Johns Hopkins
Whiting School
of Engineering

Sociology Applications

Recap: Last Class

- Language Modeling: Social Experiments
  - The reading author, Christopher Bail (2023), is a sociologist
- HW5
  - Looking forward to see your wonderful projects

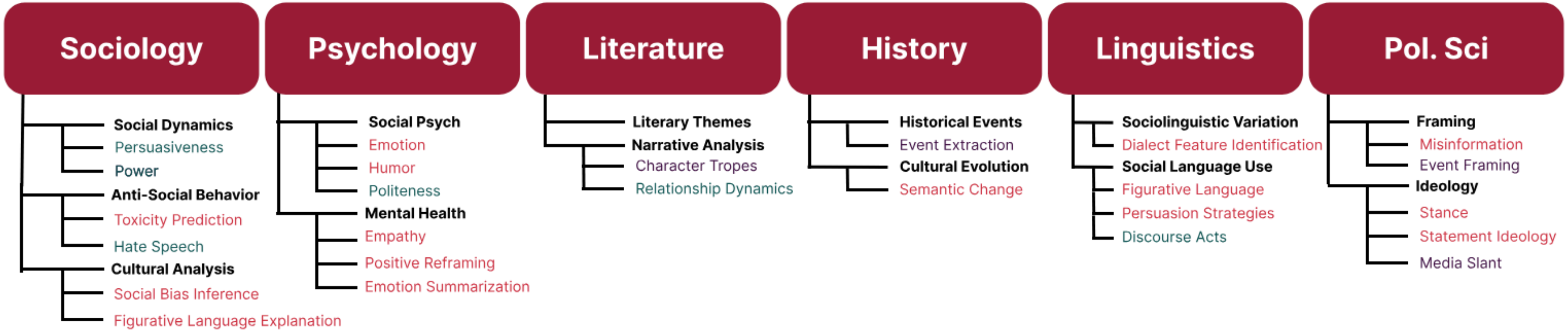
This class

- **What is sociology?**
  - History & Recent Agenda
  - Relationship with other disciplines
- **Big pictures and Examples of Computational Sociology**
  - Based on Review
  - Based on Reference + Examples

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

What is Sociology?

Core subject areas in CSS (and digital humanities)



Ziems et al. “Can Large Language Models Transform Computational Social Science?” *Computational Linguistics, 2024*

Intuition of Sociology

Reinforcement Learning: Intuition

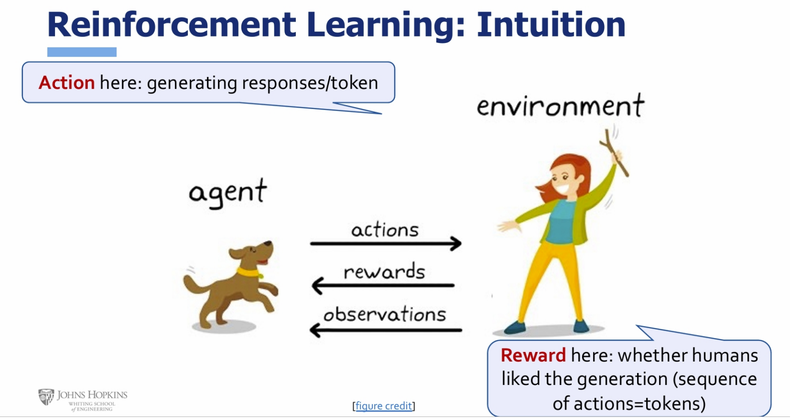
Action here: generating responses/token

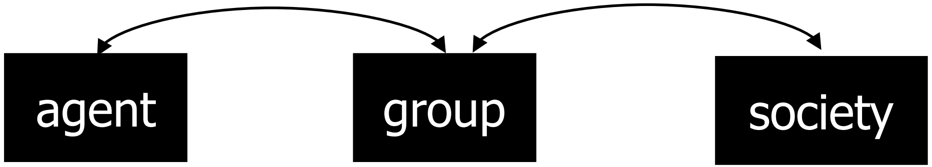
Reward here: whether humans liked the generation (sequence of actions=tokens)

[figure credit]

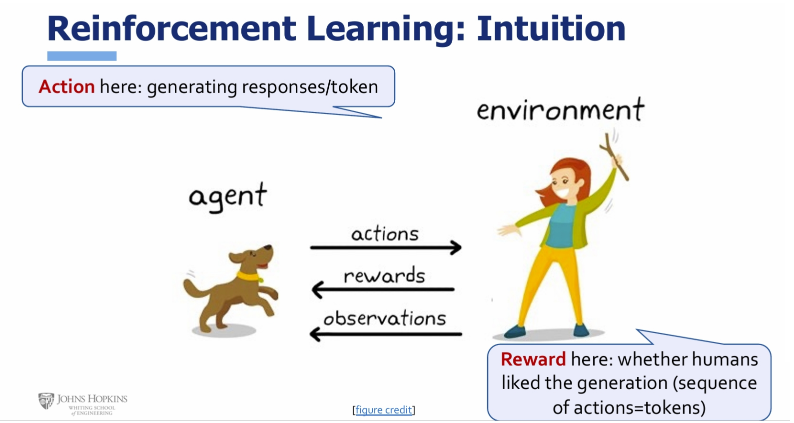
Johns Hopkins
Whiting School of Engineering

Intuition of Sociology

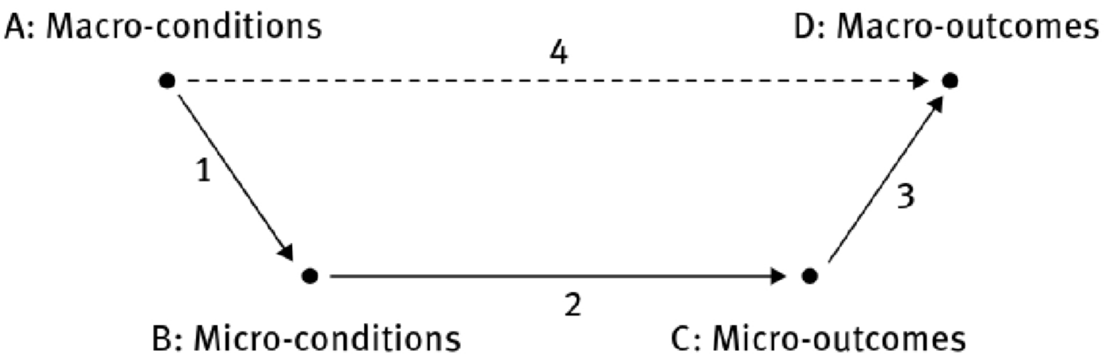




Intuition of Sociology



Coleman's boat



Period of Grand Theory

(late 19th century to 1970s)

e.g., industrialization, traditional society to capitalist society, bureaucracy, race colonialism

Modernization

Why/how societies change

Classic Social Theory

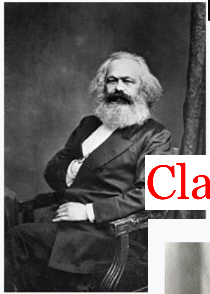
Post-Modernization

Why don't societies change

e.g., habitus, re-production, social capital, inequality, network, institutions, function, social control

Comtemporary Social Theory

Karl Marx
FRSA[1]

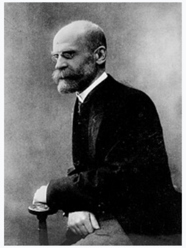


Marx in 1875

Max Weber

1918 portrait

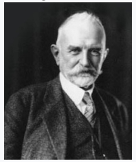
Émile Durkheim



W. E. B. Du Bois

Portrait by James E. Purdy, 1907

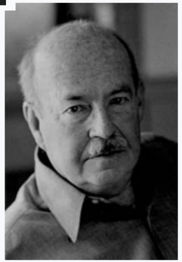
George Herbert Mead



Karl Polanyi

Polanyi, c. 1918

Talcott Parsons

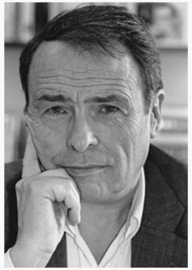


Georg Simmel

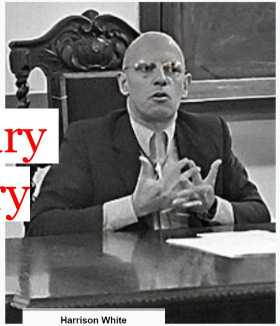
Antonio Gramsci

Gramsci in 1916

Pierre Bourdieu

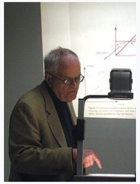


Michel Foucault



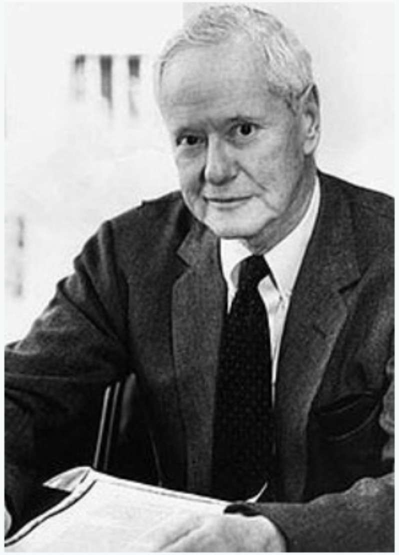
1974

Harrison White



Period of Middle Range Theory (50s to now)

Robert K. Merton



of their big brothers, some sociologists despair. They begin to ask: is a science of society really possible unless we institute a total system of sociology? But this perspective ignores the fact that between twentieth-century physics and twentieth-century sociology stand billions of man-hours of sustained, disciplined, and cumulative research. Perhaps sociology is not yet ready for its Einstein because it has not yet found its Kepler – to say nothing of its Newton, Laplace, Gibbs, Maxwell or Planck.

Einstein observed:

> The greater part of physical research is devoted to the development of the various branches in physics, in each of which the object is the theoretical understanding of more or less restricted fields of experience, and in each of which the laws and concepts remain as closely as possible related to experience.[7]

These observations might be pondered by those sociologists who expect a sound general system of sociological theory in our time – or soon after. If the science of physics, with its centuries of enlarged theoretical generalizations, has not managed to develop an all-encompassing theoretical system, then *a fortiori* the science of sociology, which has only begun to accumulate empirically grounded theoretical generalizations of modest scope, would seem well advised to moderate its aspirations for such a system.

Merton Robert 2012 (1949) "On Sociological Theories of the Middle Range" in Graig Calhoun, et al., Classical Sociological Theory (page 538)

