2f4d9ed607d1a2d6908a1d1e7	4301418	guided_auto
40e268fb9f2d008252ac25ec5debde873ac13223	72677275	mod
40e268fb9f2d008252ac25ec5debde873ac13223	422438302	mod

(Excerpted from our data)

Device identifier | Location identifier

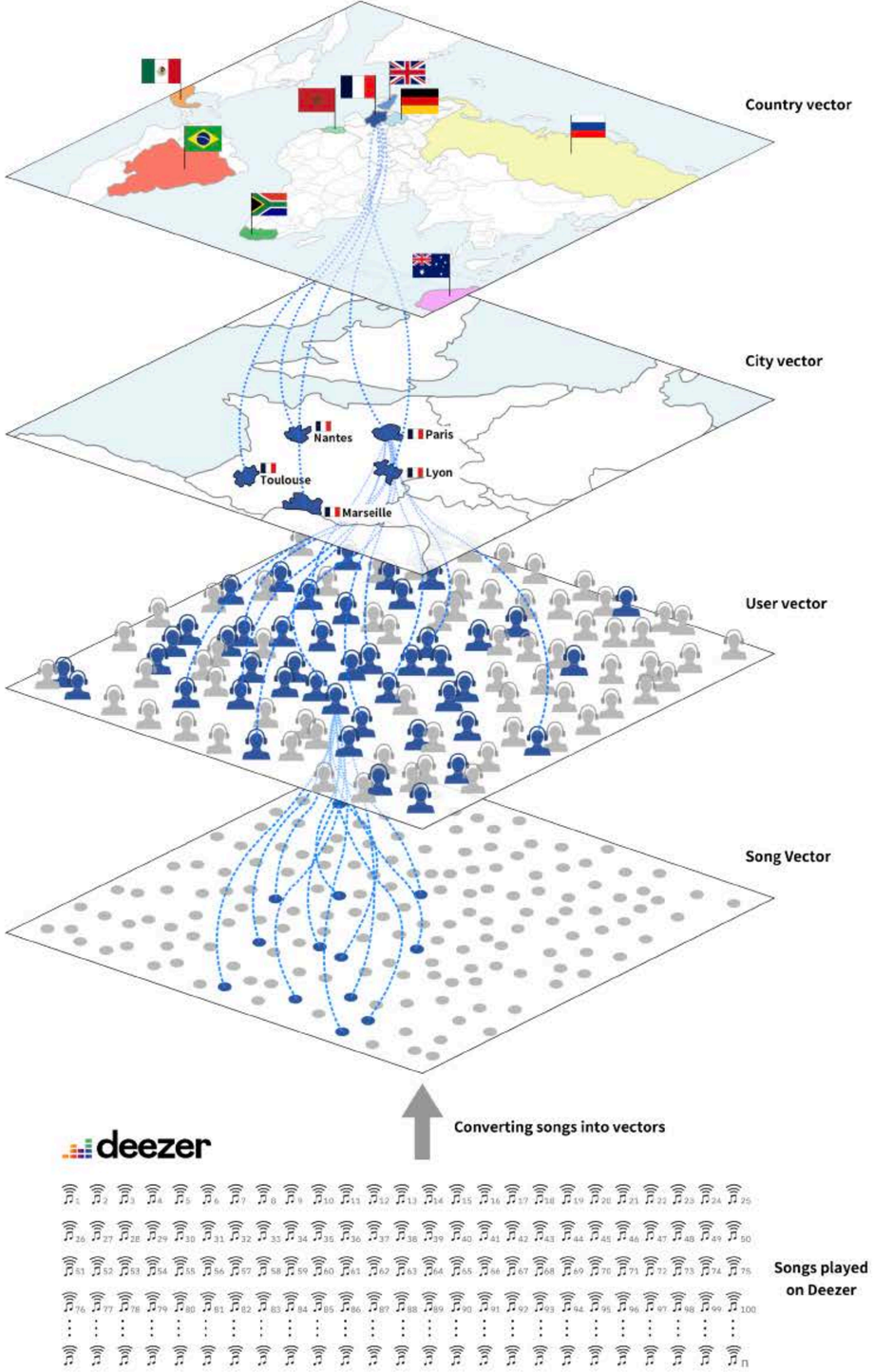
20180107	1515328485	246	0	ios	London
20180107	1515329438	231	0	ios	London
20180107	1515328757	411	0	ios	London
20180107	1515163933	210	0	ios	Cologne
20180107	1515361713	216	0	ios	Cologne

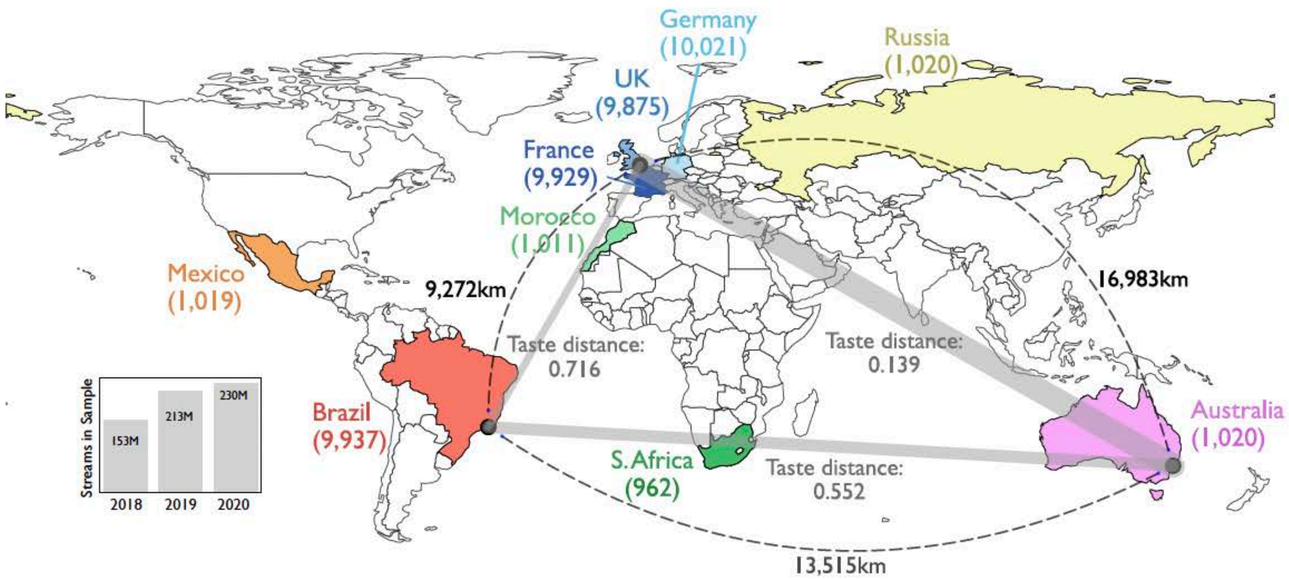
Listening day | Timestamp in Unix time | Listening duration | Identifier of skipping the track



Songs

Songs



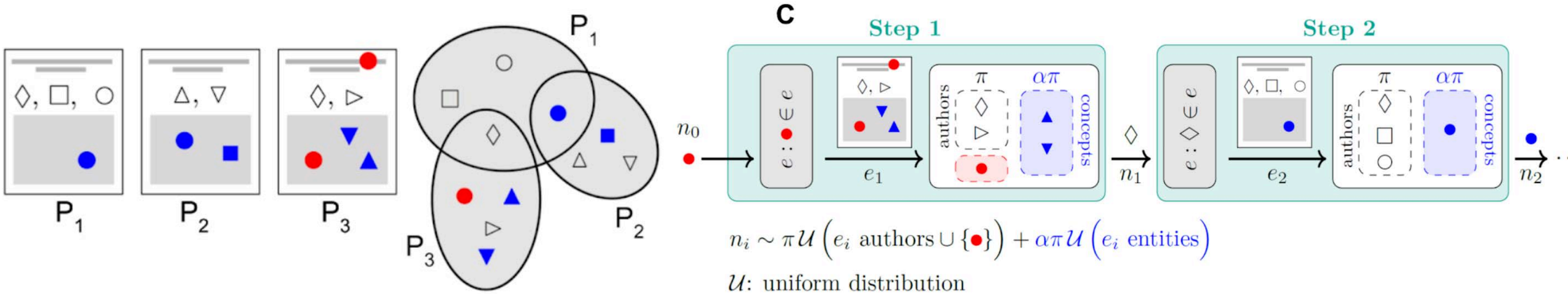


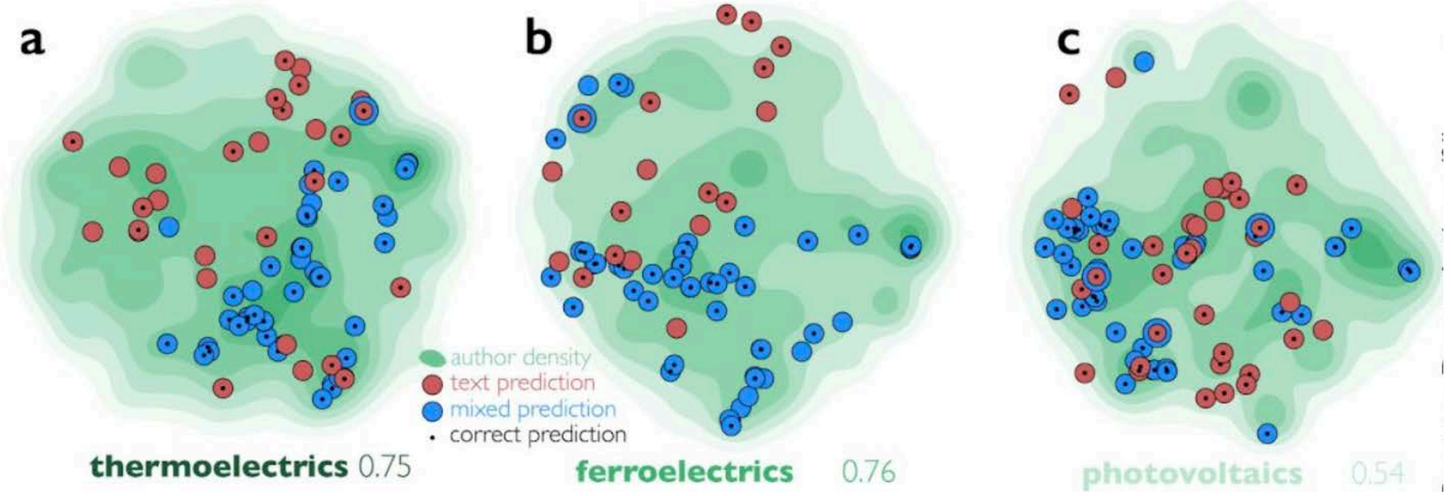
Taste Similarity Between Countries

	AU	BR	FR	DE	MX	MA	RU	ZA
BR	0.42							
FR	0.61	0.19						
DE	0.55	0.25	0.35					
MX	0.64	0.49	0.50	0.46				
MA	0.52	0.23	0.87	0.39	0.52			
RU	0.61	0.27	0.42	0.46	0.50	0.36		
ZA	0.90	0.43	0.50	0.58	0.57	0.50	0.54	
UK	0.92	0.32	0.49	0.52	0.55	0.44	0.56	0.86

$\cos\theta$: .9, .8, .7, .6, .5, .4, .3, .2

What Inferences are Most Likely **Cognitively?**





Inferences

Fig. 1: Crime data and modelling approach.

From: Event-level prediction of urban crime reveals a signature of enforcement bias in US cities

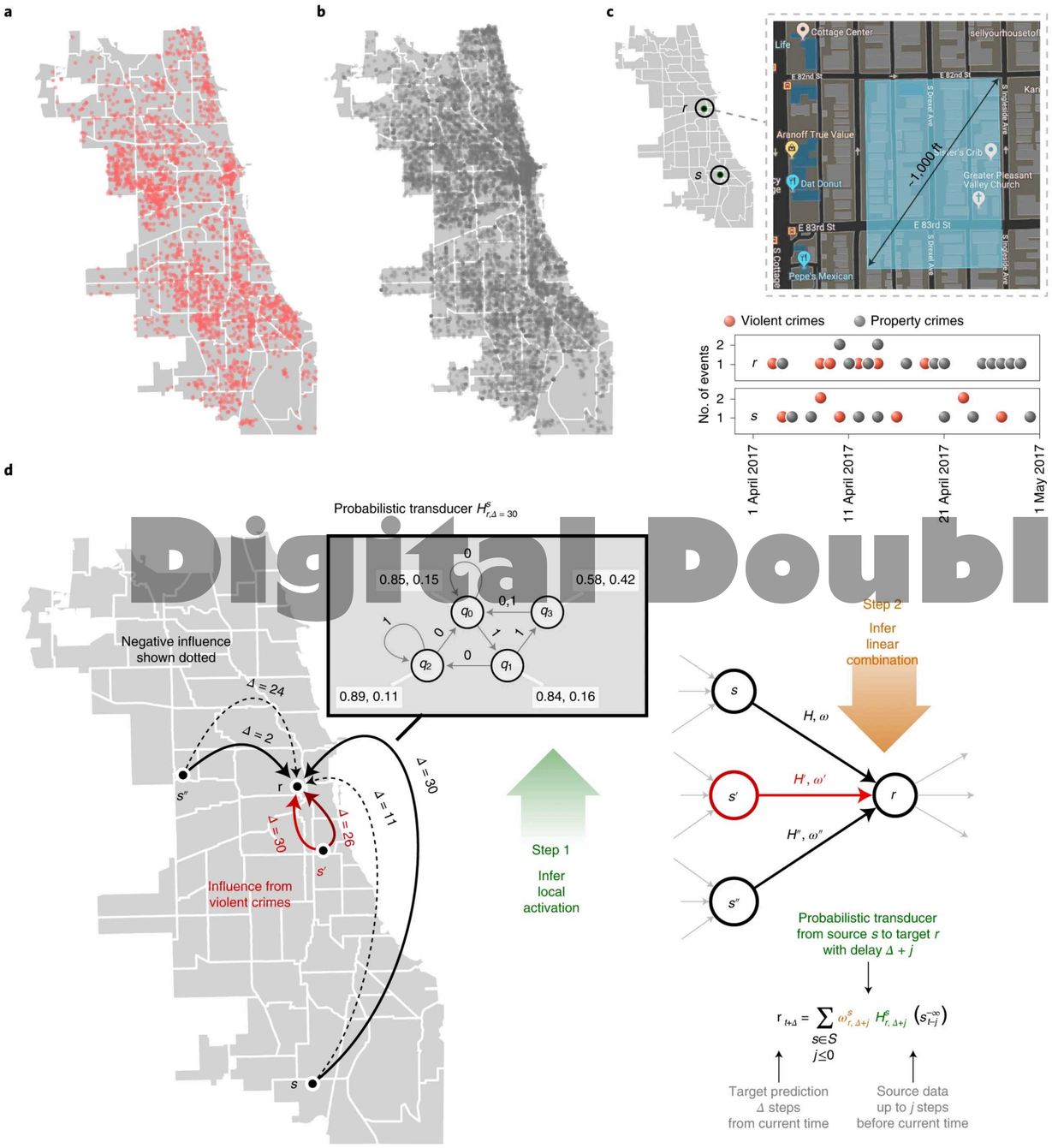


Fig. 2: Predictive performance of Granger networks.

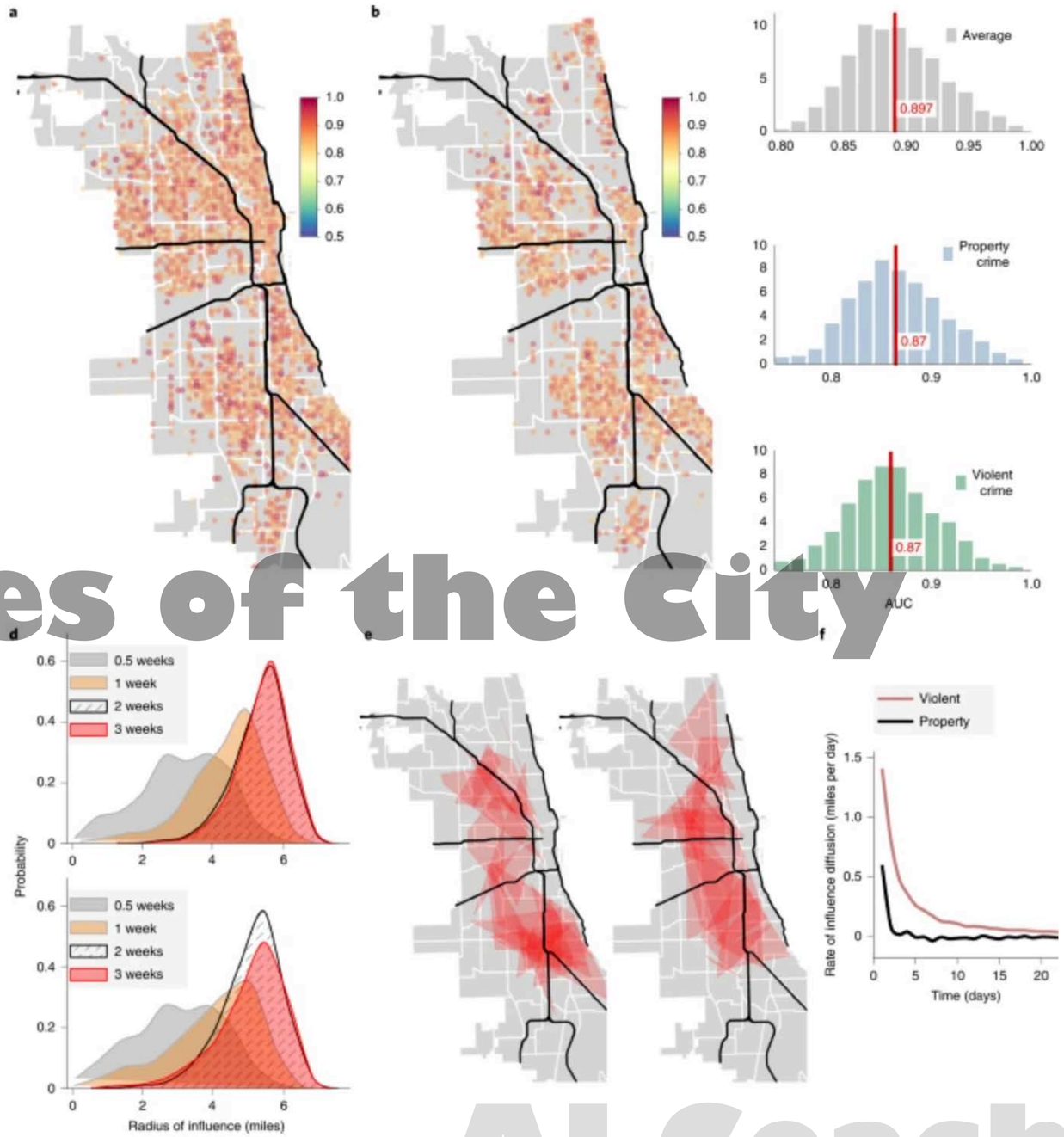


Fig. 3: Estimating bias.

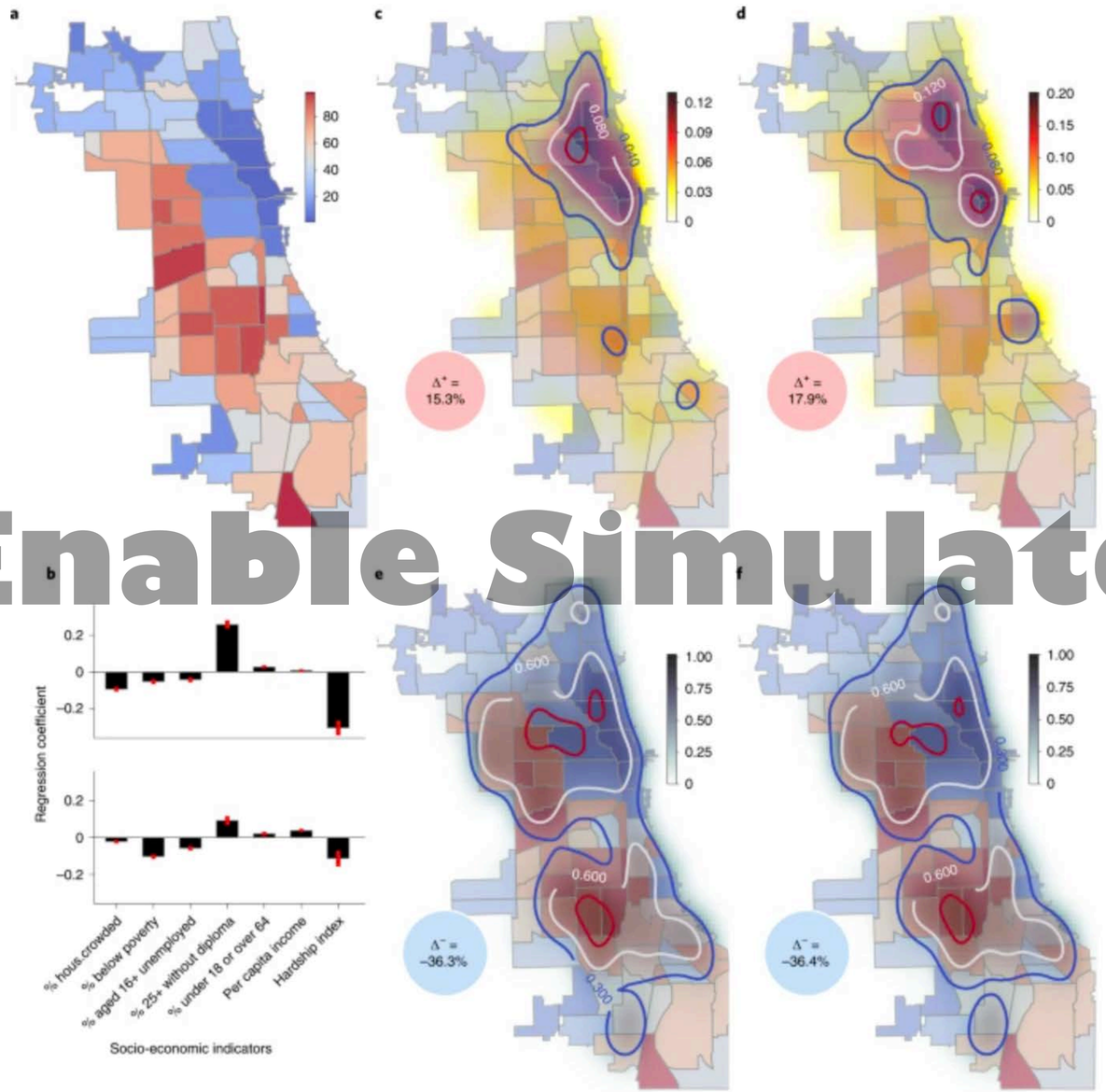
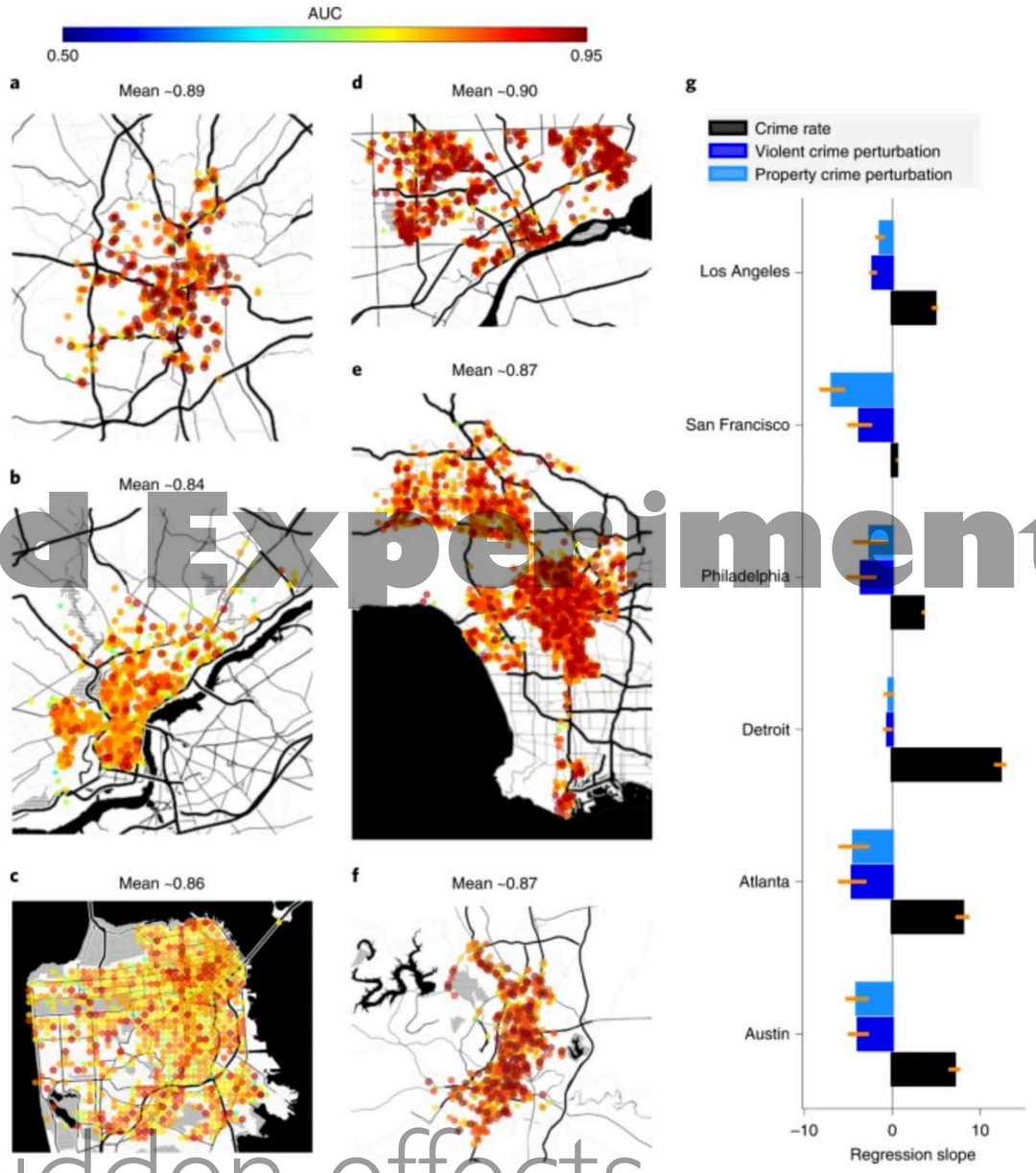


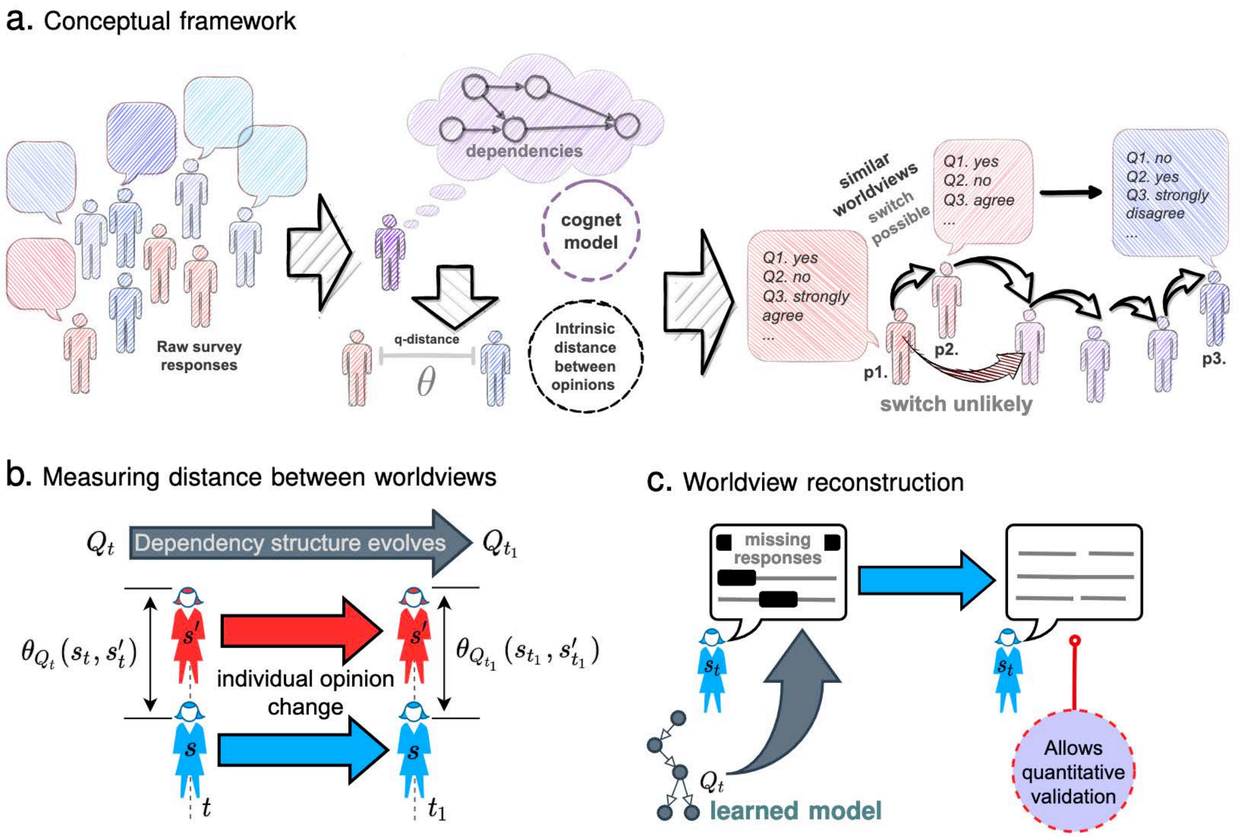
Fig. 4: Prediction of property and violent crimes across major US cities and the dependency of the perturbation response on the SES of local neighbourhoods.

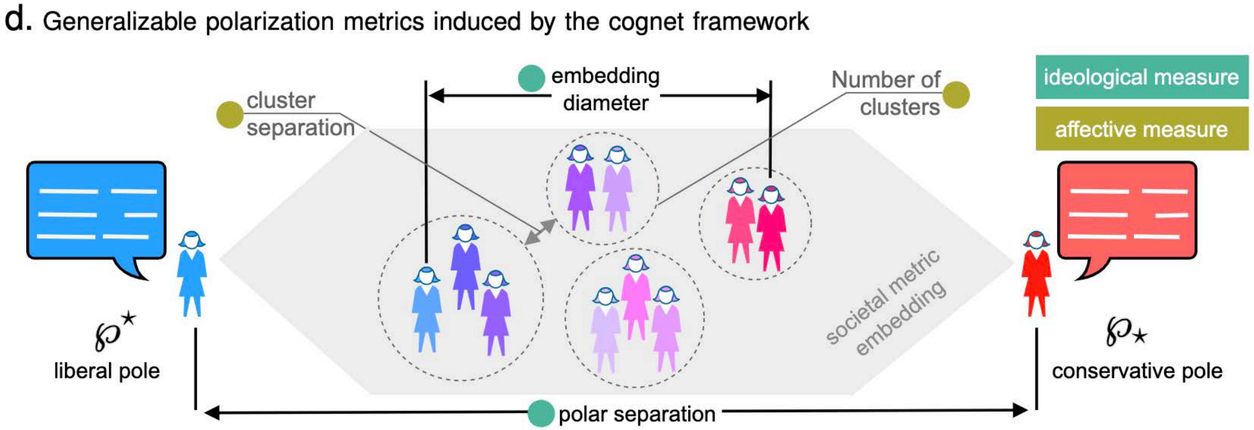


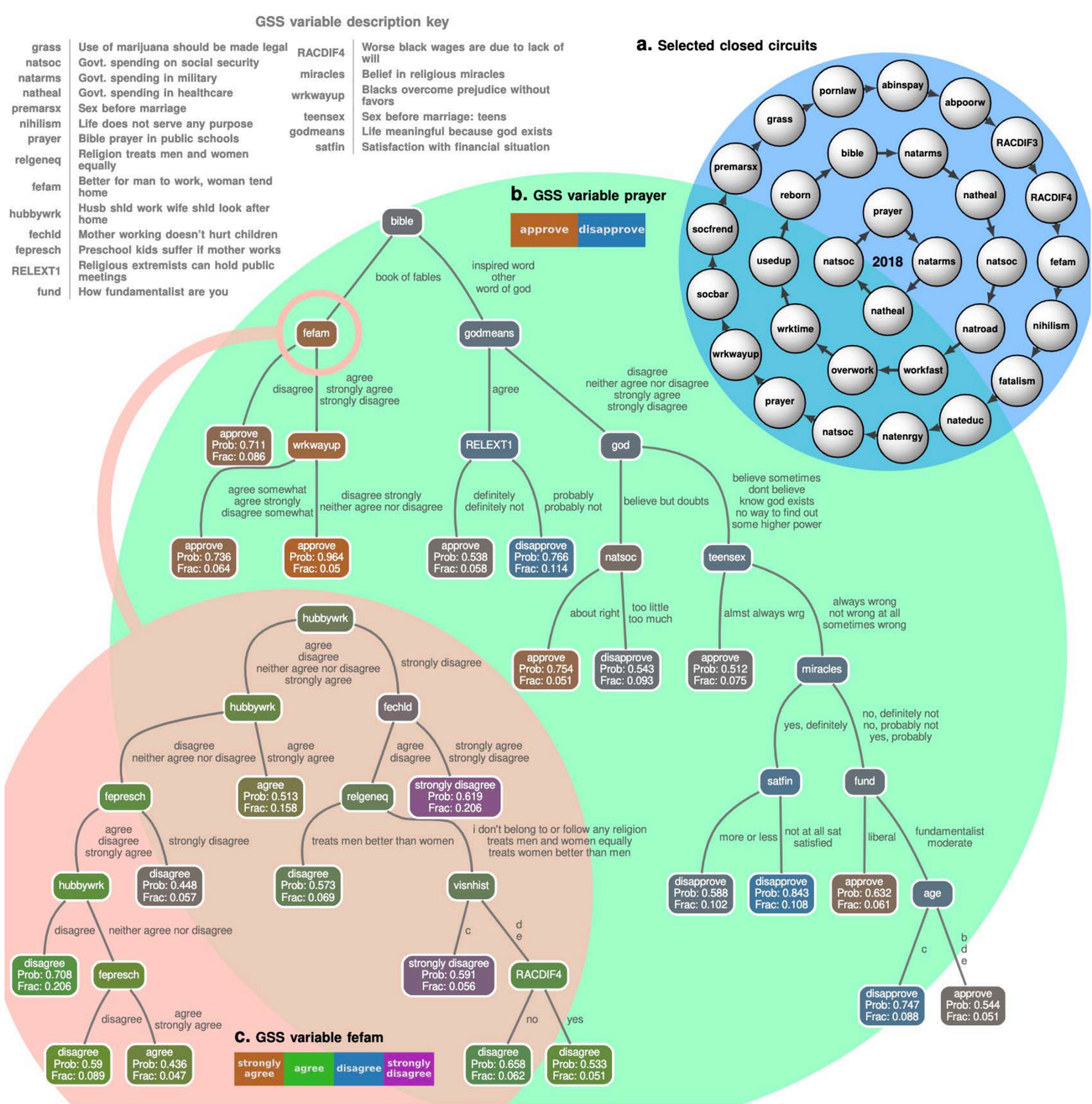
Enable Simulated Experiments

that highlight hidden effects

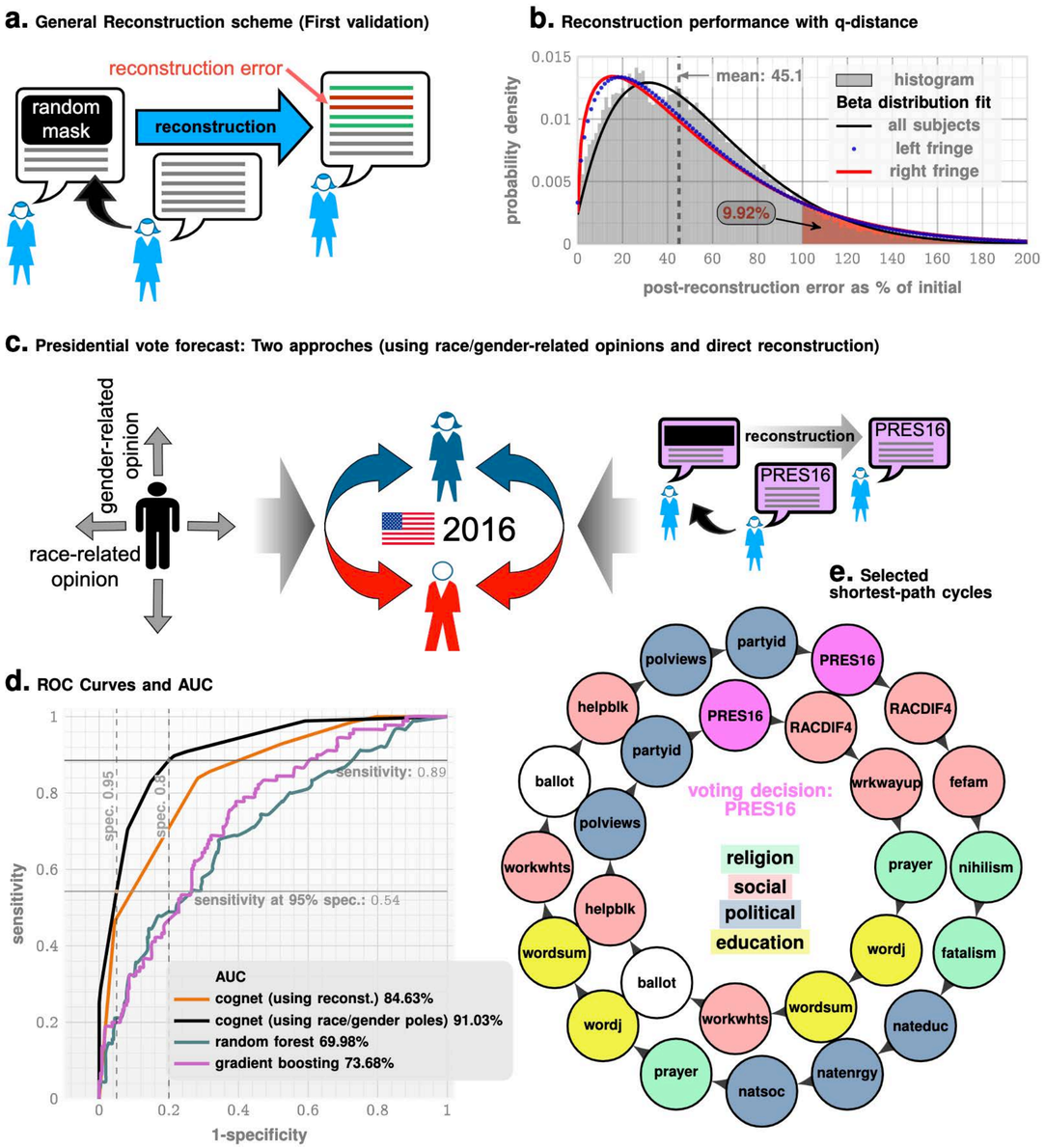
Simulated Survey Subjects

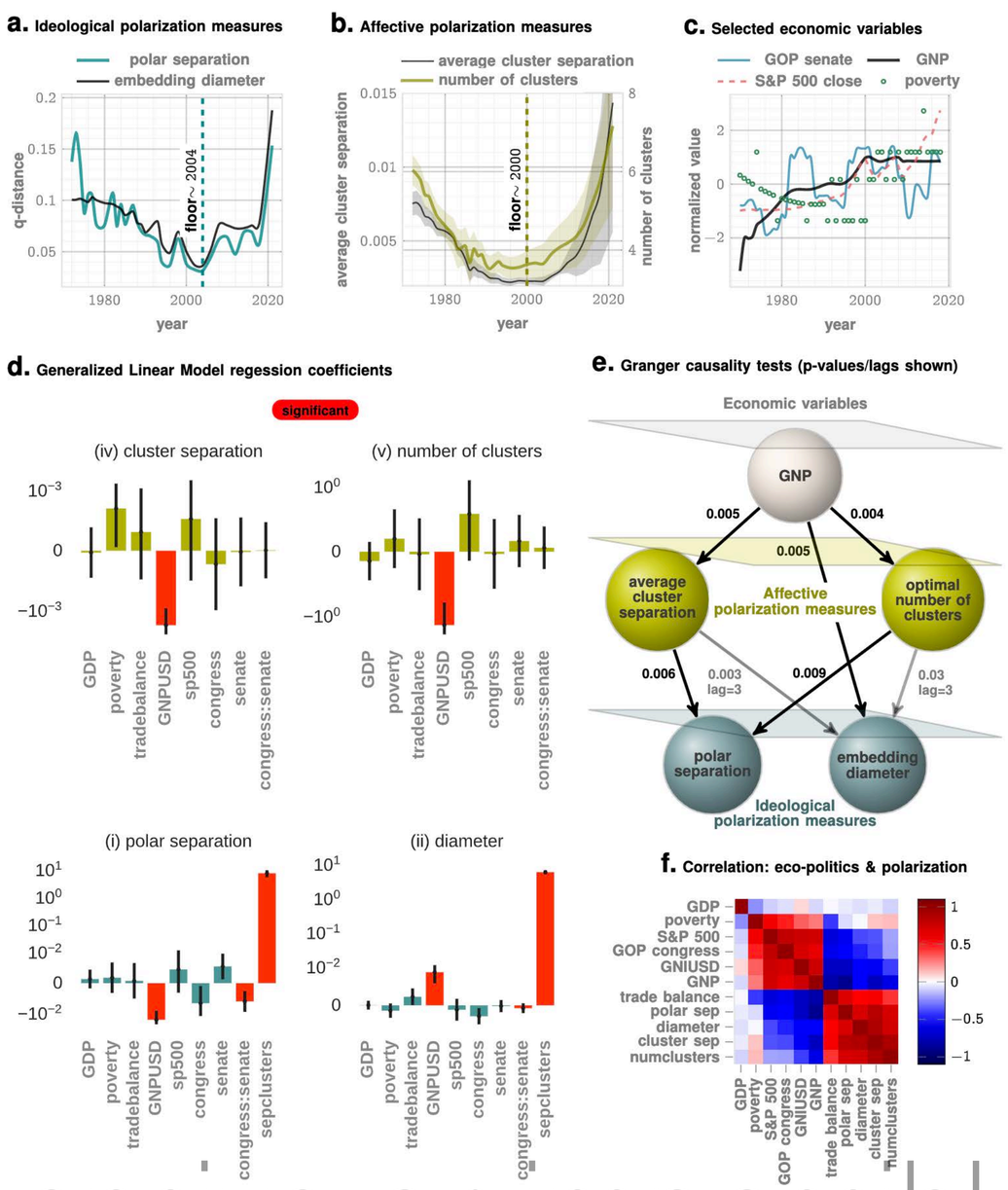






Enable Simulated Experiments





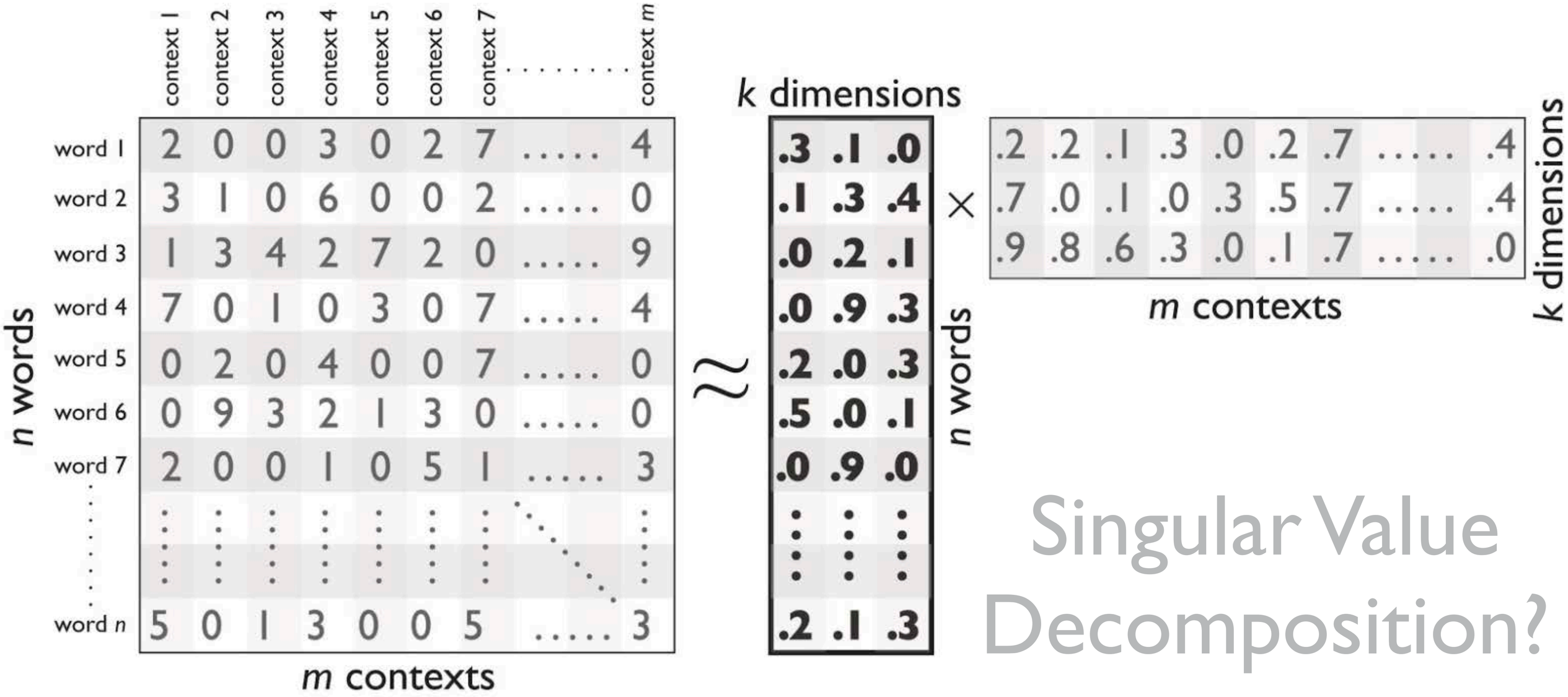
Where real ones are expensive or impossible