Section Membership Totals Comparison

Section	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2023 Low	2023 Student	2023 Regular
Aging and the Life Course	606	624	605	612	620	610	580	612	602	607	607	578	501	523	512	491	32	127	332
Altruism, Morality, and Social Solidarity	-	139	216	320	305	306	318	307	288	259	235	237	230	274	250	241	12	61	168
Animals and Society	176	180	167	172	149	160	154	141	140	146	148	149	136	145	112	102	8	22	72
Asia and Asian America	368	405	351	377	337	349	336	313	348	366	354	388	399	431	446	434	22	149	263
Biosociology and Evolutionary Sociology	203	183	159	158	174	167	150	136	123	126	96	104	92	94	82	72	2	10	60
Children and Youth	419	433	434	447	409	441	421	406	399	342	380	419	348	392	351	327	13	77	237
Collective Behavior and Social Movements	836	825	767	874	841	839	835	814	813	807	782	740	712	767	697	700	41	200	459
Communication, Information Technologies, and Media Sociology	318	317	318	323	325	375	371	331	370	350	350	365	352	437	381	398	29	119	250
Community and Urban Sociology	727	695	659	696	619	626	630	575	584	600	604	600	552	640	608	605	24	155	426
Comparative-Historical Sociology	730	765	693	708	710	810	808	815	808	714	697	681	711	721	696	669	32	209	428
Crime, Law, and Deviance	701	694	624	633	612	626	574	576	607	612	570	574	502	618	536	538	32	159	347
Decision-Making, Social Networks, and Society	152	148	151	165	162	150	135	205	137	148	124	122	112	112	104	114	5	21	88
Disability in Society	165	235	331	308	264	209	304	202	193	191	207	181	174	192	193	203	10	71	122
Drugs and Society (formerly Alcohol, Drugs, and Society)	288	255	213	226	200	195	173	171	162	157	158	140	151	161	128	122	5	33	84
Economic Sociology	780	823	836	872	823	848	808	748	782	783	731	760	701	778	716	713	30	221	462
Environmental Sociology	461	478	463	473	491	516	507	487	512	510	499	540	493	568	531	542	34	163	345
Ethnomethodology and Conversation Analysis	189	199	157	153	129	152	144	129	131	119	136	115	113	133	130	118	8	24	86
Family	817	822	801	809	754	797	786	799	721	679	678	665	573	634	617	615	18	163	434
Global and Transnational Sociology	-	-	516	649	627	703	727	713	698	697	686	684	655	724	681	701	34	217	450
History of Sociology and Social Thought	212	213	207	199	199	196	198	194	176	169	168	215	209	272	222	203	16	42	145
Inequality, Poverty, and Mobility	-	-	-	545	671	751	802	814	801	815	874	874	839	907	861	798	32	241	525
International Migration	561	607	579	630	593	680	674	625	654	620	621	629	614	657	651	609	35	176	398
Labor and Labor Movements	424	415	352	393	435	431	409	409	408	413	408	373	354	390	379	359	21	121	217
Latina/o Sociology	349	378	325	351	318	373	406	408	409	386	369	382	392	419	409	396	14	99	283
Marxist Sociology	414	401	339	343	311	306	343	307	303	303	306	308	259	323	318	307	19	97	191
Mathematical Sociology	211	225	226	231	220	216	216	214	205	206	212	215	263	312	311	312	16	93	203
Medical Sociology	1,023	1,044	1,019	1,034	1,009	1,057	1,070	1,036	1,024	986	993	937	918	965	943	901	38	248	615
Methodology	407	425	419	430	434	424	418	409	424	374	385	390	379	406	387	367	21	101	245
Organizations, Occupation, and Work	1,024	1,025	944	961	996	1,022	1,004	1,007	1,000	966	931	890	819	983	928	917	31	257	629
Peace, War, and Social Conflict	321	330	313	307	302	302	299	256	299	302	330	305	245	263	231	226	20	53	153
Political Economy of the World-System	421	431	386	430	412	414	417	409	412	373	339	306	296	339	291	285	18	84	183
Political Sociology	834	869	786	884	858	901	869	818	835	820	802	788	780	812	814	781	47	260	474
Race, Gender, and Class	943	965	900	999	942	1,004	1,006	930	908	904	977	983	965	1154	1057	1031	45	345	641
Racial and Ethnic Minorities	818	900	820	864	811	869	924	858	854	852	924	875	872	1049	972	928	21	193	714
Science, Knowledge, and Technology	495	487	477	498	496	535	582	621	612	581	594	585	569	608	570	557	29	142	386
Social Psychology	651	663	634	673	692	689	684	633	610	610	608	603	596	579	549	517	19	170	328
Sociological Practice and Public Sociology	215	334	311	332	333	328	341	317	316	303	330	301	305	362	324	308	18	79	211
Sociology of Body and Embodiment	-	302	295	307	306	309	312	321	308	307	306	317	275	252	220	222	11	72	139
Sociology of Consumers and Consumption	-	-	-	-	322	310	314	268	243	300	280	252	213	213	196	183	8	47	128
Sociology of Culture	1,198	1,227	1,132	1,228	1,181	1,209	1,219	1,115	1,079	1028	995	971	948	967	977	955	46	339	570
Sociology of Development	-	-	-	357	421	465	496	481	480	507	500	476	439	436	365	334	8	82	244
Sociology of Education	812	847	813	862	818	833	829	772	754	713	719	722	648	757	712	706	28	221	457
Sociology of Emotions	278	270	262	269	274	275	263	252	226	242	240	250	229	249	246	232	8	77	147
Sociology of Human Rights	251	321	290	302	321	287	297	266	245	233	250	261	223	228	221	195	14	40	141
Sociology of Indigenous Peoples and Native Nations													138	249	228	185	9	50	126
Sociology of Law	409	422	442	414	423	419	415	411	414	405	400	431	416	479	432	431	12	126	293
Sociology of Mental Health	408	410	418	457	428	447	338	307	313	313	382	303	277	343	285	279	10	75	194
Sociology of Population	472	512	499	520	490	514	535	549	517	474	494	494	454	484	507	476	10	114	352
Sociology of Religion	655	698	686	672	643	641	605	605	578	504	465	474	421	460	447	418	25	114	279
Sociology of Sex and Gender	1,165	1,231	1,122	1,190	1,119	1,174	1,135	1,176	1,100	1099	1121	1110	1025	1088	1050	1025	49	313	663
Sociology of Sexualities	458	486	442	507	485	558	538	580	529	504	538	506	442	481	449	458	19	169	270
Teaching and Learning in Sociology	791	813	747	804	801	767	737	675	643	620	642	657	609	690	589	599	24	122	453
Theory	838	858	818	859	802	825	856	835	824	826	802	769	754	755	735	658	33	191	434
Totals	24,994	26,329	25,464	27,827	27,417	28,410	28,312	27,358	26,921	26,271	26,347	25,994	24,692	27,275	25,647	24,863			

Largest American Sociology Association (ASA) Sections by **Number of Member in 2023**:

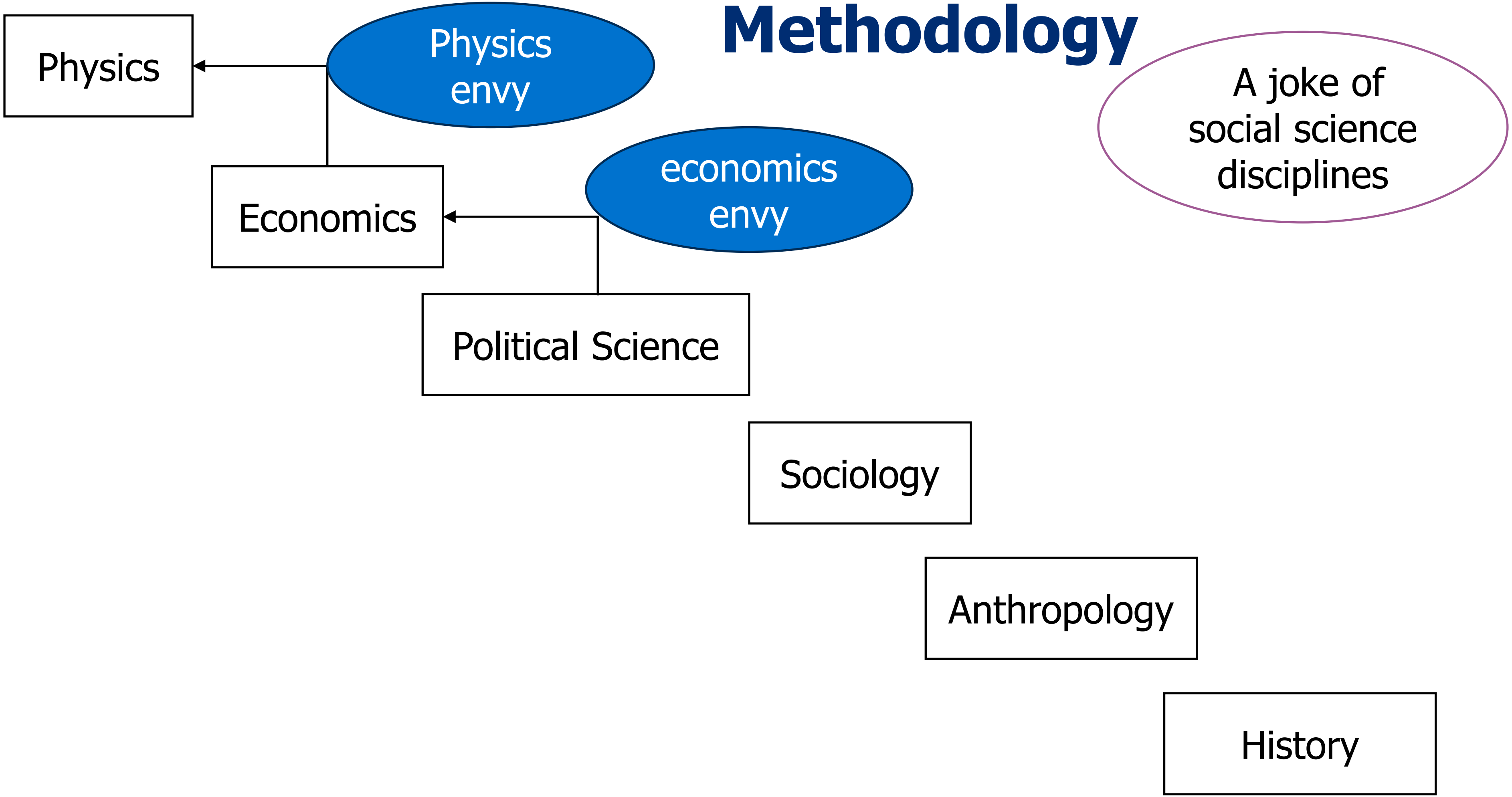
1. Race Gender and Class (1031) Sex and Gender (1025) Racial and Ethnic Minorities (928)

2. Culture (955)

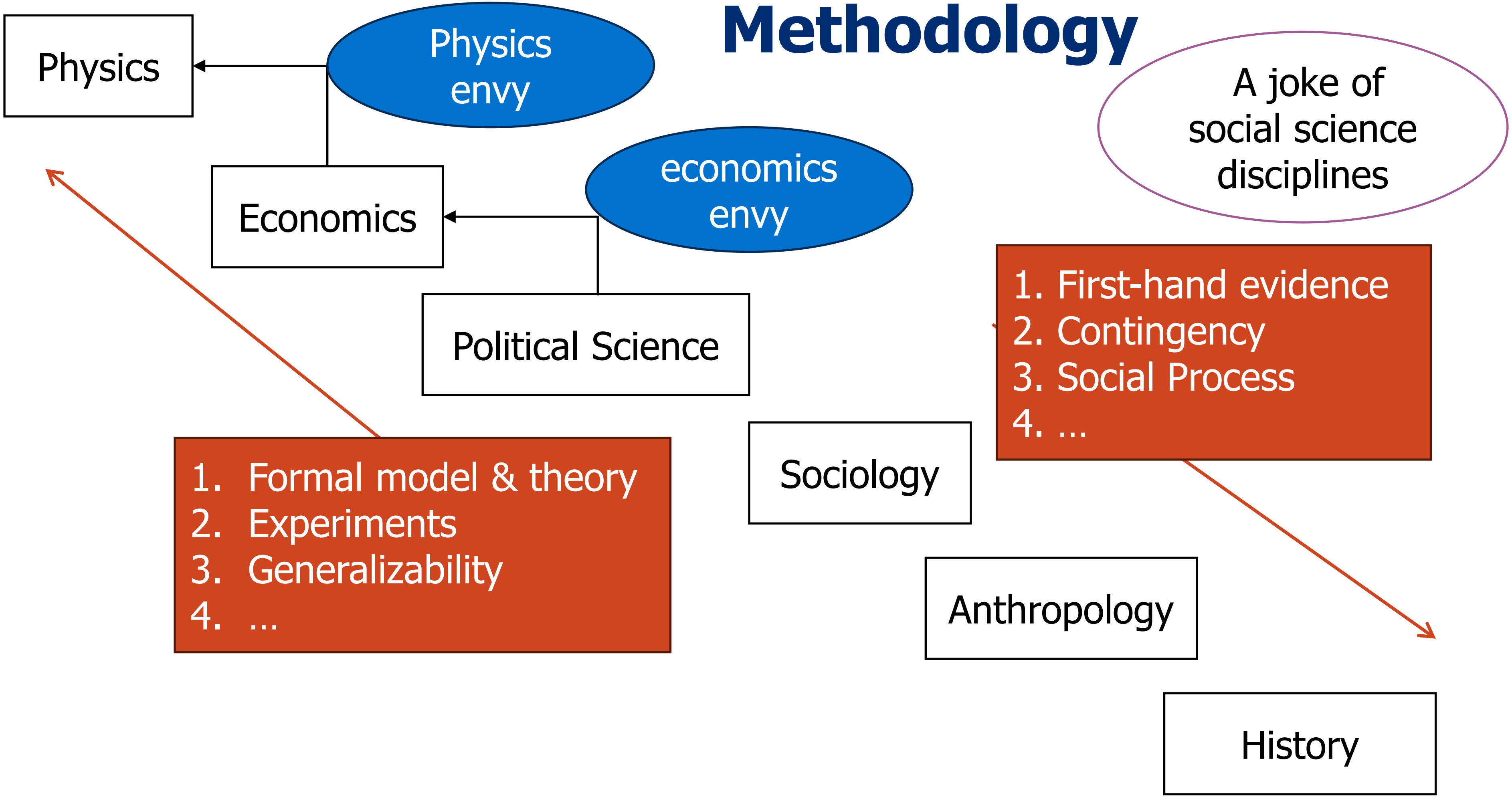
3**. Organization, Occupation, & Work (917)**

https://www.asanet.org/

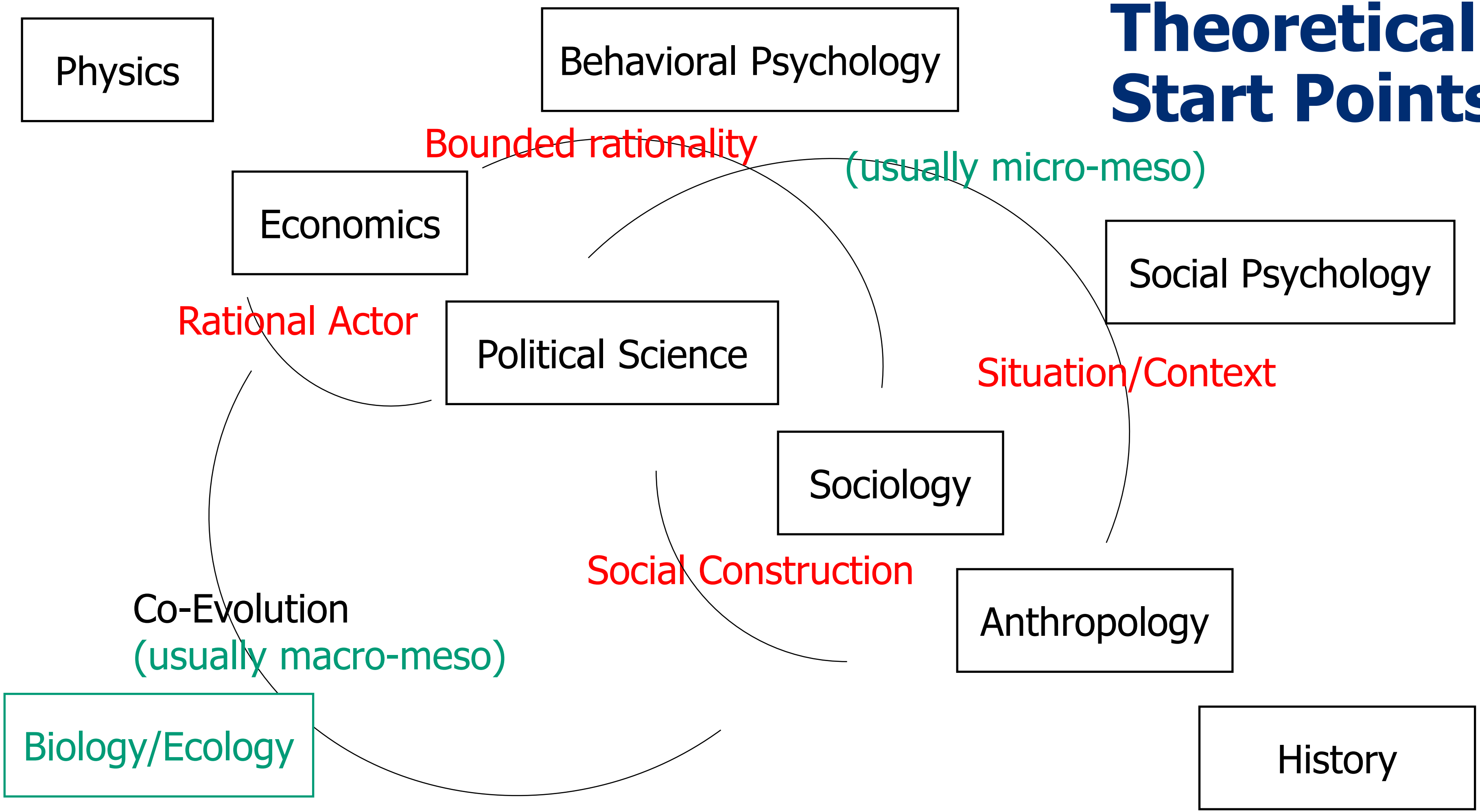
Methodology



Methodology



Theoretical Start Points



Example 1: Bounded Rationality

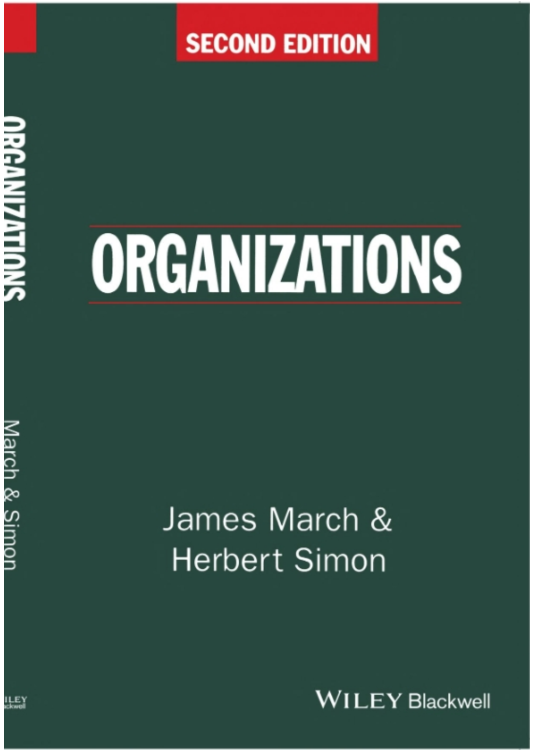
"he (Simon) helped found the Carnegie Mellon School of Computer Science, one of the first such departments in the world."

(https://en.wikipedia.org/wiki/Herbert_A._Simon)

Herbert A. Simon

Simon c. 1981

Born	Herbert Alexander Simon June 15, 1916 Milwaukee, Wisconsin, U.S.
Died	February 9, 2001 (aged 84) Pittsburgh, Pennsylvania, U.S.



James G. March

Born	January 15, 1928 Cleveland, Ohio, U.S.
Died	September 27, 2018 (aged 90) Portola Valley, California, U.S.

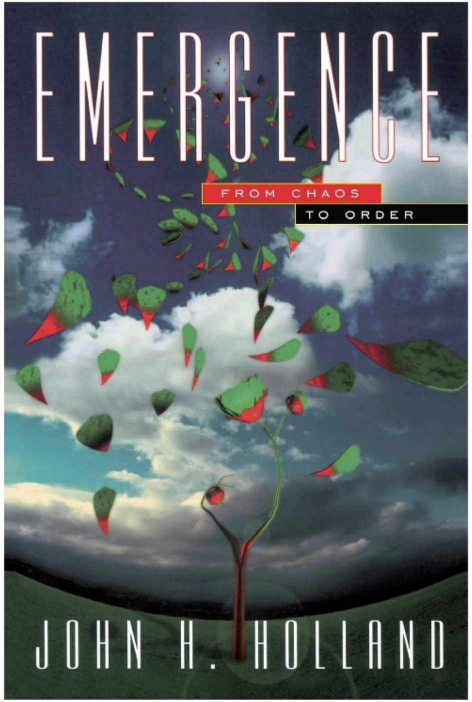
Example 2: Co-evolution

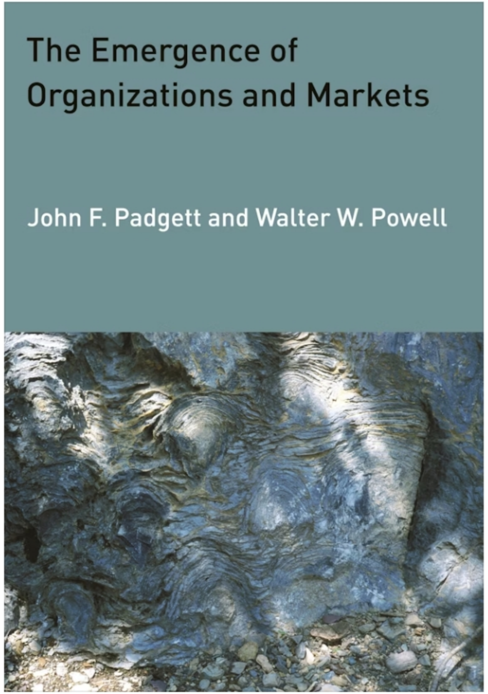
"In 1959, he received the first computer science Ph.D. from the University of Michigan"

(https://en.wikipedia.org/wiki/John_Henry_Holland)

John Henry Holland

Born	February 2, 1929 Fort Wayne, Indiana, US
Died	August 9, 2015 (aged 86) Ann Arbor, Michigan, US





JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Big Picture (1) of Computational Sociology

ANNUAL REVIEWS

Annual Review of Sociology

Computational Social Science and Sociology

Achim Edelmann,[1,2] Tom Wolff,[3] Danielle Montagne,[3] and Christopher A. Bail[3]

[1]Institute of Sociology, University of Bern, 3012 Bern, Switzerland; email: achim.edelmann@soz.unibe.ch

[2]Department of Sociology, London School of Economics and Political Science, London WC2A 2AE, United Kingdom

[3]Department of Sociology, Duke University, Durham, North Carolina 27708, USA; email: christopher.bail@duke.edu

Annu. Rev. Sociol. 2020. 46:61–81

First published as a Review in Advance on April 28, 2020

The *Annual Review of Sociology* is online at soc.annualreviews.org

Keywords

computational social science, machine learning, network analysis, text analysis, demography, social psychology, economic sociology, political sociology, cultural sociology, sociology of knowledge

- There are also annual review(s) of political science, economics, history, …

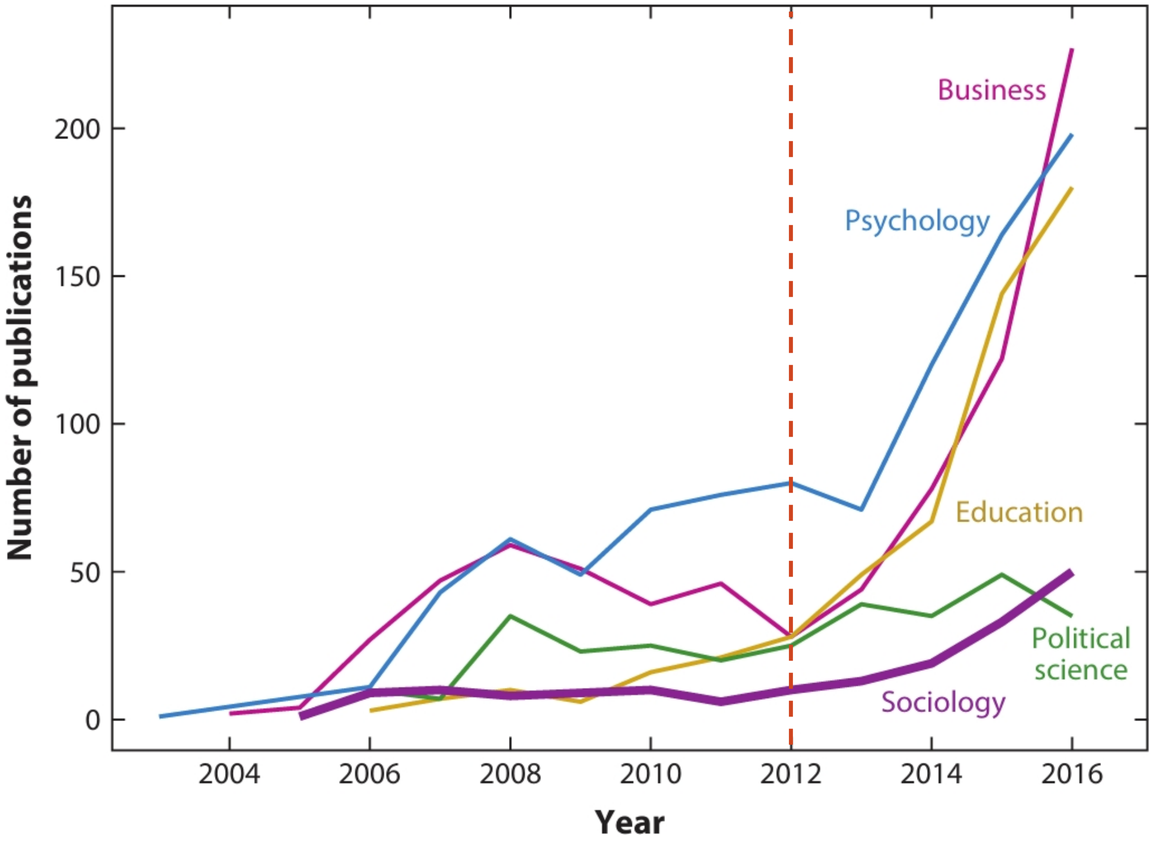
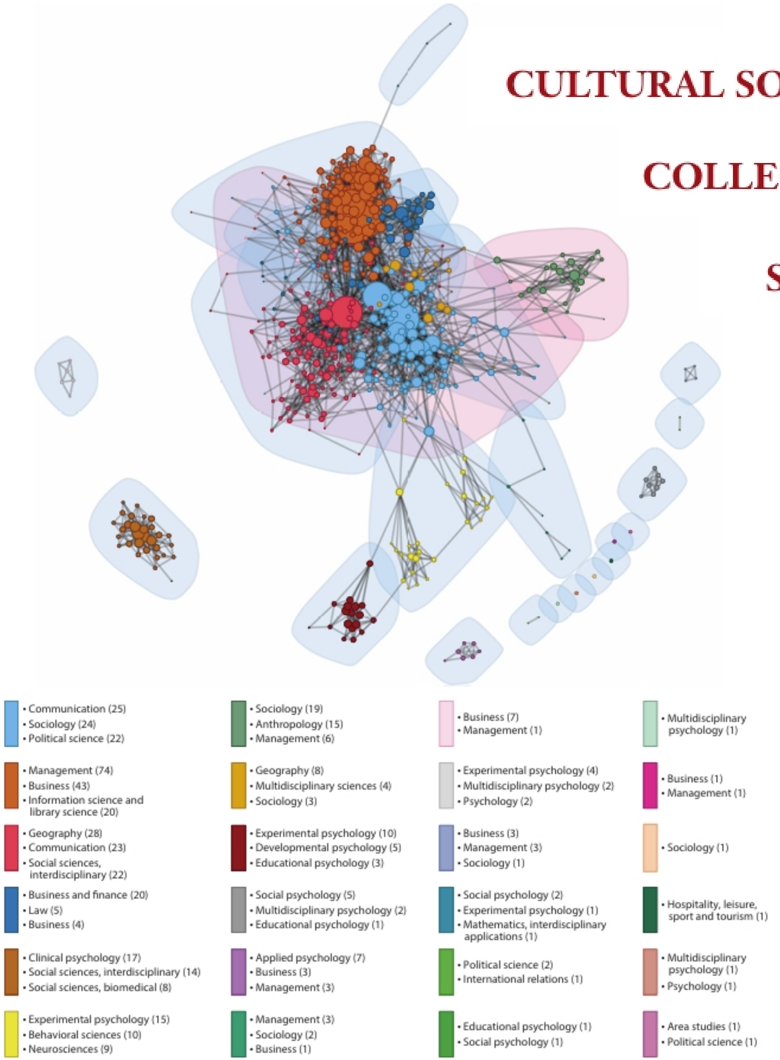


Figure 1

Number of computational social science publications by year—2003–2016—across five scholarly disciplines.