YOU in the LOOP

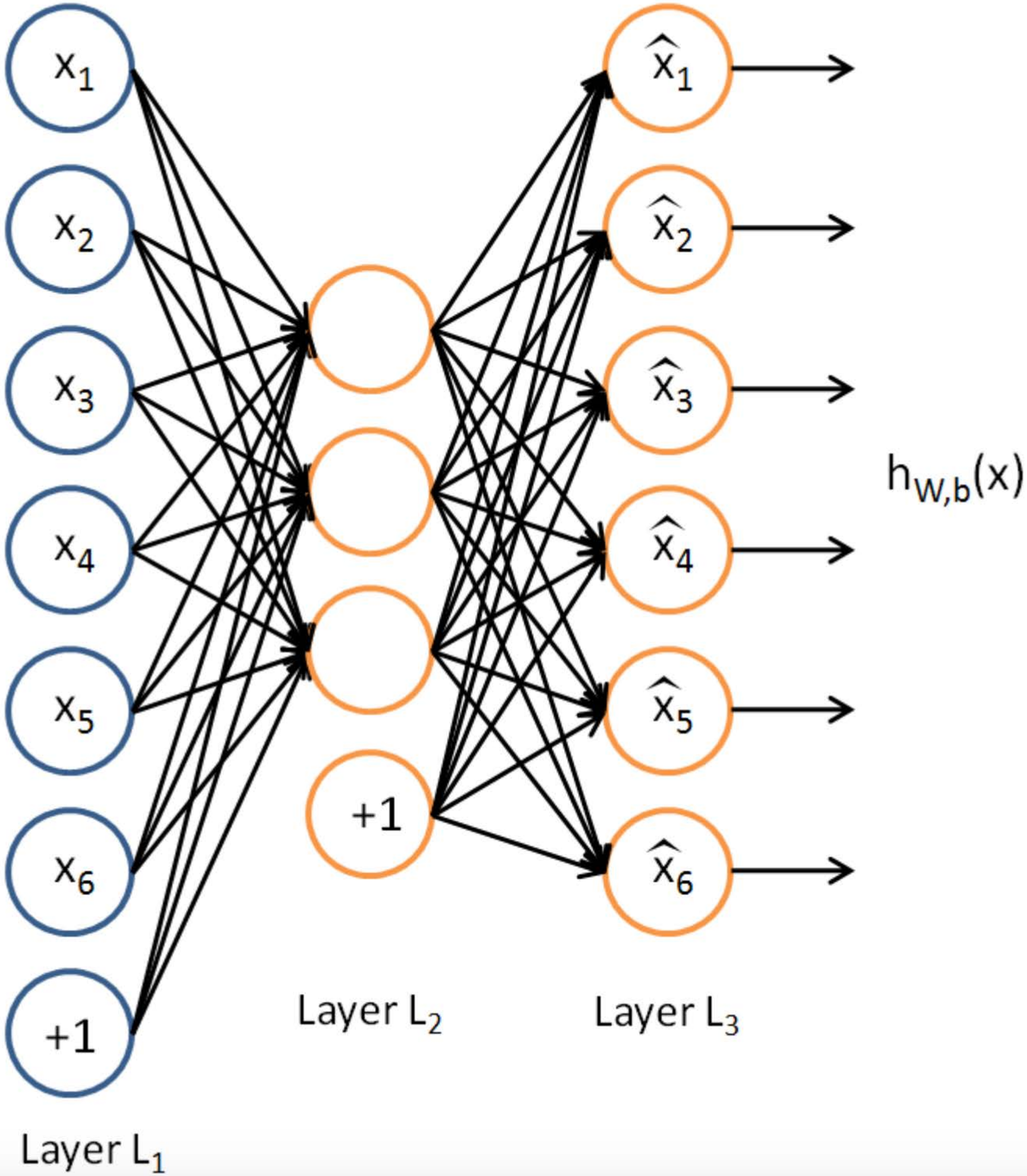
Ordinal Embedding

Is X more like Y or Z?

Factorization Strategy



Singular Value Decomposition?



Auto-encoding Neural networks

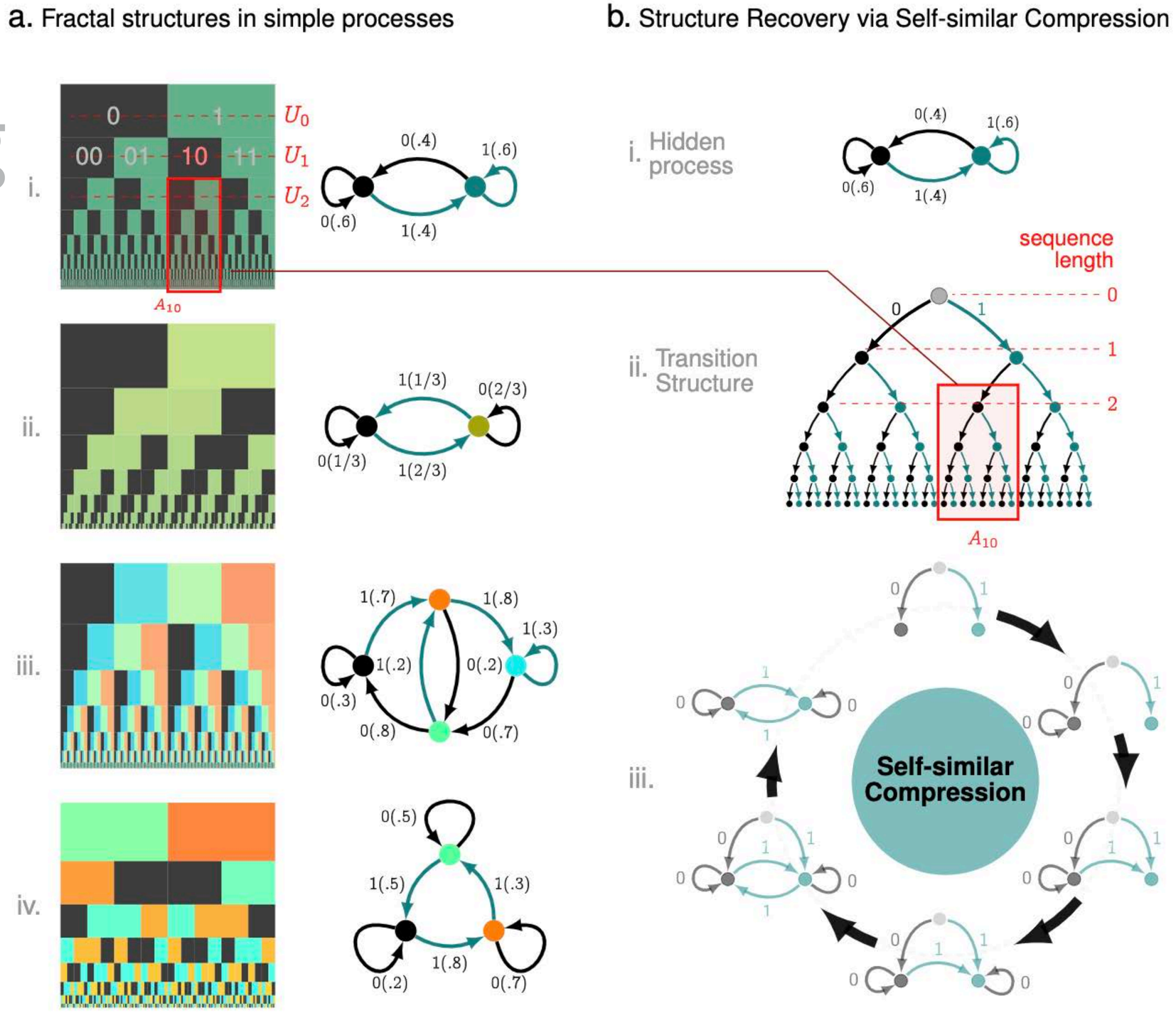
$$h_{w,\beta}(x) \approx x$$

sparsity:

$$\hat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}[a_j^{(2)}(x^{(i)})]$$

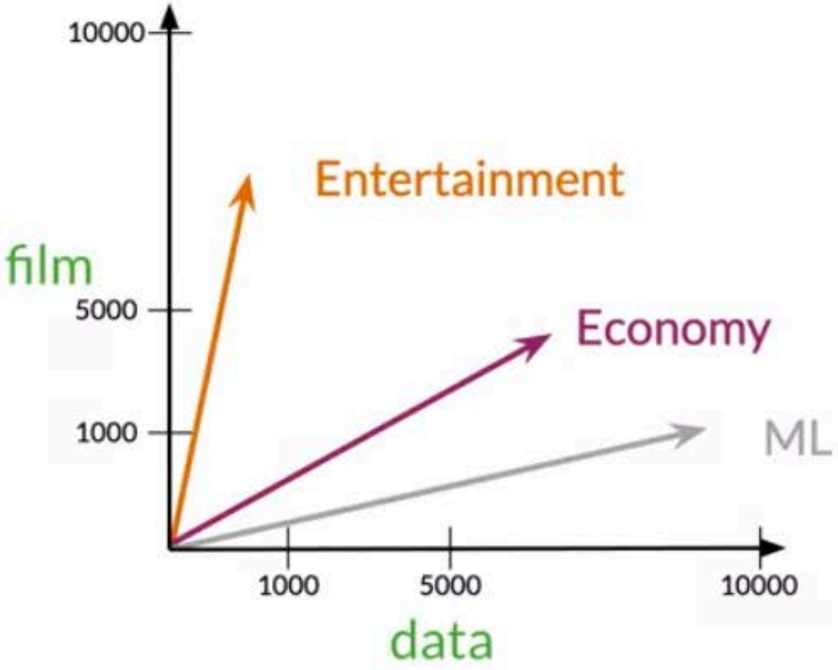
where $\hat{\rho}_j = \rho \lll 1$

Auto-encoding PFSMs



Vector Space Model

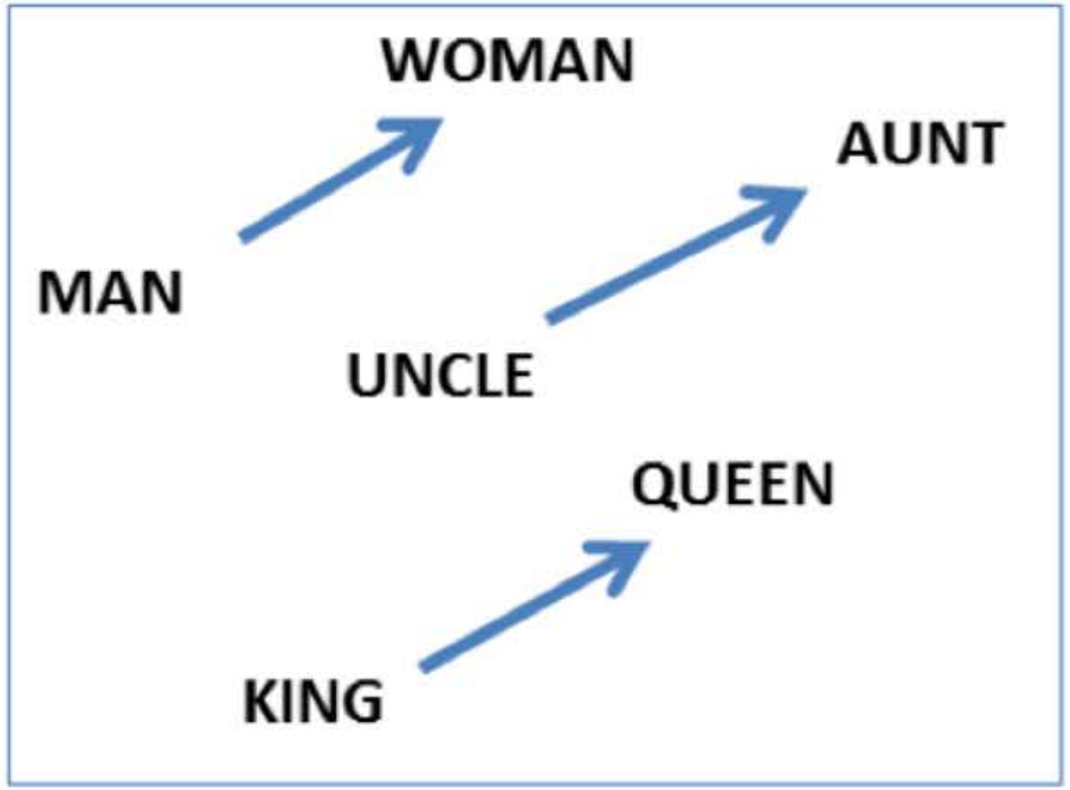
Vector Space

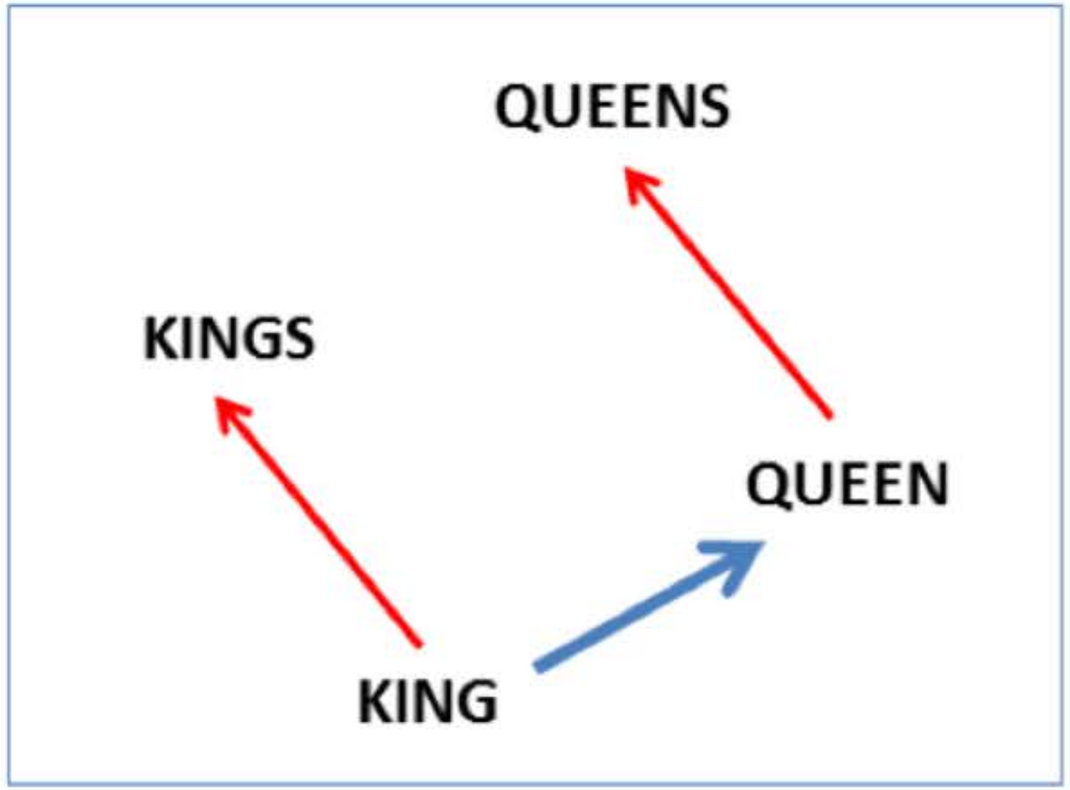


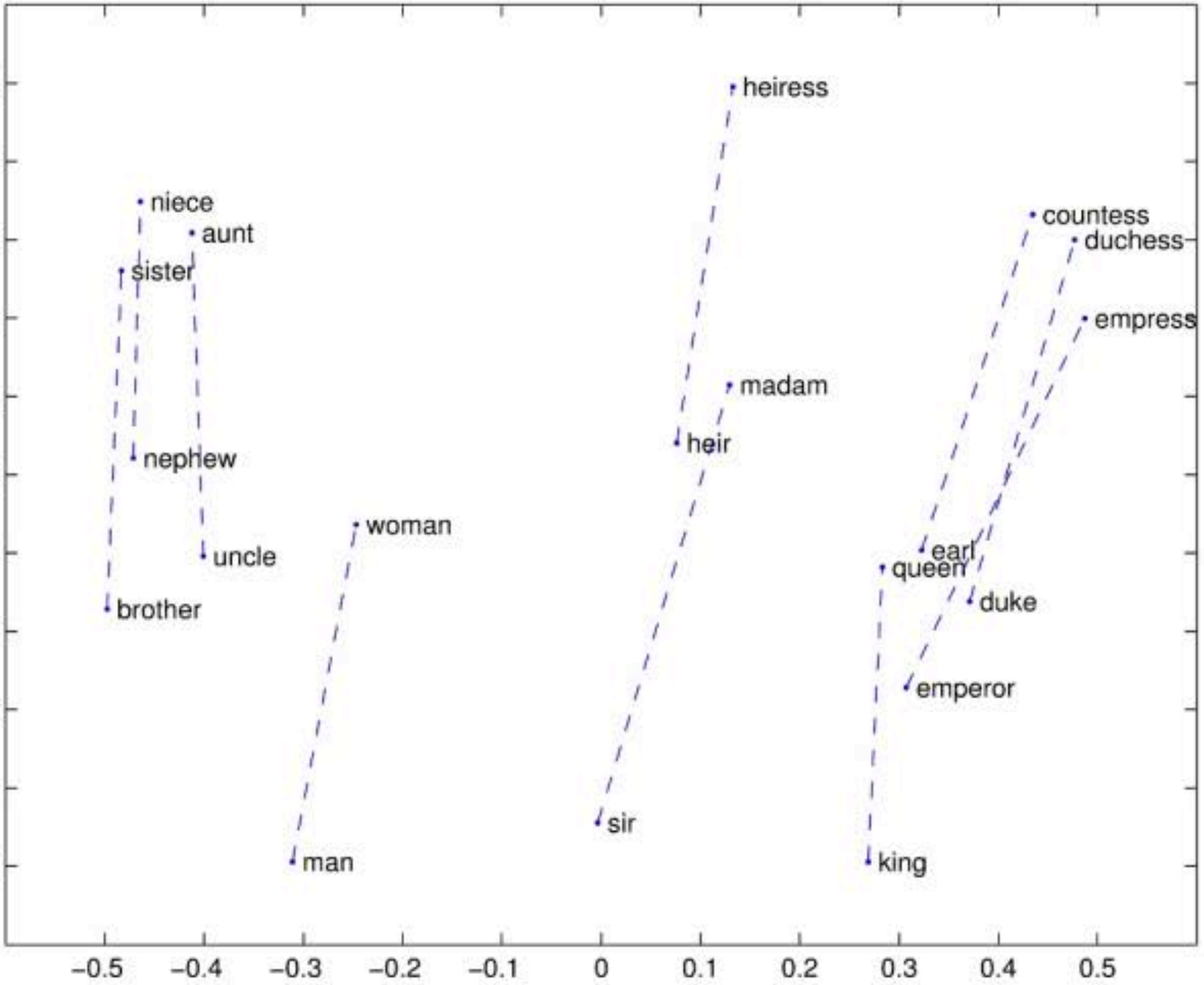
	Entertainment	Economy	ML
data	500	6620	9320
film	7000	4000	1000

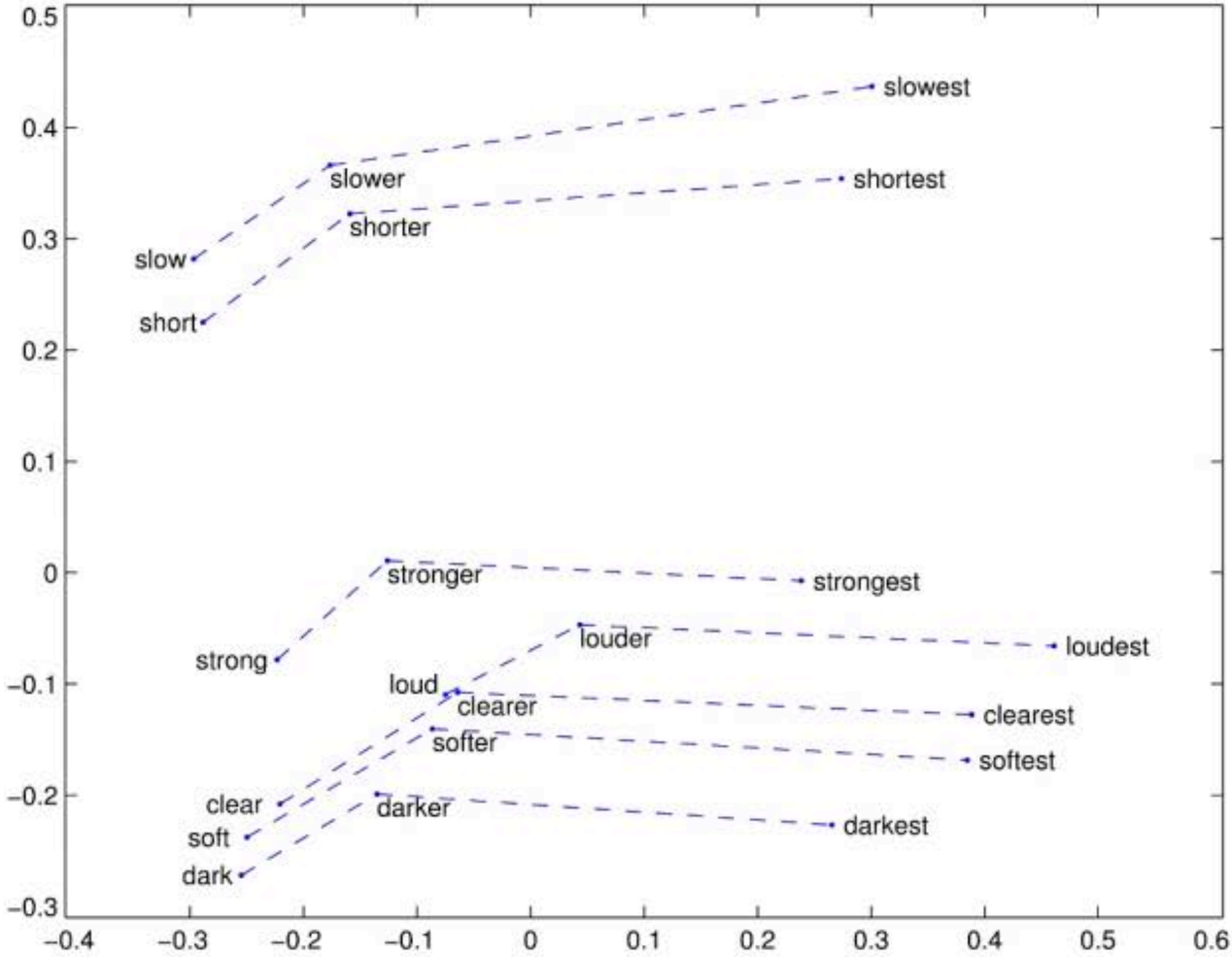
Measures of "similarity:"
Angle
Distance

Semantic relationships induce systematic shifts

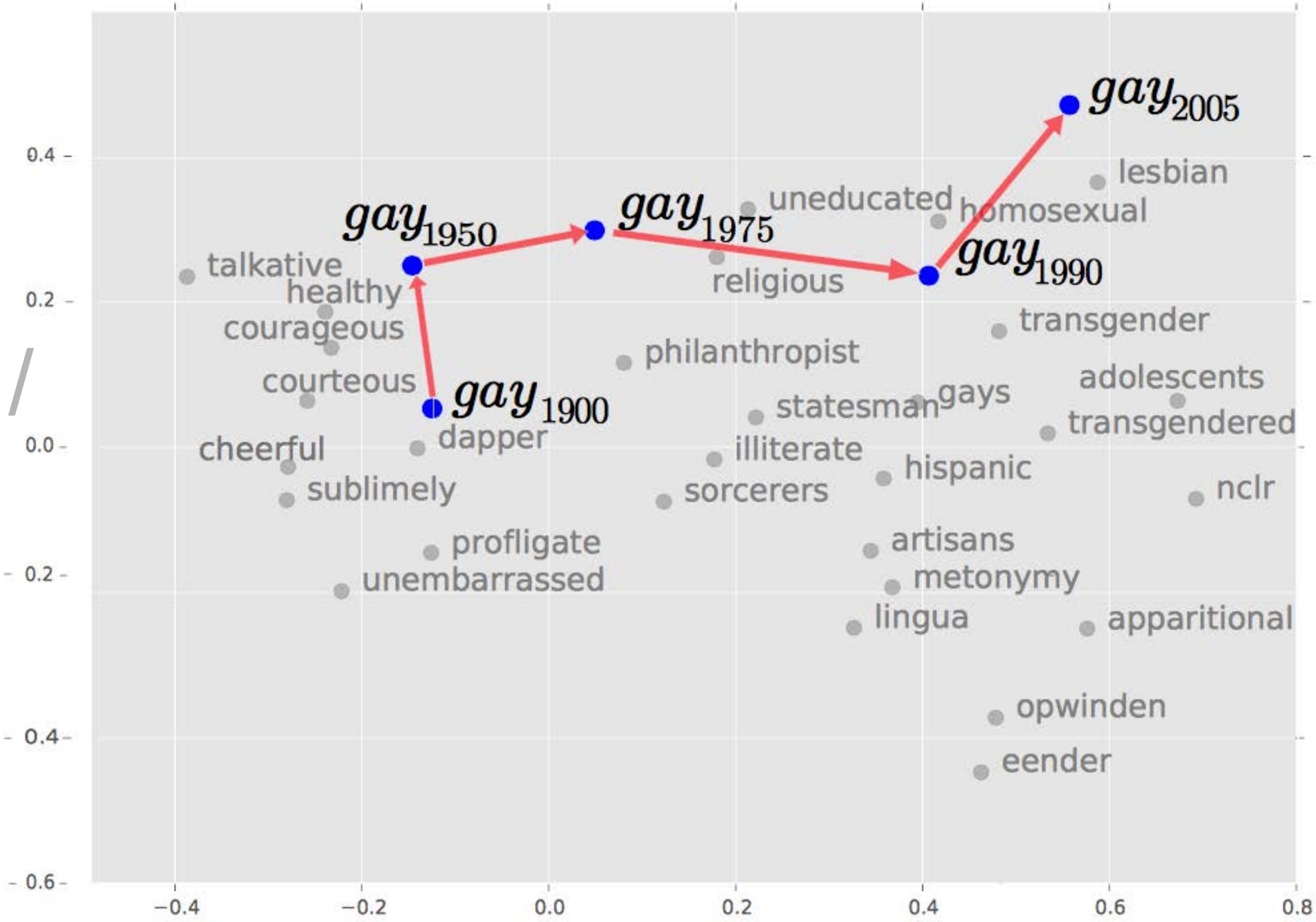





niece
aunt
sister
nephew
uncle
brother
woman
man
heiress
madam
heir
sir
countess
duchess
empress
earl
queen
duke
emperor
king
−0.5 −0.4 −0.3 −0.2 −0.1 0 0.1 0.2 0.3 0.4 0.5


slowest
slower
shortest
shorter
slow
short
stronger
strongest
louder
loudest
strong
loud
clearer
clearest
softer
softest
clear
darker
darkest
soft
dark
0.5 0.4 0.3 0.2 0.1 0 −0.1 −0.2 −0.3
−0.4 −0.3 −0.2 −0.1 0 0.1 0.2 0.3 0.4 0.5 0.6

Change over time / across cultures



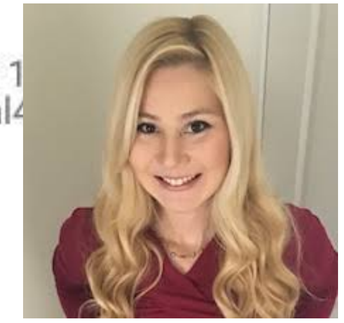
REPORT

Semantics derived automatically from language corpora contain human-like biases

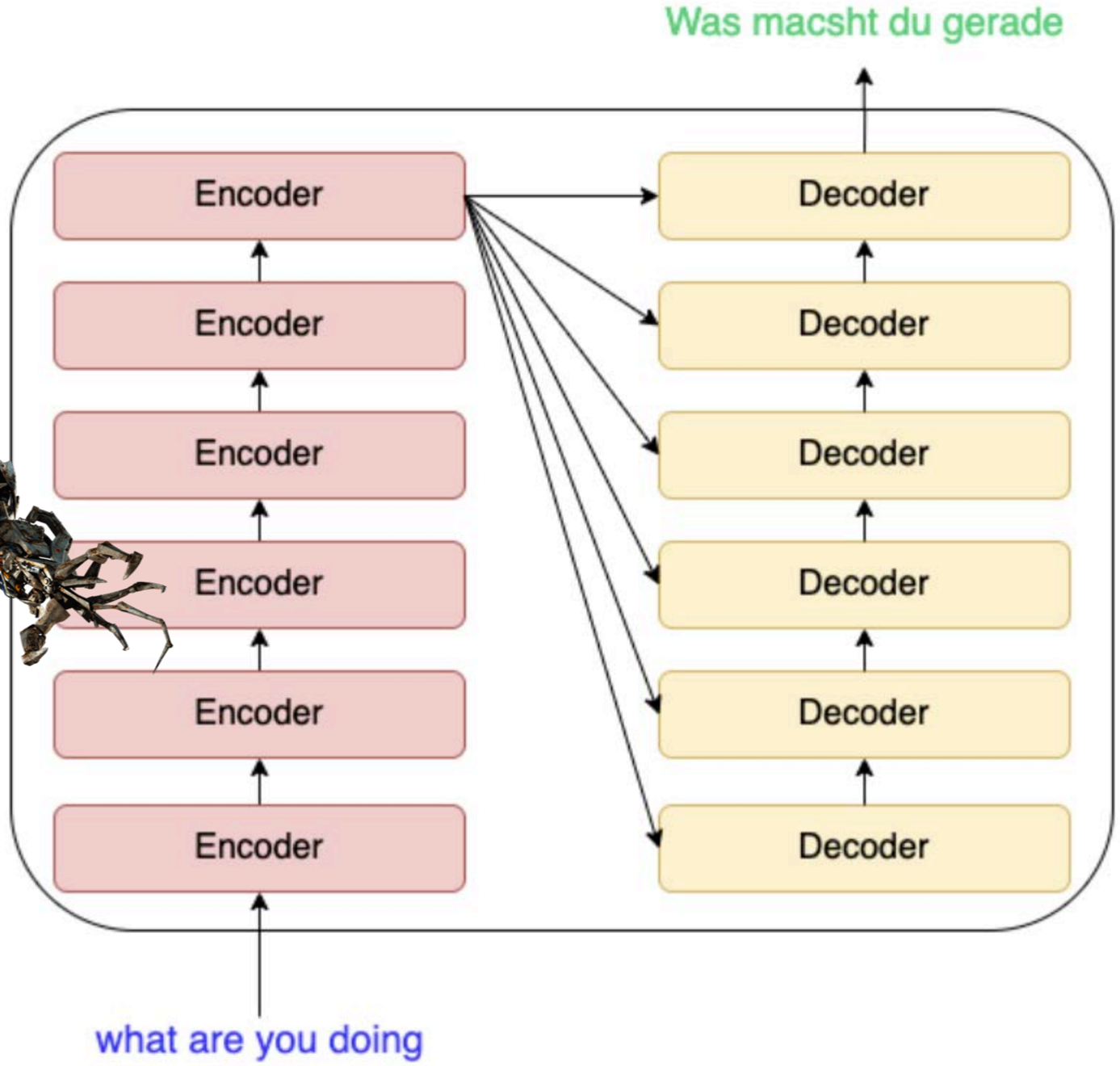
Aylin Caliskan[1,*], Joanna J. Bryson[1,2,*], Arvind Narayanan[1,*]

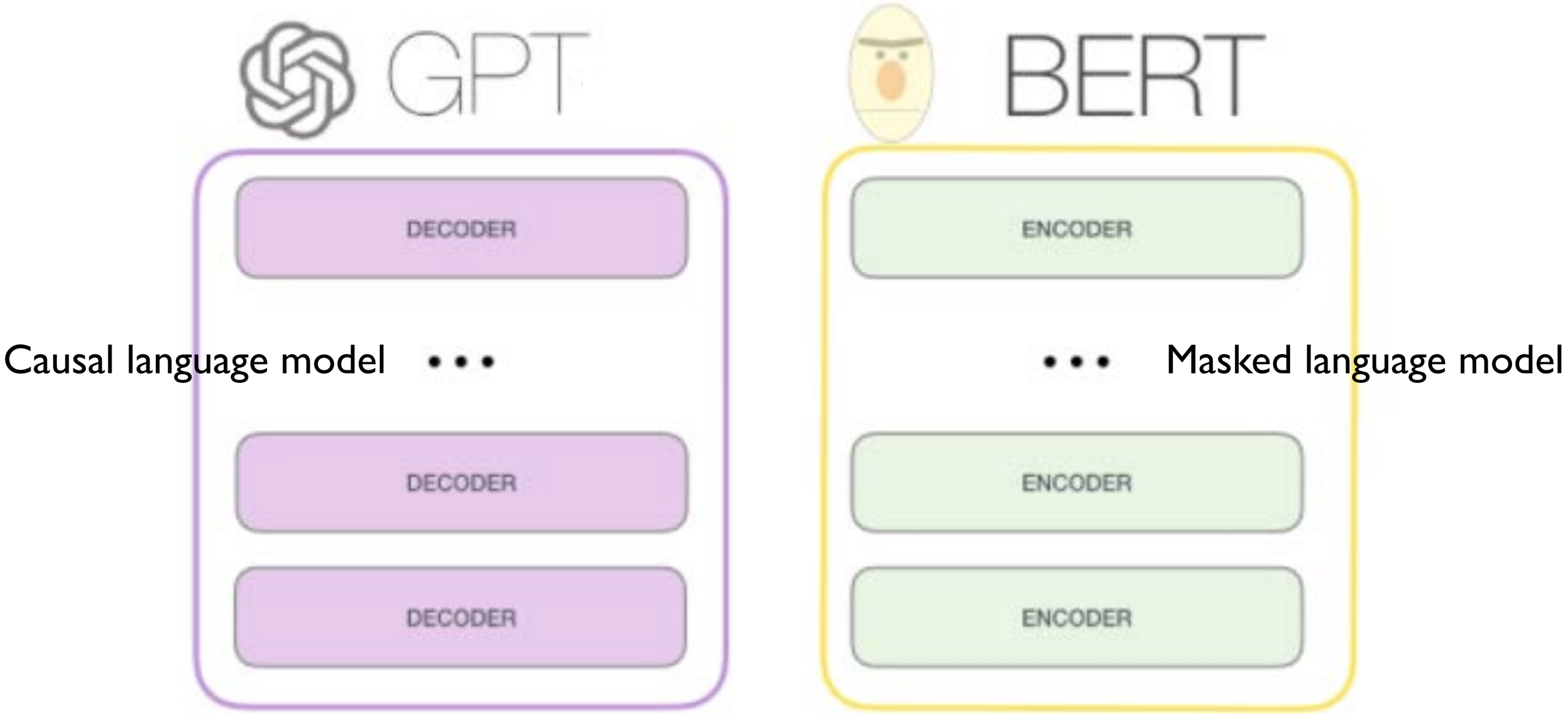
+ See all authors and affiliations

Science 14 Apr 2017:
Vol. 356, Issue 6334, pp. 1
DOI: 10.1126/science.aal4



The Transformer

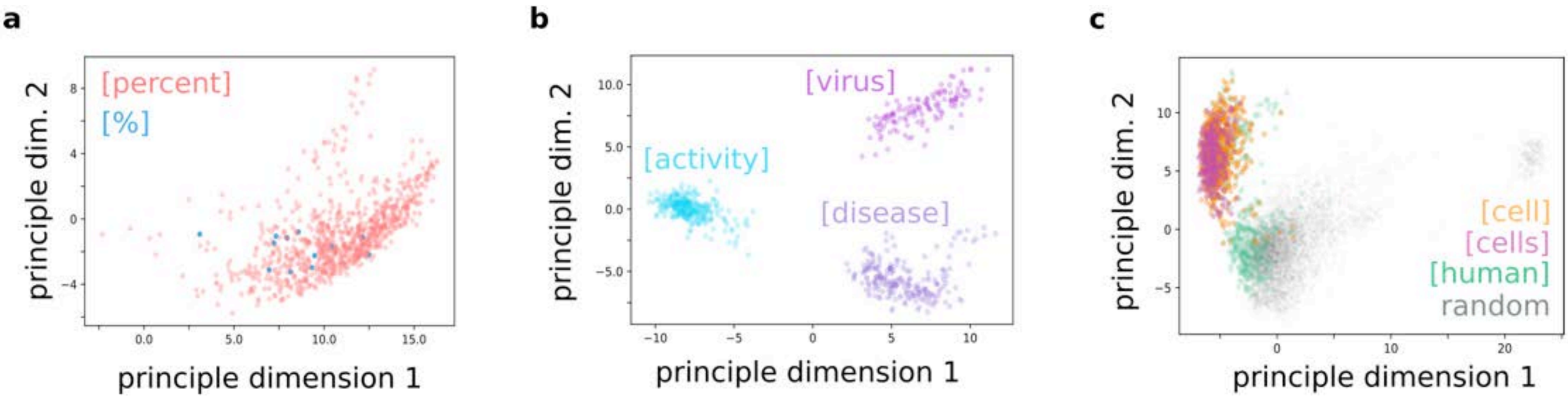




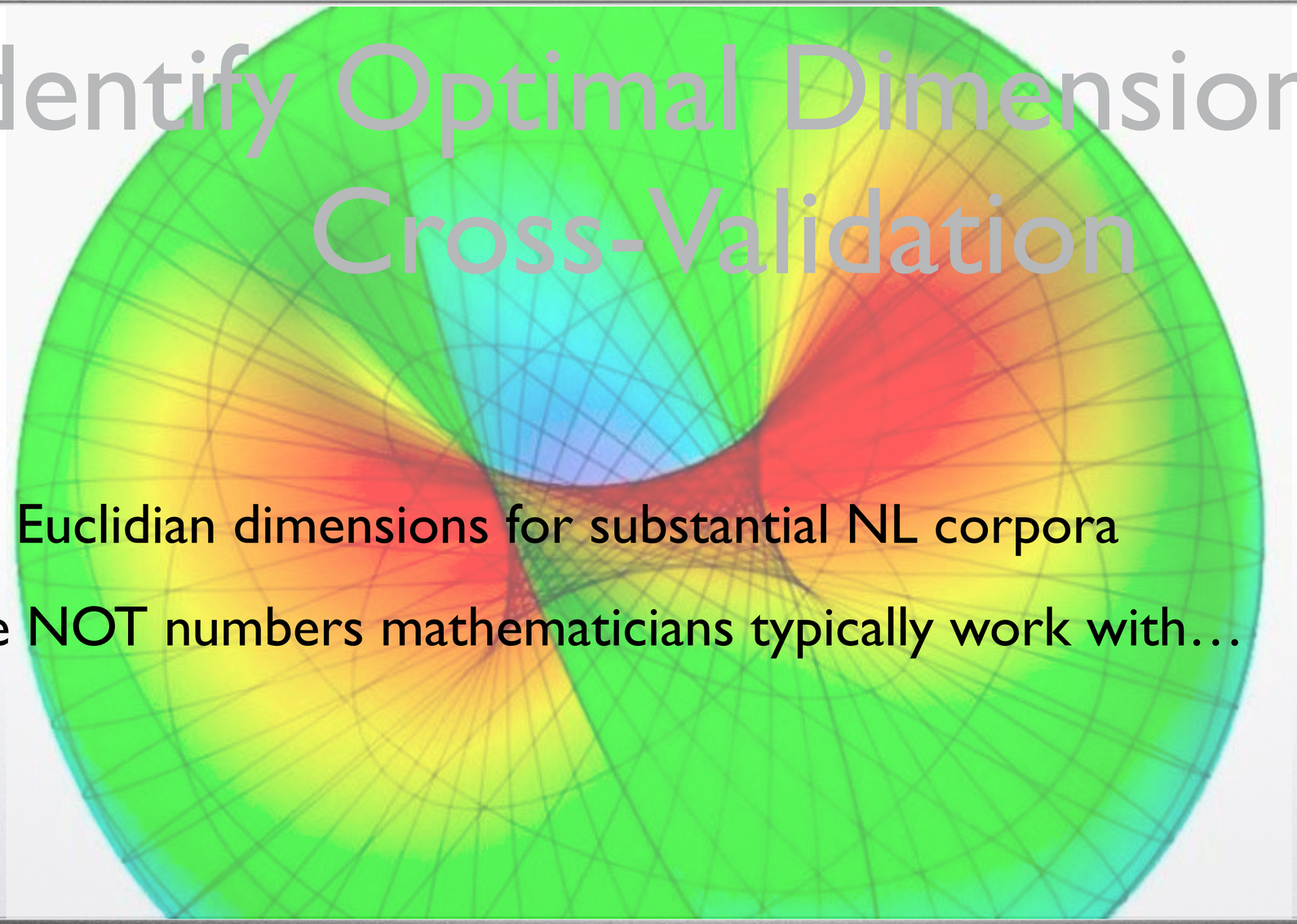
Generation

Discrimination

Each Word is a **Point Cloud**
(vector cloud)



Identify Optimal Dimensions by Cross-Validation



- 300-1000 Euclidian dimensions for substantial NL corpora
- These are NOT numbers mathematicians typically work with…

Properties of HD Geometry

Most volume is near the surface

- Most volume of the d-dimensional ball of radius r is contained within the annulus of width O(r/d) near surface;

$$\frac{\text{volume}\big((1-\epsilon)A\big)}{\text{volume}(A)} = (1-\epsilon)^d \leq e^{-\epsilon d}$$

$$A(d) = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \quad \textit{and} \quad V(d) = \frac{2\pi^{\frac{d}{2}}}{d\,\Gamma(\frac{d}{2})}$$

- Volume of the sphere is near the equator

Theorem 2.7 *For $c \geq 1$ and $d \geq 3$, at least a $1 - \frac{2}{c}e^{-c^2/2}$ fraction of the volume of the d-dimensional unit ball has $|x_1| \leq \frac{c}{\sqrt{d-1}}$.*

Entailment: Cosine/Angular Distance is a reasonable measure

Properties of HD Geometry

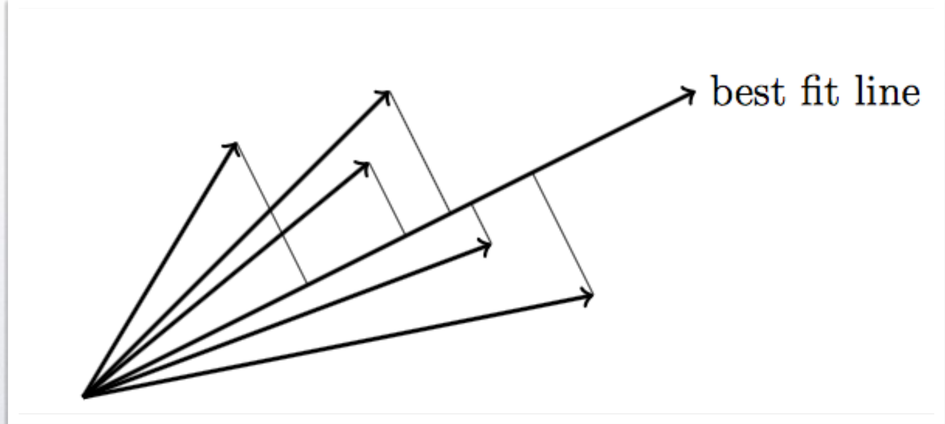
Two random points from a *d*-dimensional Gaussian with unit variance in each direction are approximately orthogonal.

Theorem 2.9 (Gaussian Annulus Theorem) *For a d-dimensional spherical Gaussian with unit variance in each direction, for any $\beta \leq \sqrt{d}$, all but at most $3e^{-c\beta^2}$ of the probability mass lies within the annulus $\sqrt{d} - \beta \leq |\mathbf{x}| \leq \sqrt{d} + \beta$, where c is a fixed positive constant.*

Arccos of cosine measure. Non-90 degree angles represent biased association

Distances are concentrated around their expected value

$$|\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$



Entailment: Right/90-degree angles = random alignment

Properties of HD Geometry

Random projection

The projection $f : \mathbf{R}^d \to \mathbf{R}^k$ that we will examine (in fact, many related projections are known to work as well) is the following. Pick k Gaussian vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_k}$ in $\mathbf{R}^d$ with unit-variance coordinates. For any vector $\mathbf{v}$, define the projection $f(\mathbf{v})$ by:

$$f(\mathbf{v}) = (\mathbf{u_1} \cdot \mathbf{v}, \mathbf{u_2} \cdot \mathbf{v}, \ldots, \mathbf{u_k} \cdot \mathbf{v}).$$

With high probablity

$$|f(\mathbf{v})| \approx \sqrt{k}|\mathbf{v}|.$$

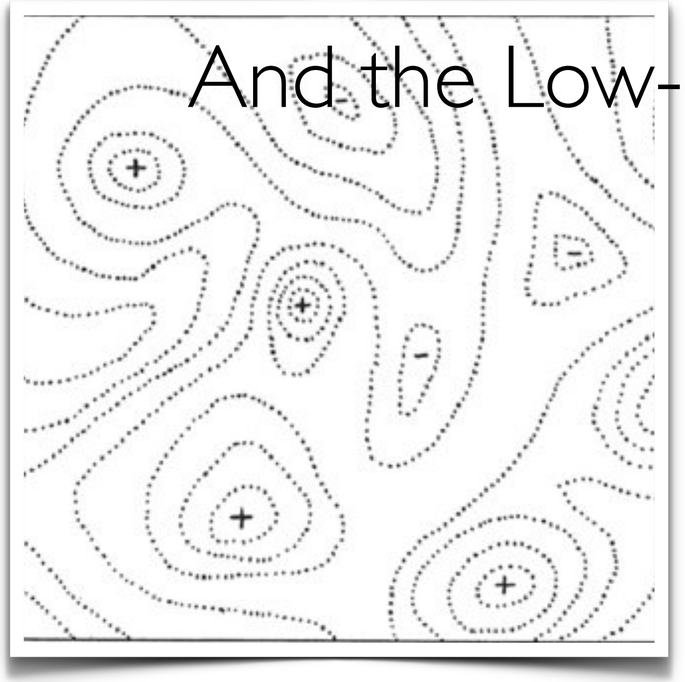
Theorem 2.11 (Johnson-Lindenstrauss Lemma) *For any $0 < \varepsilon < 1$ and any integer n, let $k \geq \frac{3}{c\varepsilon^2} \ln n$ for c as in Theorem 2.9. For any set of n points in R^d, the random projection $f : R^d \to R^k$ defined above has the property that for all pairs of points $\mathbf{v_i}$ and $\mathbf{v_j}$, with probability at least $1 - 1.5/n$,*

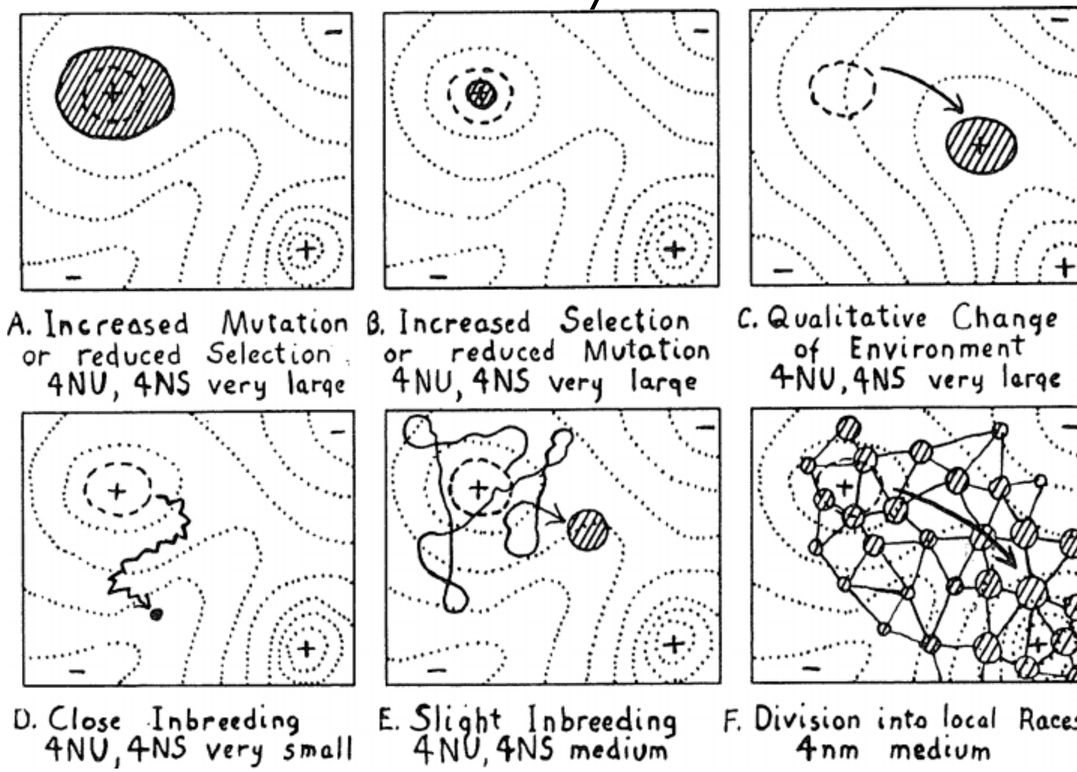
$$(1-\varepsilon)\sqrt{k}\,|\mathbf{v_i} - \mathbf{v_j}| \leq |f(\mathbf{v_i}) - f(\mathbf{v_j})| \leq (1+\varepsilon)\sqrt{k}\,|\mathbf{v_i} - \mathbf{v_j}|\,.$$

Entailment: Words close in space; many neutral pathways between meanings

Sewall Wright

And the Low-dimensional Valley of Death

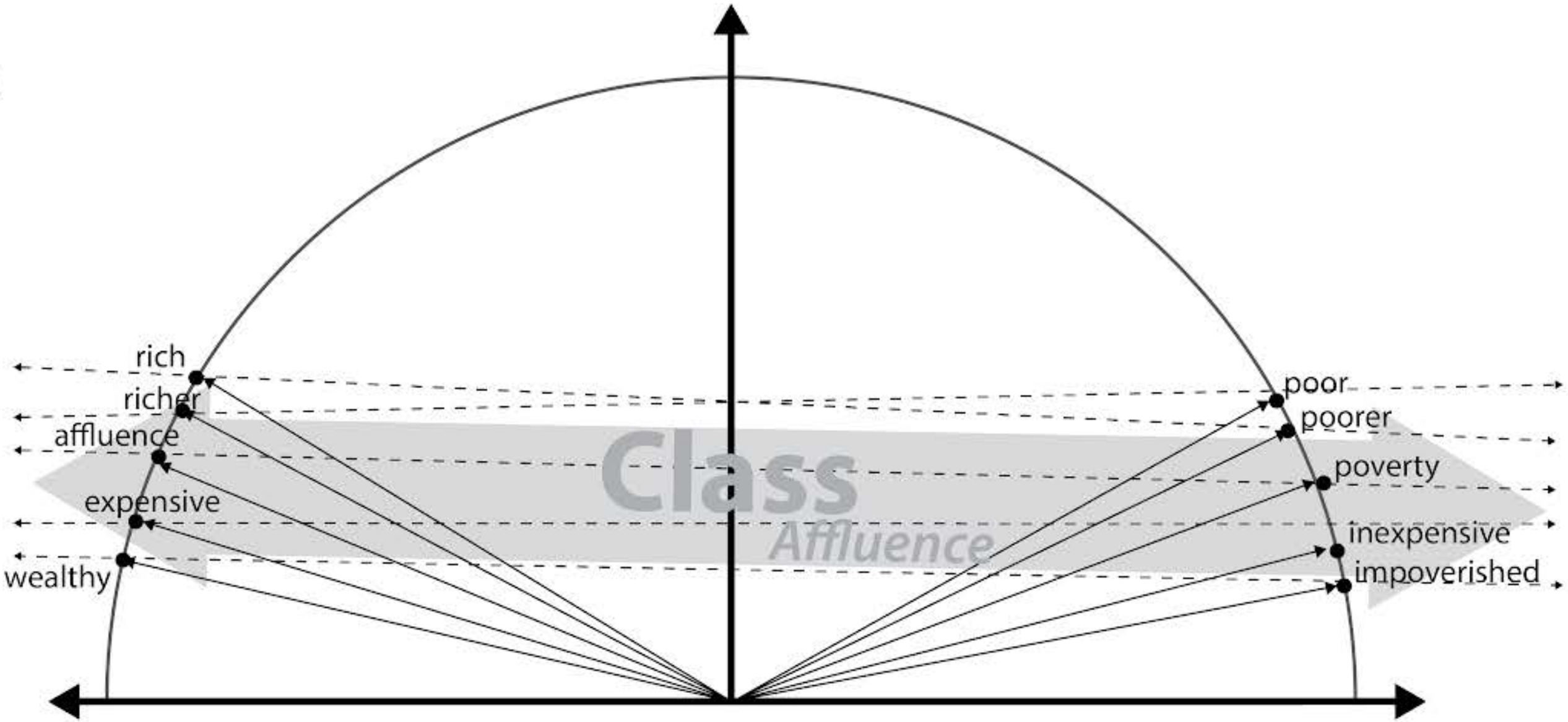


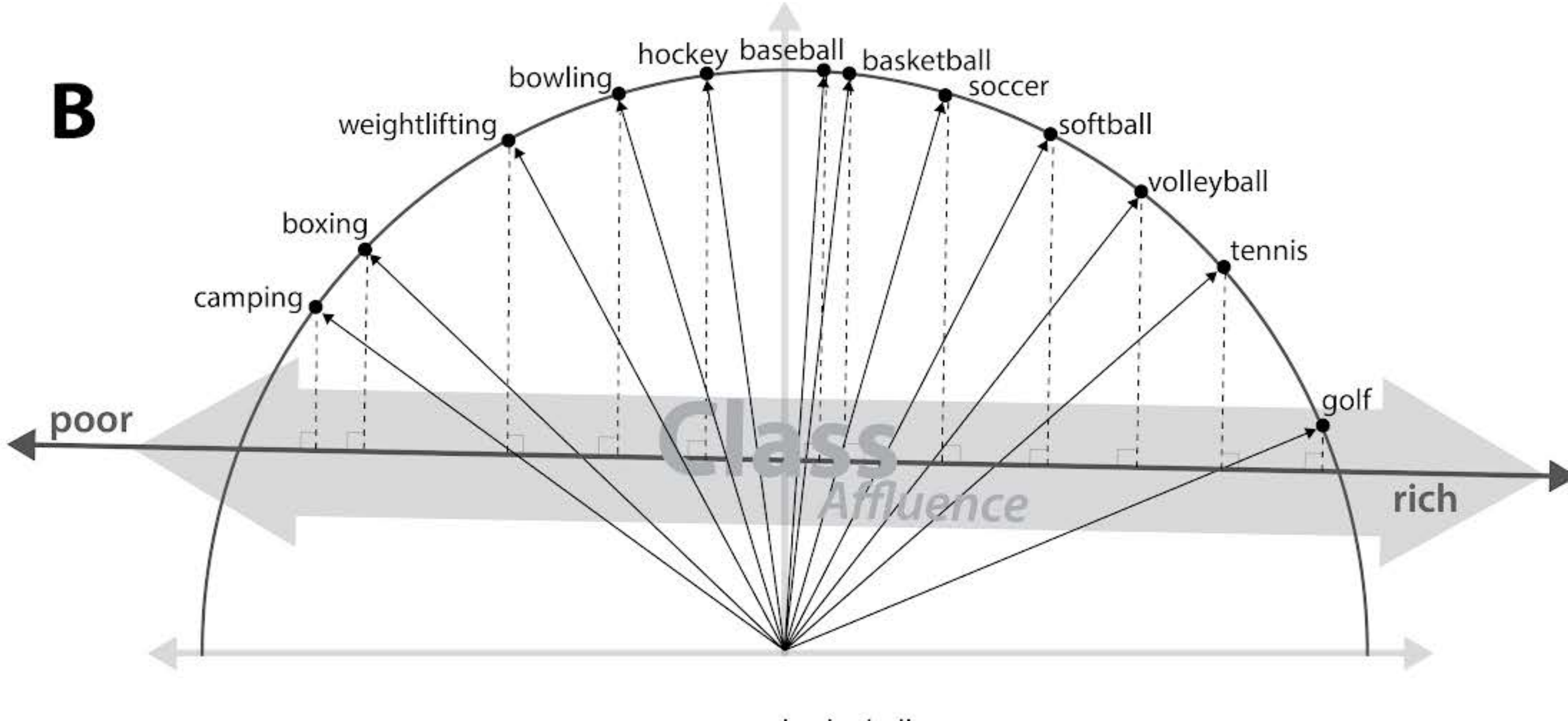


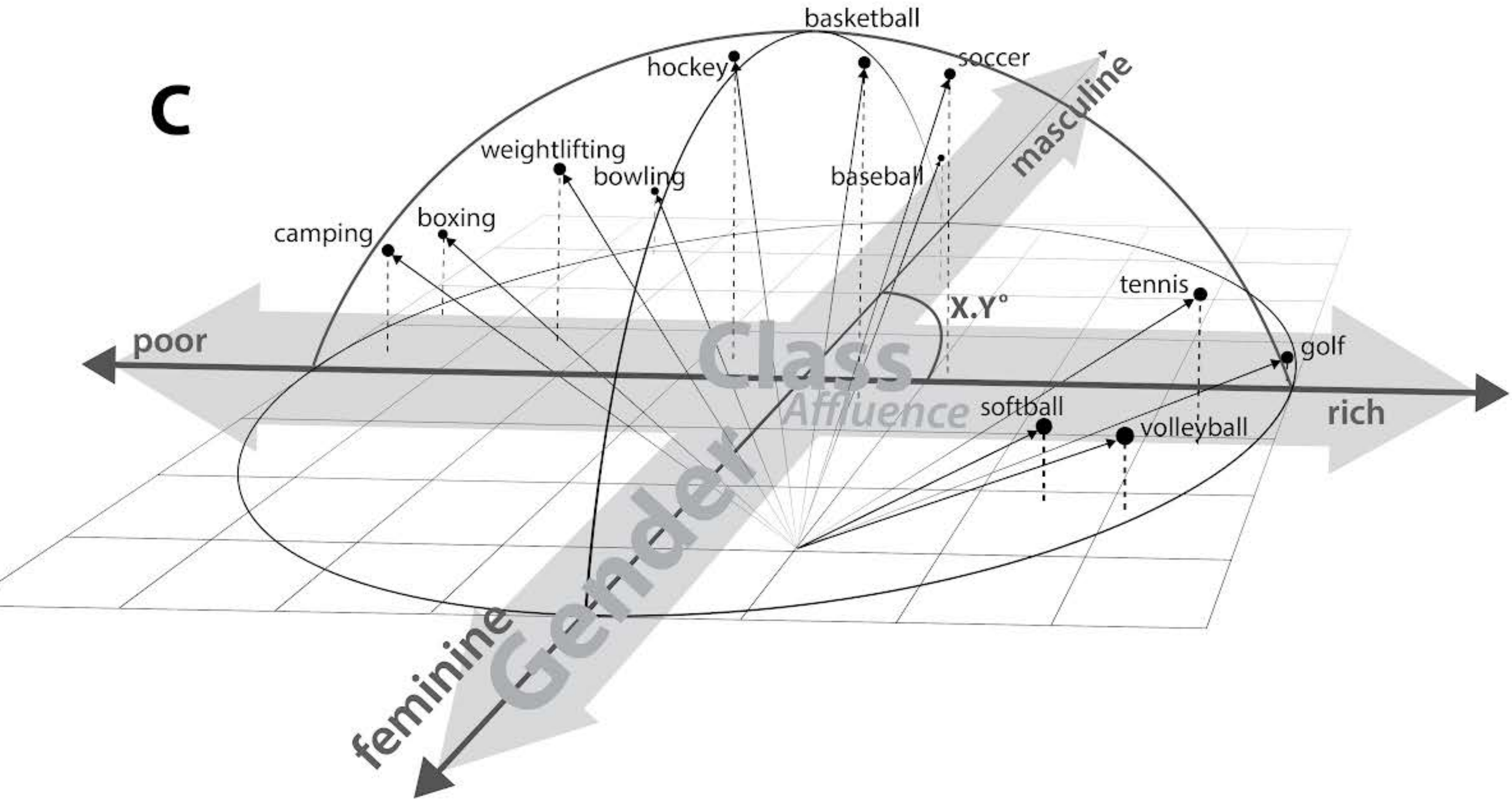
In High-dimensions, ***most evolution is neutral***

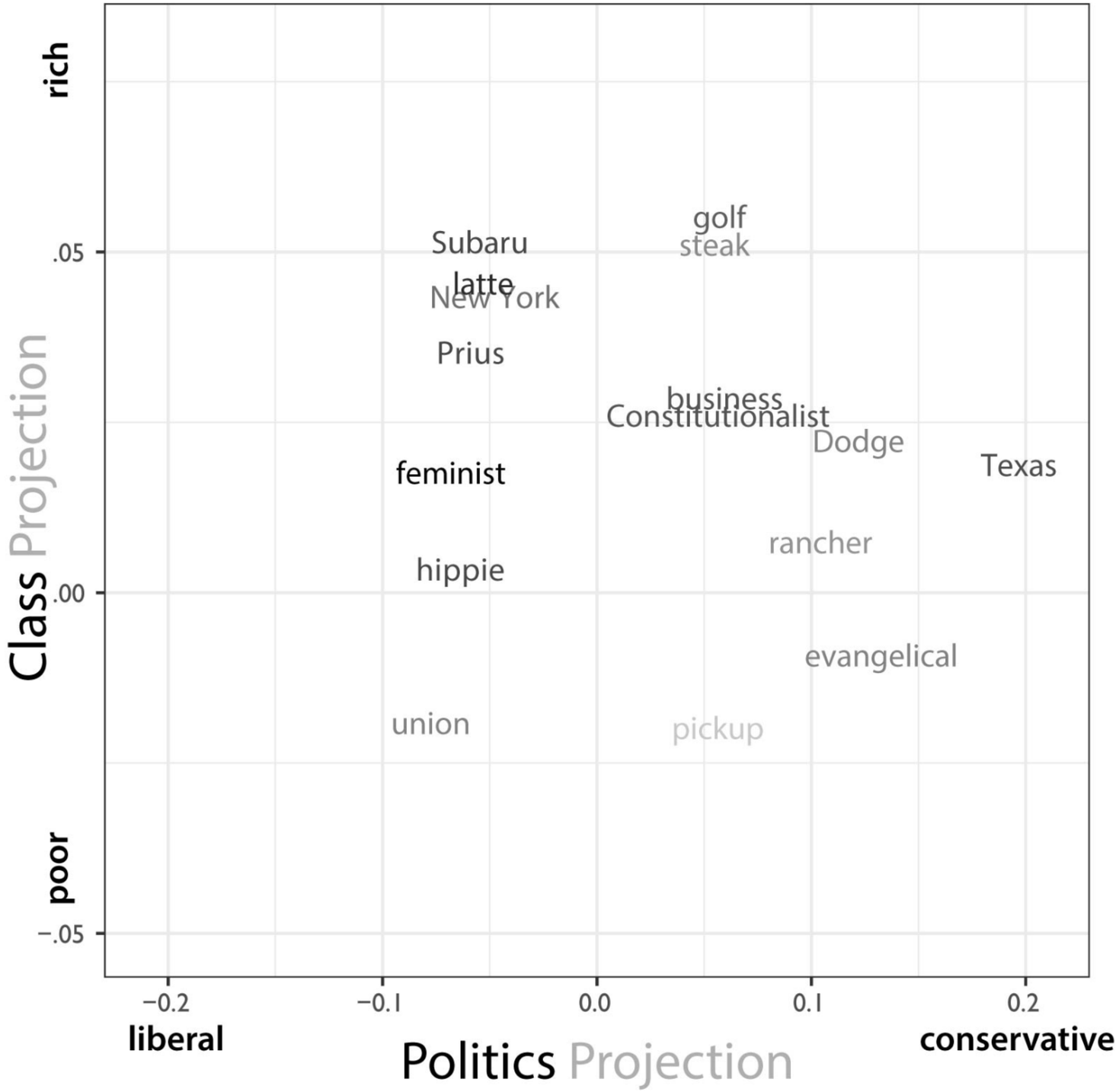
Growing a **Brain** through stories

- **“He carried his tuba”**
- **“She dropped her flute”**
- **“Steve’s pickup truck skidded to a halt”**
- **“‘Get out of the minivan, ma’am’”**


rich
richer
affluence
expensive
wealthy
poor
poorer
poverty
inexpensive
impoverished
Class
Affluence


B
camping
boxing
weightlifting
bowling
hockey
baseball
basketball
soccer
softball
volleyball
tennis
golf
poor
rich
Class
Affluence


C
basketball
hockey
soccer
weightlifting
bowling
baseball
camping
boxing
tennis
golf
softball
volleyball
X.Y°
poor
rich
Class
Affluence
masculine
feminine
Gender



Correlation between Multiple Dimensions **Class & Politics**

THE UNIVERSITY OF CHICAGO

DIVISION OF THE SOCIAL SCIENCES

Instructions: People tend to think of everyday items as being *masculine* or *feminine*. Using the sliding scales, with 0 representing "very feminine" and 100 representing "very masculine," please indicate how masculine or feminine you think each item is.

From 0 (very feminine) to 100 (very masculine), how would you rank **tennis**?

Very Feminine — Neither — Very Masculine

0 10 20 30 40 50 60 70 80 90 100

From 0 (very feminine) to 100 (very masculine), how would you rank **baseball**?

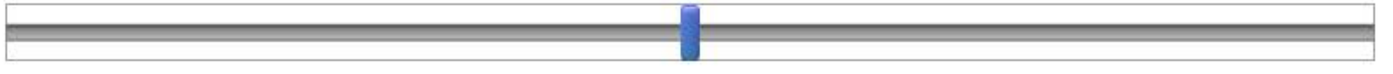
Very Feminine — Neither — Very Masculine

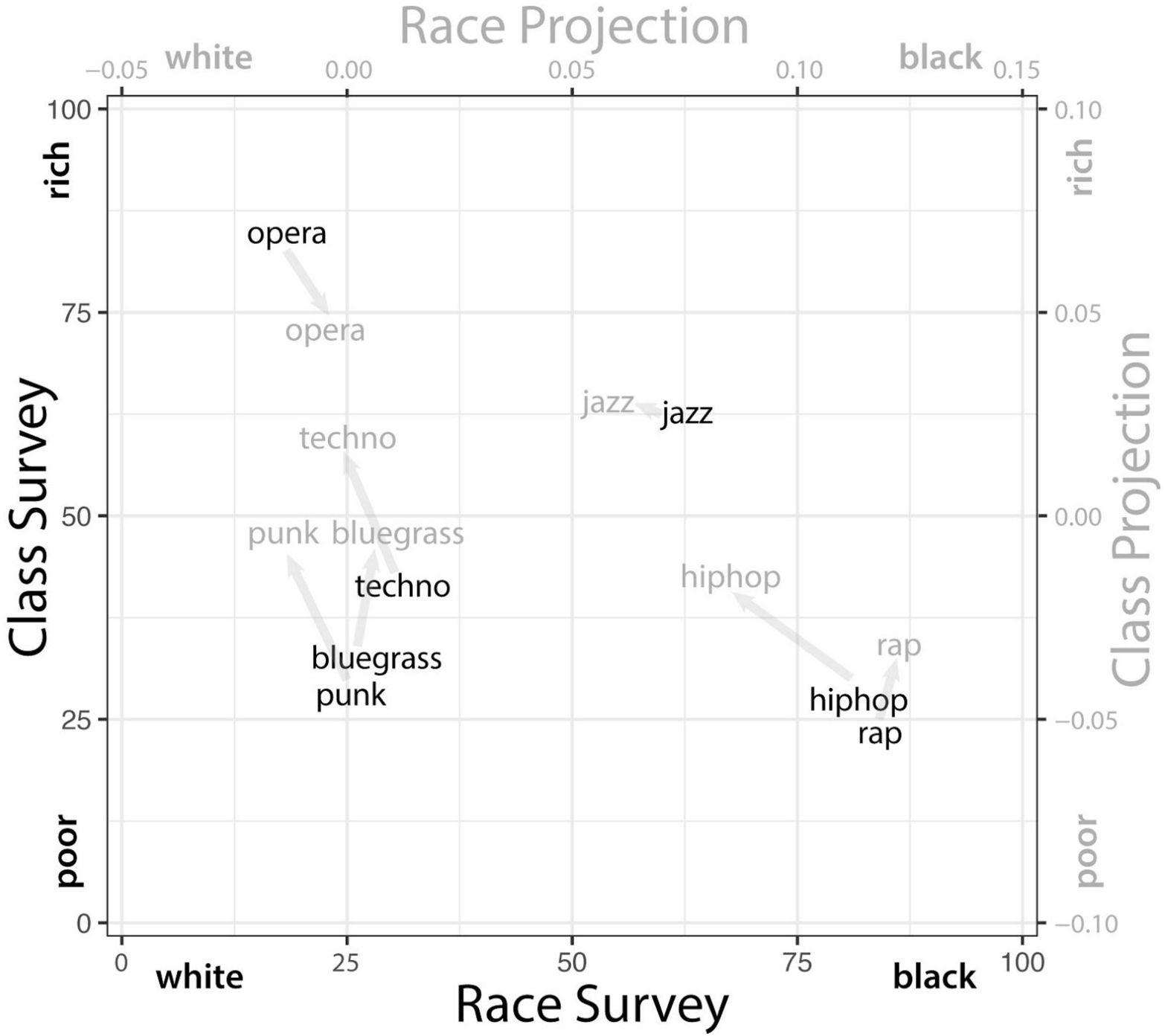
0 10 20 30 40 50 60 70 80 90 100

From 0 (very feminine) to 100 (very masculine), how would you rank **hockey**?

Very Feminine — Neither — Very Masculine

0 10 20 30 40 50 60 70 80 90 100





Correlation between **MTurk Survey** & **Contemporary Embeddings**

Testing Ecological Validity

Table 1. Pearson correlation between survey estimates and projection of word vector on cultural dimension in embedding (Google News text)

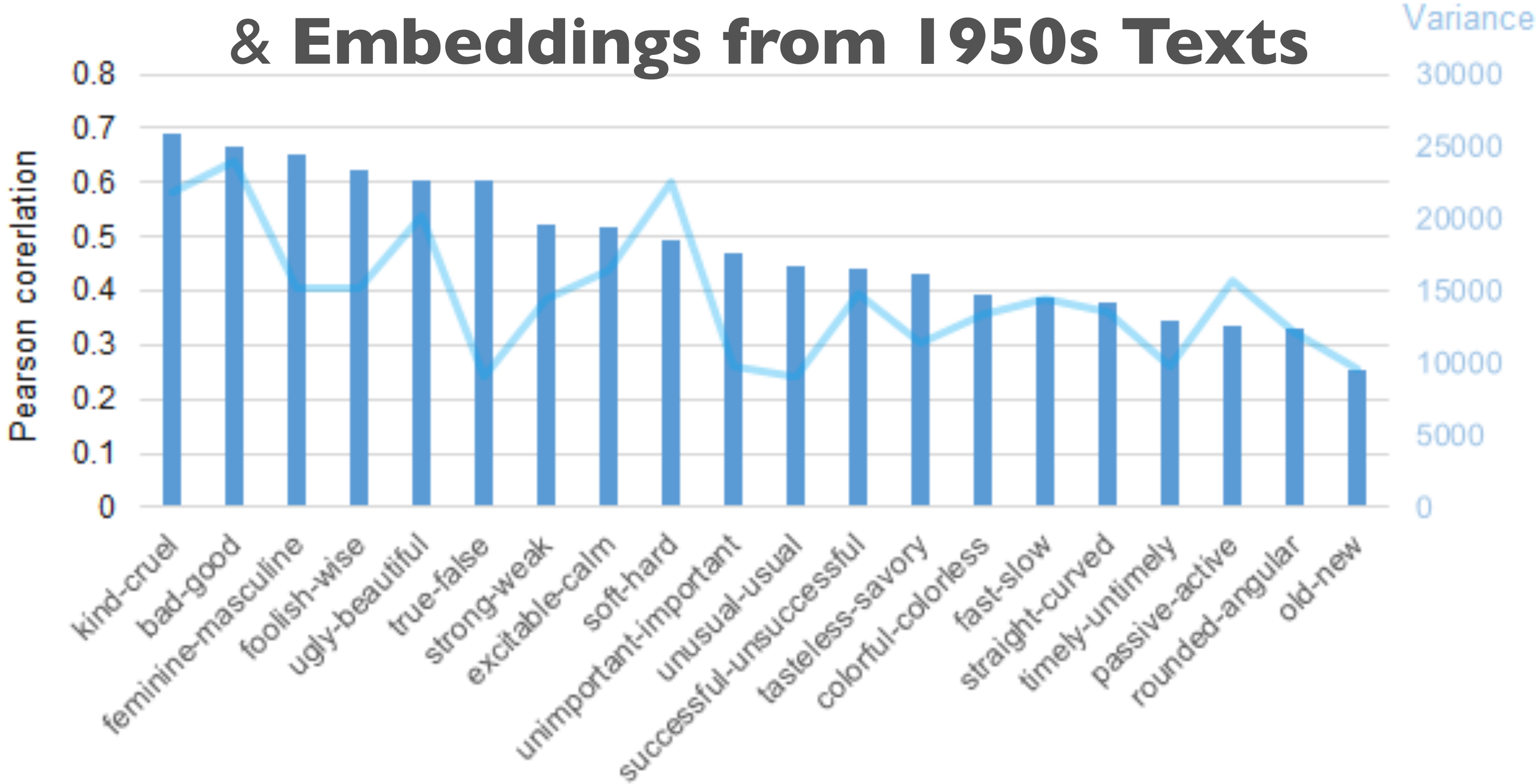
	Correlation: unweighted average	Correlation: weighted average
Gender dimension	0.869	0.928
Class dimension	0.520	0.649
Race dimension	0.699	0.813

Discovering the Most Explanatory Dimensions

Top five nearest cultural dimensions to Gender, Class, and Race

	Gender	*Class*	*Race*
1.	Maternal-Paternal	Educated-Uneducated	Outer-Inner
2.	Fashionable-Unfashionable	High-Low	Imperfect-Perfect
3.	Bisexual-Homosexual	Desirable-Undesirable	Troubled-Untroubled
4.	Emotional-Cerebral	Complex-Simple	Unsavory-Savory
5.	Rugged-Delicate	Privileged-Underprivileged	Unfriendly-Friendly

Correlation between **1958 Semantic 'Atlas'** & **Embeddings from 1950s Texts**



Classify Objects **along** Myriad Dimensions

Controlling - Passive

Honest - Dishonest

White - Black

Patient - Impatient

Safe - Dangerous

Populous - Principled

Classy - Dumpy

Confused - Clear

Male - Female

Potent - Impotent

Rich - Poor

Conservative - Liberal

Tolerant - Intolerant

Smart - Stupid

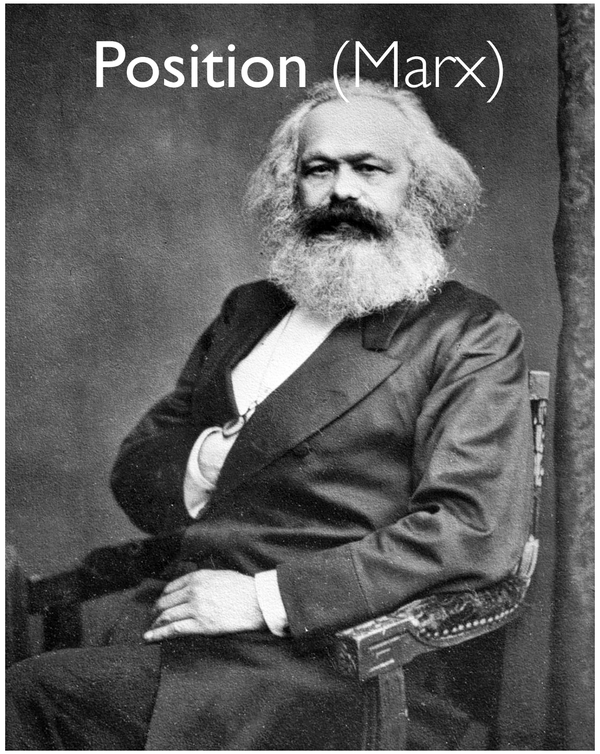
Multi-dimensional in theory / 1-2 in practice

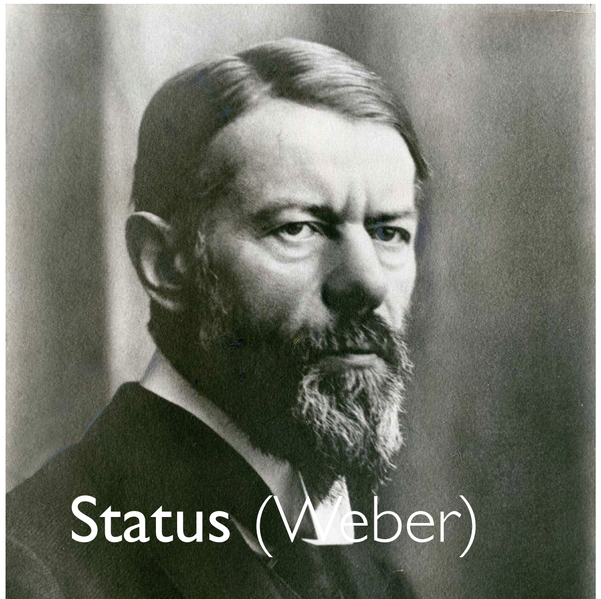
Social Class

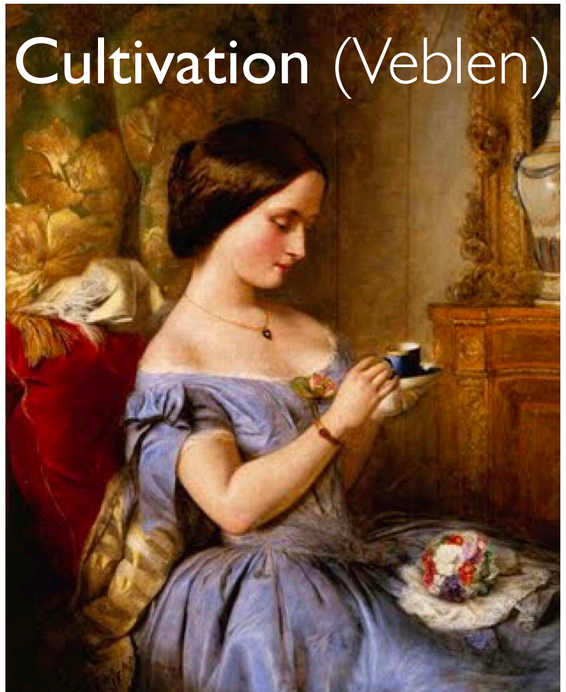
Multi-dimensional Construct

Affluence (Simmel)

Education (Fischer)





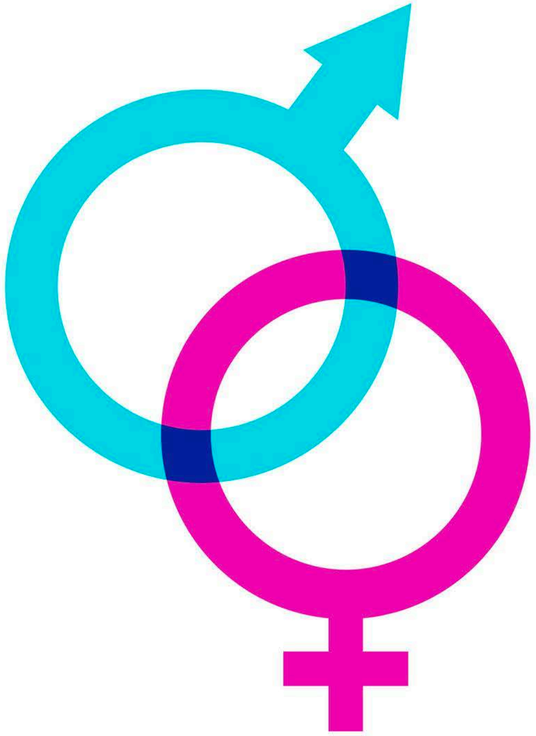


Social Class

Multi-dimensional Construct

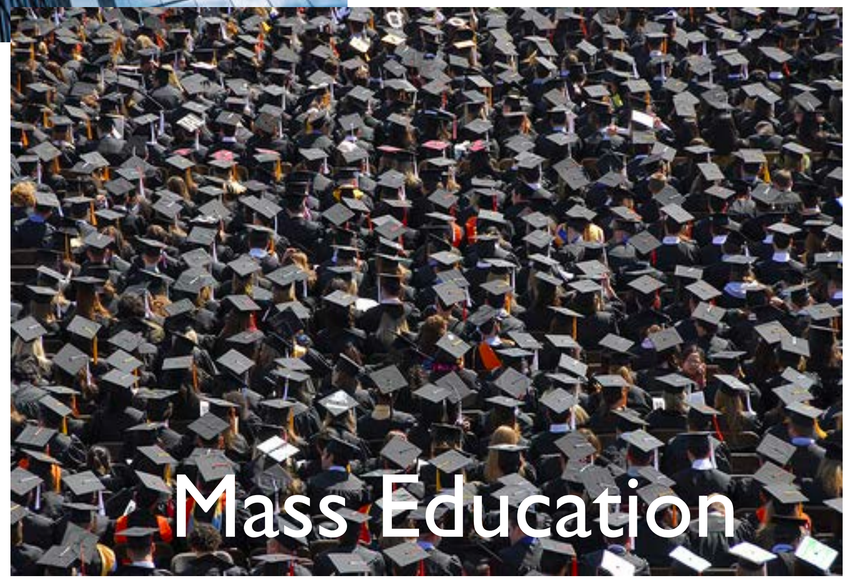
Morality (Fourcade, Lamont)