(Caption appears on following page)

CULTURAL SOCIOLOGY, SOCIAL PSYCHOLOGY, AND EMOTIONS

COLLECTIVE BEHAVIOR AND POLITICAL SOCIOLOGY

SOCIAL NETWORKS AND GROUP FORMATION

ECONOMIC SOCIOLOGY AND ORGANIZATIONS

DEMOGRAPHY AND POPULATION STUDIES

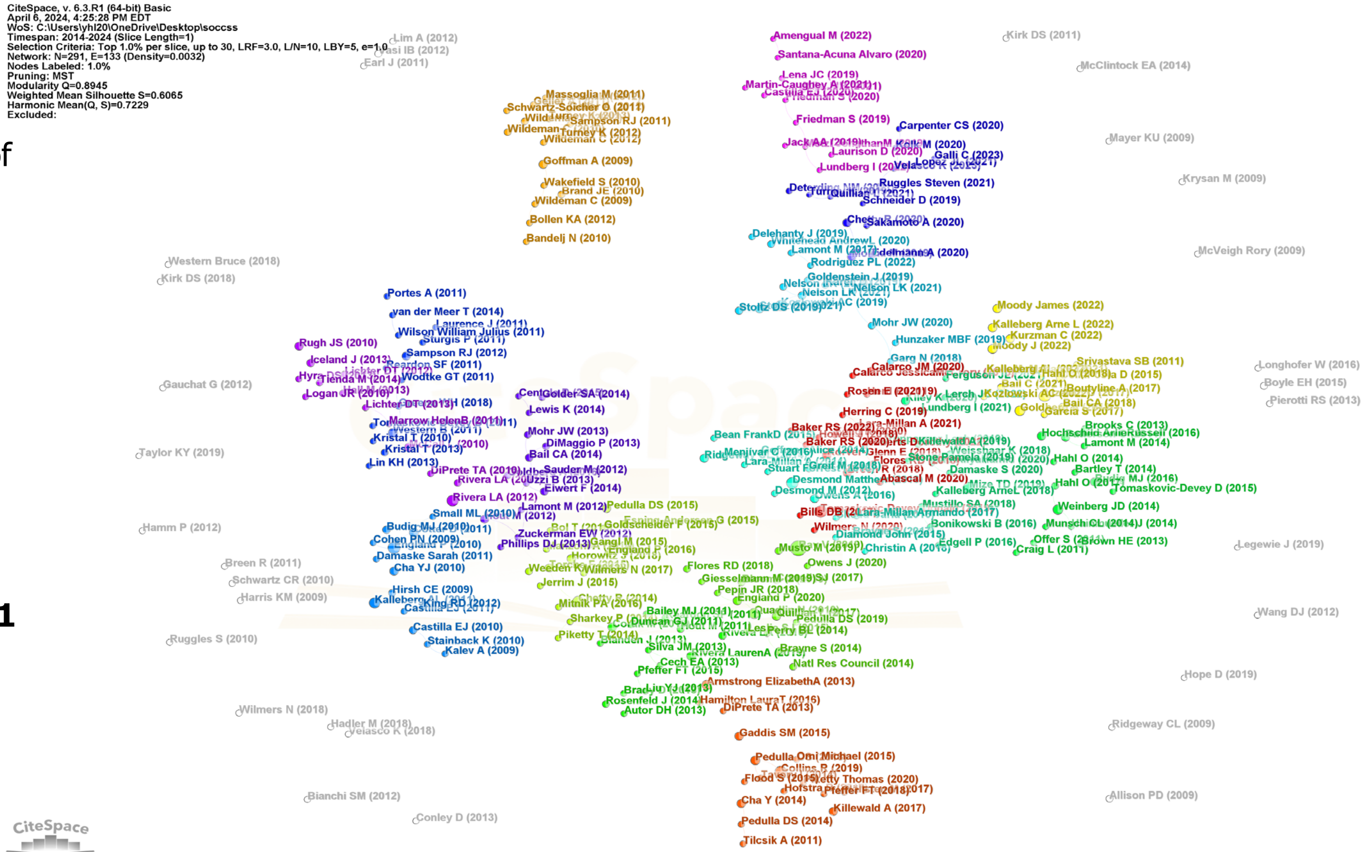
PRODUCTION OF CULTURE

SOCIOLOGY OF KNOWLEDGE

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Big Picture (2) of Computational Sociology

Co-Citation Network of Computation Related Papers from 4 Journals (2014-24)



Download references from Web of Knowledge based on:

1. Search Topic: “**computation**”, “**computational**”

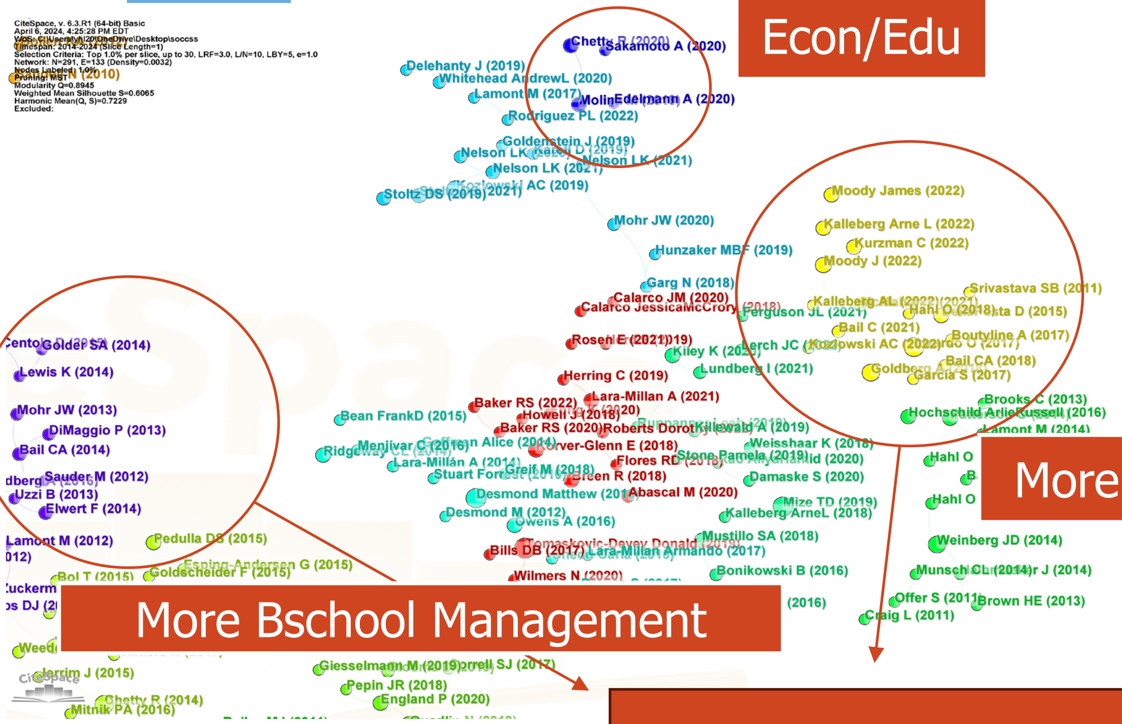
2. Search Top Sociology Journals:
American Sociological Review
American Journal of Sociology
Social Forces
Sociological Method & Research

3. Top 1% cited **references (291 in total)** in each year among these 2100+ papers

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

CiteSpace

Co-Citation network of Computation related papers from 4 journals (14-24)

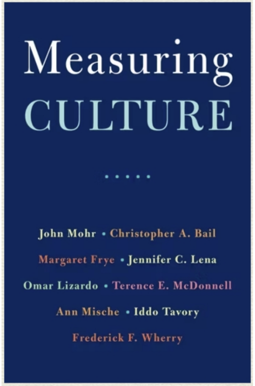


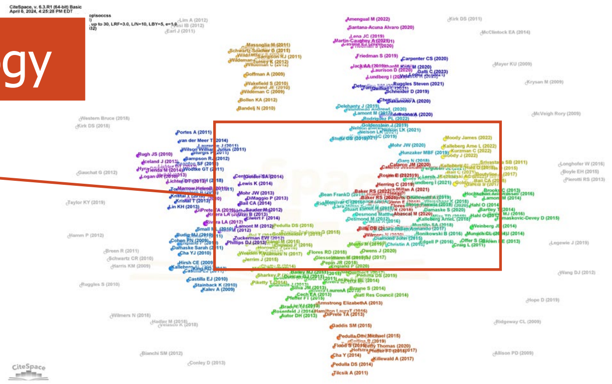
Econ/Edu

More Sociology

More Bschool Management

Network & Culture





Top 15 Co-Cited References

Traditional Soc Research about race & gender, inequality, culture, organization & occupation

CSS Applications

Count	Centra...	Year	Cited References
26	0.69	2019	Ray V, 2019, AM SOCIOL REV, V84, P26, DOI 10.1177/0003122418822335
15	0.12	2010	England P, 2010, GENDER SOC, V24, P149, DOI 10.1177/0891243210361475
15	0.69	2019	Tomaskovic-Devey Donald, 2019, RELATIONAL INEQUALITIES: AN ORGANIZATIONAL APPROACH, V0, P0
14	0.14	2016	Desmond Matthew, 2016, EVICTED POVERTY PROF, V0, P0
14	0.40	2012	Rivera LA, 2012, AM SOCIOL REV, V77, P999, DOI 10.1177/0003122412463213
13	0.07	2017	Lizardo O, 2017, AM SOCIOL REV, V82, P88, DOI 10.1177/0003122416675175
13	0.04	2019	Mize TD, 2019, SOCIOL SCI, V6, P81, DOI 10.15195/v6.a4
13	0.07	2019	Collins R, 2019, THE CREDENTIAL SOCIETY: AN HISTORICAL SOCIOLOGY OF EDUCATION AND STRATIFICATION, V0, P0
11	0.00	2011	Kalleberg AL, 2011, ROSE SER SOCIOL, V0, P1
11	0.57	2014	Chetty R, 2014, Q J ECON, V129, P1553, DOI 10.1093/qje/qju022
11	0.02	2019	Kozlowski AC, 2019, AM SOCIOL REV, V84, P905, DOI 10.1177/0003122419877135
11	0.73	2018	Quadlin N, 2018, AM SOCIOL REV, V83, P331, DOI 10.1177/0003122418762291
10	0.06	2014	Weinberg JD, 2014, SOCIOL SCI, V1, P292, DOI 10.15195/v1.a19
10	0.23	2018	Goldberg A, 2018, AM SOCIOL REV, V83, P897, DOI 10.1177/0003122418797576
9	0.17	2013	Armstrong ElizabethA, 2013, PAYING FOR THE PARTY: HOW COLLEGE MAINTAINS INEQUALITY, V0, P0

11 Social theories and research

Hiring as Cultural Matching: The Case of Elite Professional Service Firms

Lauren A. Rivera[a]

American Sociological Review
77(6) 999–1022
© American Sociological
Association 2012
DOI: 10.1177/0003122412463213
http://asr.sagepub.com
SAGE

A Theory of Racialized Organizations

Victor Ray[a]

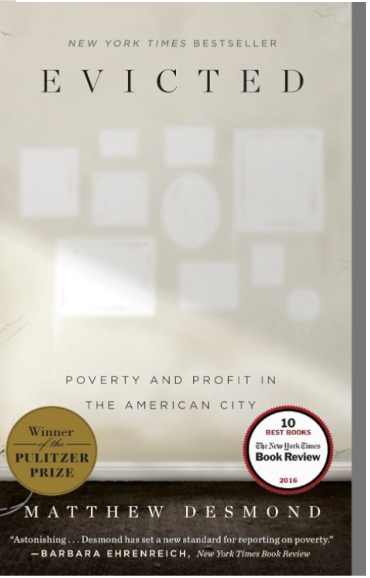
American Sociologi
2019, Vol. 84(1) 26–
© American Sociolog
Association 2019
DOI: 10.1177/00031
journals.sagepub.co
SAGE

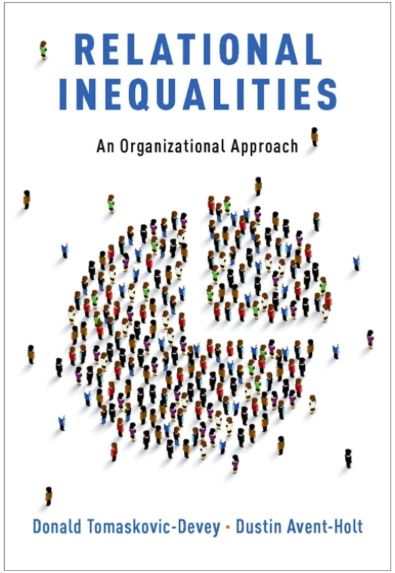
THE GENDER REVOLUTION

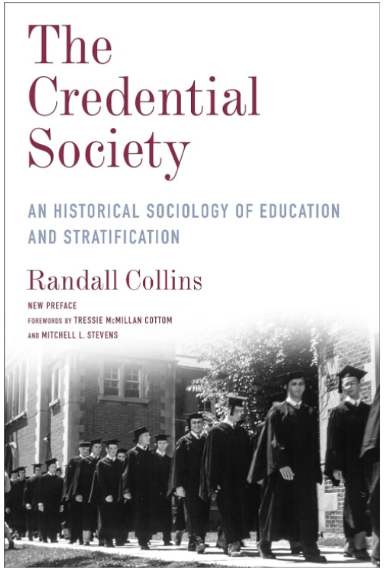
Uneven and Stalled

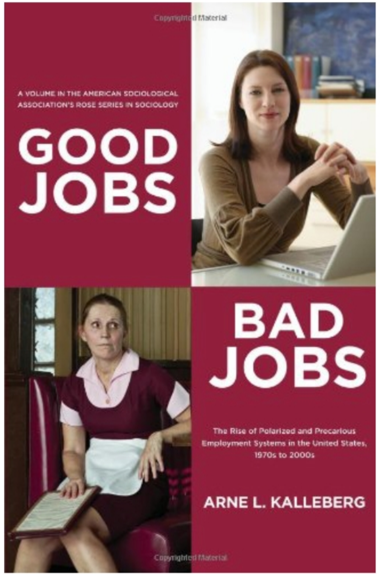
PAULA ENGLAND
Stanford University

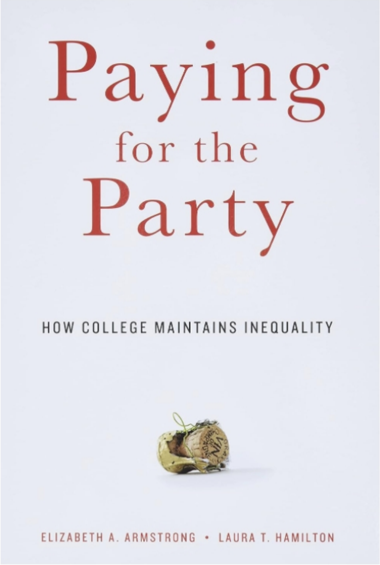
In this article, the author describes sweeping changes in the gender system and offers explanations for why change has been uneven. Because the devaluation of activities done by women has changed little, women have had strong incentive to enter male jobs, but men have had little incentive to take on female activities or jobs. The gender egalitarianism that gained traction was the notion that women should have access to upward mobility and to all areas of schooling and jobs. But persistent gender essentialism means that most people follow gender-typical paths except when upward mobility is impossible otherwise. Middle-class women entered managerial and professional jobs more than working-class women integrated blue-collar jobs because the latter were able to move up while choosing a "female" occupation; many mothers of middle-class women were already in the highest-status female occupations. The author also notes a number of gender-egalitarian trends that have stalled.