Gender (Hochschild)



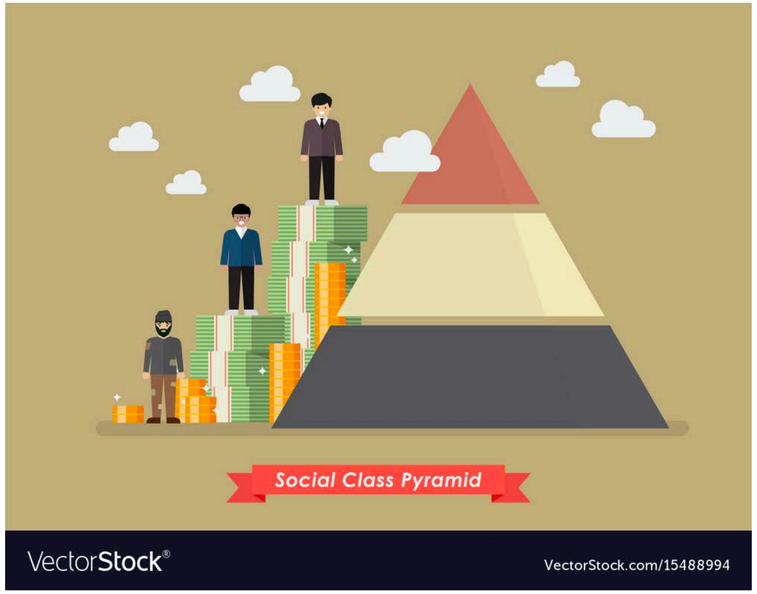
Race & Ethnicity (Du Bois)

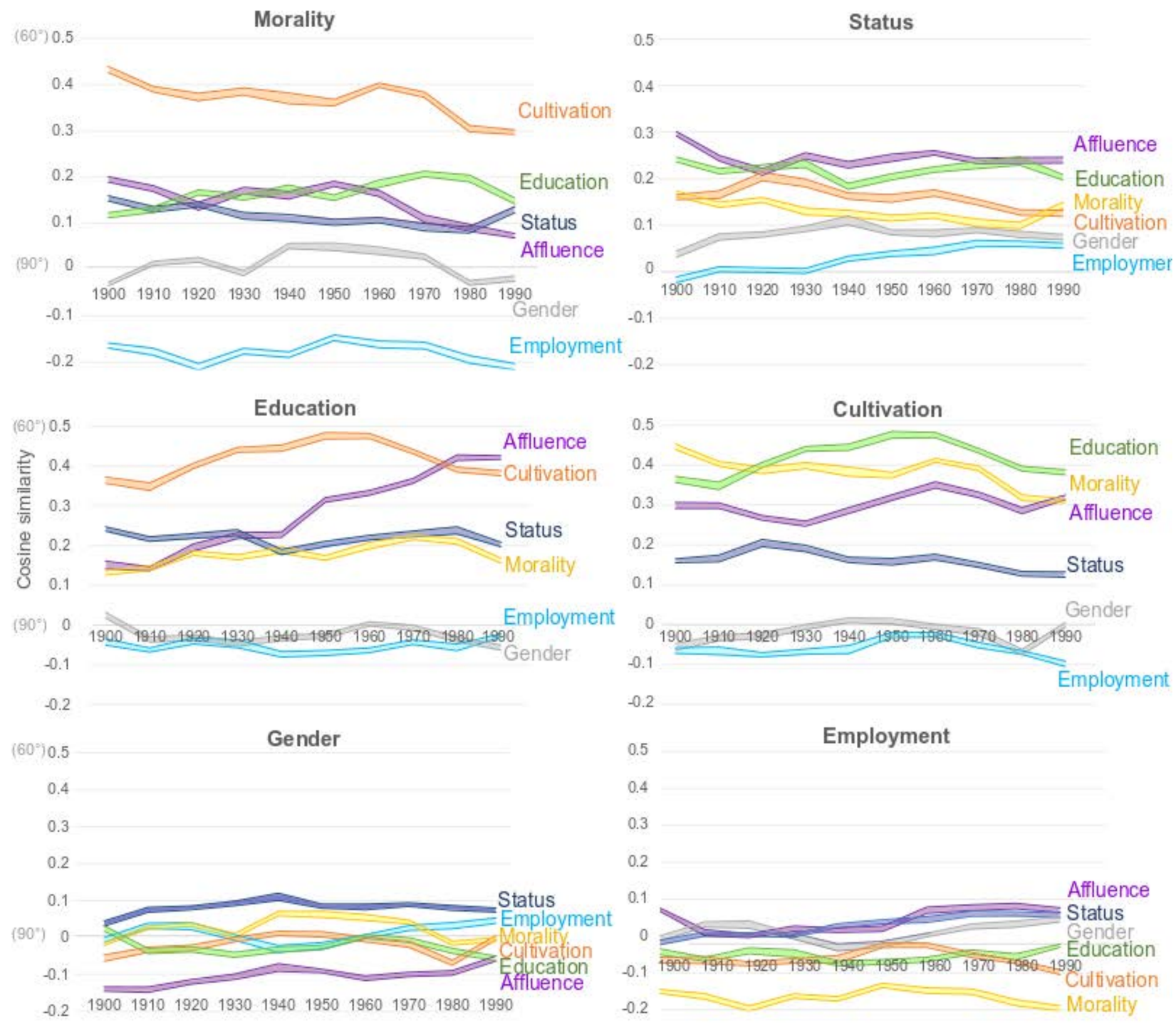
20th Century Transformations



Competing Narratives:

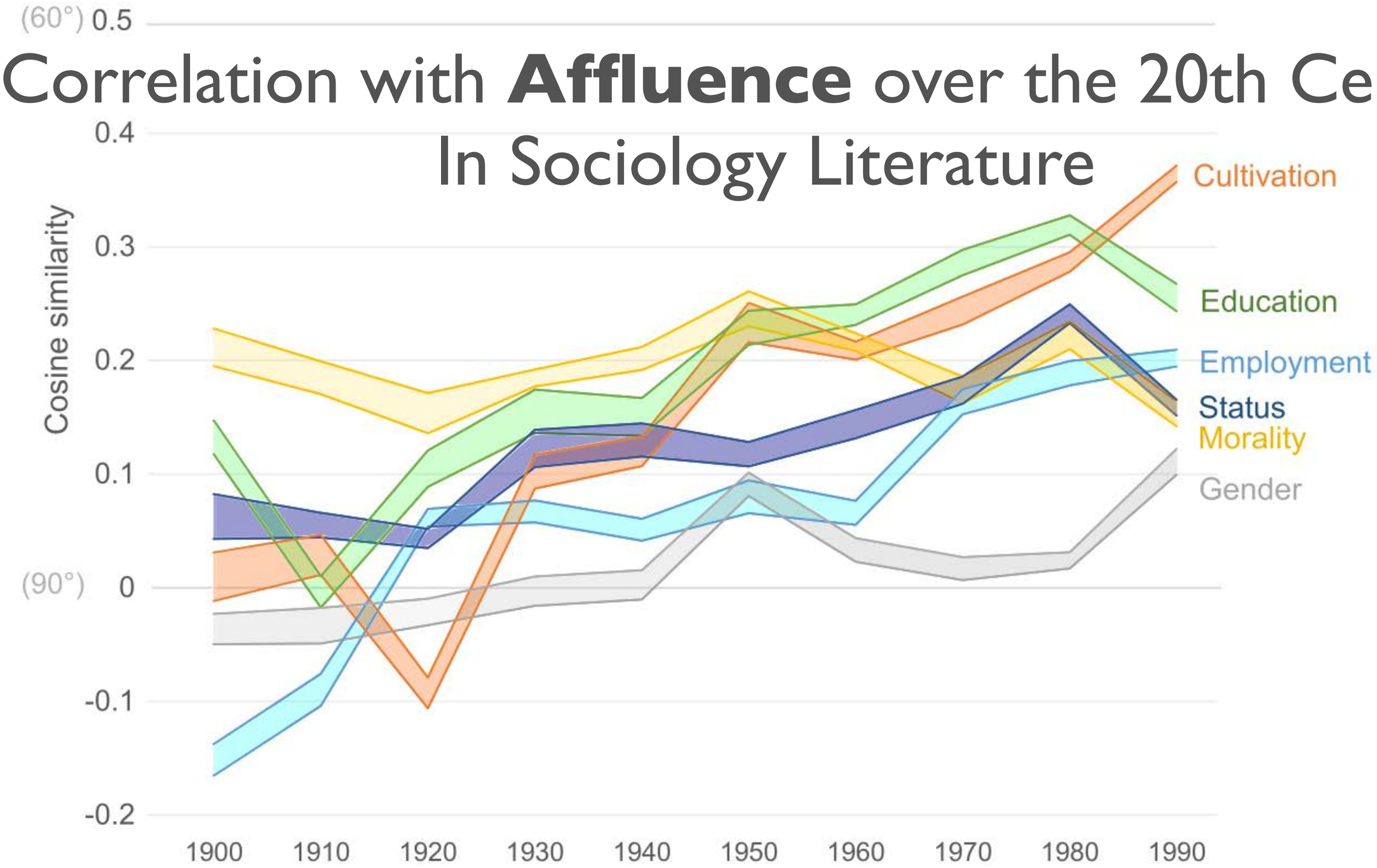
Death of Class (Clark)
Symbolic Distinction (Bourdieu)
Durable Difference (Grusky)
Persistent Multiplicity



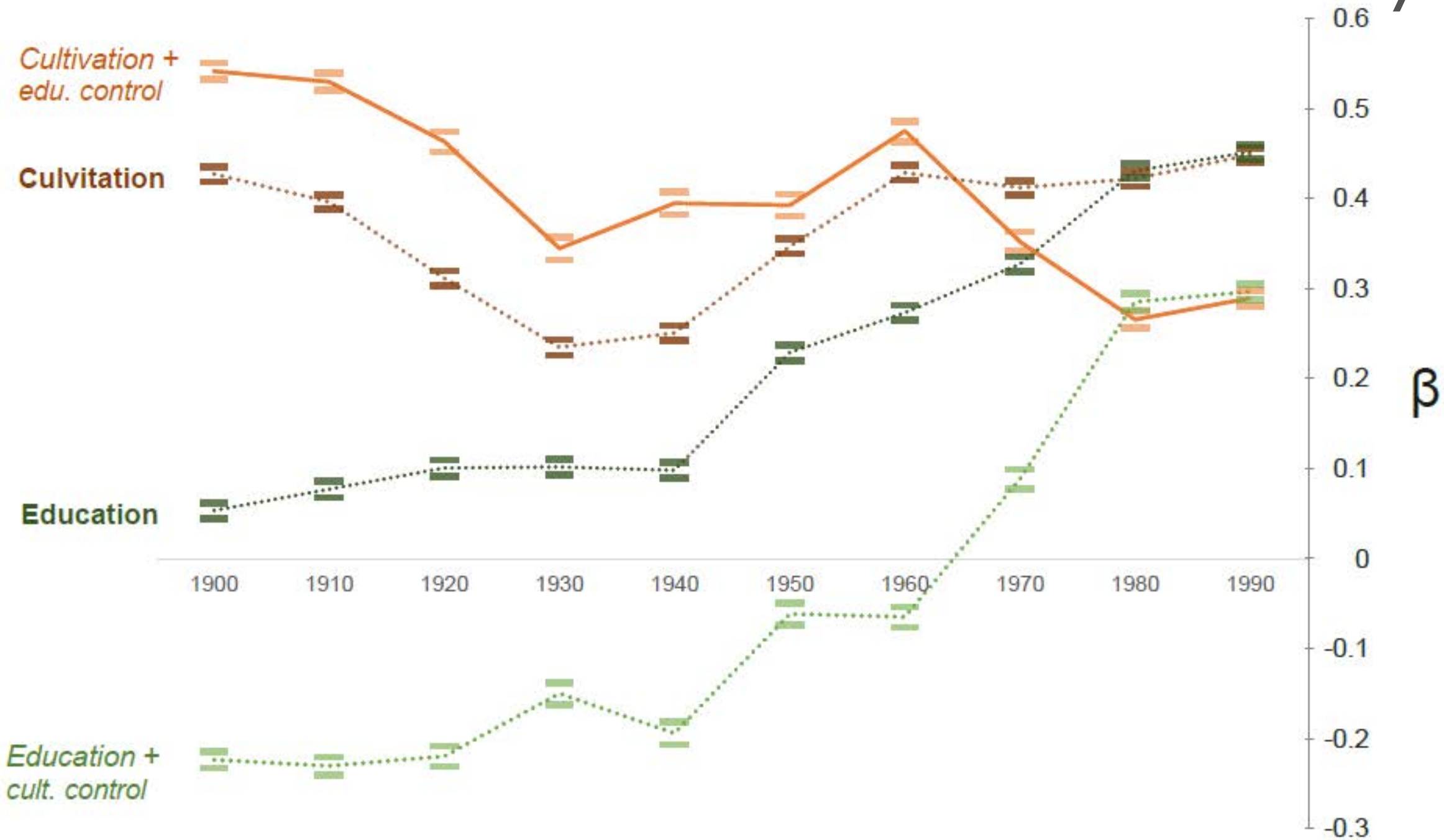


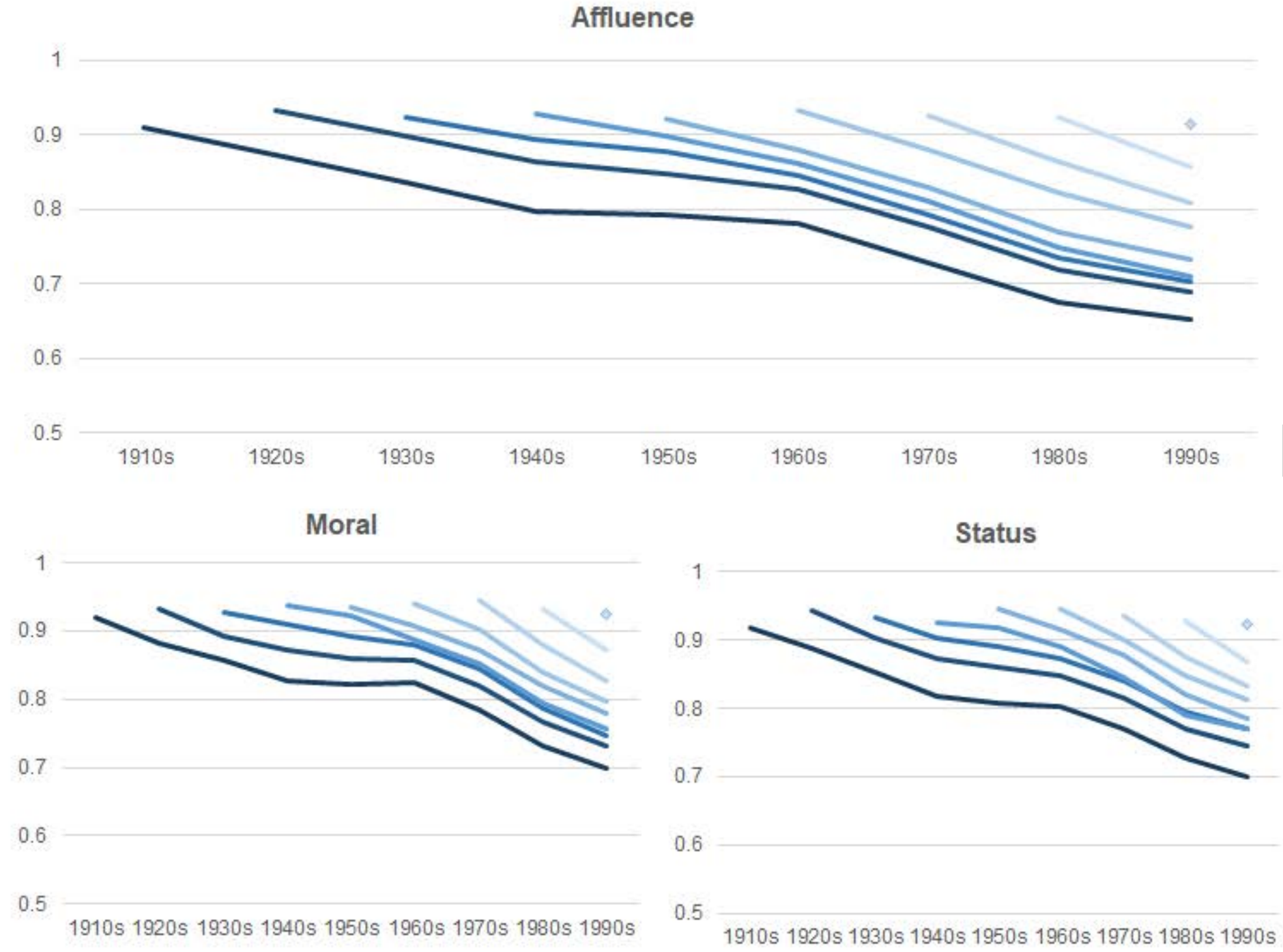
Correlations of Status over the 20th Century

Correlation with **Affluence** over the 20th Century In Sociology Literature



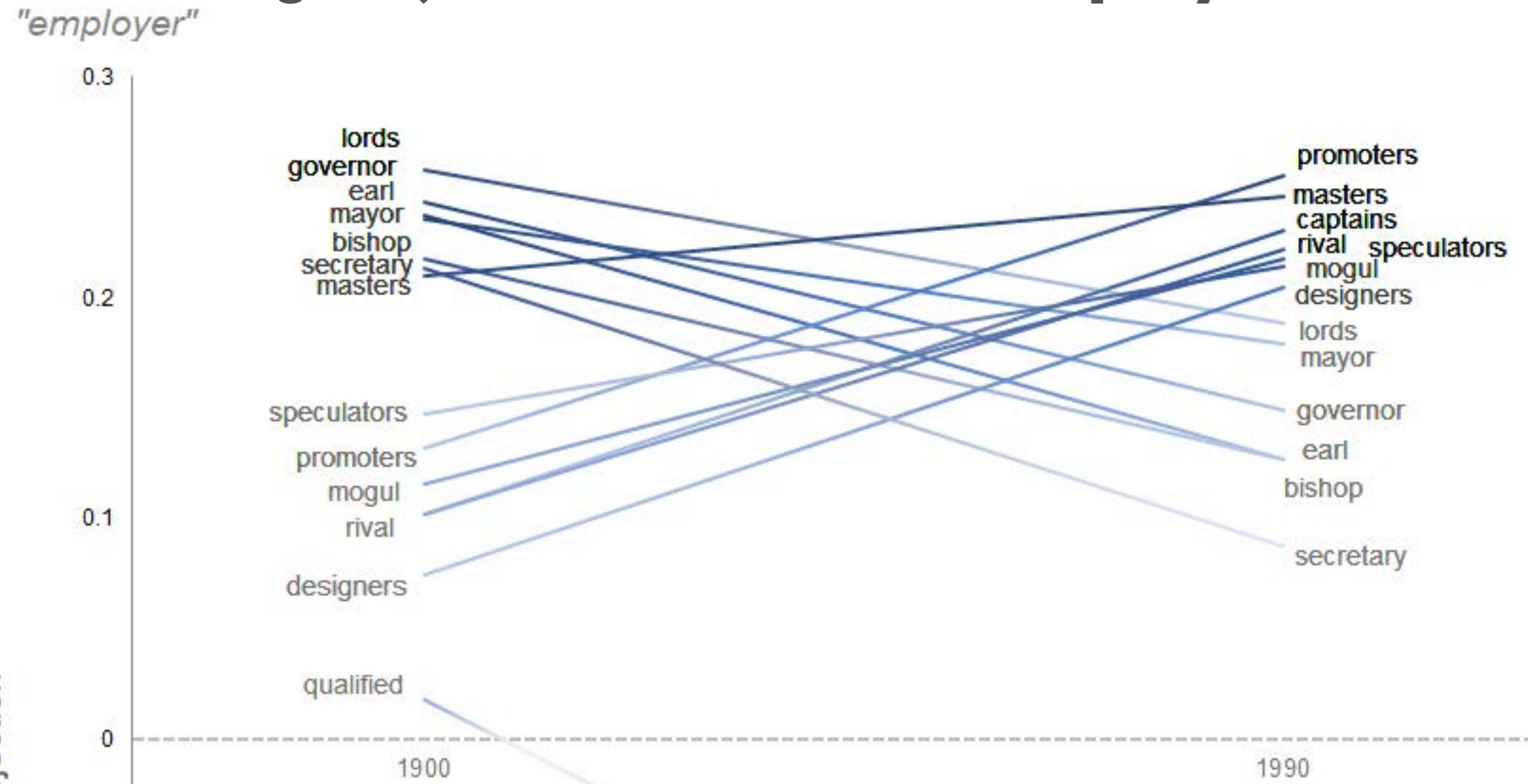
Prediction of **Affluence** over the 20th Century

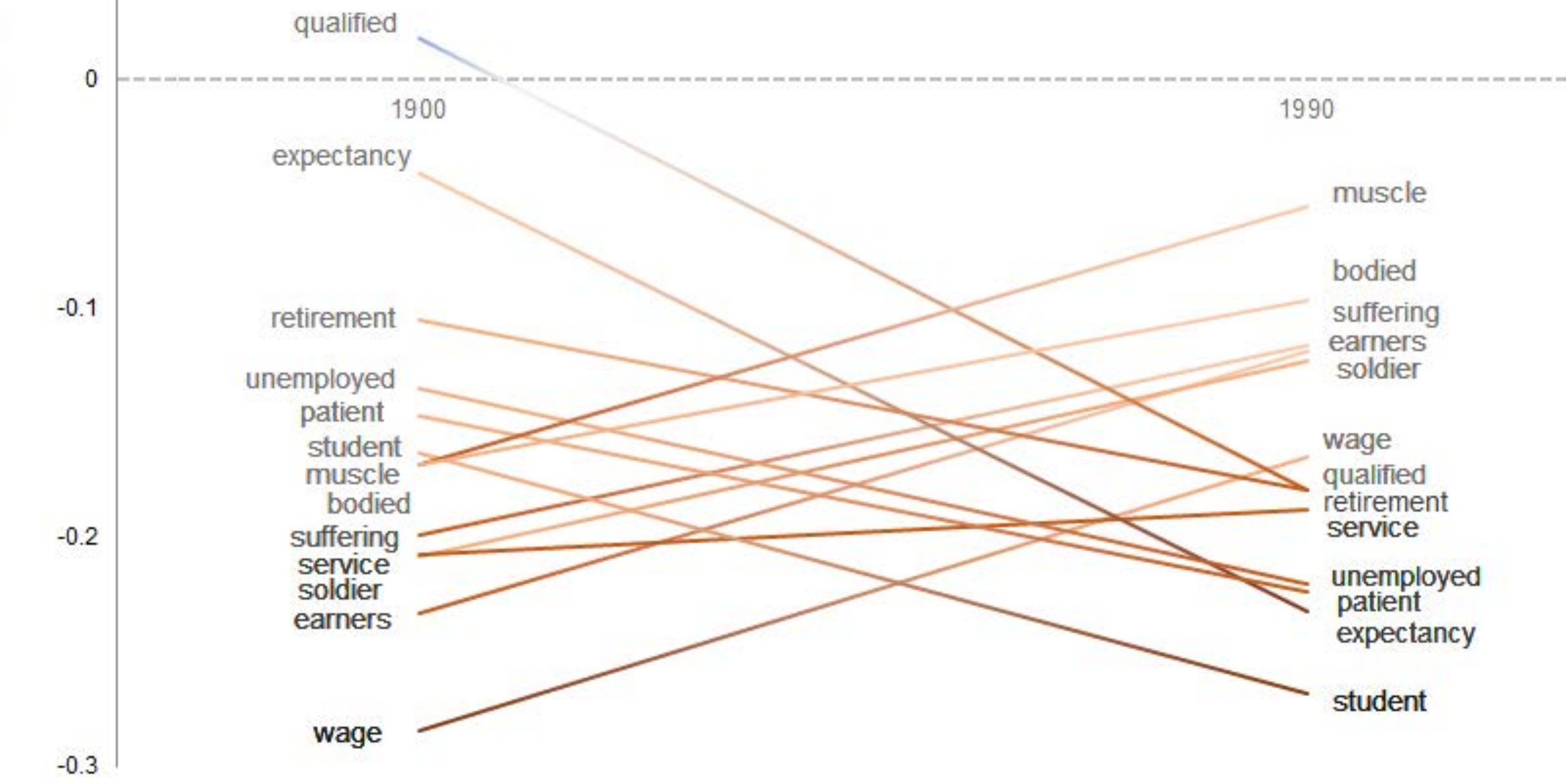




Word Projections over the 20th Century

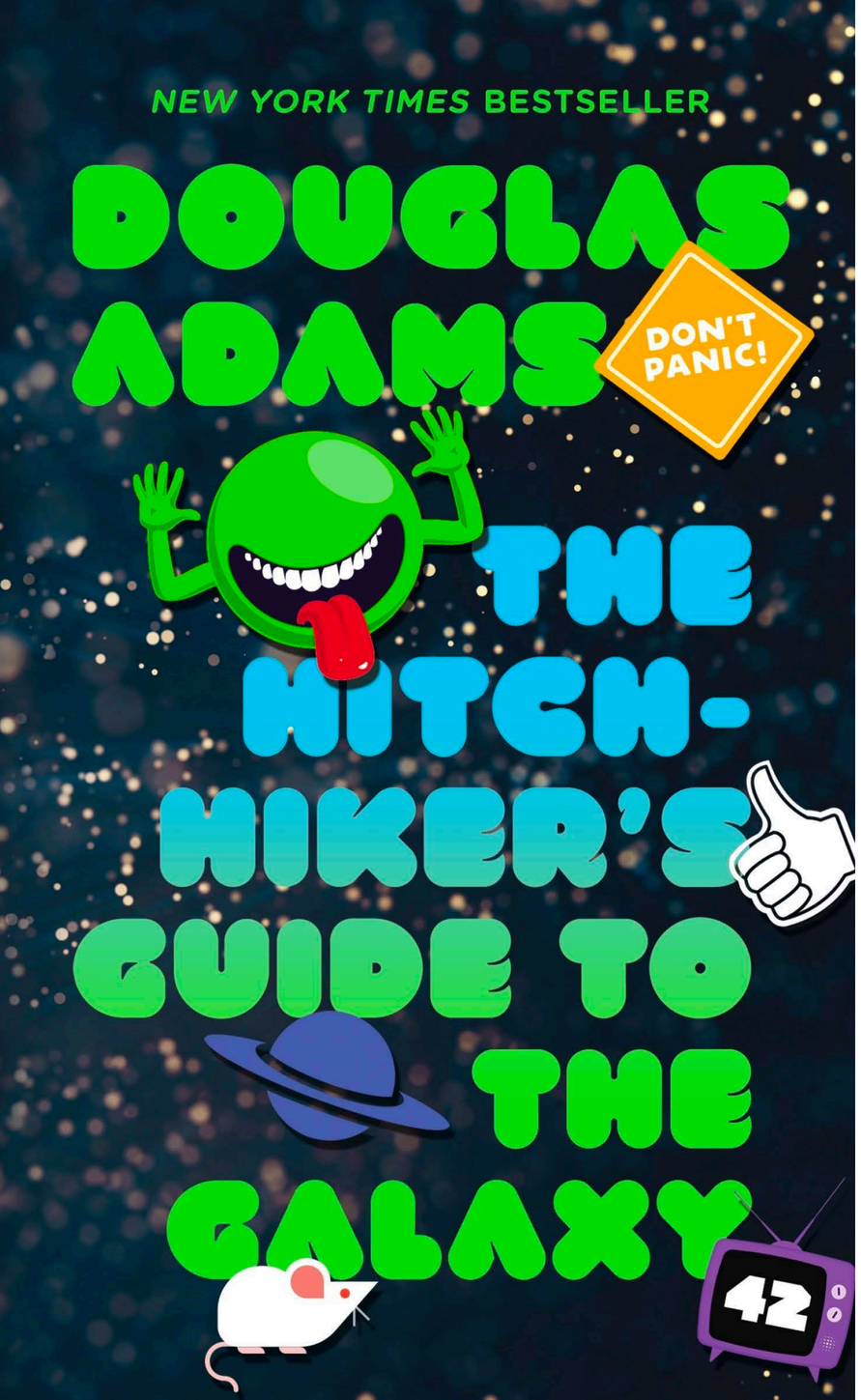
Shifting Projection of Words on **Employer/ee**





"employee" Shifting Projection of Words on **Employer/ee**

One dimension to rule them all



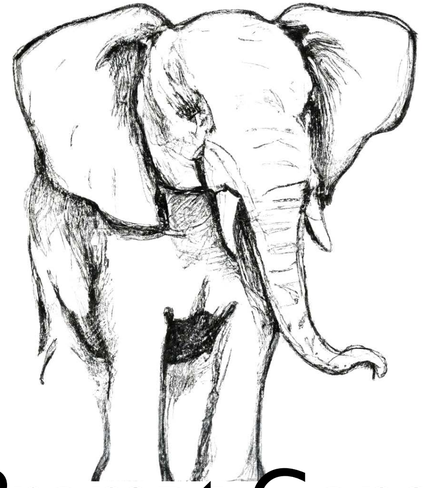
Not **Theory Free**

What is the dimension that explains
THE MOST variation in the world?

Steriodal - Nonsteriodal

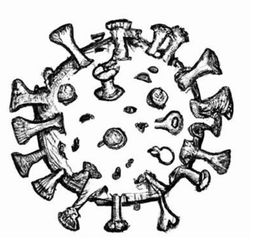
Instantiating Deep LLMs as Simulated Subjects

Predict the Future



Partisan Priming

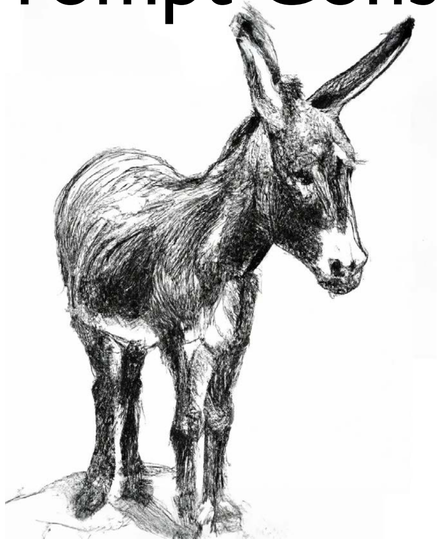
I am a strong conservative and a lifelong Republican. In 2016, I was proud to vote for Donald Trump and I think that the Democrats have been a disaster for this country.



COVID Context

Lately, one of the biggest political issues has been the COVID-19 pandemic caused by the new coronavirus. There is a lot of controversy around whether…

Prompt Construction



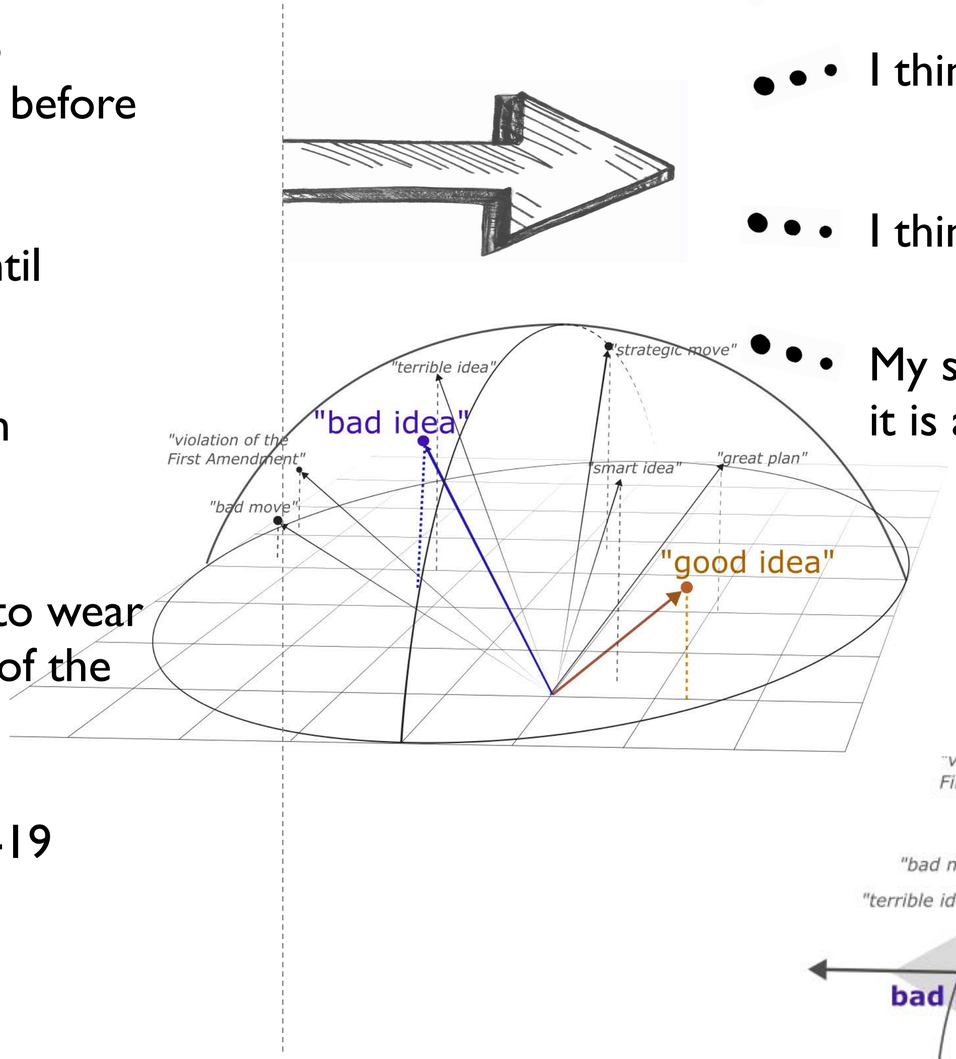
or…

I am a strong liberal and a lifelong Democrat. In 2016, I was proud to vote for Hillary Clinton and I think that the Republicans have been a disaster for this country.

Prompt Construction

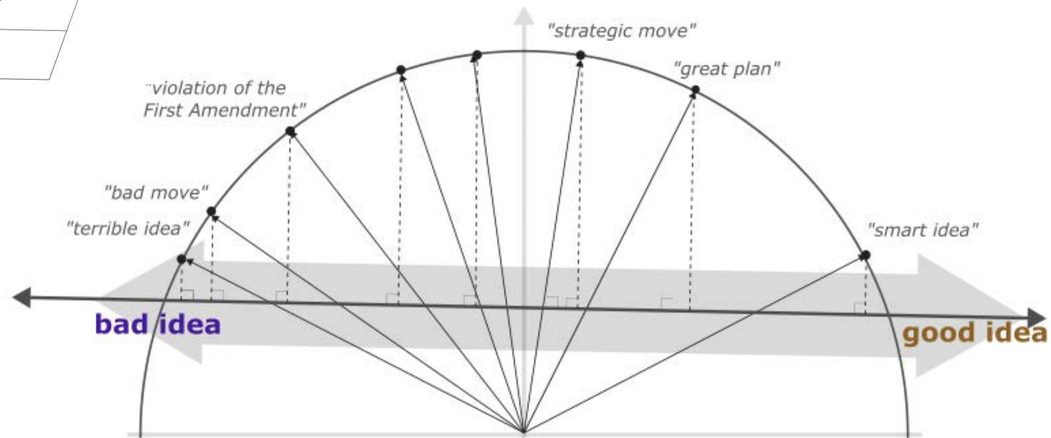
Issue

- students should be required to receive the COVID-19 vaccine before returning to school
- businesses should be closed until rates of infection go down.
- the CDC is doing a good job in handling the situation.
- individuals should be required to wear face masks to slow the spread of the virus.
- or not to get the new COVID-19 vaccine.



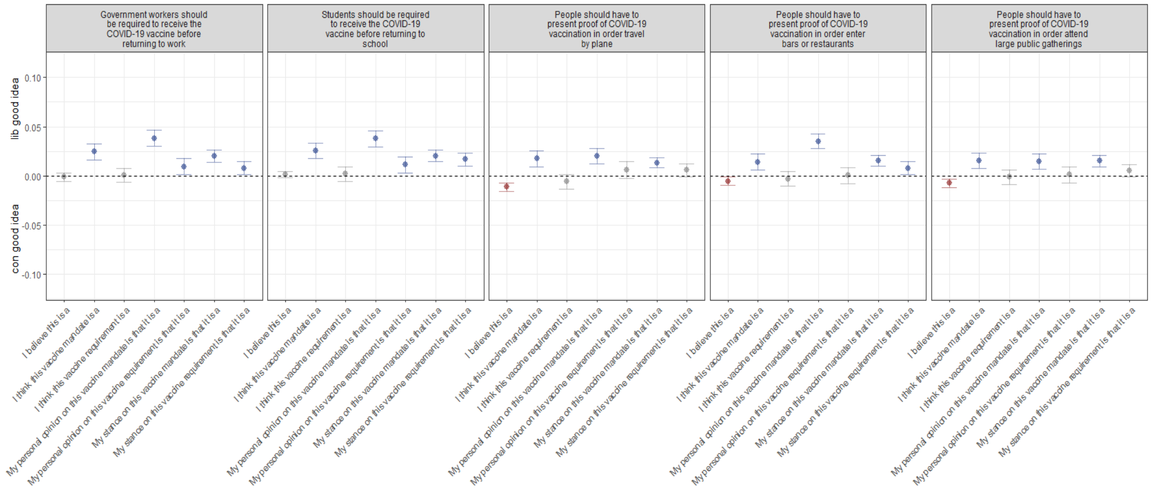
Endings

- I think this is a…
- I think this [vaccine mandate] is a…
- I think this [vaccine requirement] is a…
- My stance on this [vaccine mandate] is that it is a…

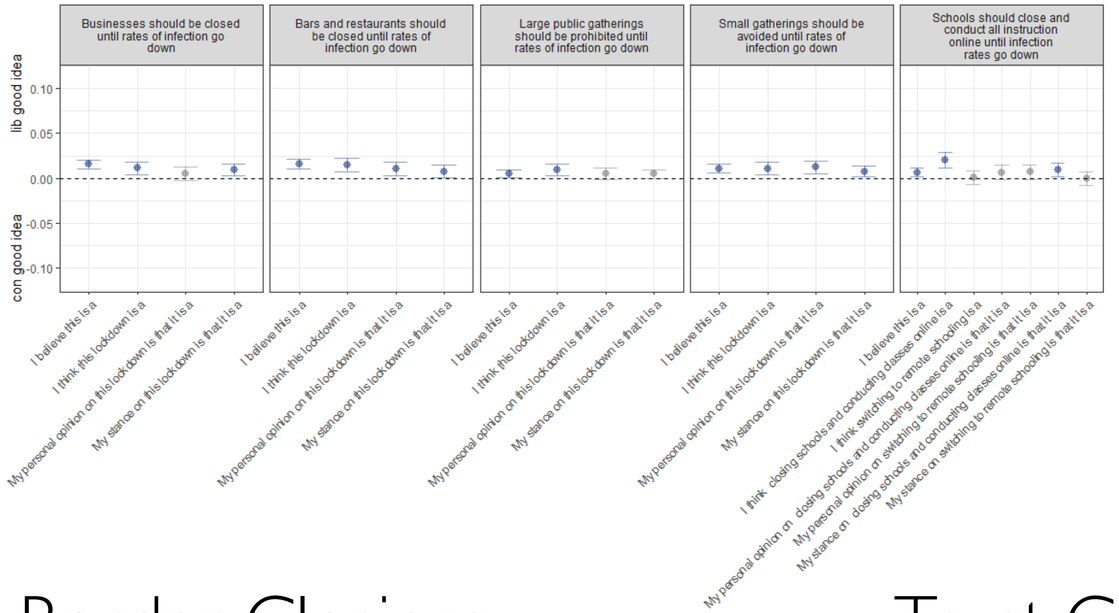


Pandemic attitudes in October 2019

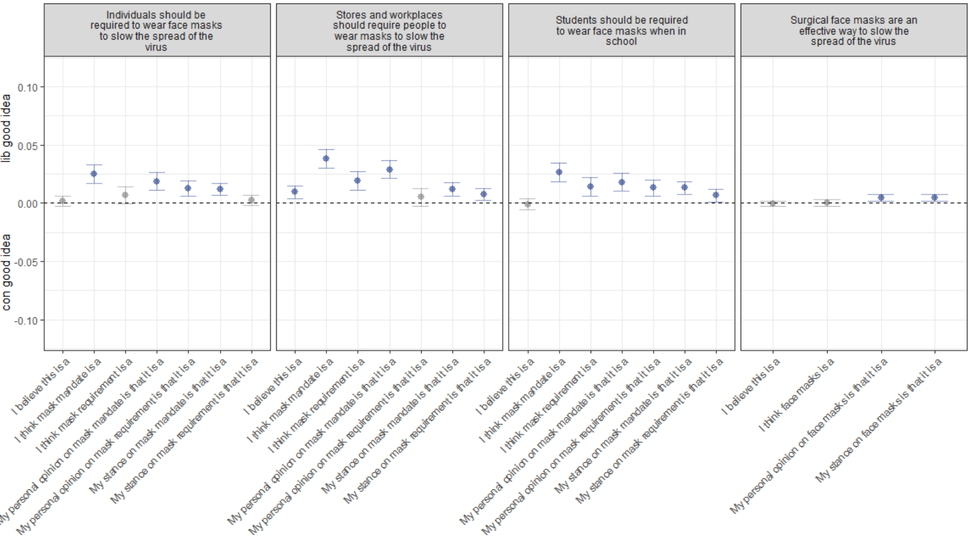
Vaccine Mandates



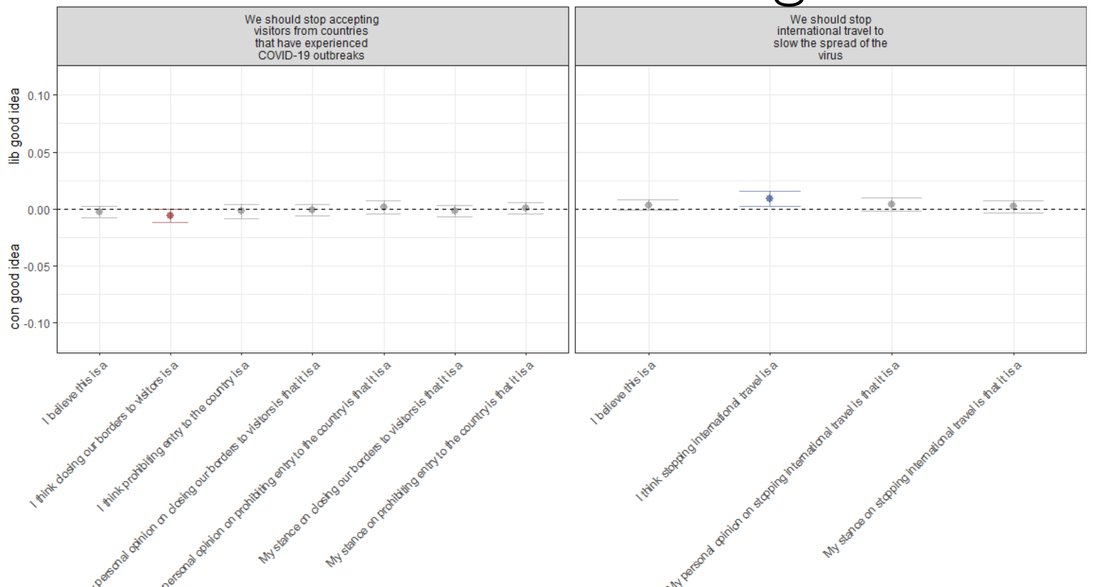
Lockdown



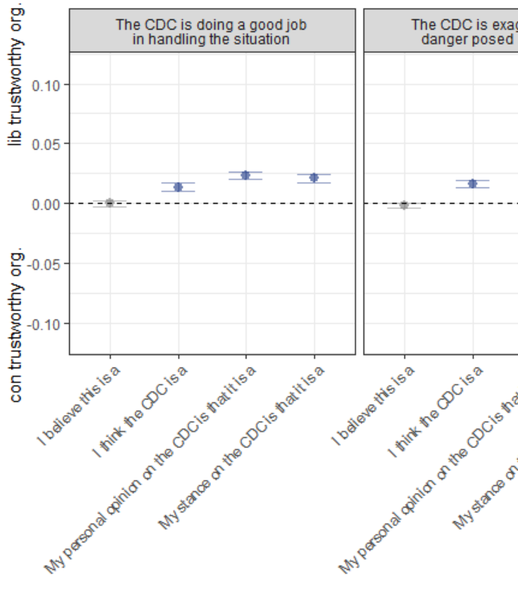
Mask Mandates



Border Closings

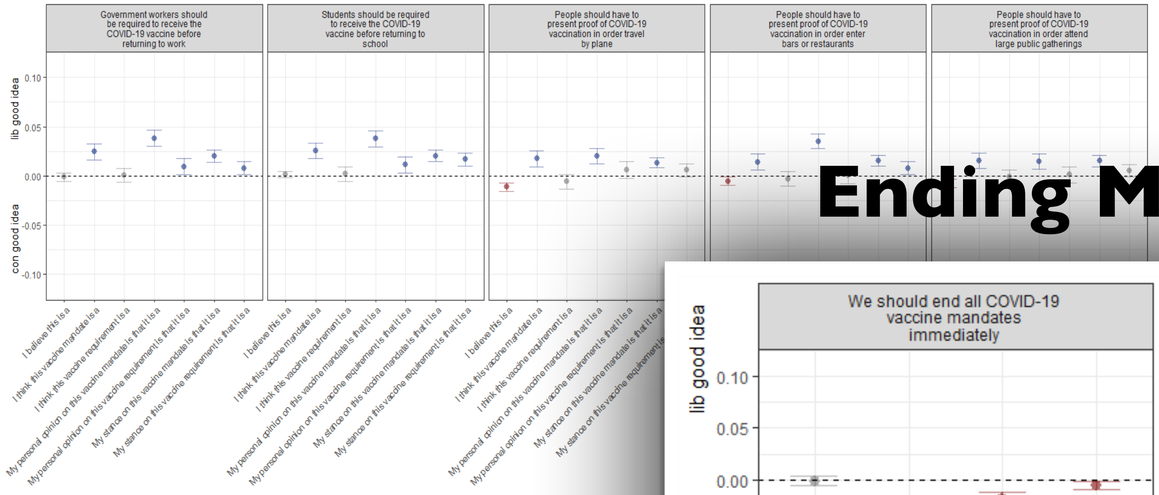


Trust CDC

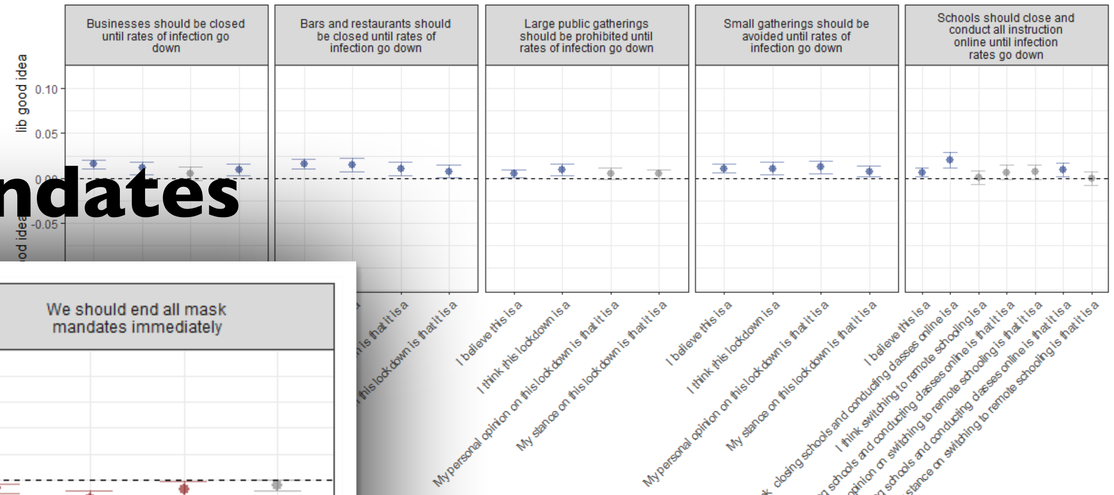


Pandemic attitudes in October 2019

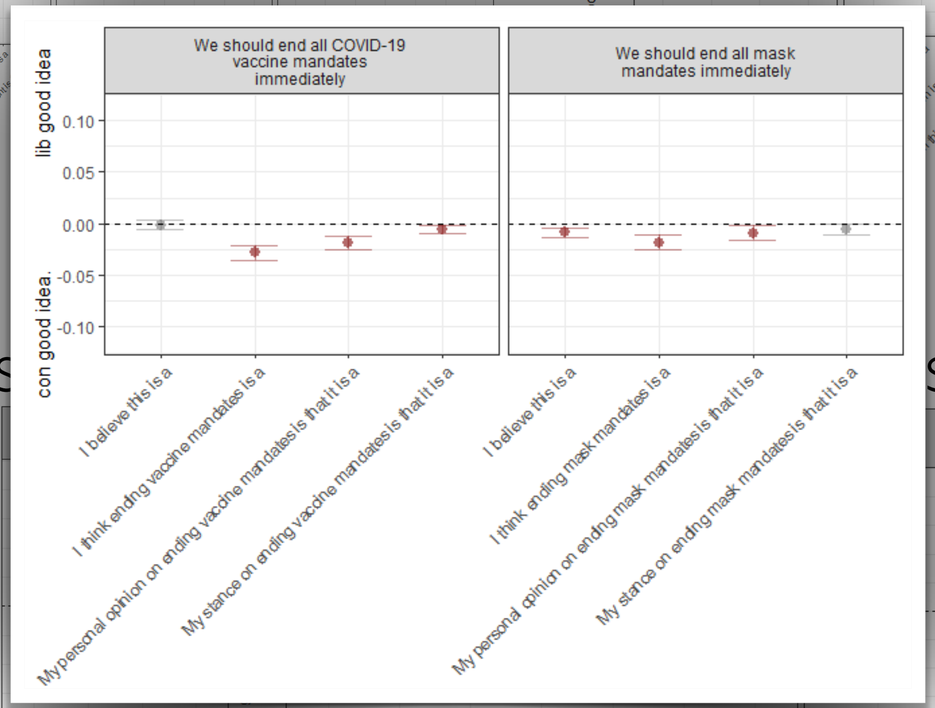
Vaccine Mandates



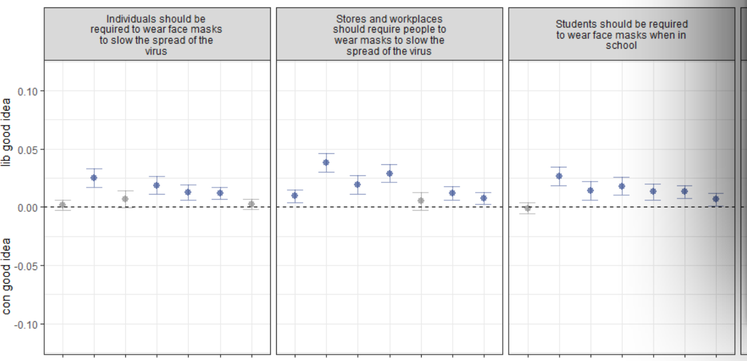
Lockdown

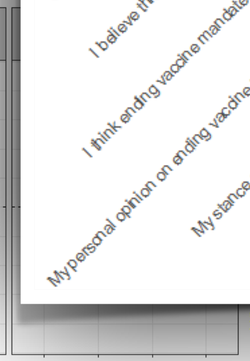


Ending Mandates

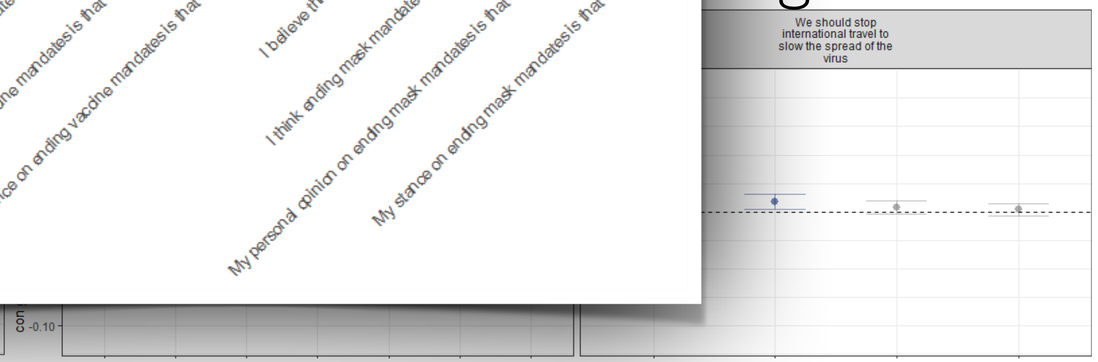


Mask Mandates

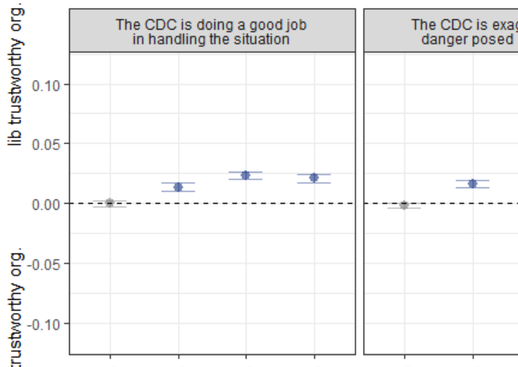




sings



Trust CDC



Interaction Models

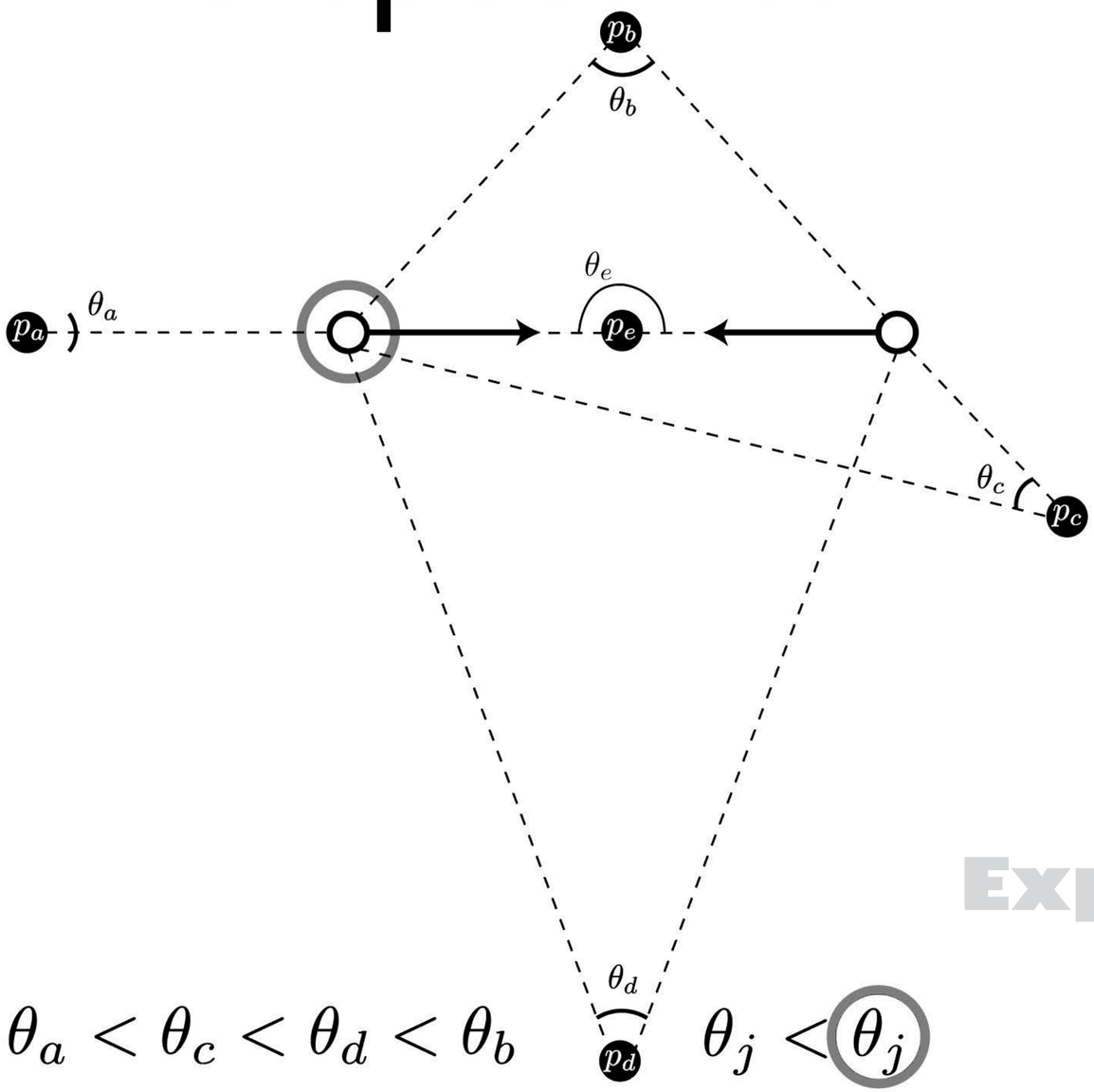
Michael Bernstein et al

- Predicting Conflict (& its content/character)
- Predicting Collaboration (& its content/character)

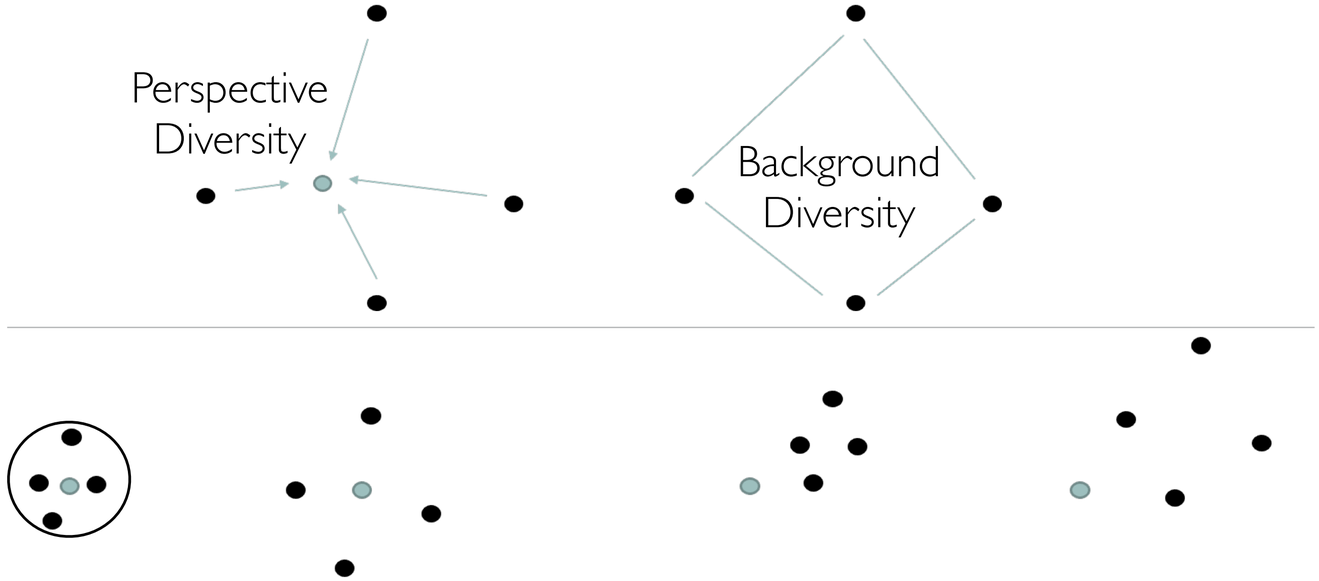
Simulating:

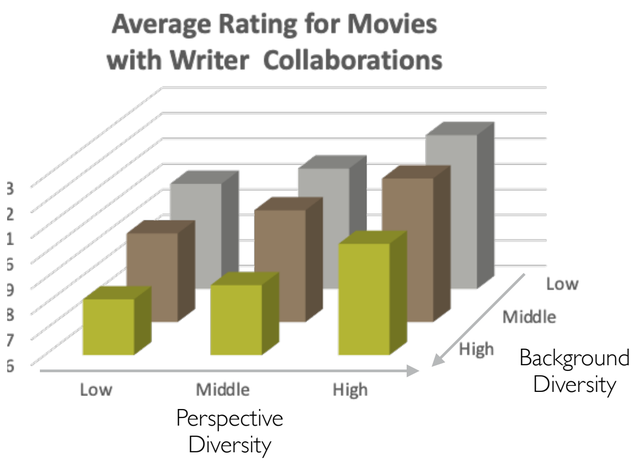
- Dates and Coupling
- Conversation & conferences that haven't or couldn't happen

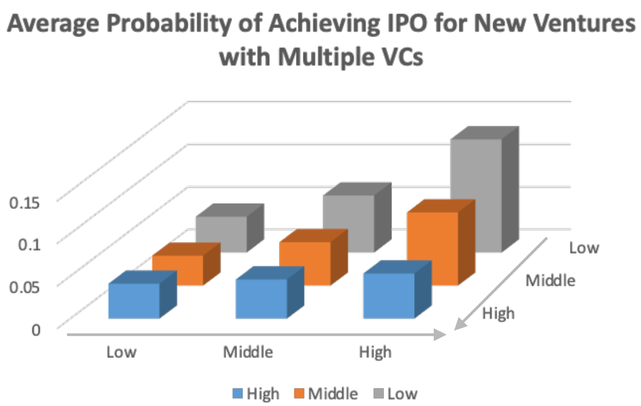
Diversity of **Perspectives**



$$\theta_a < \theta_c < \theta_d < \theta_b \qquad \theta_j < \theta_j$$





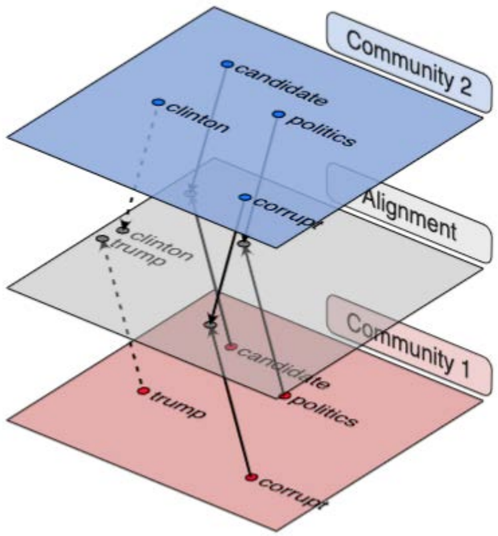


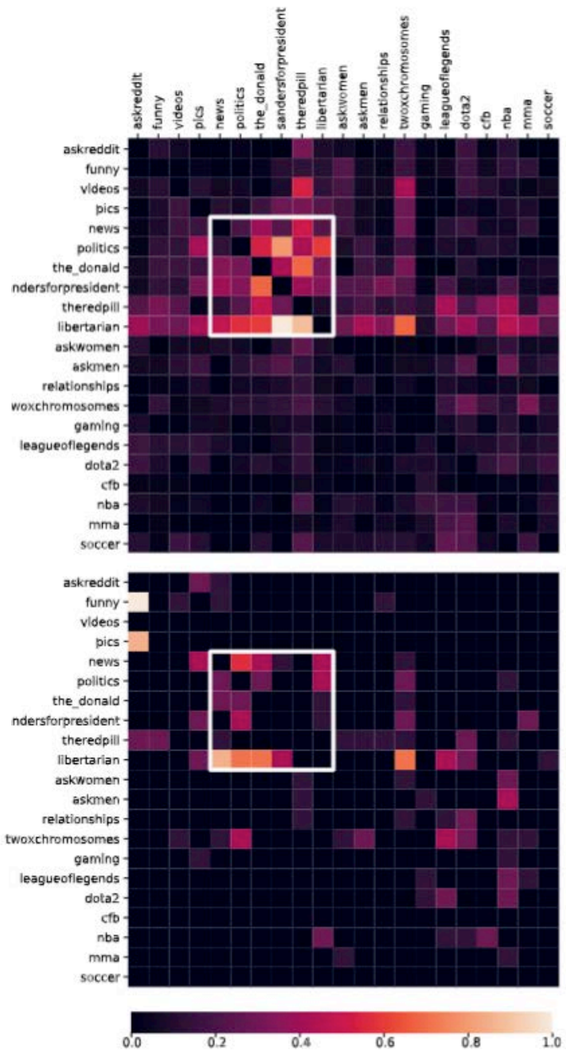
GPT4 Experiment

Optimal **Diversity**

Danish Experiment

Aligning Social Media Worldviews





r/politics	r/the_donald	Alignment
democrat	republican	0.8562
republican	democrat	0.8501
leftwing	rightwing	0.8307
socialized_medicine	universal_healthcare	0.8041
magas	libtards	0.6578

r/the_donald	r/politics	Alignment
republican	democrat	0.8570
democrat	republican	0.8527
prolife	prochoice	0.8435
foxnews	cnn	0.7960
pocahontas	elizabeth_warren	0.6694

r/askwomen	r/askmen	Alignment
son	daughter	0.7675
daughter	son	0.7621
husband	wife	0.7503
father	mother	0.7445
brother	sister	0.7145
girlfriend	boyfriend	0.7032
wife	husband	0.6941
boyfriend	girlfriend	0.6708
uncle	aunt	0.6314

r/LeagueOfLegends	r/Dota2	Alignment
/r/summonerschool	/r/learndota2	0.8420
op.gg	dotabuff	0.8396
rito	volvo	0.8378
riot	valve	0.8003
aatrox	bloodseeker	0.6473

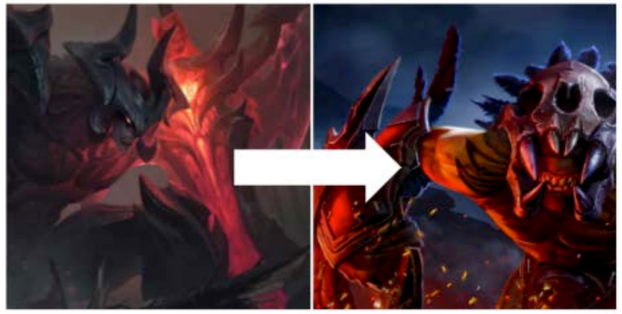


Figure 2: These are not the same character! The character on the left, "Aatrox" from *League of Legends*, projects to the character on the right, "Bloodseeker" from *Dota 2*.

Disentangling Social Media Influence

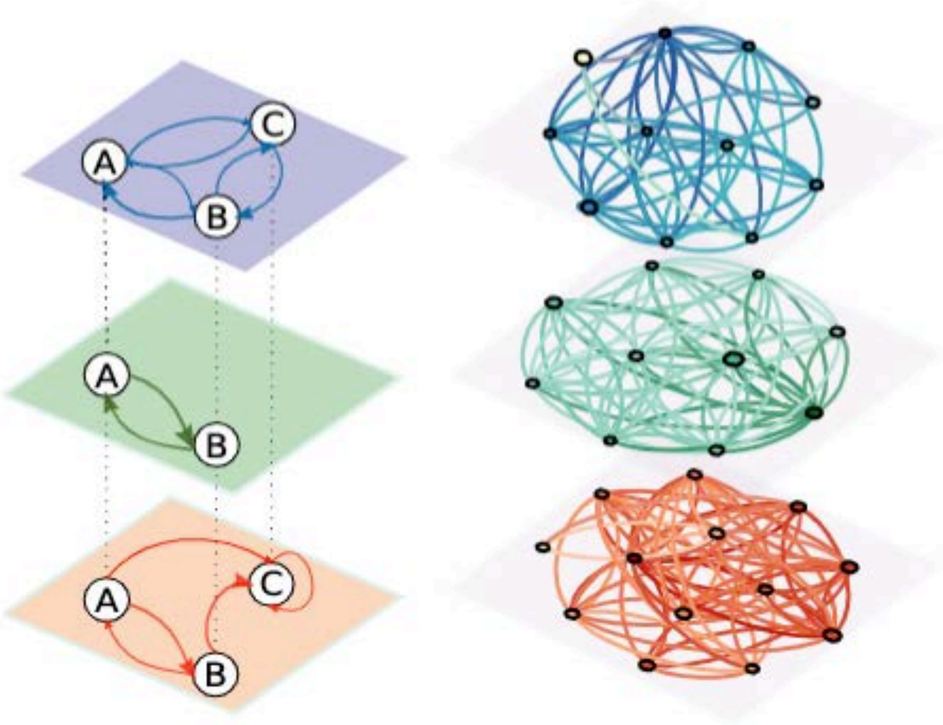


Figure 1: Representation of multi-layer interaction on Reddit. Each layer represents a subreddit.

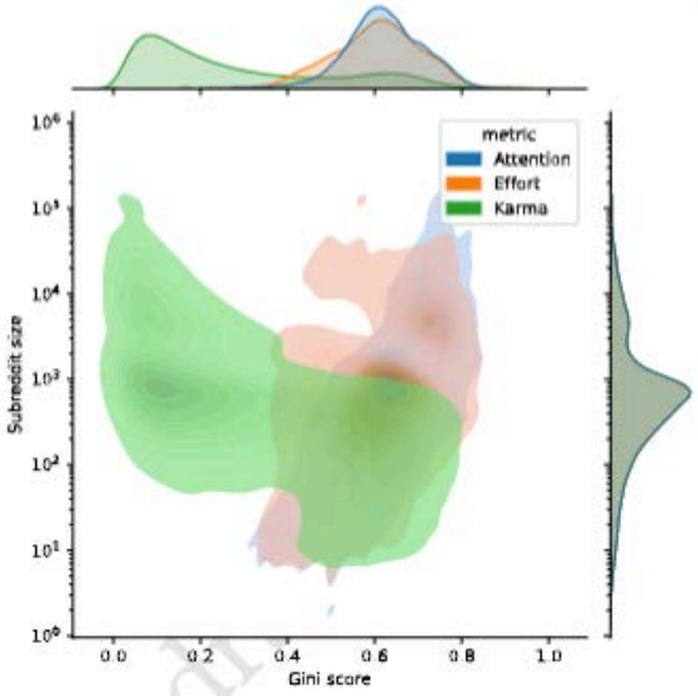
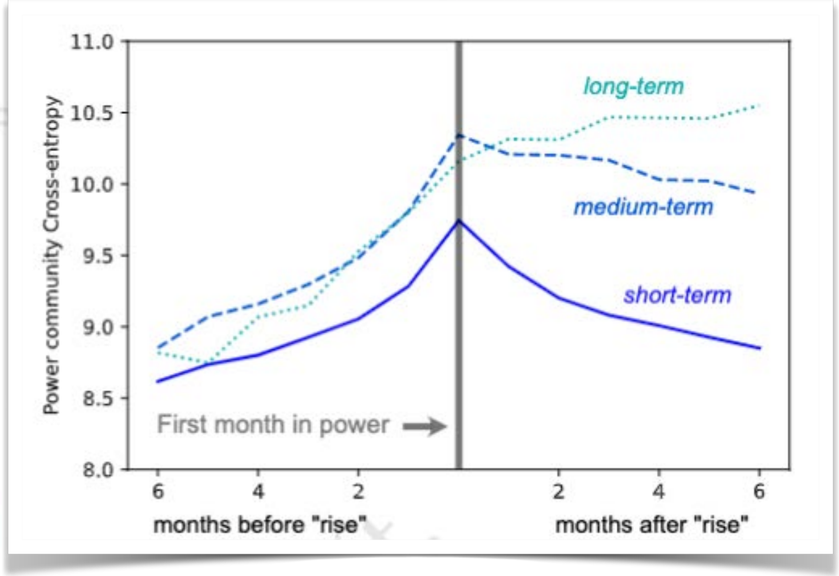
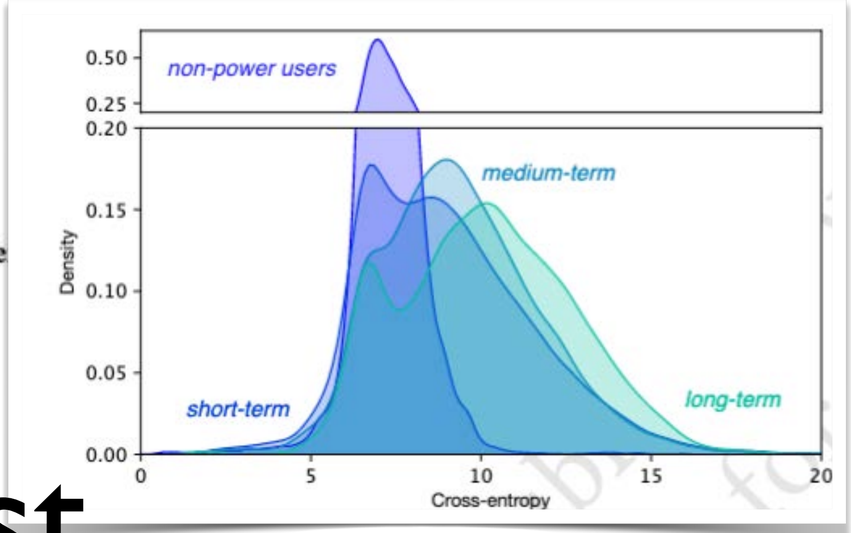


Figure 2: Distribution of Gini scores for Effort, Atte
Karma by size across all subreddits

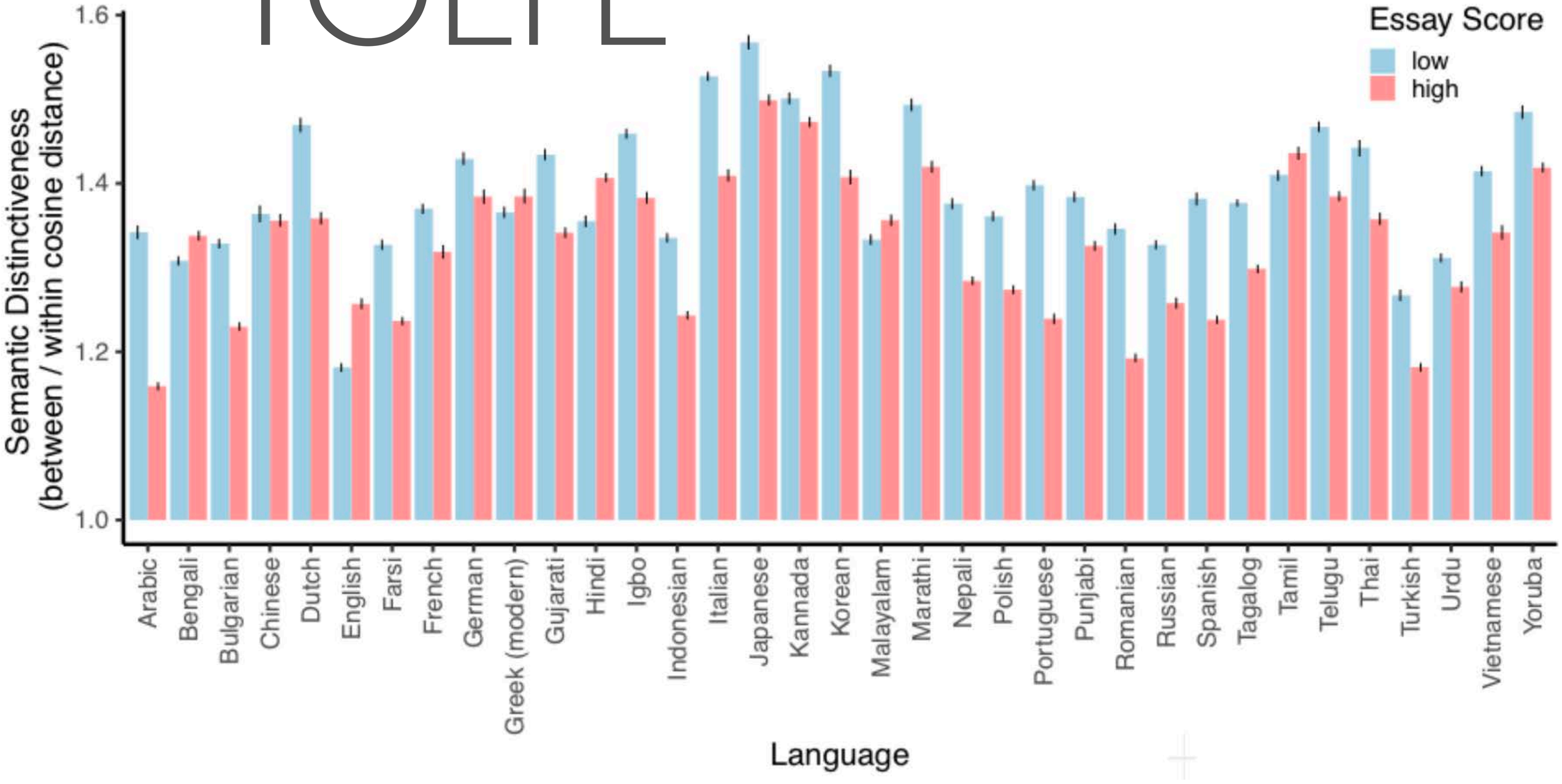


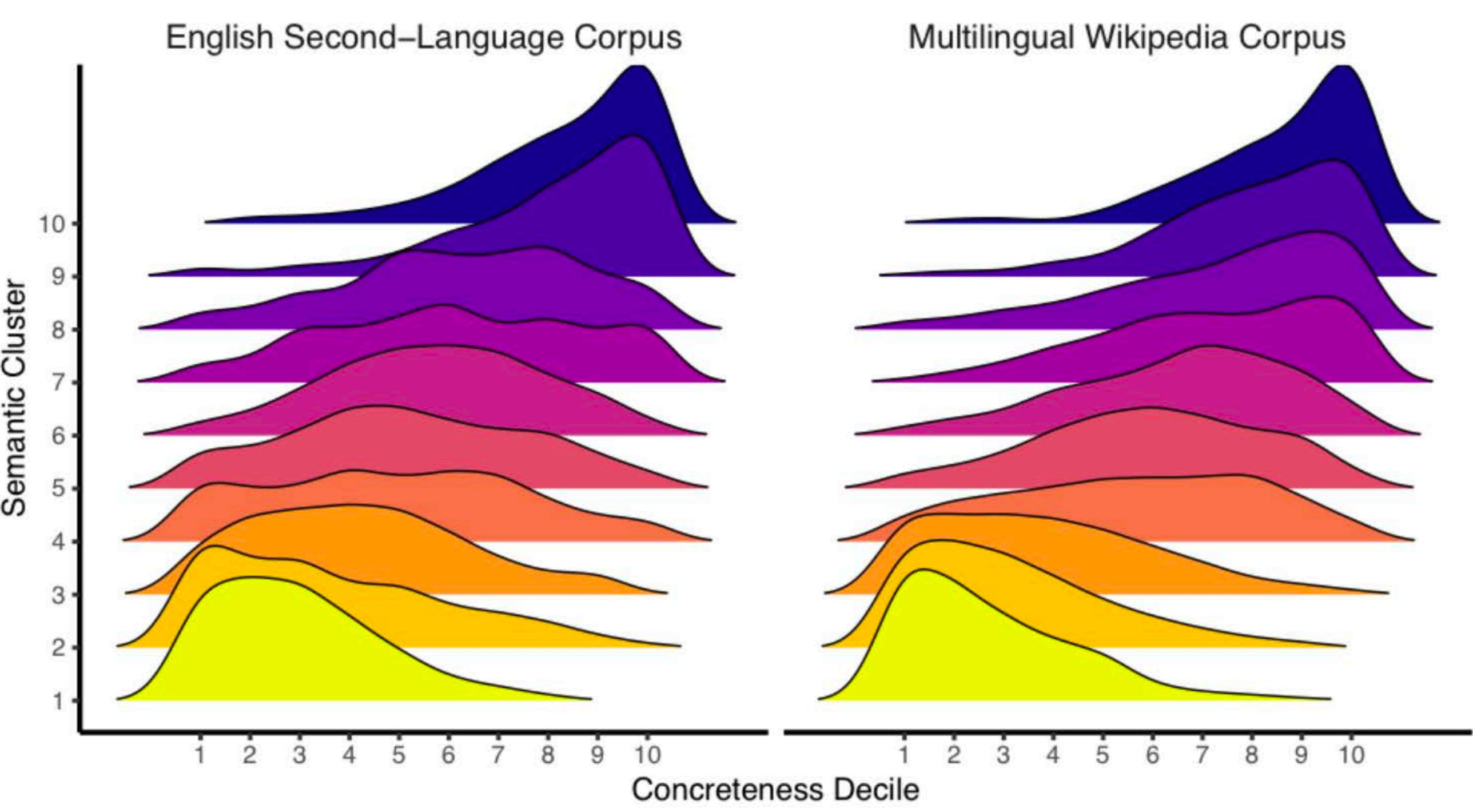


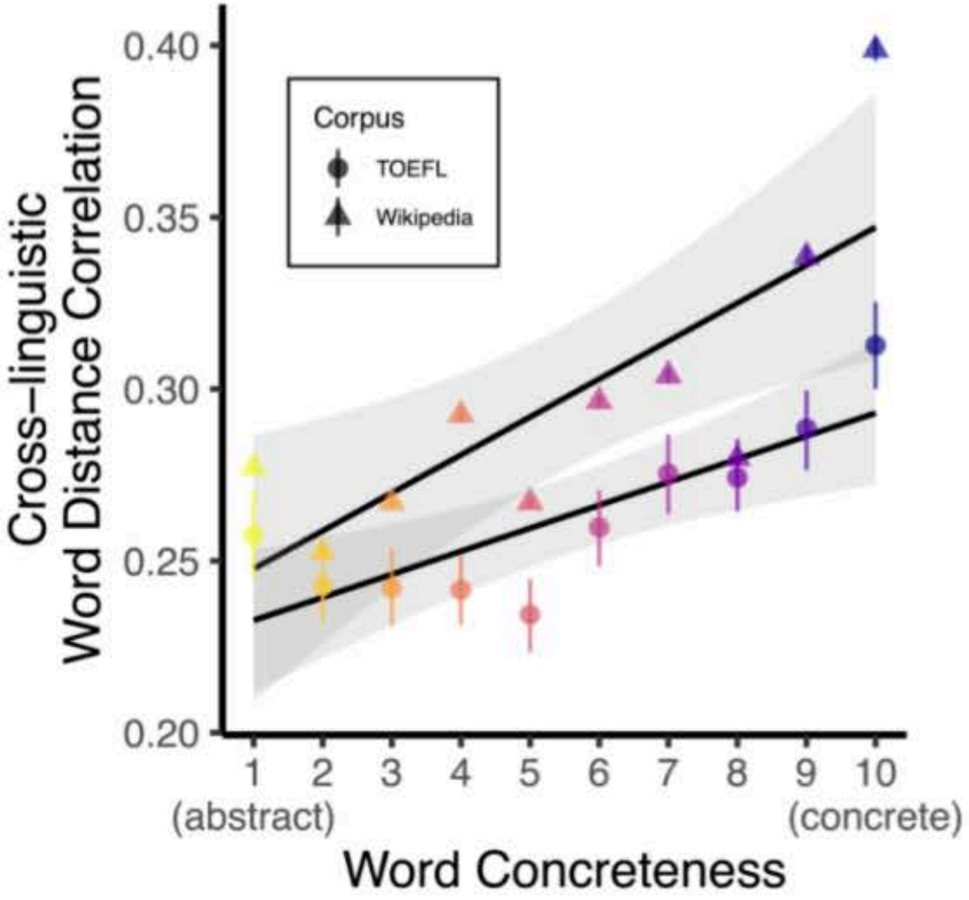
Influencers rehearse disgust

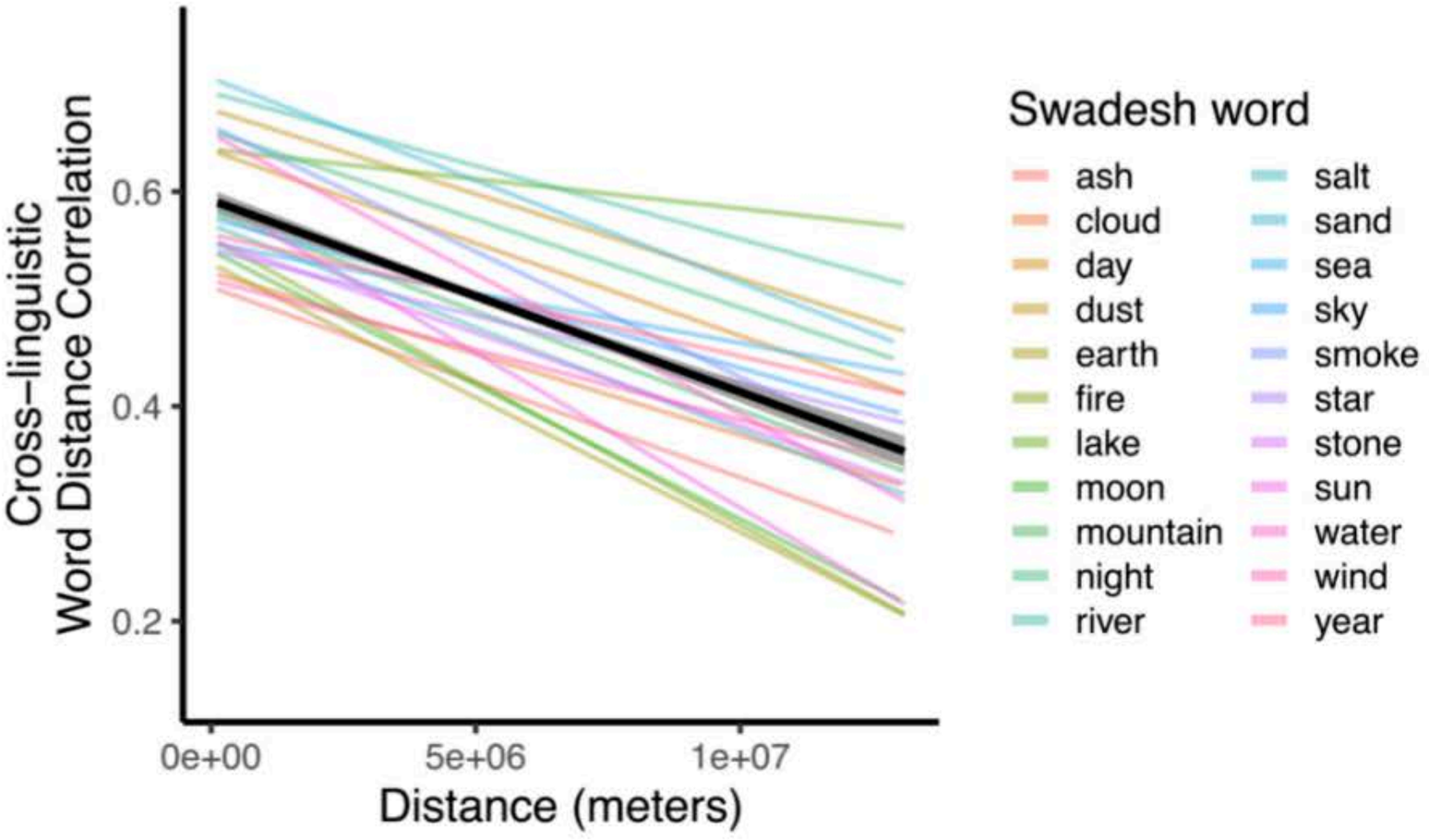
Language Reflects/Reinforces Cultural System