2 Regression papers but nonlinear/Big data

sociological science

Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects

Trenton D. Mize

Purdue University

Abstract: Many effects of interest to sociologists are nonlinear. Additionally, many effects of interest are interaction effects—that is, the effect of one independent variable is contingent on the level of another independent variable. The proper way to estimate, interpret, and present these two types of effects individually are well known. However, many analyses that combine these two—that is, tests of interaction when the effects of interest are nonlinear—are not properly interpreted or tested. The consequences of approaching nonlinear interaction effects the way one would approach a linear interaction effect are severe and can often result in incorrect conclusions. I cover both nonlinear effects in the context of linear regression, and—most thoroughly—nonlinear effects in models for categorical outcomes (focusing on binary logit/probit). My goal in this article is to synthesize an evolving methodological literature and to provide straightforward advice and techniques to estimate, interpret, and present nonlinear interaction effects.

Keywords: interaction effects; nonlinearities; categorical models; logit/probit

THE QUARTERLY JOURNAL OF ECONOMICS

Vol. 129 November 2014 Issue 4

WHERE IS THE LAND OF OPPORTUNITY? THE GEOGRAPHY OF INTERGENERATIONAL MOBILITY IN THE UNITED STATES*

RAJ CHETTY
NATHANIEL HENDREN
PATRICK KLINE
EMMANUEL SAEZ

We use administrative records on the incomes of more than 40 million children and their parents to describe three features of intergenerational mobility in the United States. First, we characterize the joint distribution of parent and child income at the national level. The conditional expectation of child income given parent income is linear in percentile ranks. On average, a 10 percentile increase in parent income is associated with a 3.4 percentile increase

Research Examples

1 NLP Paper: Word Embeddings and Culture dimensions

Dimensions

- Rich-poor
- Women-men
- Black-White
- Education

…

Words of Interest

- Sports
- Music

…

Projection and Compare

American Sociological Review
2019, Vol. 84(5) 905–949

DOI: 10.1177/0003122419877135
journals.sagepub.com/home/asr
SAGE

The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings

Austin C. Kozlowski,[a] Matt Taddy,[b]
and James A. Evans[a,c]

Abstract

We argue word embedding models are a useful tool for the study of culture using a historical analysis of shared understandings of social class as an empirical case. Word embeddings represent semantic relations between words as relationships between vectors in a high-dimensional space, specifying a relational model of meaning consistent with contemporary theories of culture. Dimensions induced by word differences (*rich* – *poor*) in these spaces correspond to dimensions of cultural meaning, and the projection of words onto these dimensions reflects widely shared associations, which we validate with surveys. Analyzing text from millions of books published over 100 years, we show that the markers of class continuously shifted amidst the economic transformations of the twentieth century, yet the basic cultural dimensions of class remained remarkably stable. The notable exception is education, which became tightly linked to affluence independent of its association with cultivated taste.

1 Simulation Paper: Associative Diffusion

How does heterogeneous culture emerge from homogeneous networks?

- Traditional Diffusion Model (such as predicting Covid-19 diffusion/infection)
- Missing Part: Perception
- Agent-Based Model
- Simulation
- Test Alternatives

American Sociological Review
2018, Vol. 83(5) 897–932

DOI: 10.1177/0003122418797576
journals.sagepub.com/home/asr
SAGE

Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation

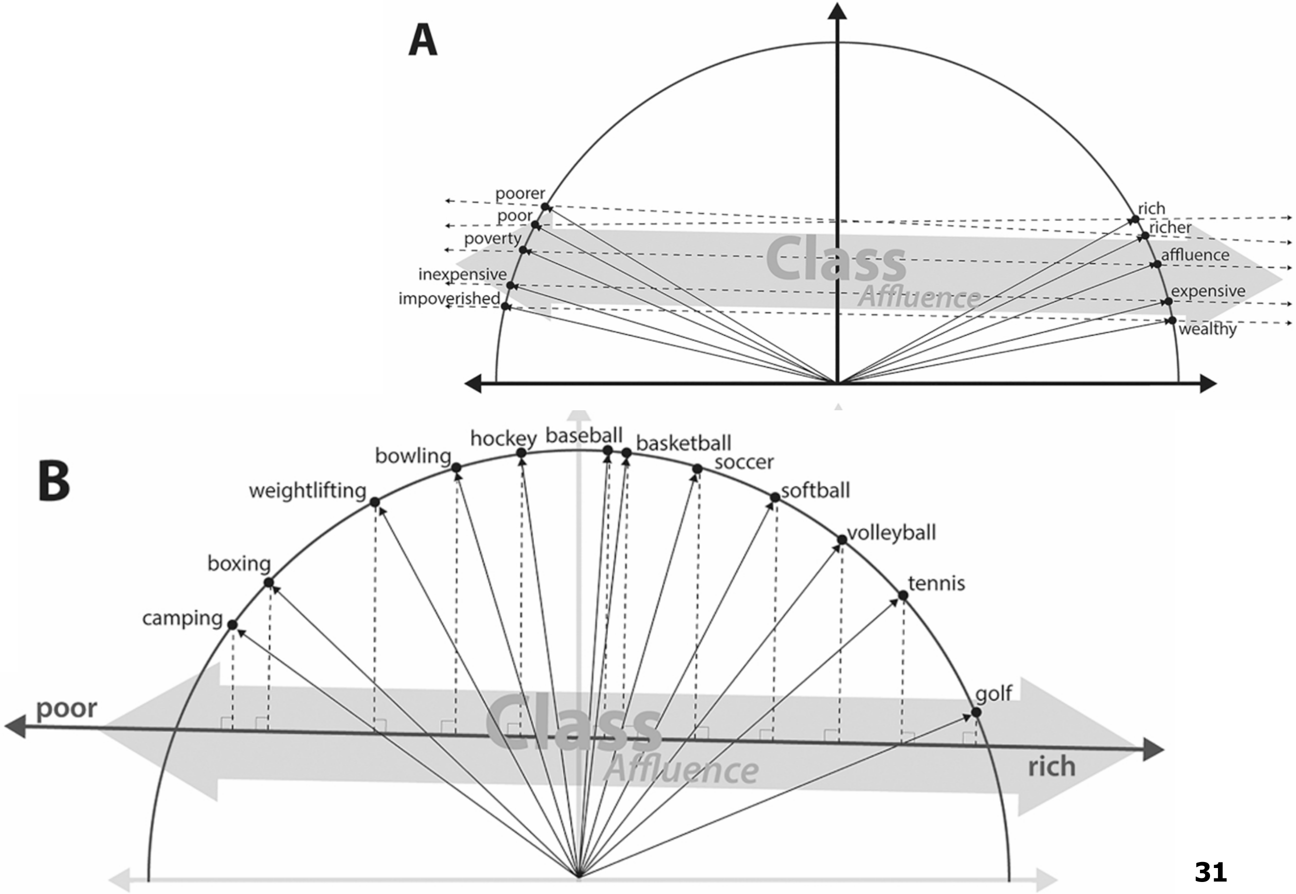
Amir Goldberg[a] and Sarah K. Stein[a]

Example 1: Embeddings

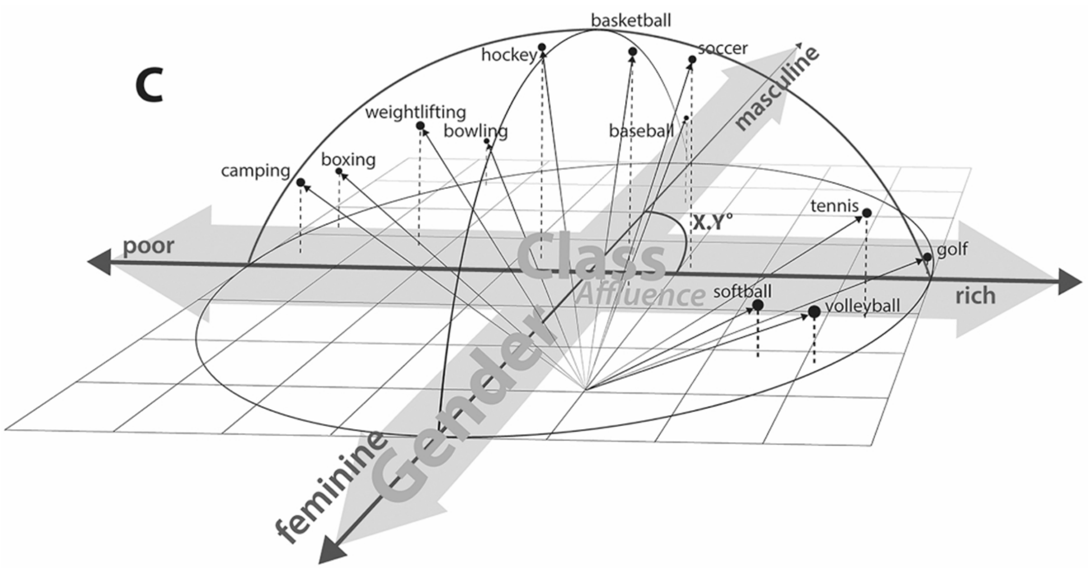
Class dimension and sports projection

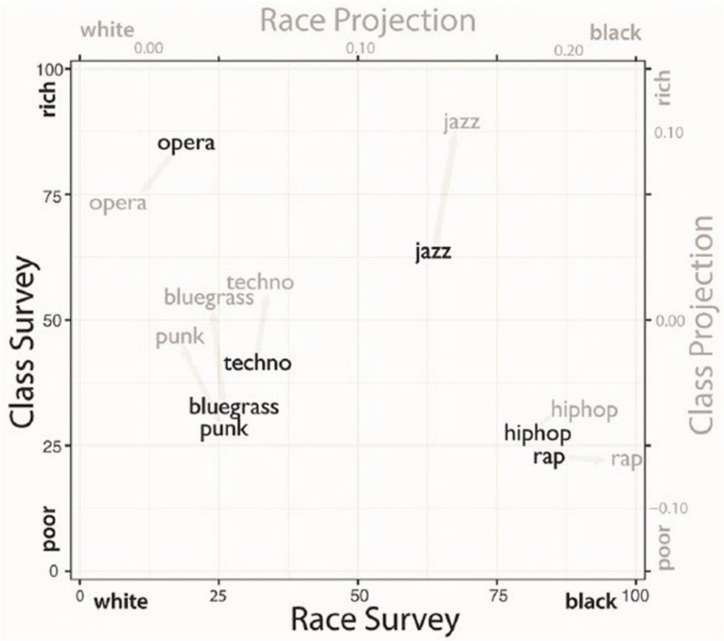
Measuring Cultural Dimensions

To identify cultural dimensions in word embedding models, we average numerous pairs of antonym words. Cultural dimensions are calculated by simply taking the mean of all word pair differences that approximate a given dimension, $\frac{\sum_{p}^{|P|} \overrightarrow{p_1} - \overrightarrow{p_2}}{|P|}$, where p are all antonym word pairs in relevant set P, and $\overrightarrow{p_1}$ and $\overrightarrow{p_2}$ are the first and second word vectors of each pair.[17] The projection of a normalized word vector onto a cultural dimension is calculated with cosine similarity, as is the angle between cultural dimensions.