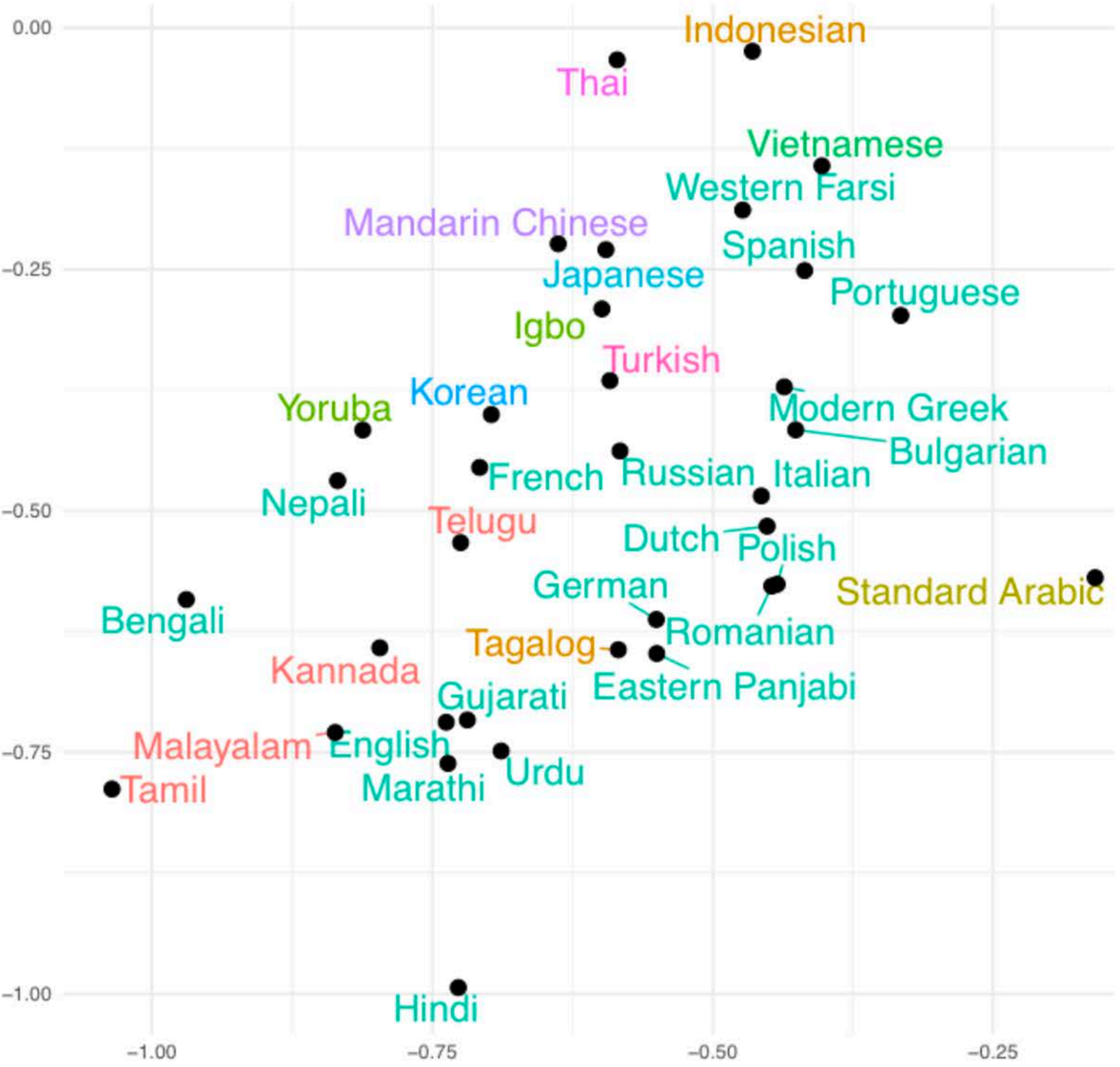
TOEFL

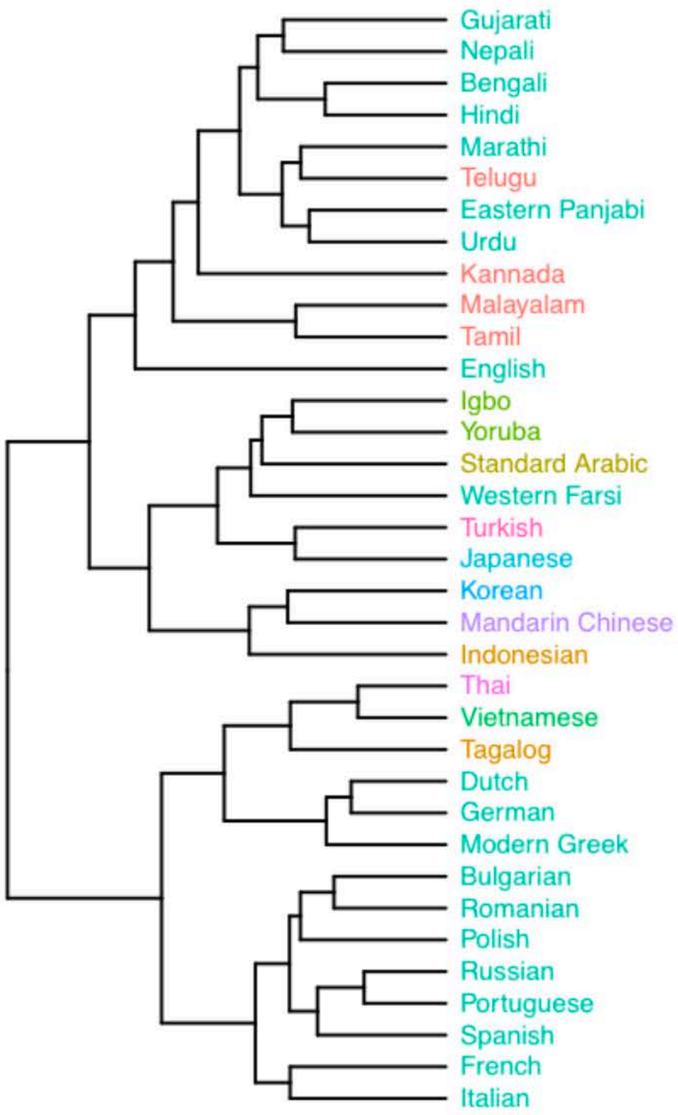

Essay Score
low
high
Semantic Distinctiveness
(between / within cosine distance)
1.6
1.4
1.2
1.0
Arabic
Bengali
Bulgarian
Chinese
Dutch
English
Farsi
French
German
Greek (modern)
Gujarati
Hindi
Igbo
Indonesian
Italian
Japanese
Kannada
Korean
Malayalam
Marathi
Nepali
Polish
Portuguese
Punjabi
Romanian
Russian
Spanish
Tagalog
Tamil
Telugu
Thai
Turkish
Urdu
Vietnamese
Yoruba
Language

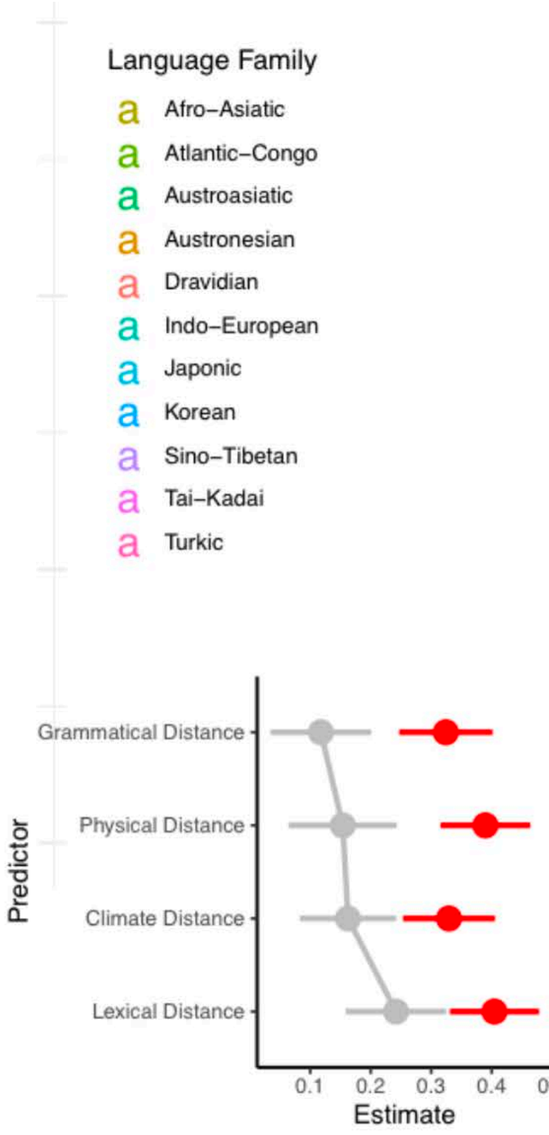

English Second–Language Corpus
Multilingual Wikipedia Corpus
Semantic Cluster
10
9
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
Concreteness Decile

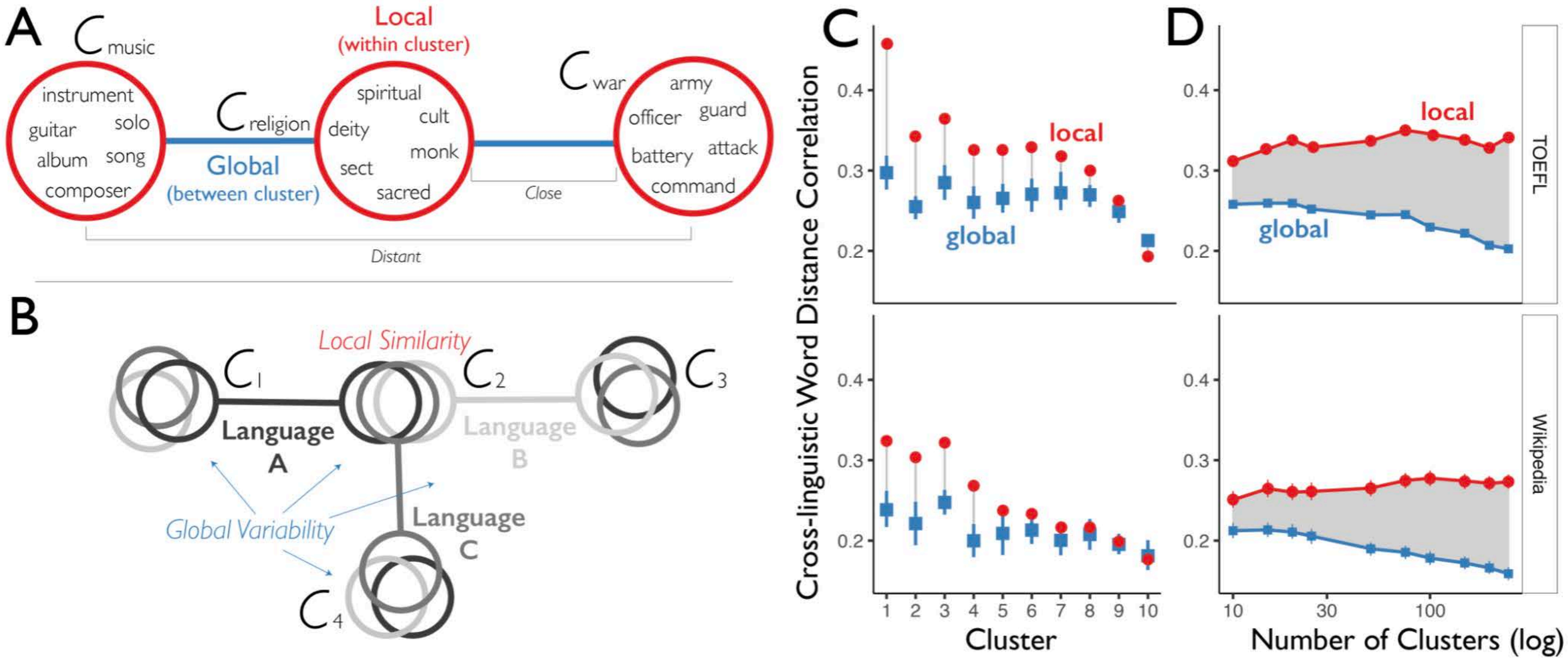

Cross–linguistic
Word Distance Correlation
0.40
0.35
0.30
0.25
0.20
Corpus
TOEFL
Wikipedia
1
2
3
4
5
6
7
8
9
10
(abstract)
(concrete)
Word Concreteness

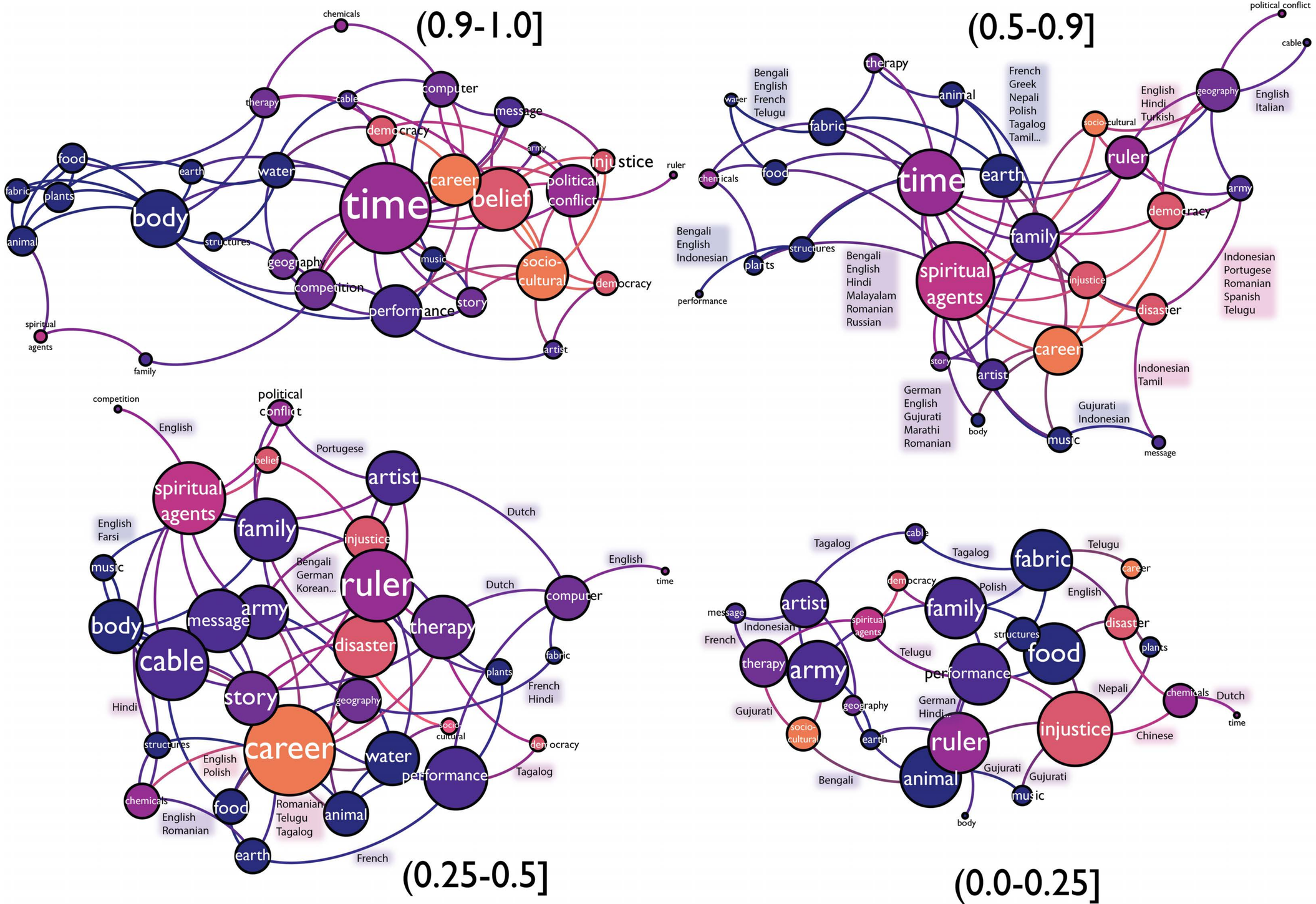

Cross–linguistic
Word Distance Correlation
0.6
0.4
0.2
0e+00
5e+06
1e+07
Distance (meters)
Swadesh word
ash
cloud
day
dust
earth
fire
lake
moon
mountain
night
river
salt
sand
sea
sky
smoke
star
stone
sun
water
wind
year


0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
Indonesian
Thai
Vietnamese
Western Farsi
Mandarin Chinese
Spanish
Japanese
Portuguese
Igbo
Turkish
Korean
Yoruba
Modern Greek
Bulgarian
French
Russian
Italian
Nepali
Telugu
Dutch
Polish
German
Standard Arabic
Bengali
Romanian
Tagalog
Kannada
Eastern Panjabi
Gujarati
Malayalam
English
Urdu
Tamil
Marathi
Hindi

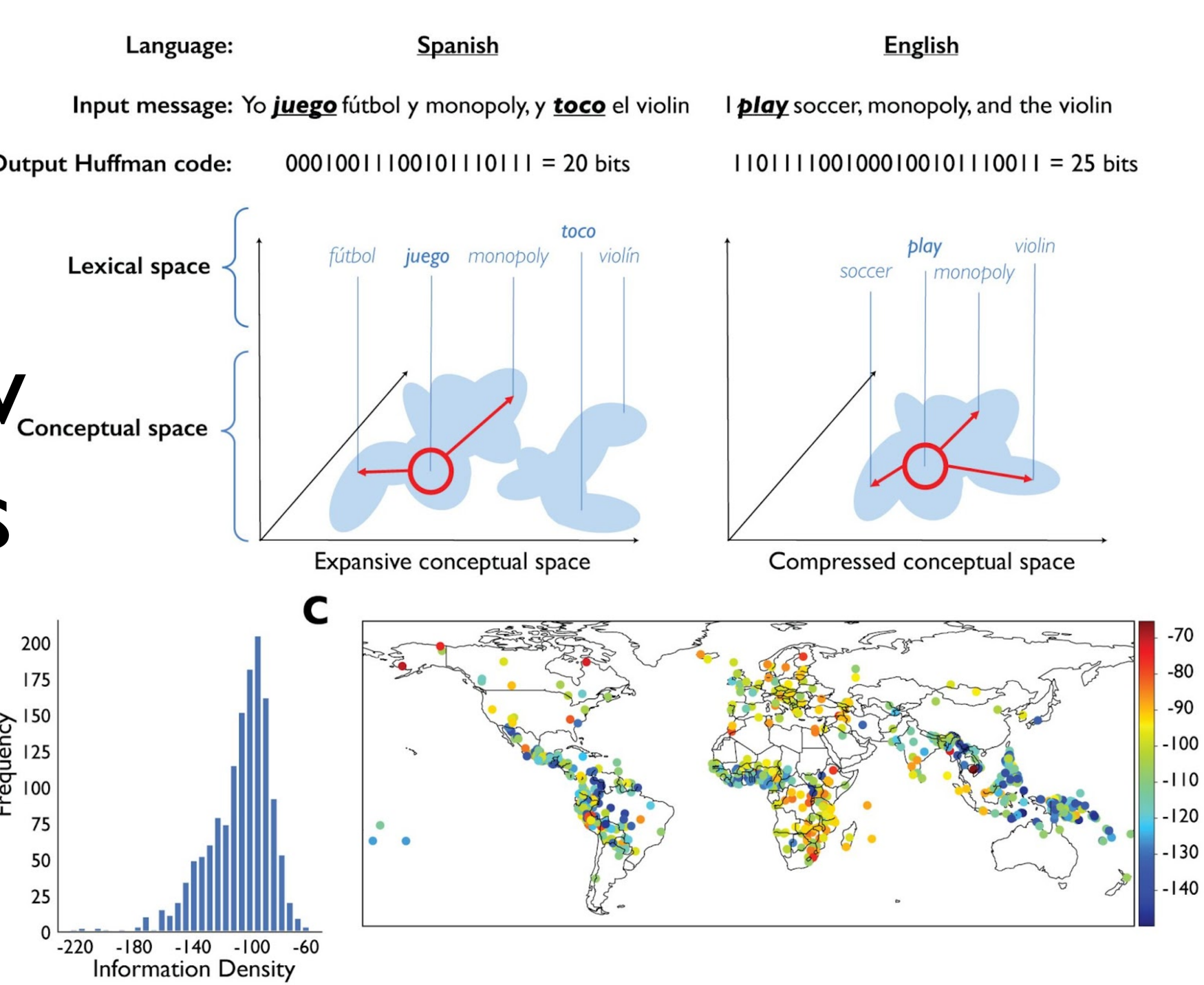

Gujarati
Nepali
Bengali
Hindi
Marathi
Telugu
Eastern Panjabi
Urdu
Kannada
Malayalam
Tamil
English
Igbo
Yoruba
Standard Arabic
Western Farsi
Turkish
Japanese
Korean
Mandarin Chinese
Indonesian
Thai
Vietnamese
Tagalog
Dutch
German
Modern Greek
Bulgarian
Romanian
Polish
Russian
Portuguese
Spanish
French
Italian

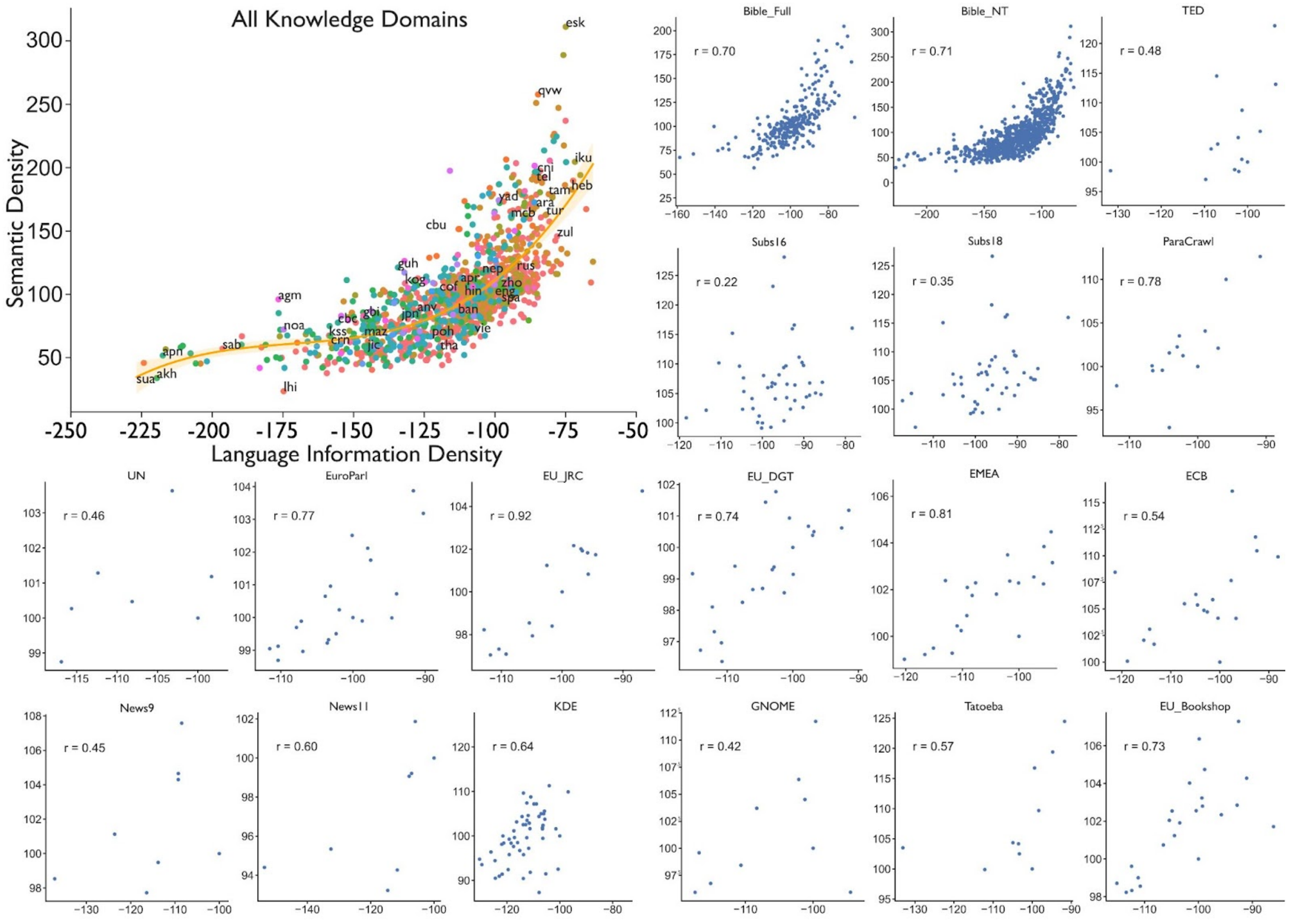

Language Family
Afro−Asiatic
Atlantic−Congo
Austroasiatic
Austronesian
Dravidian
Indo−European
Japonic
Korean
Sino−Tibetan
Tai−Kadai
Turkic
Predictor
Grammatical Distance
Physical Distance
Climate Distance
Lexical Distance
0.1
0.2
0.3
0.4
Estimate

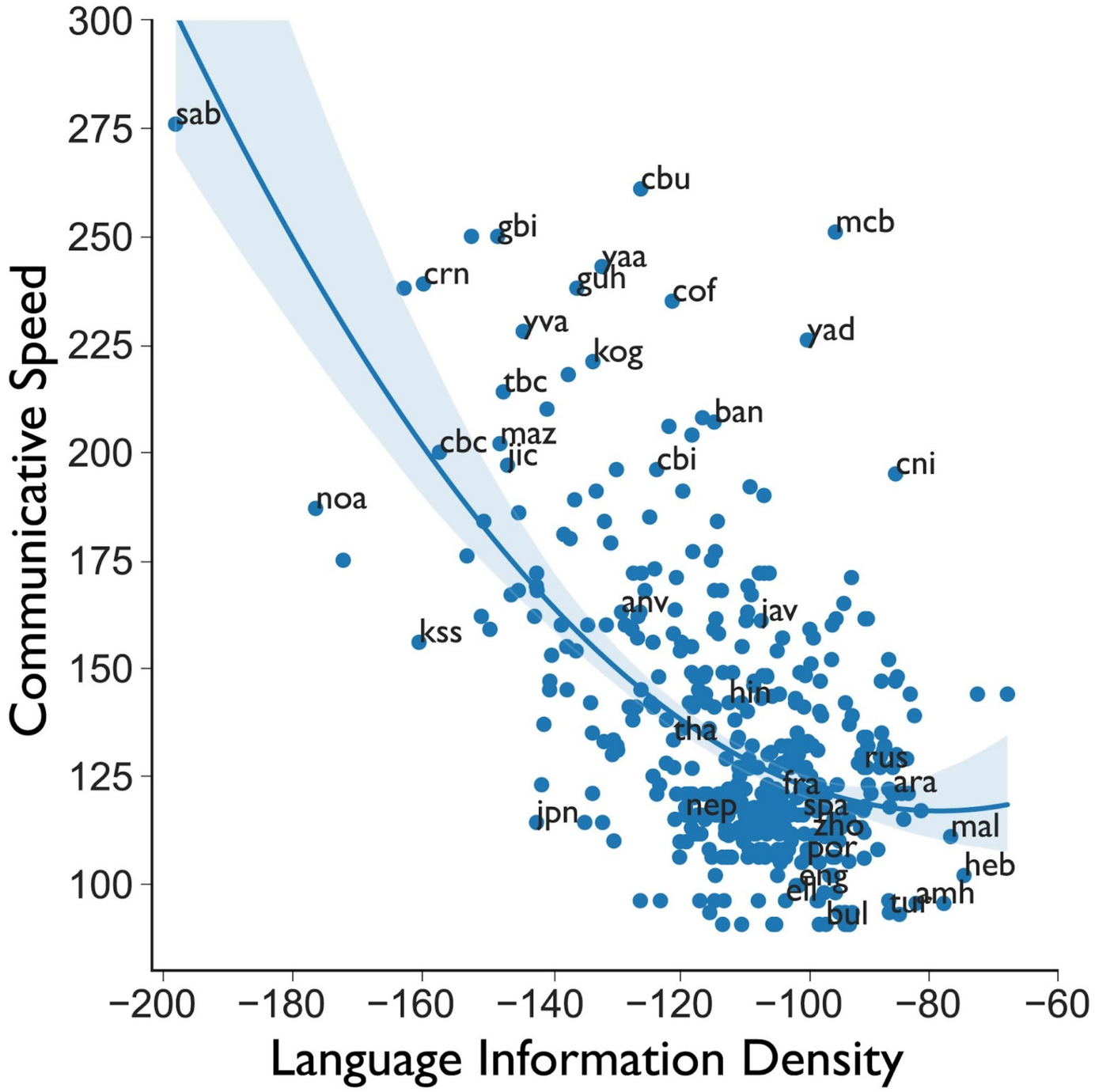

A
music
instrument
guitar
solo
album
song
composer
religion
Global
(between cluster)
Local
(within cluster)
spiritual
deity
cult
monk
sect
sacred
war
army
officer
guard
battery
attack
command
Close
Distant
B
Local Similarity
Language A
Language B
Language C
Global Variability
C
D
Cross-linguistic Word Distance Correlation
local
global
TOEFL
Wikipedia
Cluster
Number of Clusters (log)


(0.9-1.0]
chemicals
computer
therapy
cable
message
democracy
army
injustice
ruler
food
earth
water
career
belief
political conflict
fabric
plants
body
time
animal
structures
geography
music
socio-cultural
competition
democracy
story
performance
spiritual agents
artist
family
(0.5-0.9]
political conflict
cable
therapy
Bengali English French Telugu
water
animal
French Greek Nepali Polish Tagalog Tamil...
English Hindi Turkish
geography
English Italian
fabric
socio-cultural
ruler
chemicals
food
time
earth
army
democracy
Bengali English Indonesian
structures
family
plants
Bengali English Hindi Malayalam Romanian Russian
spiritual agents
injustice
Indonesian Portugese Romanian Spanish Telugu
performance
disaster
career
story
artist
Indonesian Tamil
German English Gujurati Marathi Romanian
Gujurati Indonesian
body
music
message
(0.25-0.5]
competition
political conflict
English
Portugese
belief
artist
spiritual agents
Dutch
English Farsi
family
injustice
music
Bengali German Korean...
ruler
English
time
Dutch
computer
body
message
army
therapy
disaster
cable
fabric
plants
French Hindi
story
geography
Hindi
socio-cultural
structures
career
water
democracy
English Polish
performance
Tagalog
chemicals
food
Romanian Telugu Tagalog
animal
English Romanian
earth
French
(0.0-0.25]
Tagalog
cable
Tagalog
fabric
Telugu
career
democracy
Polish
English
artist
family
message
Indonesian
spiritual agents
disaster
structures
French
food
plants
therapy
army
Telugu
performance
Nepali
chemicals
Dutch
Gujurati
geography
German Hindi...
time
socio-cultural
earth
ruler
injustice
Chinese
Gujurati
Gujurati
Bengali
animal
music
body

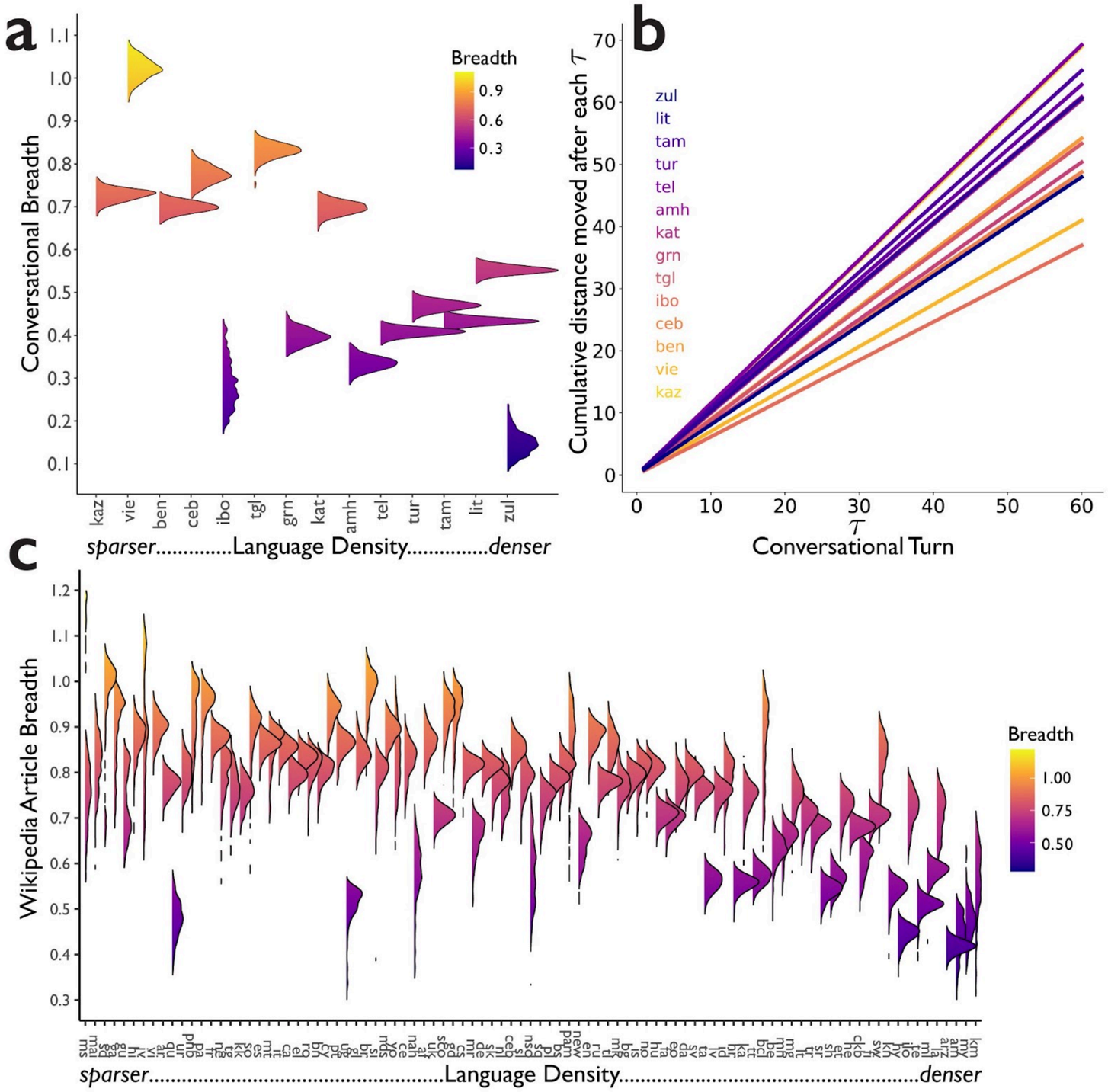
Embedding Language Allows New Dimensions of Discovery



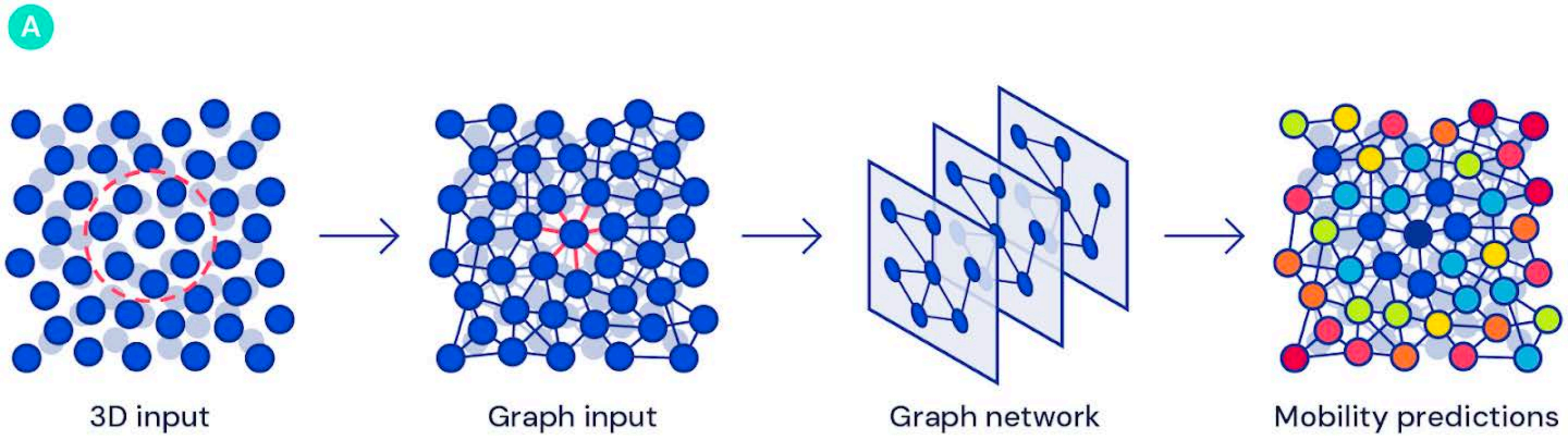

All Knowledge Domains
Semantic Density
Language Information Density
Bible_Full
r = 0.70
Bible_NT
r = 0.71
TED
r = 0.48
Subs16
r = 0.22
Subs18
r = 0.35
ParaCrawl
r = 0.78
UN
r = 0.46
EuroParl
r = 0.77
EU_JRC
r = 0.92
EU_DGT
r = 0.74
EMEA
r = 0.81
ECB
r = 0.54
News9
r = 0.45
News11
r = 0.60
KDE
r = 0.64
GNOME
r = 0.42
Tatoeba
r = 0.57
EU_Bookshop
r = 0.73



Association between information density and communicative speed.
Informationally denser languages that pack information into fewer bits communicate more quickly.