Gender, Class & Sports / Race, Music

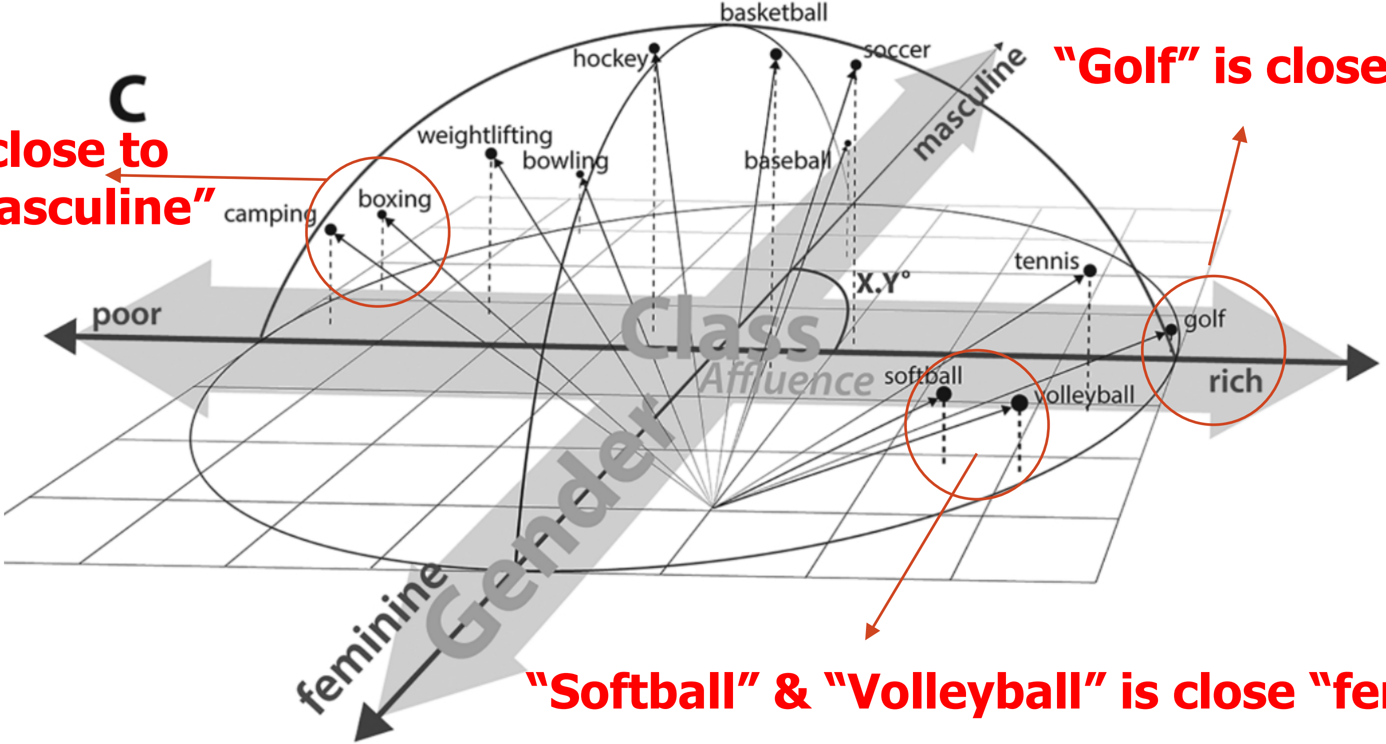




Example: Sports by class & gender (Kozlowski, Taddy and Evans 2019: 913)

“Boxing” is close to “poor” & “masculine”

“Golf” is close to “rich”



“Softball” & “Volleyball” is close “feminine”

Geometry of Culture: Validation using Amazon Turk Survey

In the survey, respondents were asked to rate 59 different items on scales representing association along class, race, and gender lines. All questions followed the format, "On a scale from 0 to 100, with 0 representing *very working class* and 100 representing *very upper class*, how would you rate a *steak*?" For measuring race and gender associations, the survey posed similarly worded questions, replacing "working class" and "upper class" with "white" and "African American," or "feminine" and "masculine," respectively. A full list of items asked on the survey is available in Appendix Table B1. Words were

Table D1. Word Pairs Used to Construct Affluence, Gender, and Race Dimensions for Amazon Mechanical Turk Survey Validation

Affluence		Gender	Race
rich-poor	precious-cheap	man-woman	black-white
richer-poorer	priceless-worthless	men-women	blacks-whites
richest-poorest	privileged-underprivileged	he-she	Black-White
affluence-poverty	propertied-bankrupt	him-her	Blacks-Whites
affluent-destitute	prosperous-unprosperous	his-her	African-European
advantaged-needy	developed-underdeveloped	his-hers	African-Caucasian
wealthy-impoverished	solvency-insolvency	boy-girl	Afro-Anglo
costly-economical	successful-unsuccessful	boys-girls	
exorbitant-impecunious	sumptuous-plain	male-female	
expensive-inexpensive	swanky-basic	masculine-feminine	
exquisite-ruined	thriving-disadvantaged		
extravagant-necessitous	upscale-squalid		
flush-skint	valuable-valueless		
invaluable-cheap	classy-beggarly		
lavish-economical	ritzy-ramshackle		
luxuriant-penurious	opulence-indigence		
luxurious-threadbare	solvent-insolvent		
luxury-cheap	moneyed-moneyless		
moneyed-unmonied	rich-penniless		
opulent-indigent	affluence-penury		
plush-threadbare	posh-plain		
luxuriant-penurious	opulence-indigence		

Geometry of Culture: Validation using Amazon Turk Survey

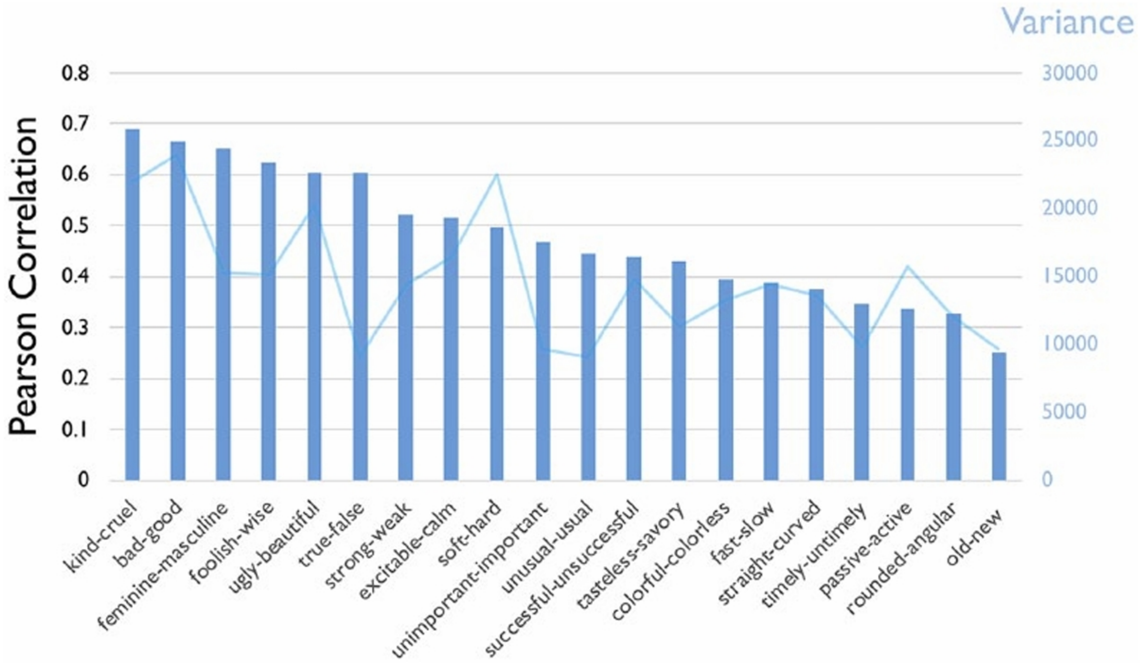


Figure 4. Correlations between Word Embedding Projections and Human-Rated Associations on 20 Semantic Dimensions, Alongside Variance of Average Human-Ratings on Those Dimensions; 1950 to 1959 Google Ngrams Corpus

Geometry of Culture: Validation using Amazon Turk Survey

Table B1. List of Words Rated in Cultural Associations Survey

Occupations	Clothing	Sports	Music Genres	Vehicles	Food	First Names
Banker	Blouse	Baseball	Bluegrass	Bicycle	Beer	Aaliyah
Carpenter	Briefcase	Basketball	Hip hop	Limousine	Cheesecake	Amy
Doctor	Dress	Boxing	Jazz	Minivan	Hamburger	Connor
Engineer	Necklace	Golf	Opera	Motorcycle	Pastry	Jake
Hairdresser	Pants	Hockey	Punk	Skateboard	Salad	Jamal
Journalist	Shirt	Soccer	Rap	SUV	Steak	Molly
Lawyer	Shorts	Softball	Techno	Truck	Wine	Shanice[a]
Nanny	Socks	Tennis				Tyrone
Nurse	Suit	Volleyball				
Plumber	Tuxedo					
Scientist						

Table B3. Percentage of Statistically Significant ($p < .01$) Survey Differences Correctly Classified in Google News Word Embedding Model

	Sports	Food	Music	Occupations	Vehicles	Clothes	Names	All Domains
Gender	87.9%	88.2%	72.2%	93.6%	82.4%	74.4%	95.2%	84.8%
Class	96.3%	93.8%	88.9%	60.9%	94.1%	90.0%	77.3%	75.3%
Race	90.0%	68.8%	100%	51.5%	87.5%	55.0%	94.7%	69.1%

Geometry of Culture: Validation using Historical Survey IPUMS

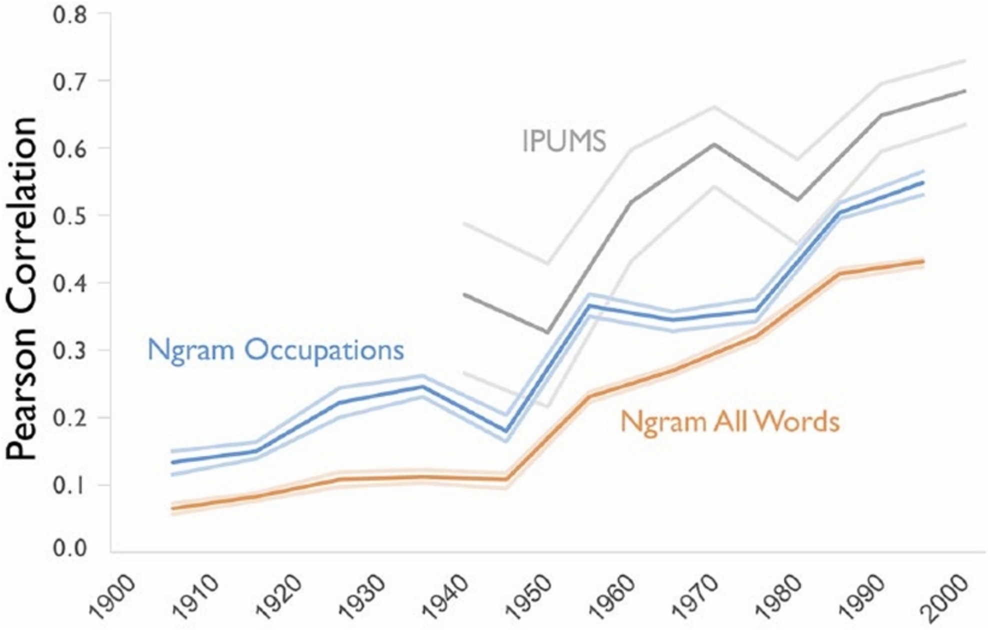


Figure I1. Correlations of Affluence and Education from IPUMS Surveys and Google Ngrams Text
Note: Correlation of occupations' average income and average education by decade; correlation of occupation names' projections on affluence and education dimensions; and correlation of all words' projections on affluence and education dimensions.

Geometry of Culture: Change/Sustain of Cosine Similarity between Dimensions

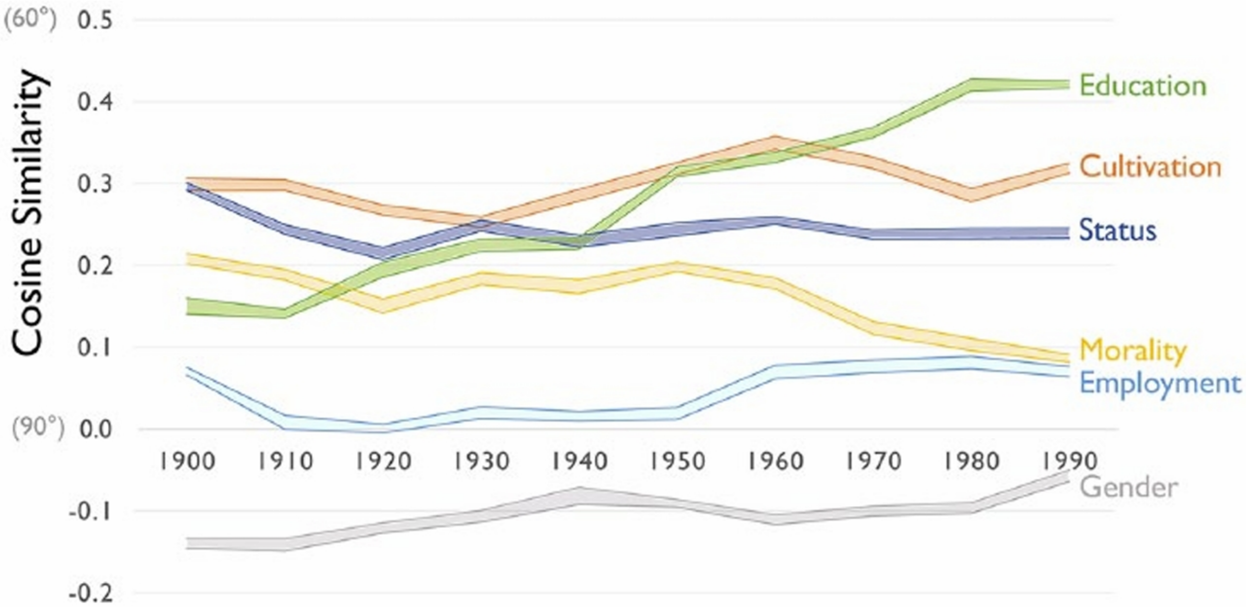


Figure 5. Cosine Similarity between the Affluence Dimension and Six Other Cultural Dimensions of Class by Decade; 1900 to 1999 Google Ngrams Corpus
Note: Bands represent 90 percent bootstrapped confidence intervals produced by subsampling.