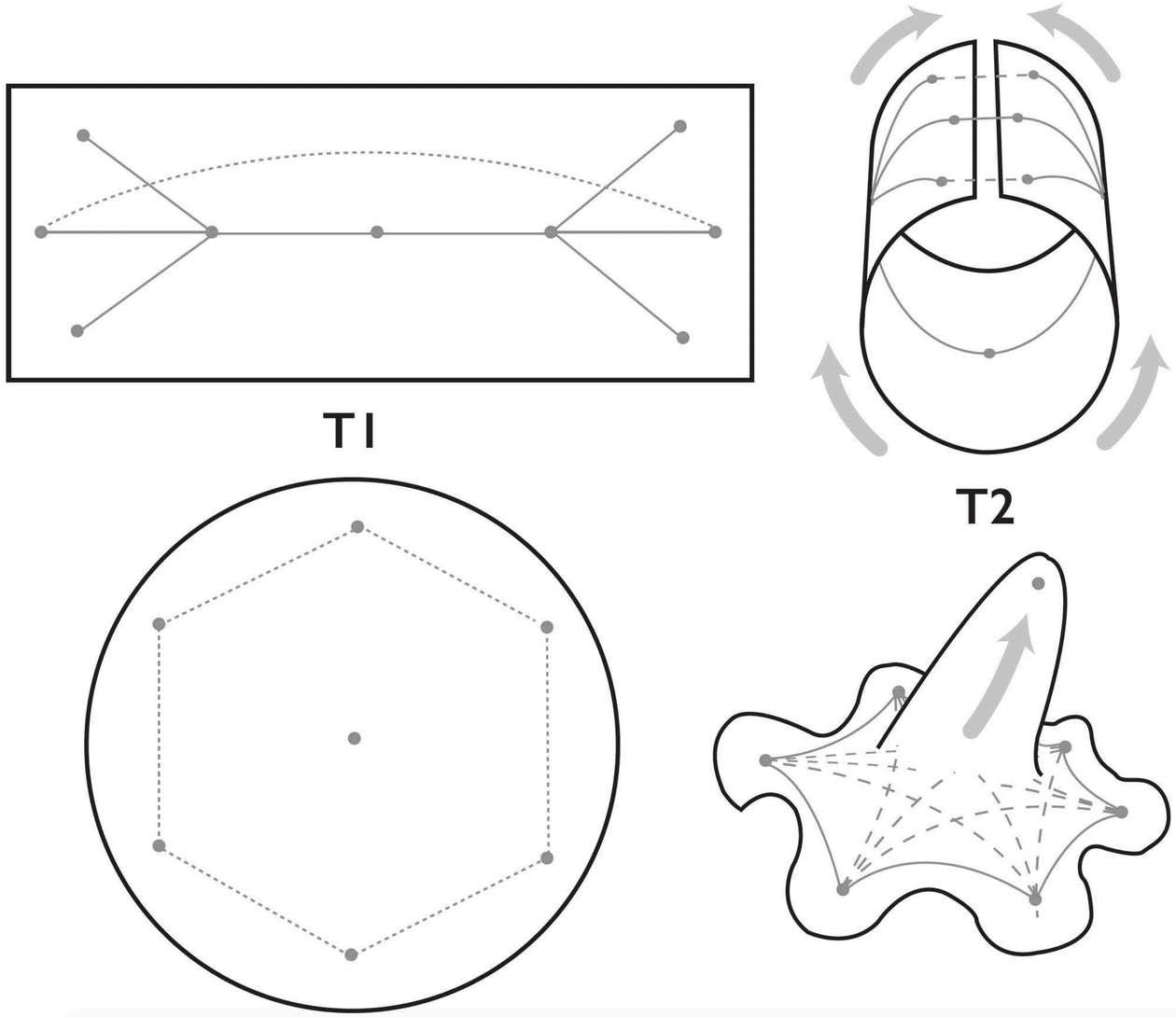

a
Breadth
0.9
0.6
0.3
Conversational Breadth
kaz
vie
ben
ceb
ibo
tgl
grn
kat
amh
tel
tur
tam
lit
zul
sparser.....Language Density.....denser
b
Cumulative distance moved after each τ
zul
lit
tam
tur
tel
amh
kat
grn
tgl
ibo
ceb
ben
vie
kaz
τ
Conversational Turn
c
Wikipedia Article Breadth
Breadth
1.00
0.75
0.50
sparser.....Language Density.....denser

Auto-Encoding Social Networks



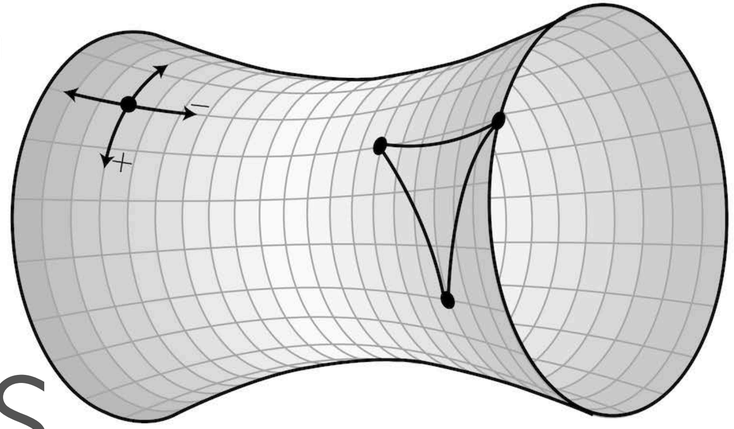
Continuous Social Spaces Predict Future Ties

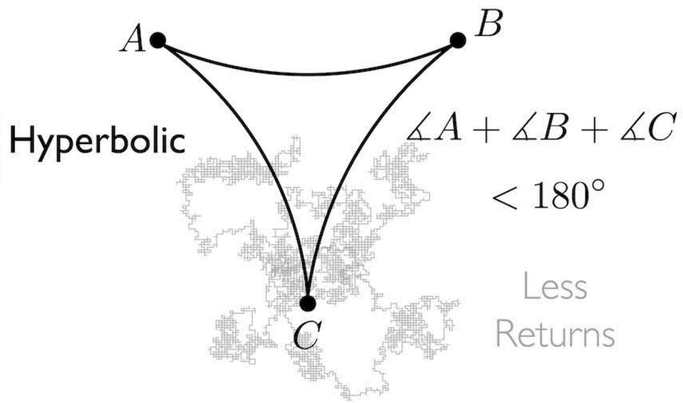
Raise the Issue of ***Curvature***



Overlapping connections create a continuous space

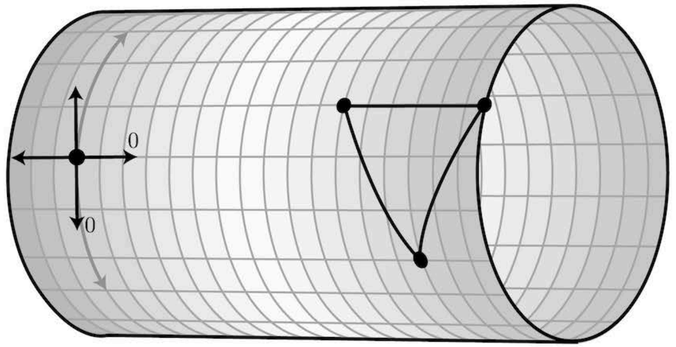
Any simply connected, complete Riemannian manifold of constant sectional curvature is either Euclidean, spherical, or hyperbolic:

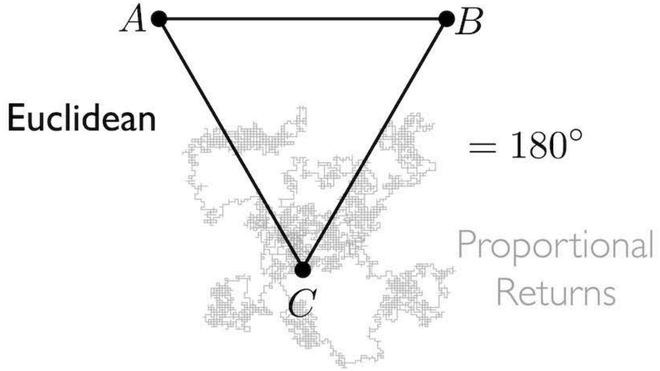


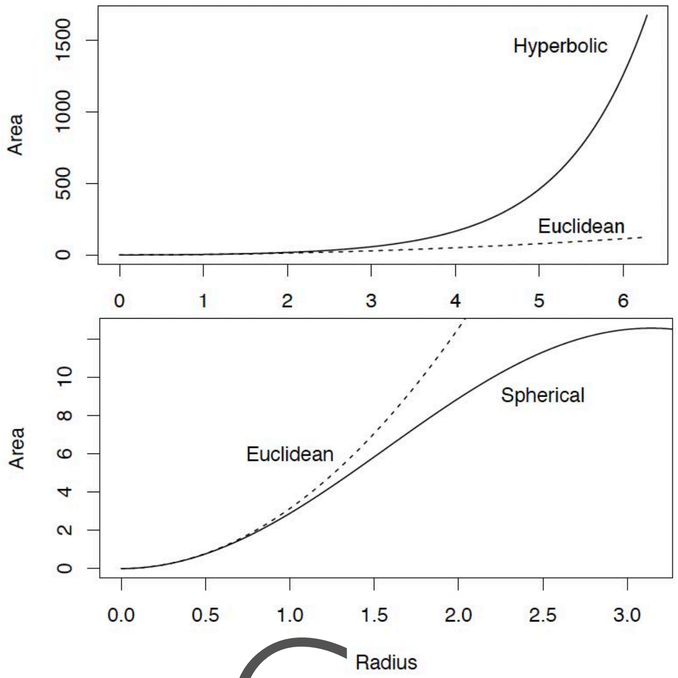


Pluralities

Tree-like Hierarchies

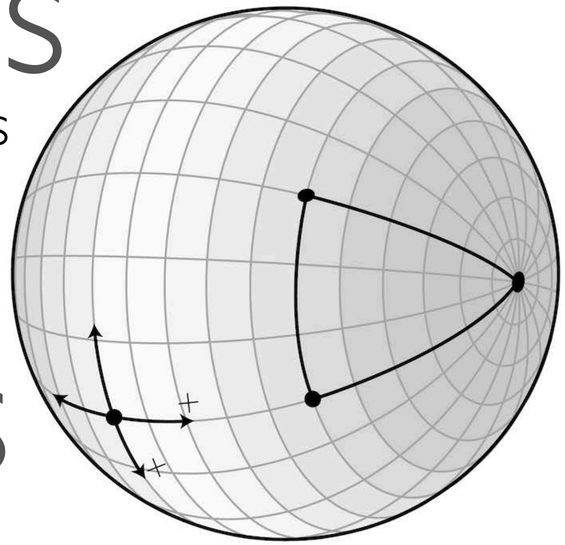


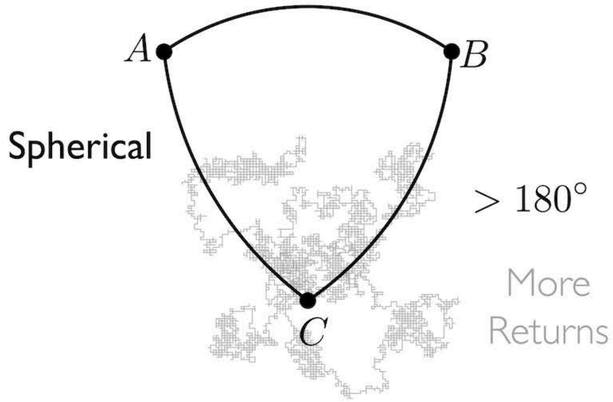




Spaces

Flat Dimensions





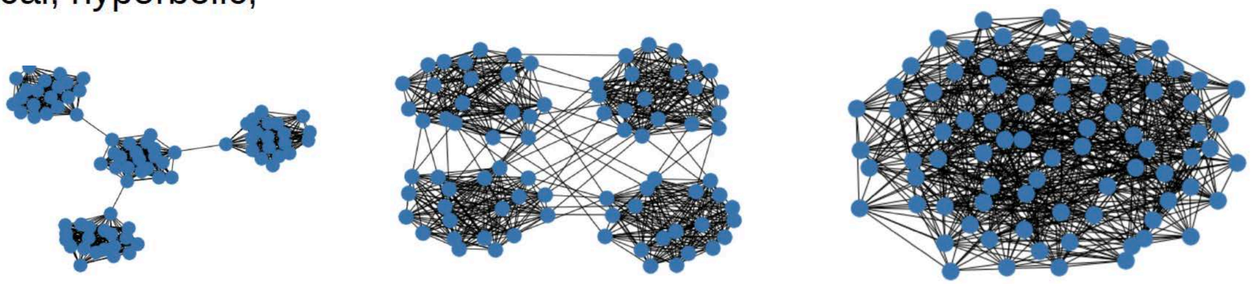
Communities

Self-reinforcing Cycles

Statistical Test for Sectional Curvature Sign

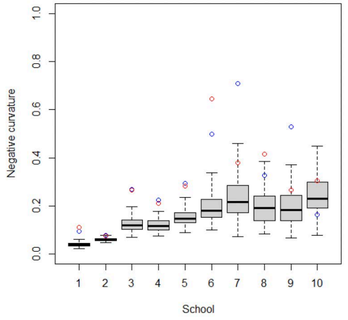
- Model: $P(y_{ij} = 1|\eta, z) = exp(\eta_i + \eta_j - d(z_i, z_j))$
- Ties form independently conditional on positions
- Solve for (noisy) distance matrix, infer most likely latent geometry (spherical, hyperbolic, Euclidean) that generated the observed network and its curvature

Simulation

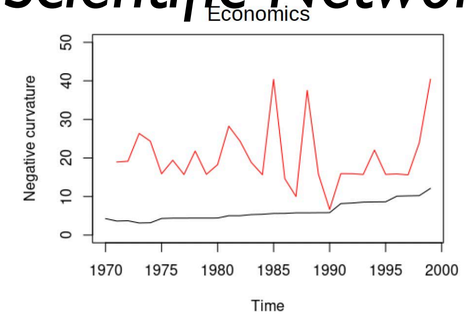


Manifold Estimation of Sectional Curvature

High School Networks

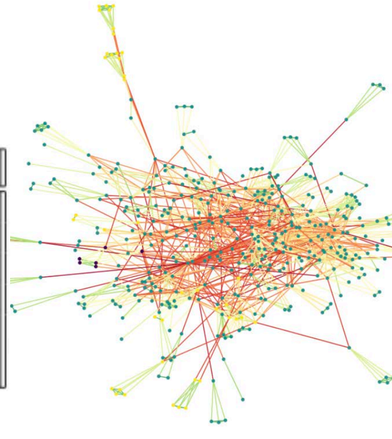


Scientific Networks



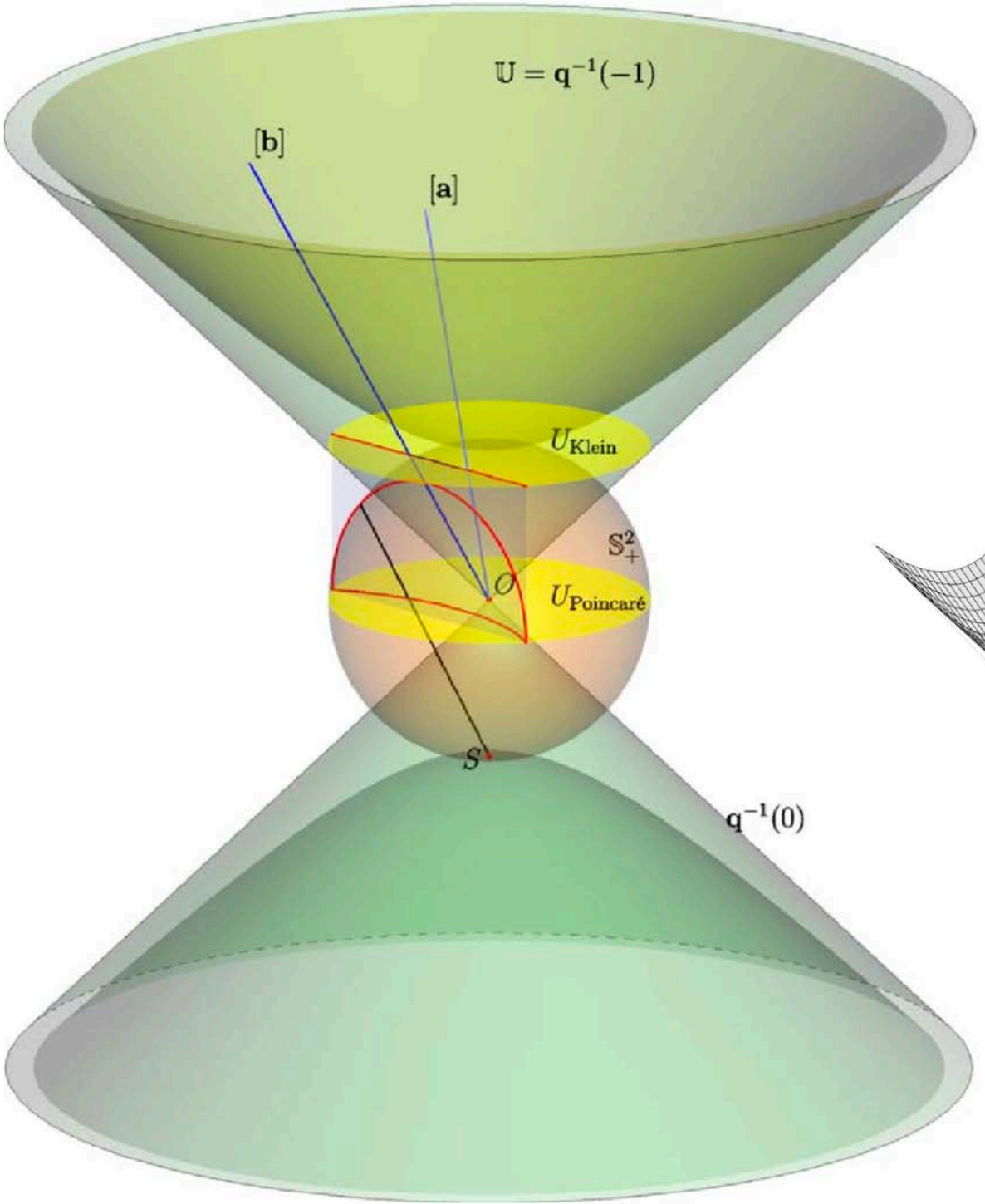
Indian Villages

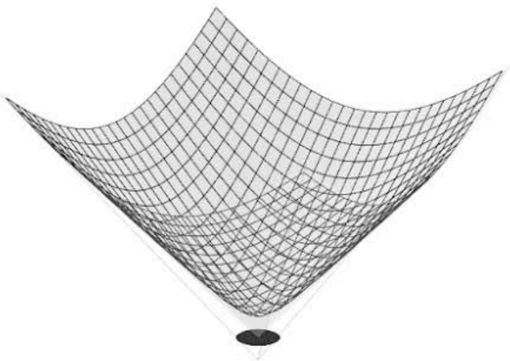
Tie type	Percentage ties negatively curved
Hinduism - Christianity	100%
Islam - Christianity	100%
Hinduism - Islam	97.3%
Hinduism - Hinduism	49.7%
Islam - Islam	33.2 %
Christianity - Christianity	9.6%

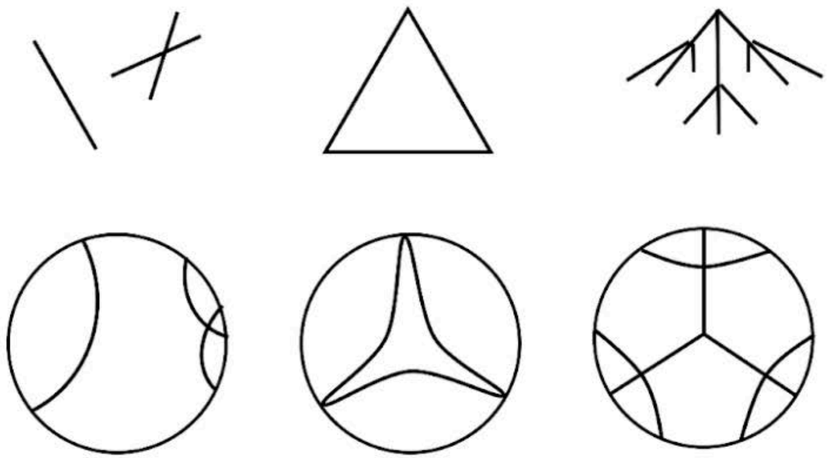


Olivier-Ricci Estimation of Edge Curvature

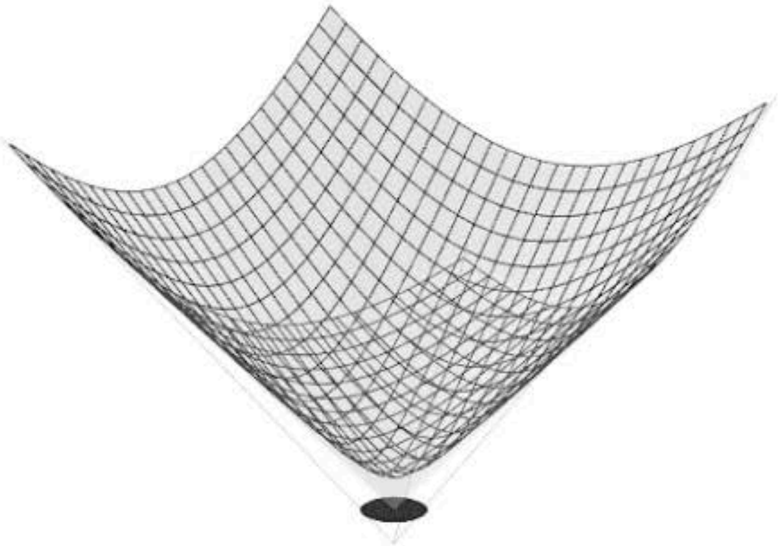
Hyperbolic Embeddings On a Disk

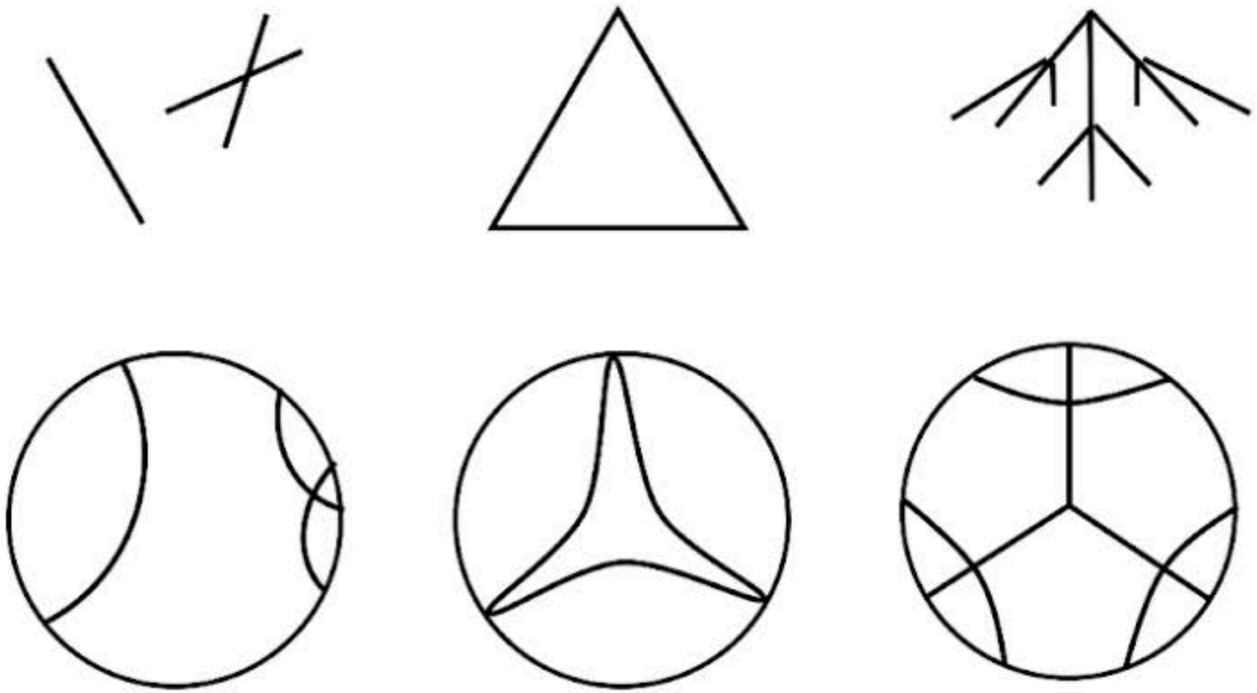






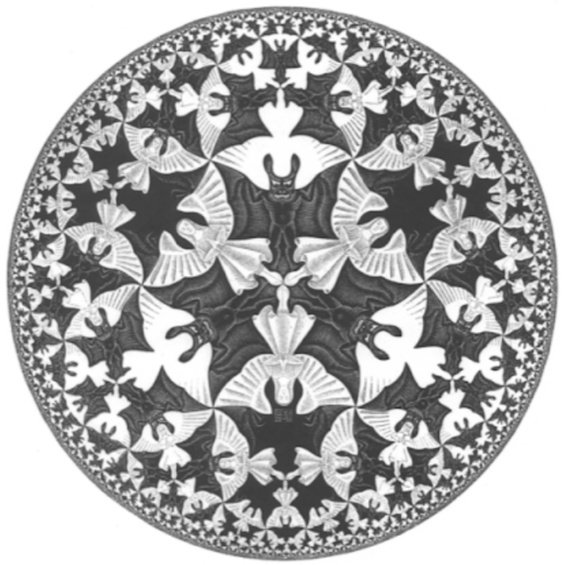
Hyperbolic Geometry on a Disk

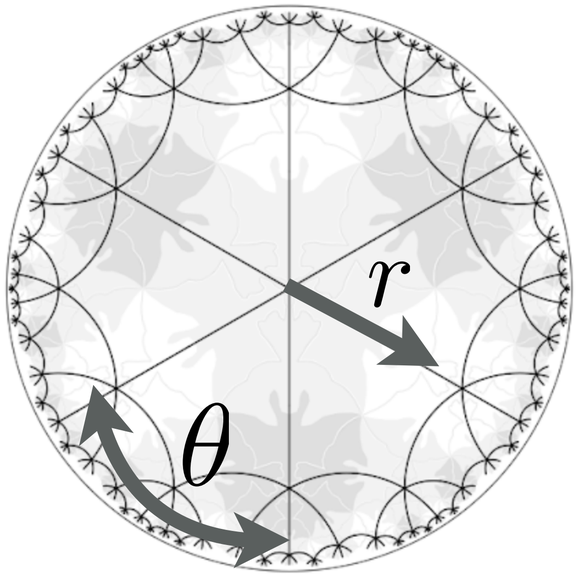


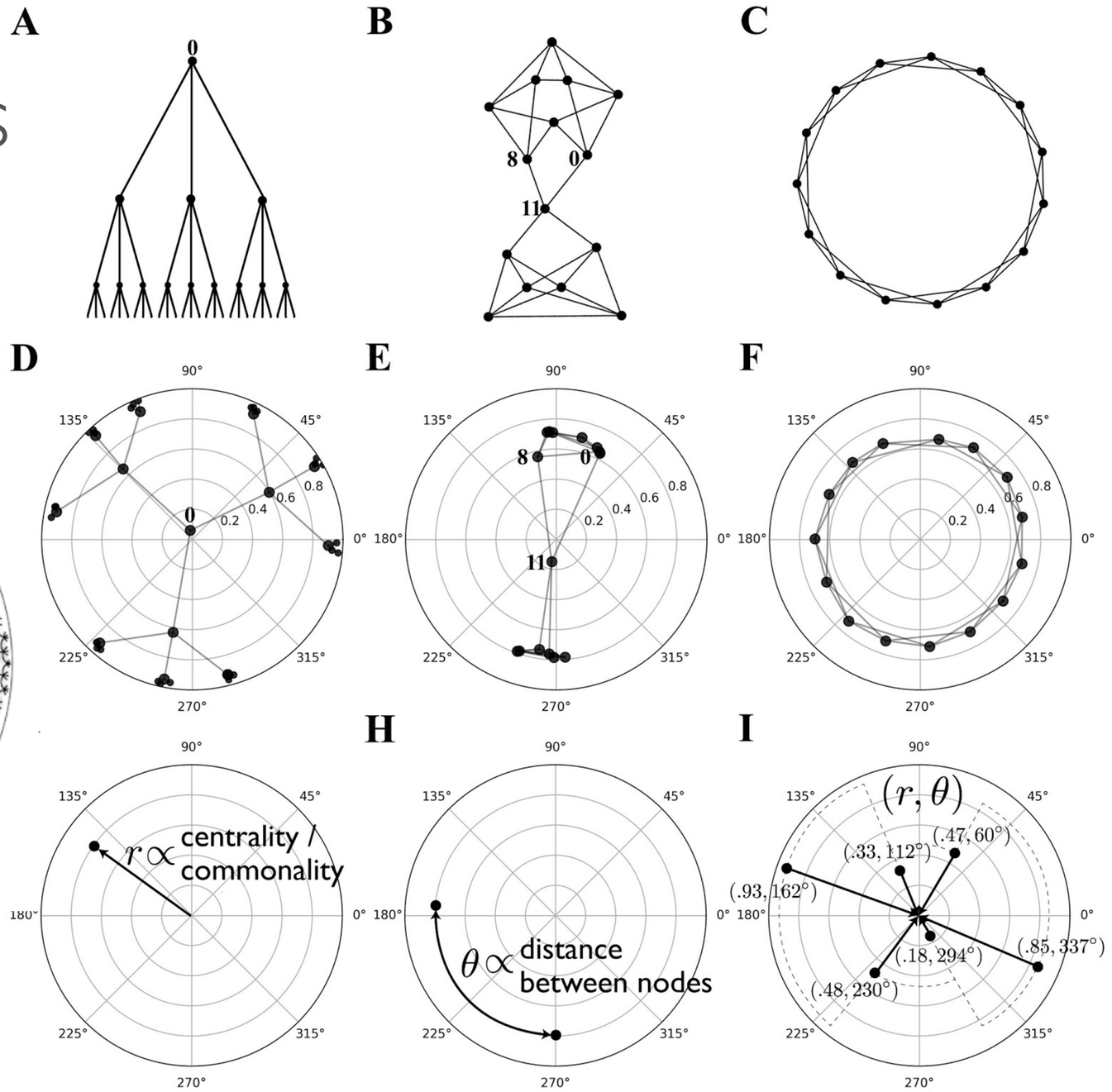


Hyperbolic Embeddings

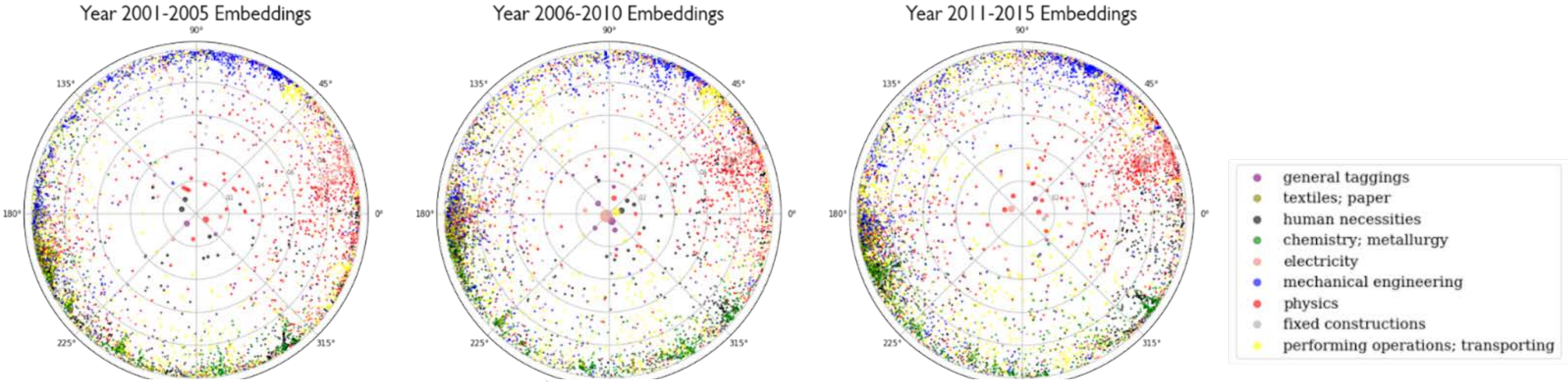
Trace hierarchy







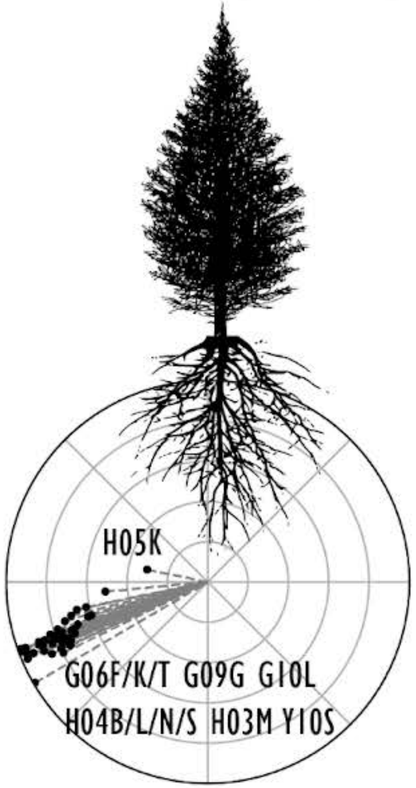
Poincaré Embedding for AI Technologies, 2001-2015

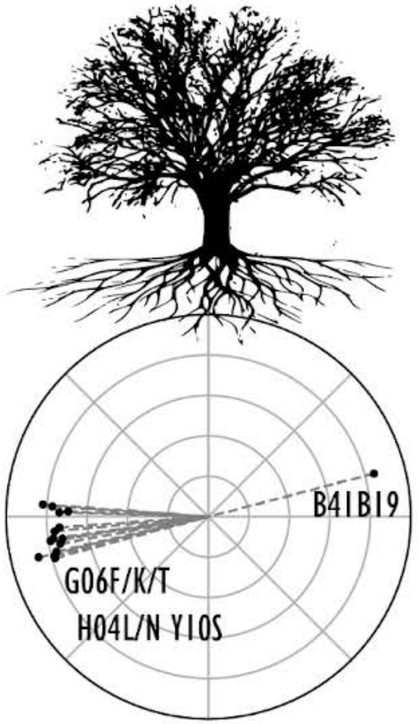


Patent Number US8179974
Multi-level representation of reordered transform coefficients

Techniques and tools for encoding and decoding a block of frequency coefficients are presented.

H05K:
Printed Circuits; manufacture of assemblages of electrical components
H03M:
Coding; decoding; code conversation

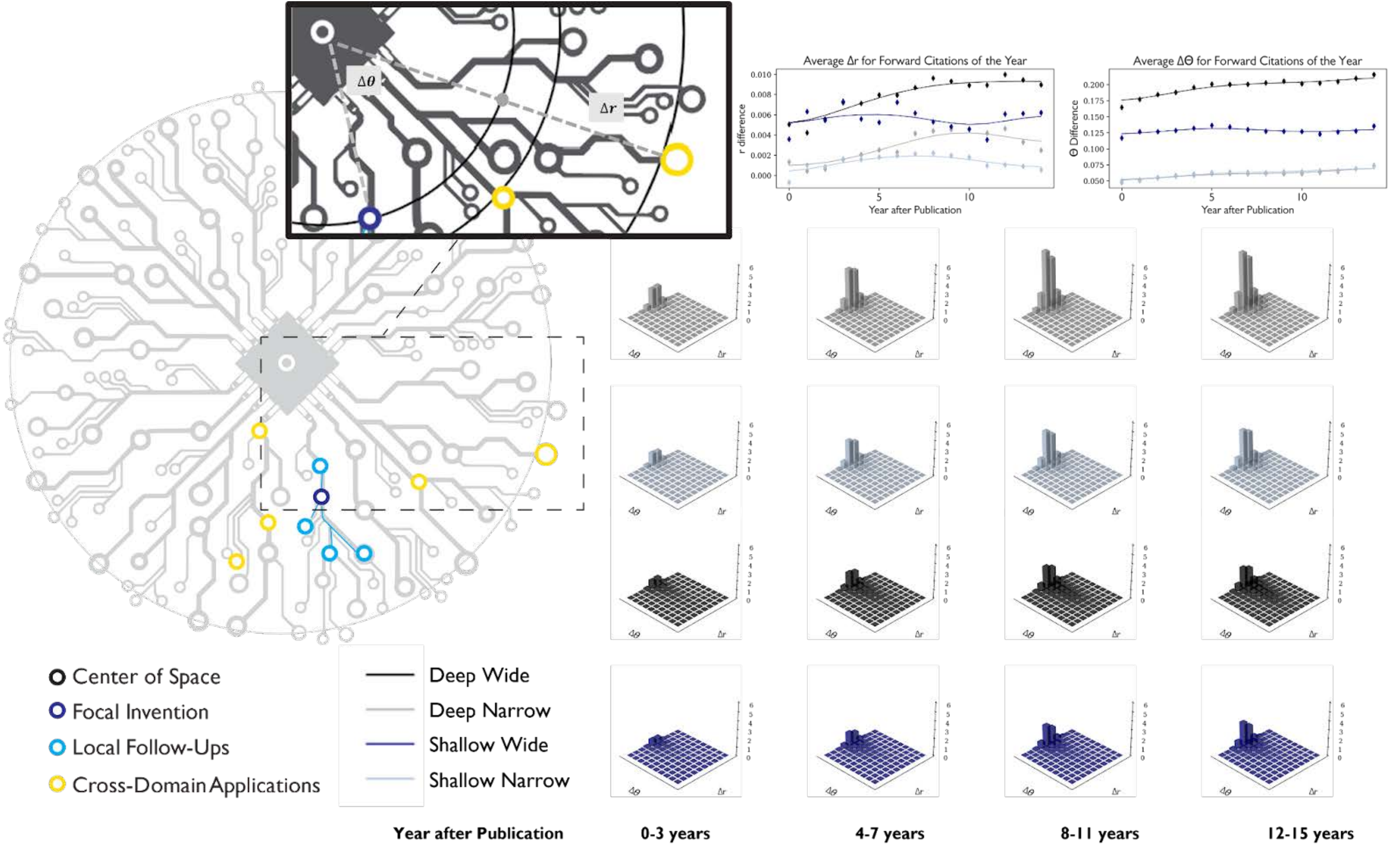




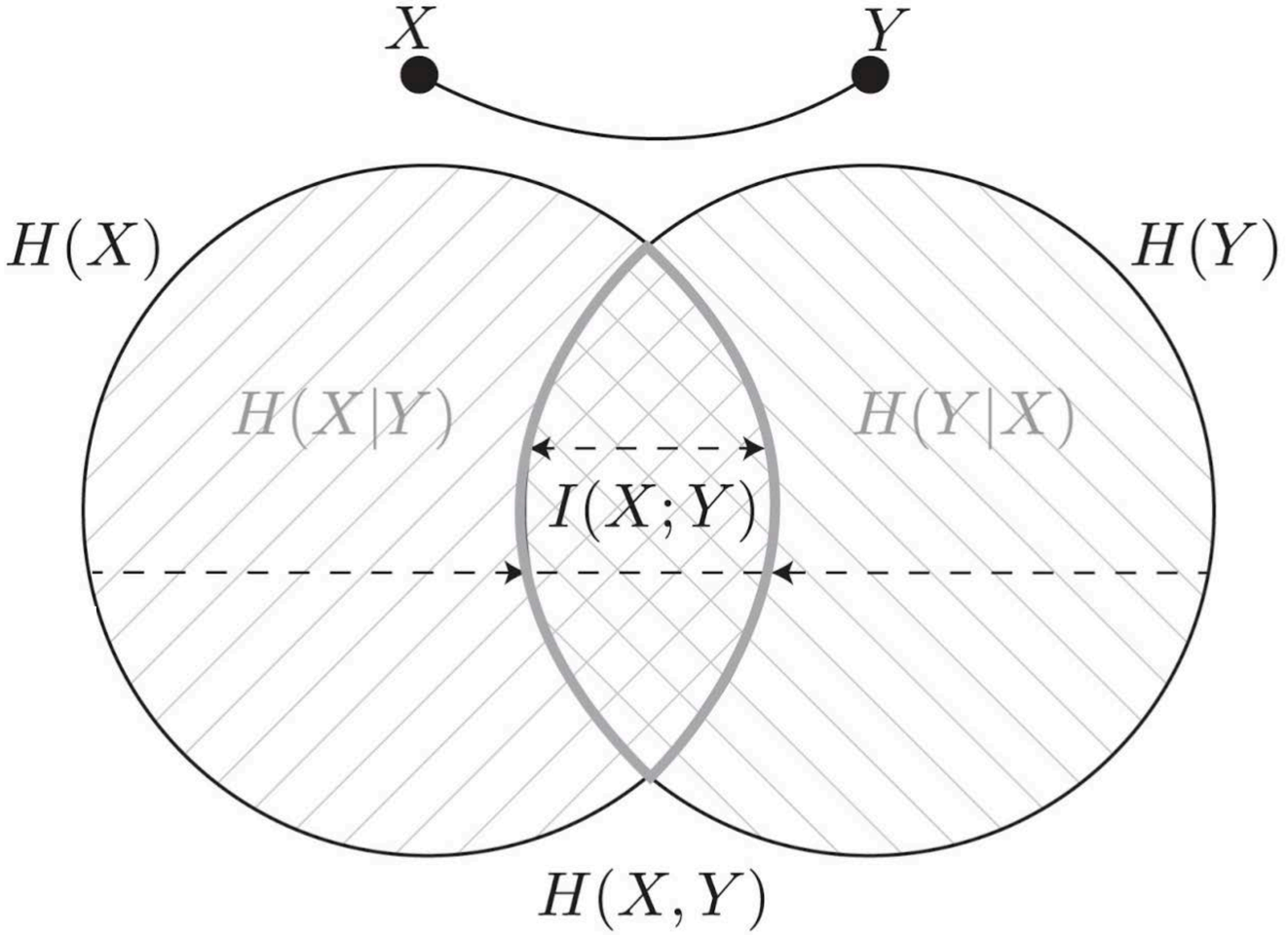
Patent Number US6362895
PDF to PostScript conversion of graphic image files

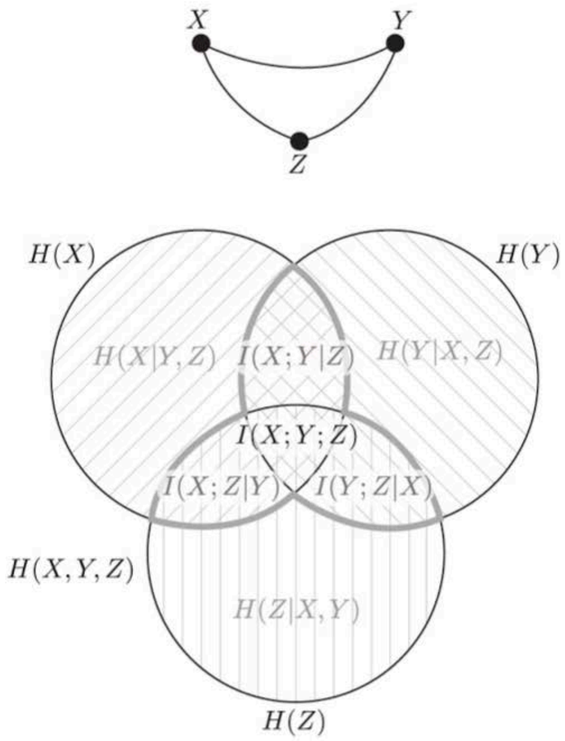
An on-line automated printing system quickly produces consistent printed materials.