Geometry of Culture : Ave Projections Overtime

Figure 9 displays the stability of projections for the 50,000 most common words on each class dimension. The first line represents the average correlation of word projections in the 1900s with their projections in the 1910s, 1920s, and so on through the 1990s. Similarly, the second line shows the correlation between projections in the 1920s with those in the 1930s, 1940s, and so on. For each decade, a word's projection is highly correlated with its projection the following decade, in most cases greater than .9. This correlation diminishes by

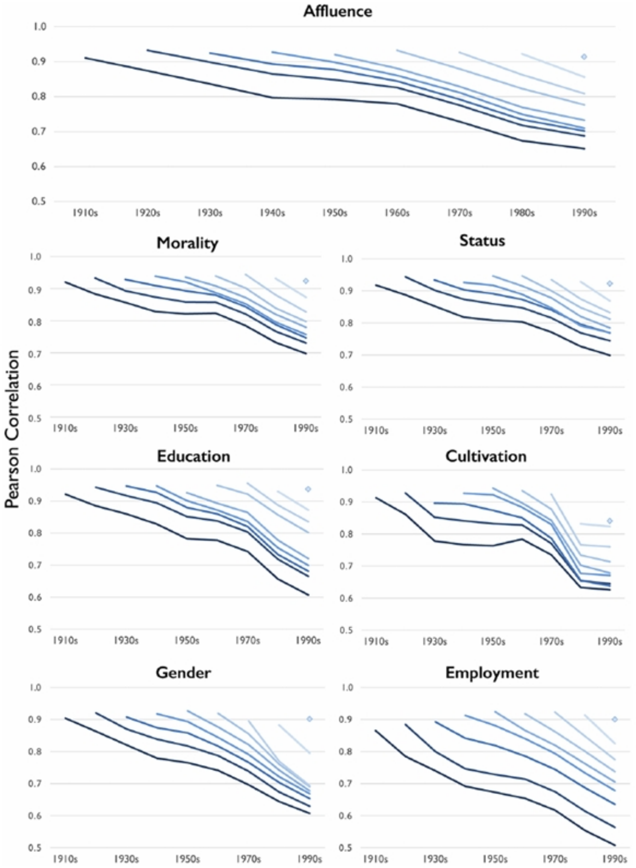


Figure 9. Correlation of 50,000 Most Common Words' Projection in One Decade with Their Projection in Each Subsequent Decade for Seven Cultural Dimensions of Class; 1900 to 1999 Google Ngrams Corpus

Related development 1: Construct Dimensions

Forging Better Axes: Evaluating and Improving the Measurement of Semantic Dimensions in Word Embeddings *

Andrei Boutyline

Department of Sociology
University of Michigan
boutyl@umich.edu

Ethan Johnston

Department of Sociology
University of Michigan
ethjoh@umich.edu

August 15, 2023

PREPRINT

Related development 1: Three ways to evaluate dimensions

analogy-solving algorithms—*PairDir*, *3CosAdd* and *3CosMul* (Mikolov, Chen, et al. 2013; Levy, Goldberg, and Dagan 2015)—and used them as the basis for three candidate metrics of anchor set reliability. These metrics reflect three more basic characteristics of anchor sets: PairDir measures how parallel are the within-anchor-pair offsets across the different anchor pairs (*parallelism*), whereas 3CosAdd and 3CosMul reflect both how synonymous are the different terms at one axis endpoint (*synonymy*) and how antonymous are the opposing terms within the same anchor pair (*antonymy*).

Related development 1: 3CosAdd

Mikolov, Yih, and Zweig (2013) thus proposed an algorithm later dubbed *3CosAdd*, which iterates through all other words known to the embedding to find one with the greatest cosine similarity to (4):

$$\text{solution}_{\text{3CosAdd}}(a_1 : z_1 :: a_2 : \underline{z_2}) = \underset{z_2}{\operatorname{argmax}}\, sim(\vec{z}_2, \vec{a}_2 + (\vec{z}_1 - \vec{a}_1)). \tag{5}$$

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Related development 1: PairDir

We will assume that S *perfectly defines* the semantic axis $\vec{X}$ only if the word vectors in each pair are identical aside from their difference along the target semantic axis, $\forall (a_i, z_i) \in S, [\![\vec{z_i} - \vec{a_i}]\!] = [\![\vec{X}]\!]$. Thus, for any two pairs $\{(a_i, z_i), (a_j, z_j)\} \subseteq S$, their normalized offset vectors would be identical, $[\![\vec{z_i} - \vec{a_i}]\!] = [\![\vec{X}]\!] = [\![\vec{z_j} - \vec{a_j}]\!]$. For any two given anchor pairs $(a_i, z_i), (a_j, z_j)$, the degree to which this offset parallelism holds empirically can be measured via cosine similarity. This yields a measure called PairDir:

$$PairDir(a_i : z_i :: a_j : z_j) = sim(\vec{z_i} - \vec{a_i}, \vec{z_j} - \vec{a_j}) = \frac{(\vec{z_i} - \vec{a_i}) \cdot (\vec{z_j} - \vec{a_j})}{\|\vec{z_i} - \vec{a_i}\| \|\vec{z_j} - \vec{a_j}\|} \tag{7}$$

which varies from 1 (perfectly parallel offsets) to 0 (perfectly orthogonal offsets) to -1 (perfectly opposite offsets). This measure is illustrated visually in Figure 1, where PairDir(woman:man::girl:boy)

Related development 1: 3cOSmul

quality. A more common response to this critique, however, has been to replace 3CosAdd with 3CosMul (Levy, Goldberg, and Dagan 2015), which is composed from the same three components as 3CosAdd but weighs them more equally by using multiplication and division in place of addition and subtraction. As an analogy-solving algorithm, it equals:

$$\text{solution}_{\text{3CosMul}}(a_1 : z_1 :: a_2 : \underline{z_2}) = \operatorname*{argmax}_{z_2} \frac{sim(z_2, a_2) * sim(z_2, z_1)}{sim(z_2, a_1)} \tag{12}$$

We transform 3CosMul into an anchor set-level metric analogous to eq. (10). This yields:

$$3CosMul(S) = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{sim(\vec{z}_j, \vec{a}_j) * sim(\vec{z}_j, \vec{z}_i)}{sim(\vec{z}_j, \vec{a}_i)} + \frac{sim(\vec{a}_i, \vec{z}_i) * sim(\vec{a}_i, \vec{a}_j)}{sim(\vec{a}_i, \vec{z}_j)} \right) \tag{13}$$

To provide further insight into what characteristics of anchor pairs may affect axis quality, we will also examine anchor set synonymy and antonymy directly:

$$synonymy(S) = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(sim(a_i, a_j) + sim(z_i, z_j) \right) \tag{14}$$

$$antonymy(S) = \frac{1}{n} \sum_{i=1}^{n} sim(a_i, z_i) \tag{15}$$

Related development 1: Correlation with Human Rate

Table 2: Characteristics of N=23 longer anchor sets. Columns contain the human-derived ratings of accuracy (Acc.), PairDir (PairD), 3CosAdd (CAdd), 3CosMul (CMul), synonymy (Syn.), and antonymy (Ant.) metrics calculated using the Google News and HathiTrust embeddings.

	Google News embedding						HathiTrust embedding					
Axis	Acc.	PairD	CAdd	CMul	Syn.	Ant.	Acc.	PairD	CAdd	CMul	Syn.	Ant.
feminine-masc.	0.653	0.165	0.524	0.851	0.216	0.618	0.744	0.160	0.597	0.877	0.247	0.685
soft-hard	0.544	0.066	0.232	0.682	0.217	0.307	0.653	0.143	0.331	0.744	0.269	0.387
unusual-usual	0.340	0.089	0.309	0.731	0.236	0.393	0.425	0.117	0.360	0.758	0.302	0.439
rounded-angular	0.369	0.050	0.211	0.664	0.188	0.286	0.537	0.086	0.297	0.720	0.238	0.377
foolish-wise	0.547	0.191	0.381	0.775	0.325	0.421	0.620	0.184	0.412	0.790	0.336	0.466
important-un...	0.422	0.132	0.263	0.703	0.225	0.298	0.492	0.111	0.329	0.741	0.270	0.404
fast-slow	0.450	0.119	0.281	0.713	0.245	0.340	0.548	0.122	0.320	0.736	0.265	0.385
kind-cruel	0.691	0.175	0.294	0.724	0.290	0.310	0.767	0.239	0.377	0.774	0.346	0.378
straight-curved	0.476	0.040	0.207	0.663	0.182	0.291	0.523	0.077	0.282	0.713	0.215	0.364
timely-untimely	0.429	0.102	0.292	0.722	0.219	0.369	0.318	0.100	0.345	0.750	0.243	0.429
tasteless-savory	0.593	0.161	0.326	0.743	0.305	0.375	0.621	0.134	0.328	0.742	0.312	0.390
excitable-calm	0.526	0.119	0.242	0.691	0.252	0.285	0.670	0.157	0.266	0.704	0.290	0.282
passive-active	0.514	0.068	0.214	0.671	0.165	0.285	0.610	0.107	0.277	0.710	0.240	0.336
bad-good	0.548	0.202	0.352	0.759	0.312	0.376	0.652	0.172	0.368	0.767	0.296	0.414
strong-weak	0.425	0.092	0.232	0.682	0.212	0.287	0.499	0.126	0.286	0.717	0.250	0.336
true-false	0.515	0.143	0.305	0.730	0.263	0.355	0.623	0.134	0.387	0.773	0.328	0.465
successful-un...	0.670	0.149	0.338	0.751	0.213	0.392	0.660	0.121	0.374	0.765	0.246	0.457
old-new	0.363	0.060	0.232	0.681	0.154	0.308	0.395	0.066	0.305	0.726	0.235	0.402
ugly-beautiful	0.676	0.151	0.281	0.713	0.308	0.314	0.750	0.183	0.338	0.748	0.339	0.368
colorful-colorless	0.568	0.066	0.212	0.668	0.271	0.281	0.585	0.128	0.270	0.705	0.314	0.311
E	0.757	0.179	0.387	0.779	0.255	0.441	0.742	0.195	0.435	0.805	0.292	0.488
P	0.547	0.070	0.303	0.726	0.181	0.401	0.526	0.098	0.333	0.744	0.210	0.418
A	0.463	0.062	0.296	0.721	0.192	0.395	0.555	0.086	0.326	0.739	0.212	0.415
Average	0.525	0.115	0.292	0.719	0.236	0.353	0.588	0.132	0.345	0.750	0.274	0.409
cor(X, *Accuracy*)		0.623	0.508	0.522	0.475	0.376		0.775	0.425	0.447	0.483	0.197

Related development 2: Dimension and Topic Modeling

Integrating topic modeling and word embedding to characterize violent deaths

Alina Arseniev-Koehler[a,b,1], Susan D. Cochran[b,c,d], Vickie M. Mays[b,e,f], Kai-Wei Chang[b,g], and Jacob G. Foster[a,b,1]

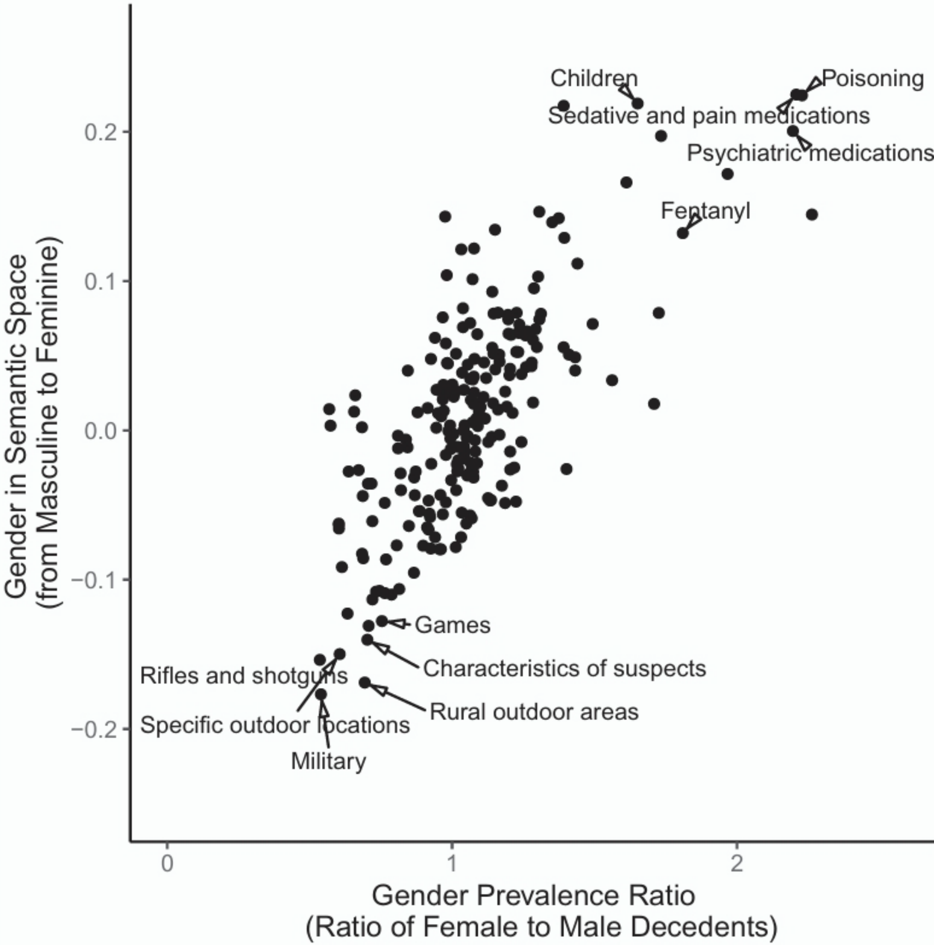
[a]Department of Sociology, University of California, Los Angeles, CA 90095; [b]Bridging Research Innovation, Training and Education for Science, Research & Policy Center, University of California, Los Angeles, CA 90095; [c]Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA 90095; [d]Department of Statistics, University of California, Los Angeles, CA 90095; [e]Department of Psychology, University of California, Los Angeles, CA 90095; [f]Department of Health Policy and Management, Fielding School of Public Health, University of California, Los Angeles, CA 90095; and [g]Department of Computer Science, University of California, Los Angeles, CA 90095

Edited by Sanjeev Arora, Computer Science Department, Princeton University, Princeton, NJ; received June 26, 2021; accepted January 14, 2022

Table 1. Sample topics within narratives of violent death

Topic label	Seven most representative terms
Physical aggression	Tackled, lunged_toward, began_attacking, advanced_toward, attacked, slapped, intervened
Causal language	Sparked, preceded, triggered, precipitated, led, prompted, culminated
Preparation for death	Disposal, deeds, prepaid_funeral, burial, worldly, miscellaneous, pawning
Cleanliness	Unkempt, messy, disorganized, cluttered, dirty, tidy, filthy
Everything seemed fine	Fell_asleep, everything_seemed_fine, seemed_fine, wakes_up, ran_errands, ate_breakfast, watched_television
Suspicion and paranoia	Conspiring_against, plotting_against, restraining_order_filed_against, belittled, please_forgive, making_fun, reminded
Reclusive behavior and chronic illness	Recluse, heavy_drinker, very_ill, chronic_alcoholic, bedridden, reclusive, recovering_alcoholic

Most representative terms are listed in order of highest to lowest cosine similarity to each topic's atom vector. Topic labels are manually assigned. As part of preprocessing the narratives, we transformed commonly occurring phrases into single terms (29).