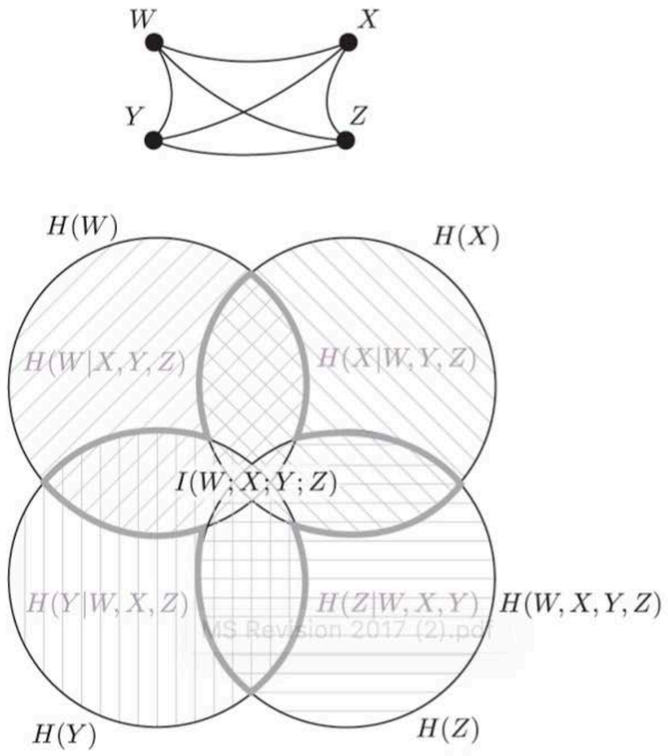
B41B19:
Photoelectronic composing machines
G06:
Computing; calculating; counting.
H04:
Electric Communication Technique


Δθ
Δr
Average Δr for Forward Citations of the Year
r difference
Year after Publication
Average ΔΘ for Forward Citations of the Year
Θ Difference
Year after Publication
Center of Space
Focal Invention
Local Follow-Ups
Cross-Domain Applications
Deep Wide
Deep Narrow
Shallow Wide
Shallow Narrow
Year after Publication
0-3 years
4-7 years
8-11 years
12-15 years

Social Connection & Cultural Collapse

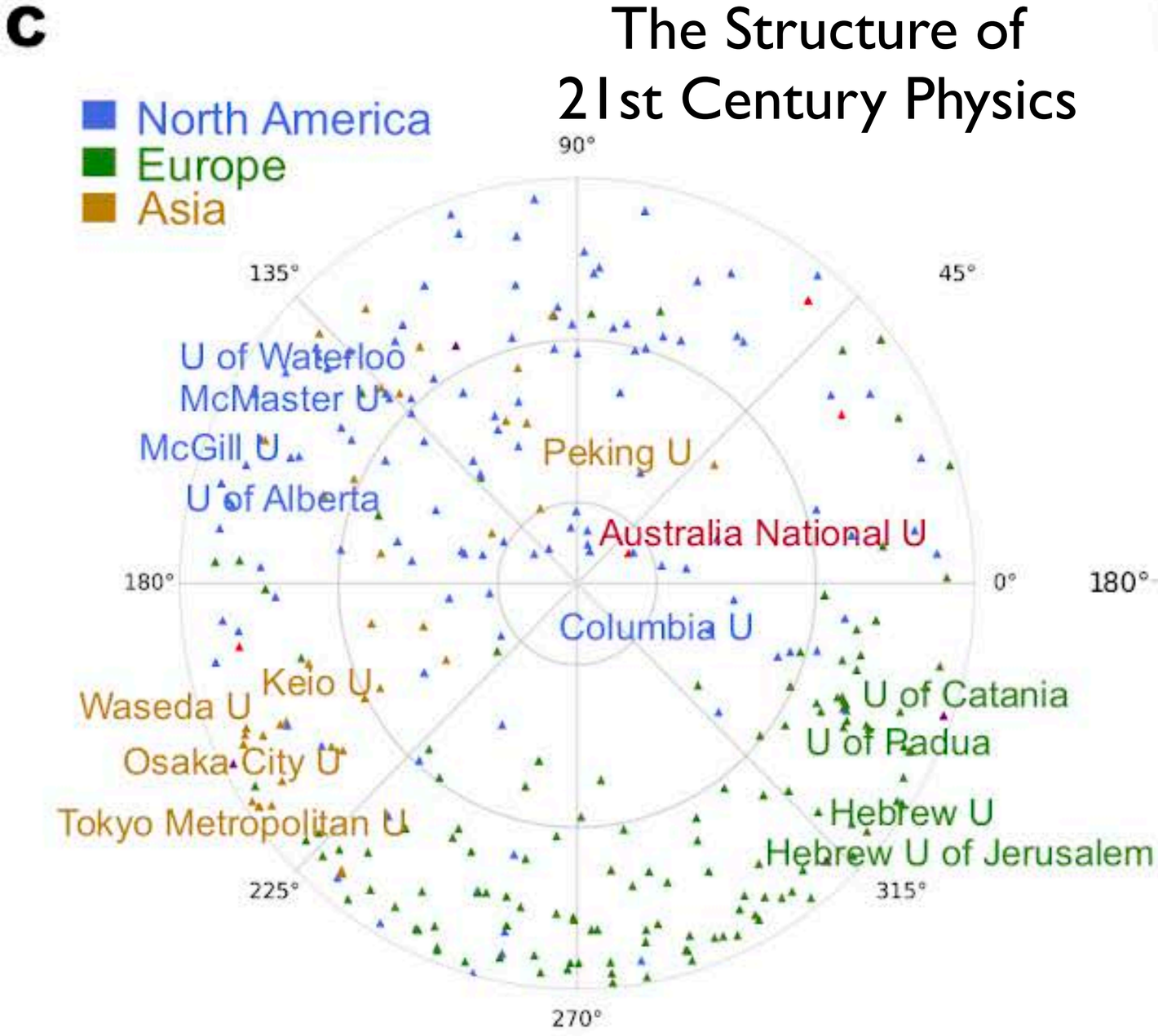


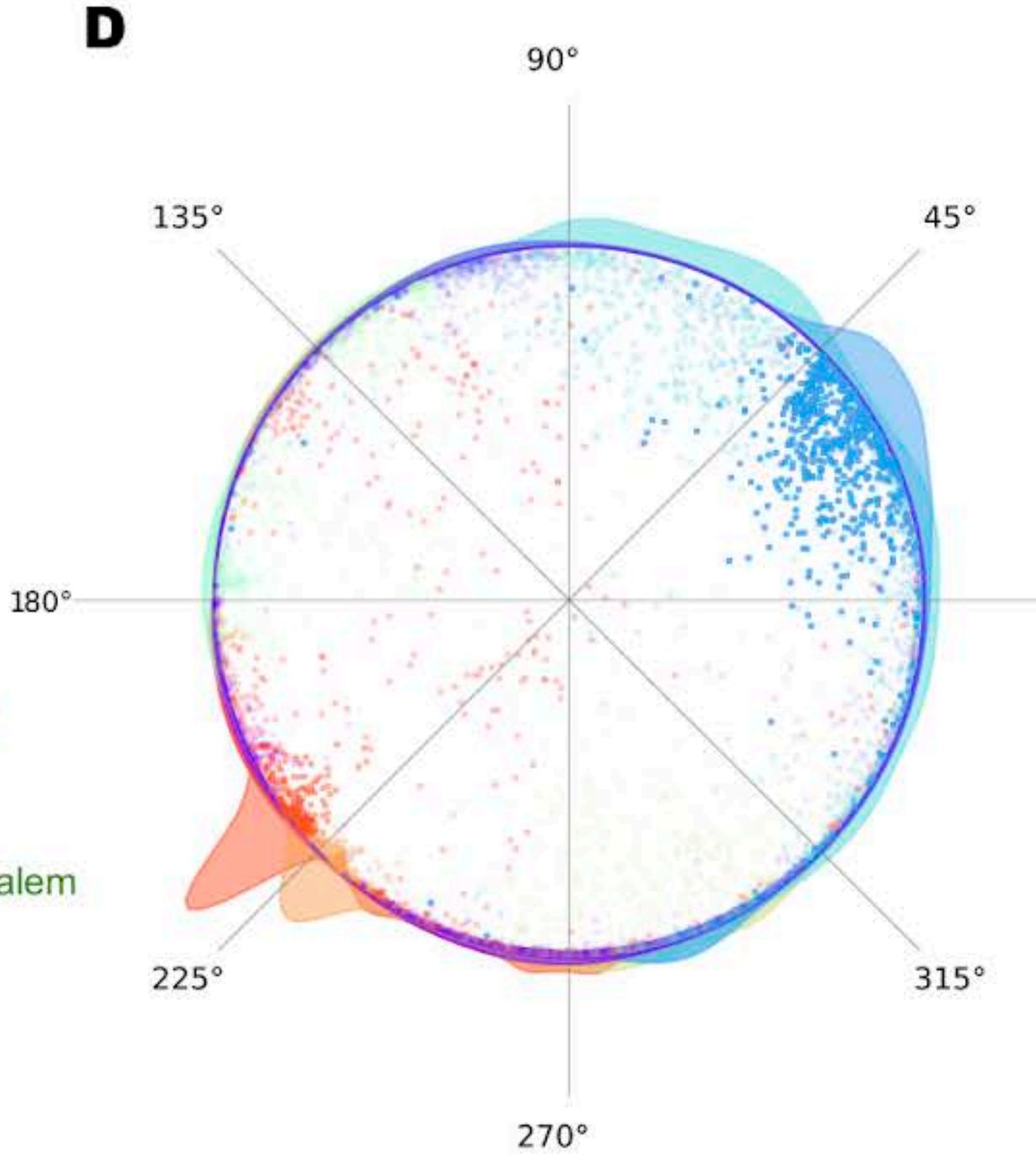




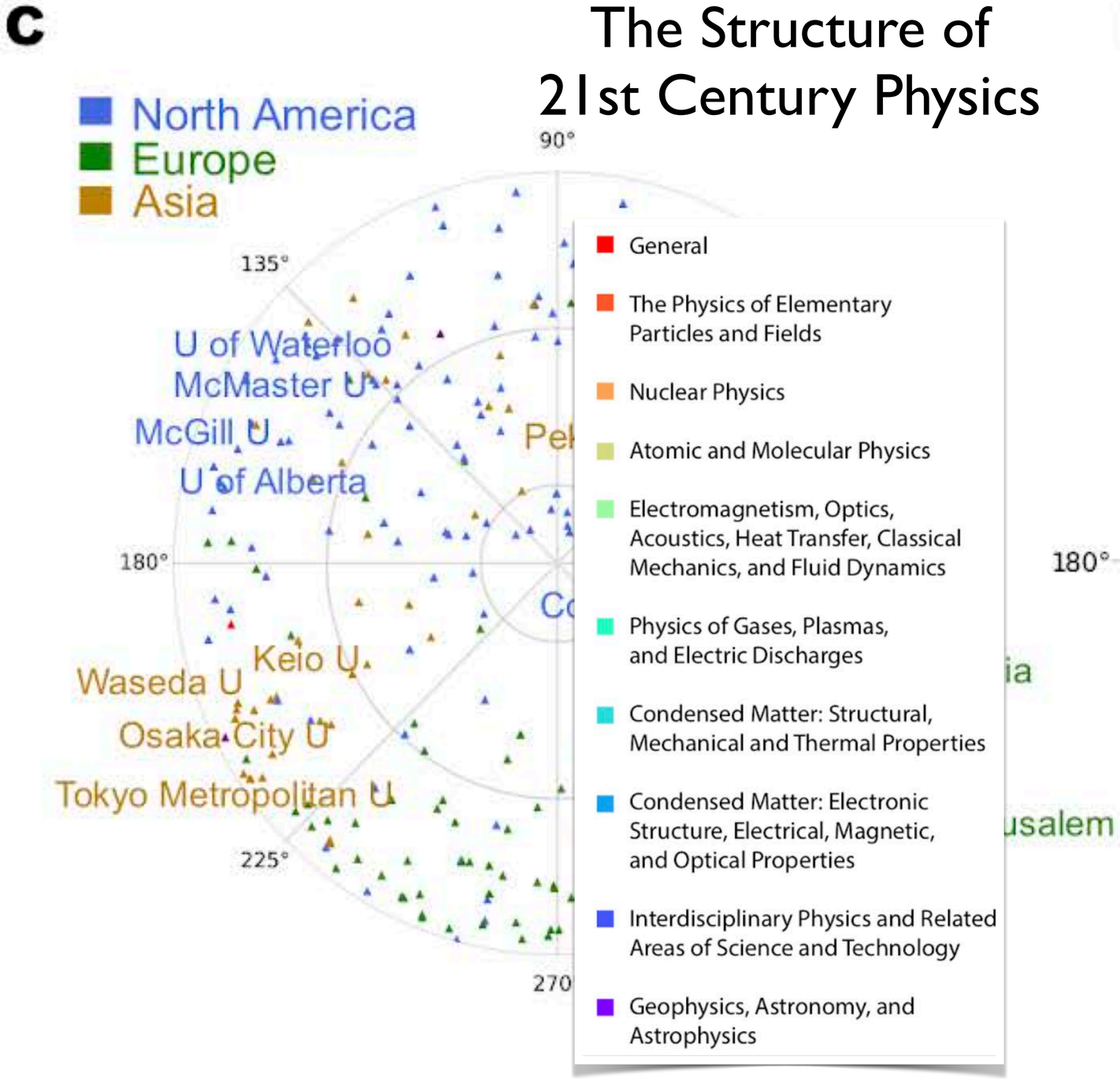
Mutual information I(X;Y) rises;
Joint entropy H(X,Y) shinks

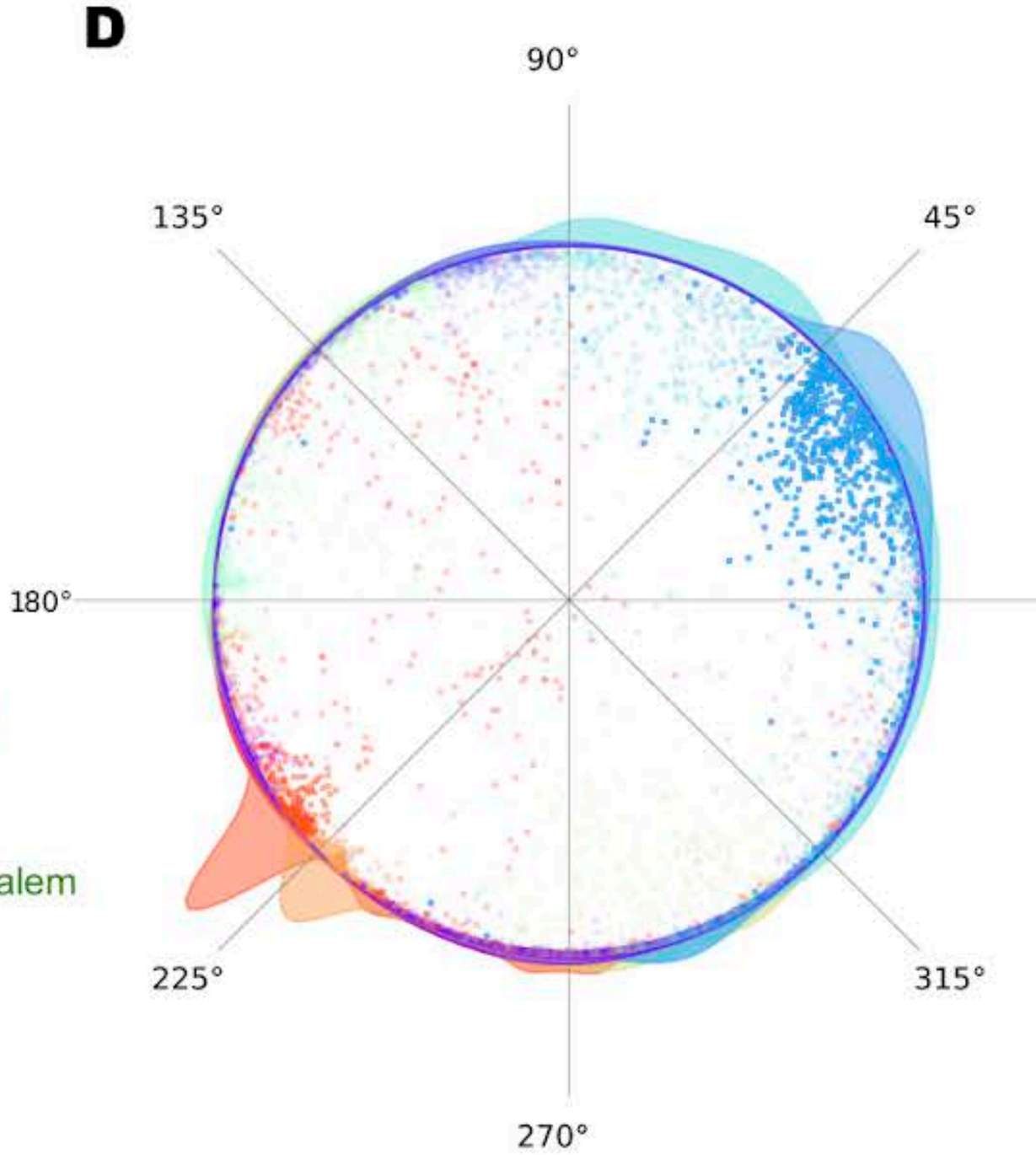
The Structure of 21st Century Physics

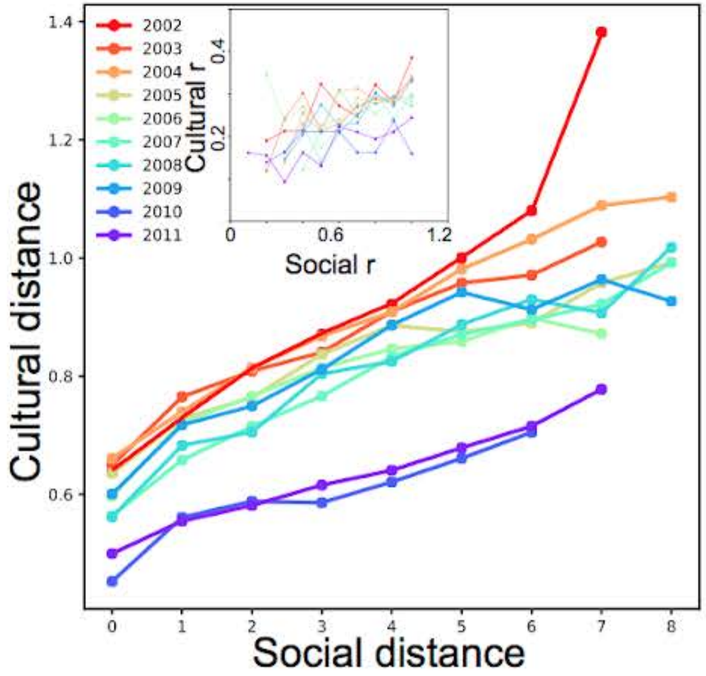


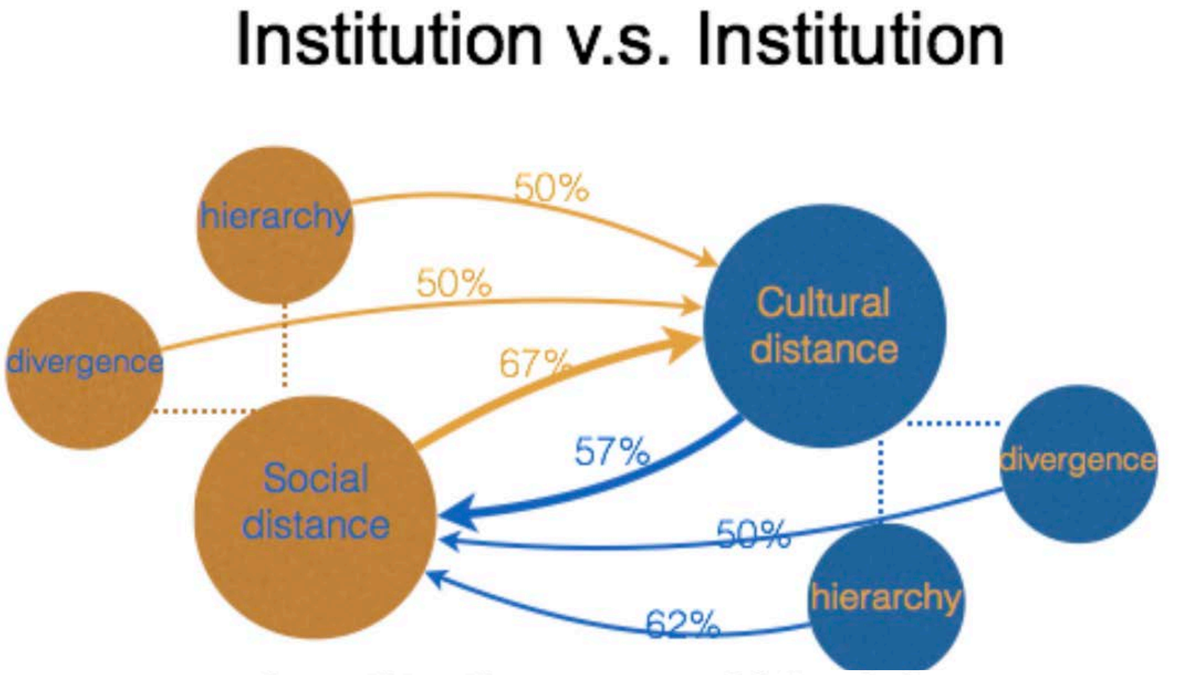


The Structure of 21st Century Physics

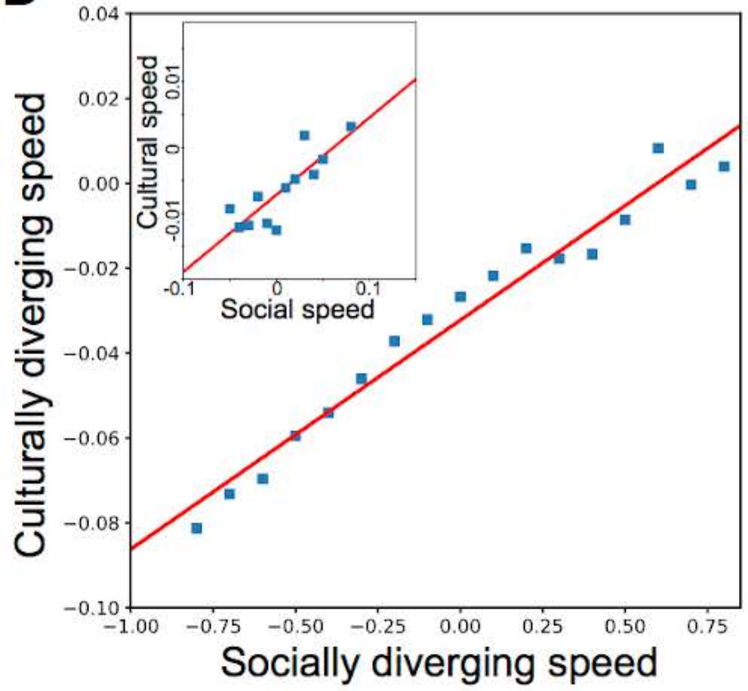


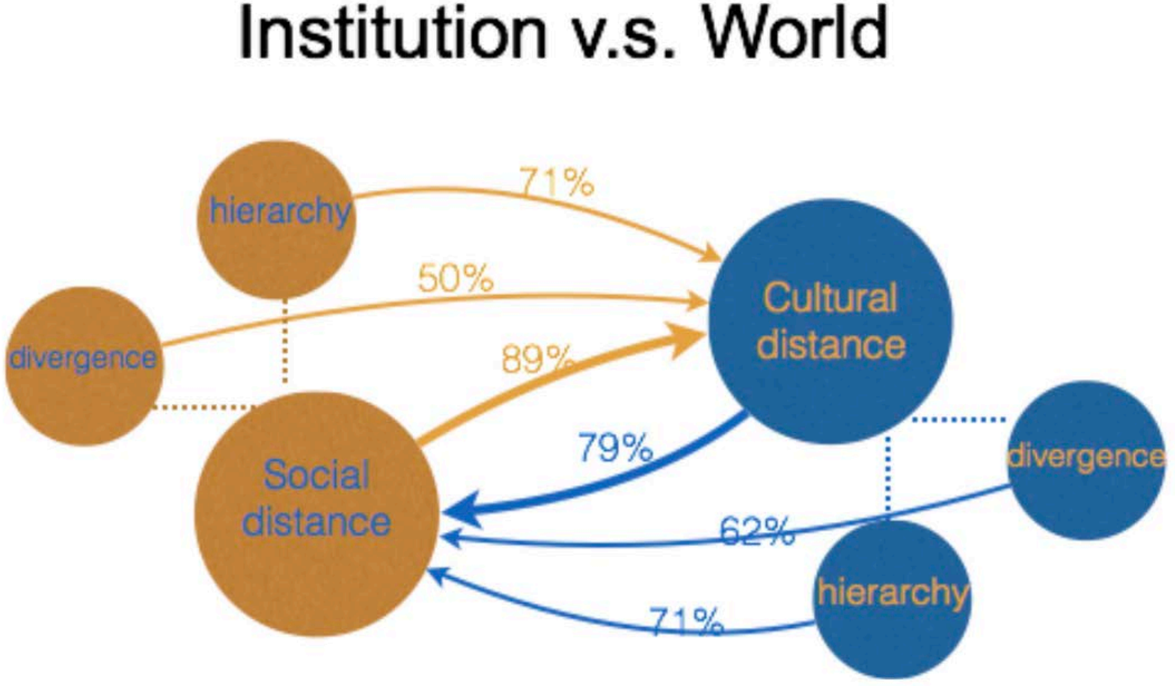



Cultural distance
Social distance
Cultural r
Social r
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011


Institution v.s. Institution
hierarchy
divergence
Cultural distance
Social distance
divergence
hierarchy
50%
50%
67%
57%
50%
62%

D


Culturally diverging speed
Socially diverging speed
Cultural speed
Social speed


Institution v.s. World
hierarchy
divergence
Cultural distance
Social distance
divergence
hierarchy
71%
50%
89%
79%
62%
71%

Diversity Collapse
In other Domains

- Cultural objects and the Interwebz
- Language extinction with social contact
- Biological extinction with ecological contact
- Science and Scholarship with crowding

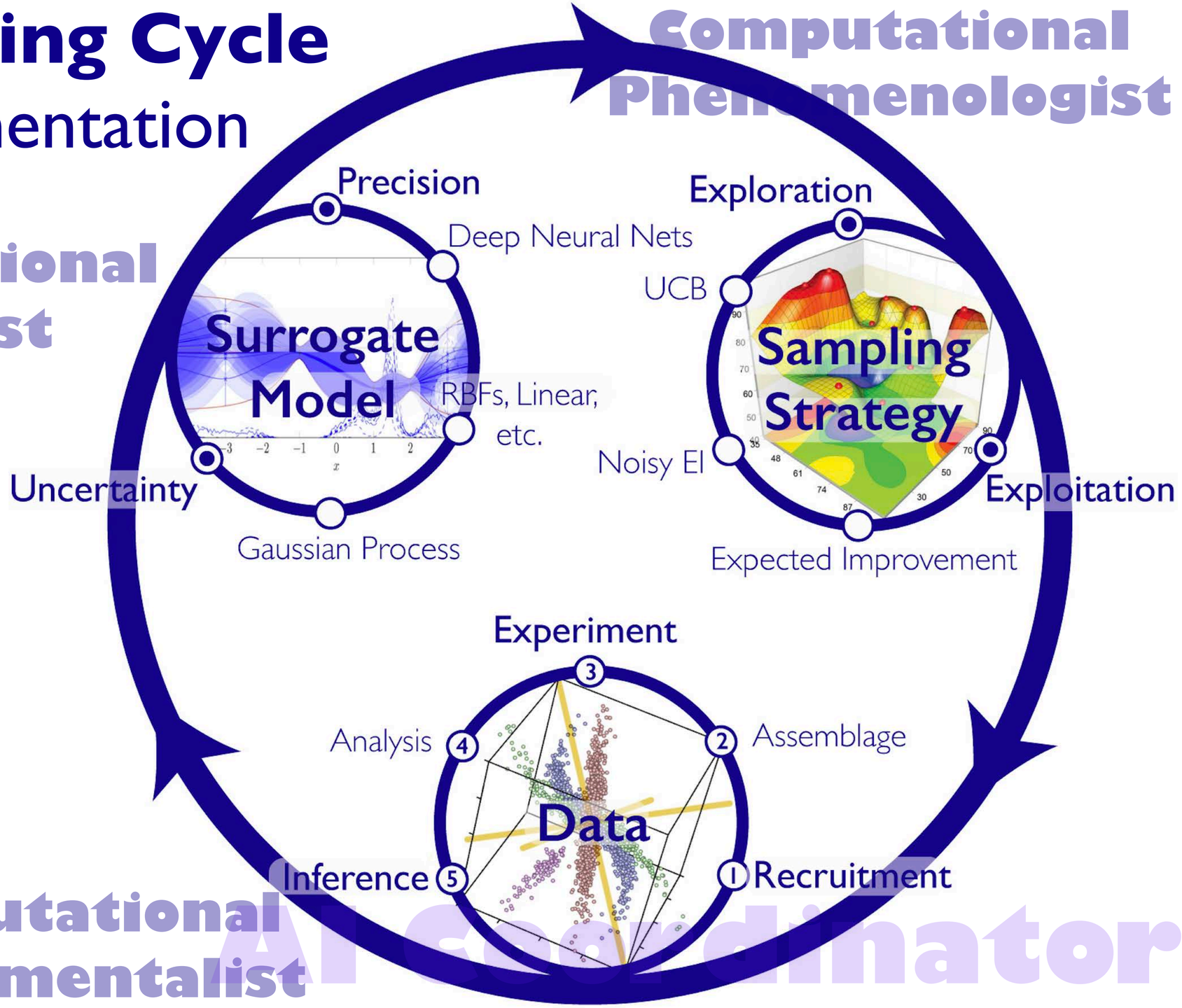
The **Active Learning Cycle** for Social Experimentation

Computational Theorist

Computational Phenomenologist

Computational Experimentalist

AI Coordinator

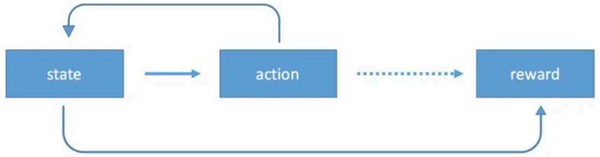


FLAVORS...

Multi-Armed Bandits

Bayesian Optimization

Contextual Bandits



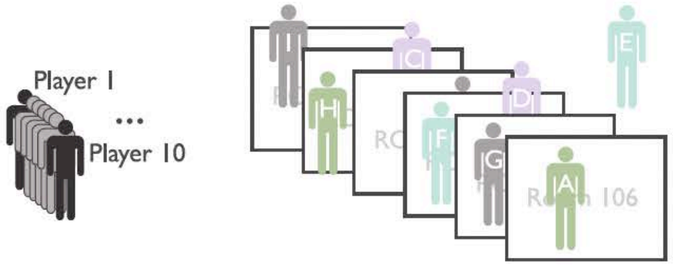
Reinforcement learning


Optimal Team
For Specific Tasks
Size
Social Perception
Cognitive Style:
Visual Imagery
Internal Verbalization
Orthographic Imagery
Representational Manipuation
1000s
of Configurations

Room Assignment Task (Low Complexity)

Process of Solving

N = 9 M = 6 Q = 8



Students constraints:

B and E must be neighbors

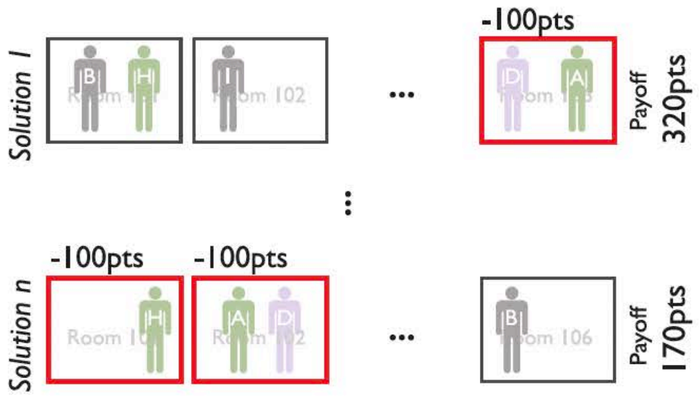
⋮

C and E can't live in the same room

Payoff:

Rooms	101	•••	106
Student A	10	•••	20
⋮	⋮		⋮
Student F	22	•••	40

Grading

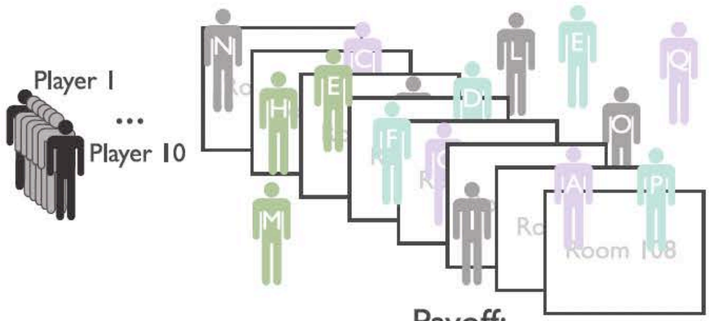


$$\sum_{(s,r)\in S} (\text{payoff}(s,r)) - 100 \times \text{violation}(S)$$

Room Assignment Task (High Complexity)

Process of Solving

N = 18 M = 8 Q = 18



Students constraints:

A and B must be neighbors

A and J can't live in the same room

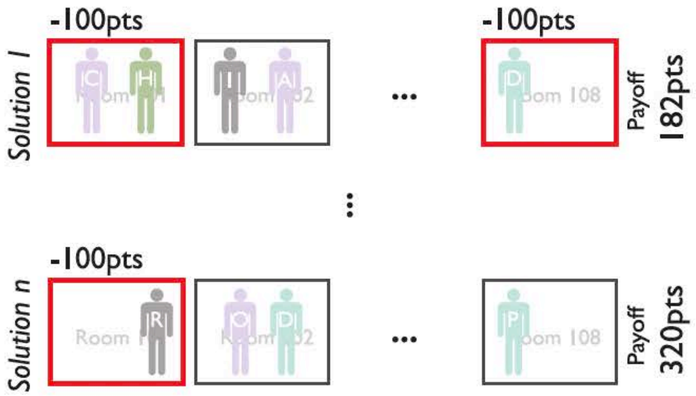
⋮

L and Q can't be neighbors

Payoff:

Rooms	101	•••	108
Student A	15	•••	32
⋮	⋮		⋮
Student R	10	•••	21

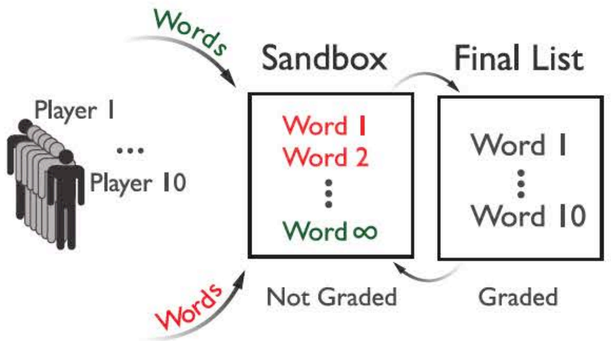
Grading



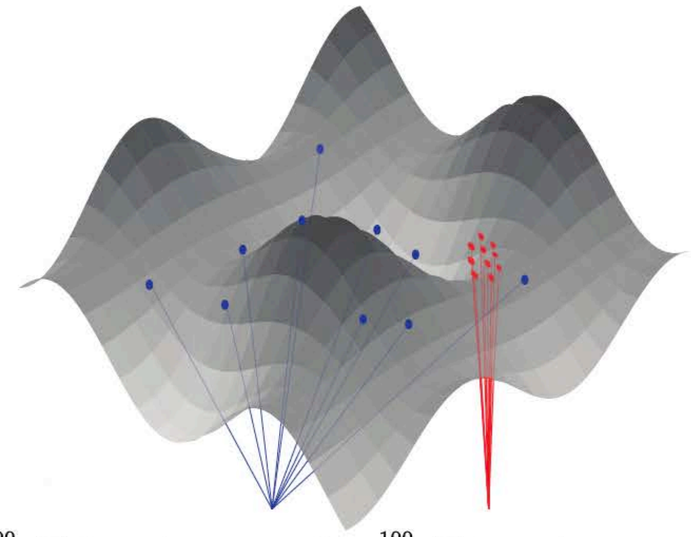
$$\sum_{(s,r)\in S} (\text{payoff}(s,r)) - 100 \times \text{violation}(S)$$

Divergent Association Task

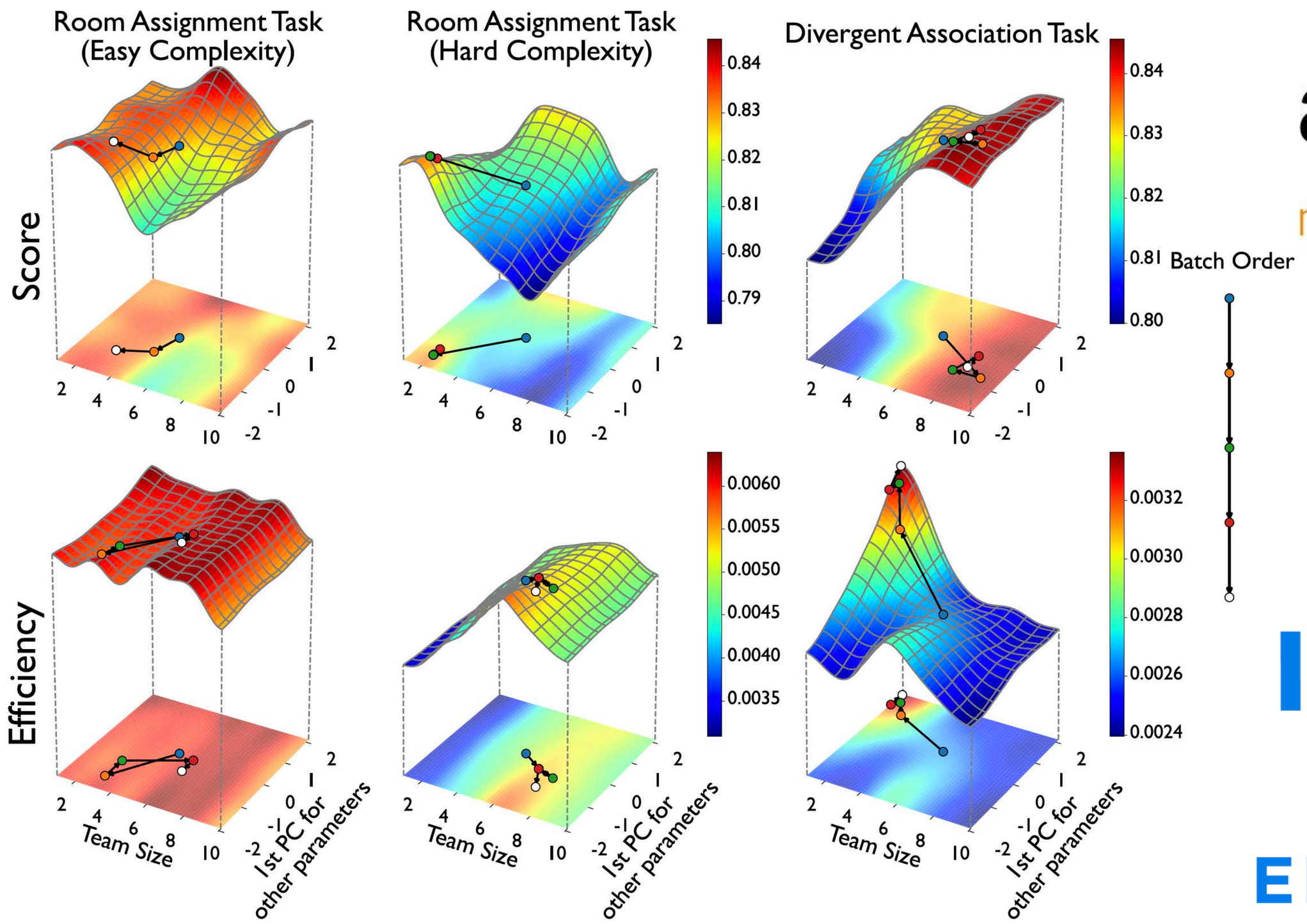
Process of Solving



Grading

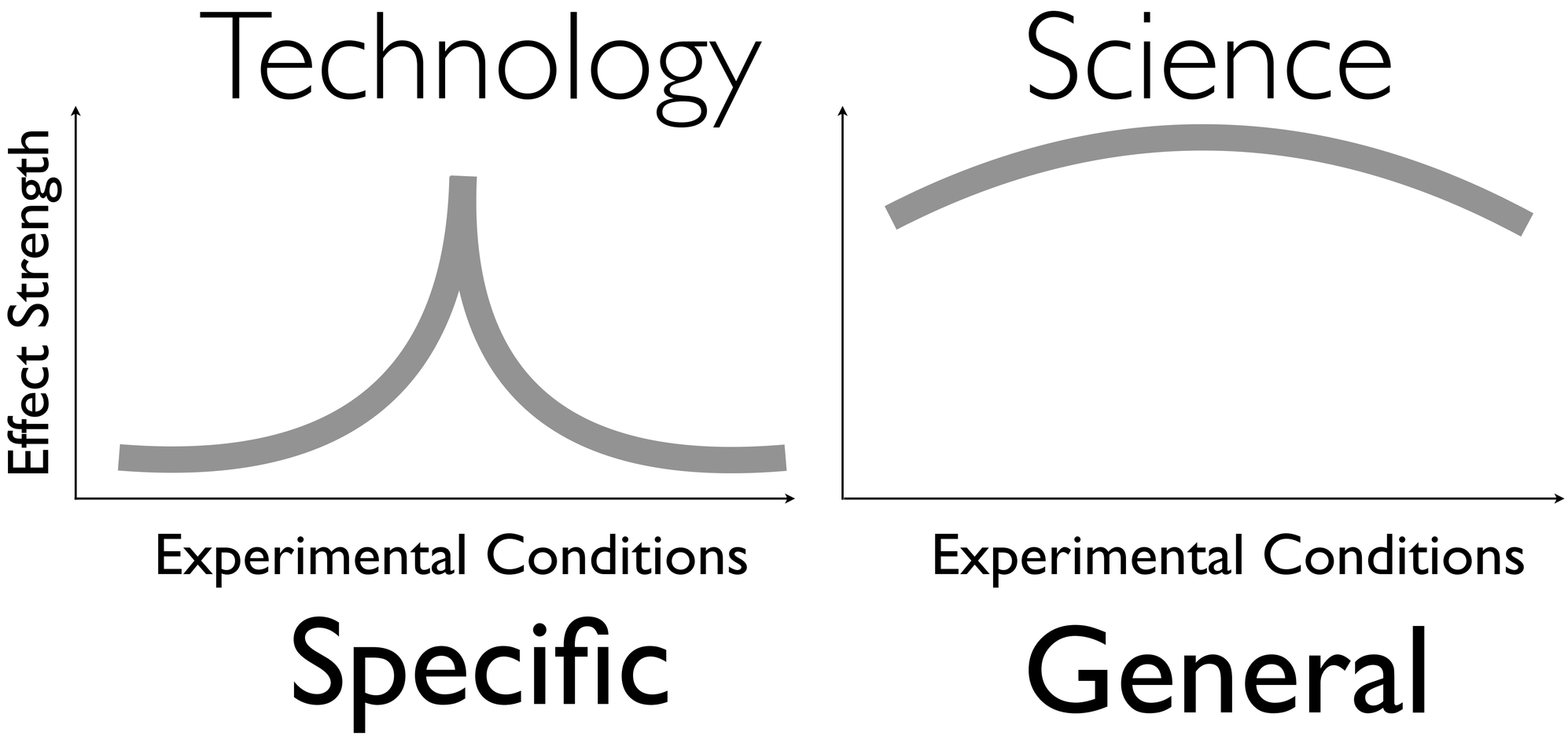


$$\frac{100}{|H|}\sum_{(i,j)\in H}(1-\cos(\text{GloVe}_i,\text{GloVe}_j)) > \frac{100}{|L|}\sum_{(i,j)\in L}(1-\cos(\text{GloVe}_i,\text{GloVe}_j))$$



amazon mechanical turk

EMPIRICA



Reproducible

Conclusion

- Auto-encoders produce powerful, high-dimensional cultural & social representations

  (not just text, but networks, exposures, etc.)

- NOT Text & Networks as Data / as Simulations
- Characterize and engage calculus of meaning & relationship
- Generate Social Science Fiction - Counterfactuals