Related development 2: Dimension and Topic Modeling

Table 2. Characteristics of violent deaths with two selected topics

Characteristic	Topic: Rifles and shotguns: AOR (95% CI)	Topic: Sedative and pain medications: AOR (95% CI)
Female decedent*	0.49 (0.48 to 0.51)	2.52 (2.47 to 2.58)
Decedent race/ethnicity†		
American Indian/Alaska Native, NH	1.31 (1.20 to 1.42)	0.46 (0.41 to 0.52)
Asian/Pacific Islander, NH	0.48 (0.43 to 0.54)	0.64 (0.59 to 0.70)
Black or African American, NH	0.88 (0.85 to 0.91)	0.54 (0.51 to 0.56)
Hispanic	0.59 (0.56 to 0.62)	0.63 (0.60 to 0.67)
Two or more races, NH	1.01 (0.92 to 1.10)	0.80 (0.73 to 0.88)
Unknown race, NH	0.70 (0.56 to 0.87)	0.70 (0.56 to 0.87)
Decedent age, y‡		
20 to 29	0.96 (0.91 to 1.00)	1.37 (1.29 to 1.46)
30 to 39	0.90 (0.86 to 0.95)	1.74 (1.64 to 1.85)
40 to 49	0.93 (0.88 to 0.98)	1.97 (1.86 to 2.10)
50 to 59	1.03 (0.98 to 1.08)	2.17 (2.04 to 2.30)
60+	1.40 (1.33 to 1.47)	1.68 (1.58 to 1.79)
Manner of death§		
Homicide	0.79 (0.77 to 0.82)	0.14 (0.13 to 0.15)
Legal intervention	1.09 (1.01 to 1.17)	0.22 (0.19 to 0.26)
Undetermined	0.06 (0.06 to 0.07)	2.01 (1.95 to 2.07)
Unintentional	3.16 (2.84 to 3.51)	0.13 (0.10 to 0.19)
Multiple decedents in incident¶	1.76 (1.68 to 1.84)	0.40 (0.37 to 0.43)
Word count#	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)

Related development 3: Embedding Regression

- Apply to situation of rare words (even appear once is enough)

Embedding Regression: Models for Context-Specific Description and Inference

PEDRO L. RODRIGUEZ *New York University, United States*
ARTHUR SPIRLING *New York University, United States*
BRANDON M. STEWART *Princeton University, United States*

Social scientists commonly seek to make statements about how word use varies over circumstances—including time, partisan identity, or some other document-level covariate. For example, researchers might wish to know how Republicans and Democrats diverge in their understanding of the term "immigration." Building on the success of pretrained language models, we introduce the à la carte on text (conText) embedding regression model for this purpose. This fast and simple method produces valid vector representations of how words are used—and thus what words "mean"—in different contexts. We show that it outperforms slower, more complicated alternatives and works well even with very few documents. The model also allows for hypothesis testing and statements about statistical significance. We demonstrate that it can be used for a broad range of important tasks, including understanding US polarization, historical legislative development, and sentiment detection. We provide open-source software for fitting the model.

Related development 3: Embedding Regression

- Apply to situation of rare words (even appear once is enough)

A La Carte Embedding:
Cheap but Effective Induction of Semantic Feature Vectors

Mikhail Khodak*, Nikunj Saunshi*
Princeton University
{mkhodak,nsaunshi}@princeton.edu

Yingyu Liang
University of Wisconsin-Madison
yliang@cs.wisc.edu

Tengyu Ma
Facebook AI Research
tengyuma@stanford.edu

Brandon Stewart, Sanjeev Arora
Princeton University
{bms4,arora}@princeton.edu

Related development 3: Embedding Regression

- Weight vectors based on sentences/closest 6-10 words/…

1. *The debate lasted hours, but finally we* [voted on the $\boxed{\text{bill}}$ and it passed] *with a large majority.*
2. *At the restaurant we ran up* [a huge wine $\boxed{\text{bill}}$ to be paid] *by our host.*

Related development 3: Embedding Regression

- Weight vectors based on sentences/closest 6-10 words/…

$$\underbrace{\begin{bmatrix}-1.22\\1.33\\0.53\end{bmatrix}}_{\text{voted}}\underbrace{\begin{bmatrix}1.83\\0.56\\-0.81\end{bmatrix}}_{\text{on}}\underbrace{\begin{bmatrix}-0.06\\-0.73\\0.82\end{bmatrix}}_{\text{the}}\ \text{bill}\ \underbrace{\begin{bmatrix}1.81\\1.86\\1.57\end{bmatrix}}_{\text{and}}\underbrace{\begin{bmatrix}-1.50\\-1.65\\0.48\end{bmatrix}}_{\text{it}}\underbrace{\begin{bmatrix}-0.12\\1.63\\-0.17\end{bmatrix}}_{\text{passed}}.$$

Obtaining $\mathbf{u}_w$ for "bill: simply requires averaging these vectors and thus

$$\mathbf{u}_{\text{bill}_1} = \begin{bmatrix}0.12\\0.50\\0.40\end{bmatrix},$$

Related development 3: Embedding Regression

- Weight vectors based on sentences/closest 6-10 words/…

$$\widehat{\mathbf{A}} = \begin{bmatrix} 0.81 & 3.96 & 2.86 \\ 2.02 & 4.81 & 1.93 \\ 3.14 & 3.81 & 1.13 \end{bmatrix}.$$

Taking inner products, we have

$$\widehat{\mathbf{v}}_{\text{bill}_1} = \widehat{\mathbf{A}} \cdot \mathbf{u}_{\text{bill}_1} = \begin{bmatrix} 3.22 \\ 3.42 \\ 2.73 \end{bmatrix} \text{ and } \widehat{\mathbf{v}}_{\text{bill}_2} = \widehat{\mathbf{A}} \cdot \mathbf{u}_{\text{bill}_2} = \begin{bmatrix} -1.91 \\ -1.58 \\ -0.62 \end{bmatrix}.$$

Related development 3: Embedding Regression

- Weight vectors based on sentences/closest 6-10 words/…

Down weight
Common words

$$\widehat{\mathbf{A}} = \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{w=1}^{W} \alpha(n_w) \|\mathbf{v}_w - \mathbf{A}\mathbf{u}_w\|_2^2. \qquad (1)$$

$$\widehat{\mathbf{A}} = \begin{bmatrix} 0.81 & 3.96 & 2.86 \\ 2.02 & 4.81 & 1.93 \\ 3.14 & 3.81 & 1.13 \end{bmatrix}.$$

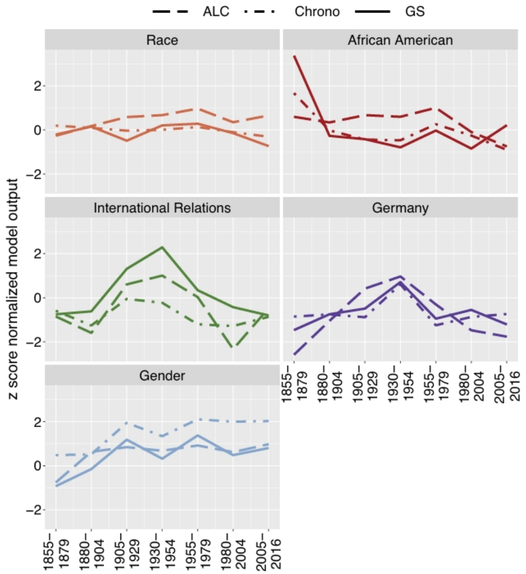
Taking inner products, we have

$$\widehat{\mathbf{v}}_{\text{bill}_1} = \widehat{\mathbf{A}} \cdot \mathbf{u}_{\text{bill}_1} = \begin{bmatrix} 3.22 \\ 3.42 \\ 2.73 \end{bmatrix} \text{ and } \widehat{\mathbf{v}}_{\text{bill}_2} = \widehat{\mathbf{A}} \cdot \mathbf{u}_{\text{bill}_2} = \begin{bmatrix} -1.91 \\ -1.58 \\ -0.62 \end{bmatrix}.$$

Related development 3: Embedding Regression

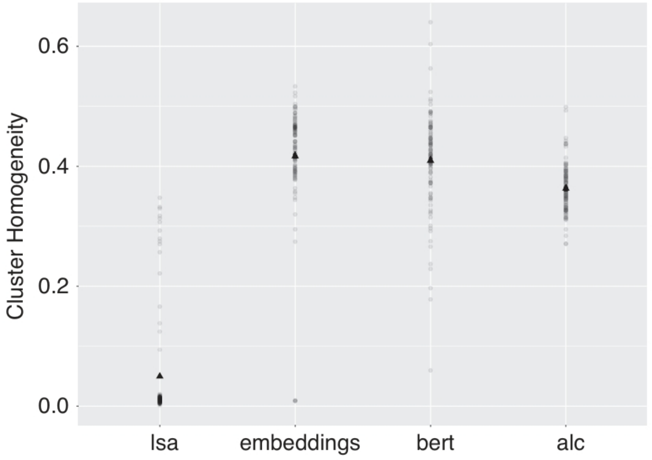
- Compared with other computational expensive methods

FIGURE 1. Replication of Figure 3 in Rodman (2020) Adding ALC Results



Note: ALC = ALC model, CHR = chronological model, and GS = gold standard.

FIGURE 3. Cluster Homogeneity



Note: Cluster homogeneity (in terms of `Trump` vs. `trump`) of *k*-means with two clusters of individual term instances embedded using different methods.

My work: Racial Triangulation

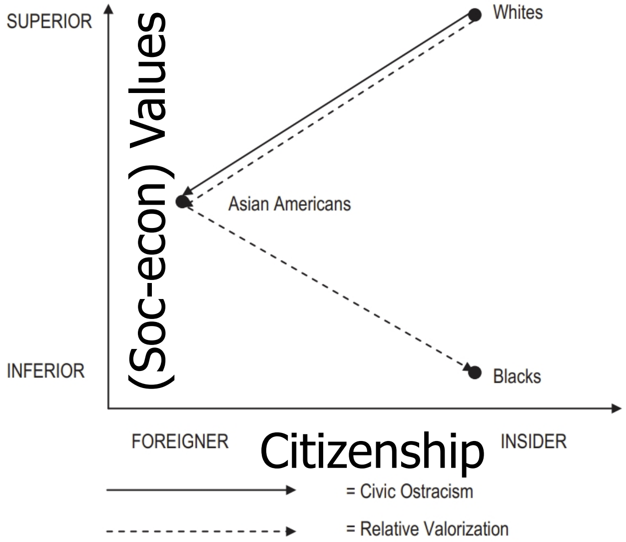
- **Post Civil Rights Movement**
  - From **explicit** (e.g., formal segregation) to **implicit racism** (e.g., stereotypes; claimed "colorblindness")

- **Development of Racial Theories**
  - Go beyond a static and consistent single dimension (e.g., color-line, class, nation-based theories) (Du Bois 1899; Marx 1972; Omi and Winant 2014)
  - Increasingly aware of the **dynamic nature of race** in **racialized unequal systems** (Doane and Bonilla-Silva 2003; Omi and Winant 2014)

Racial Triangulation Theory (RTT)

- Kim (1999:107) categorizes three racial groups - **Blacks, Whites, and Asian Americans** - into two dimensions:
  - **Relative Valorization** The dominant group (Whites) valorizes one subordinate group (Asian Americans) relative higher to the other (Blacks);
  - **Civic Ostracism** The dominant group constructs Asian Americans as unassimilable foreigners to ostracize them from politic and citizenship.
  - **To Control both Subordinate Groups**

Figure 1. The Field of Racial Positions in Racial Triangulation (reproduced from C. Kim 1999:108)



Source: Claire Kim, *Politics & Society* 27(1):105-38, copyright 1999 by Sage Publications; reprinted by permission of Sage Publications.

Word Vectors Social Application 1

- Use **word vectors** to represent **Social Entities' Positions** in constructed **Social Space** (e.g., black, white)
- Use **pairs of word vectors** to measure **Social Dimensions** (Boutyline and Johnston 2023: 7; Kozlowski, Taddy and Evans 2019)

A Gender Dimension — Female - Male →

Higher value
=>More female

Lower value
=>More male

$$\overrightarrow{\text{"}gender\text{"}} = \left[\!\left[\overrightarrow{\text{"}femininity\text{"}} - \overrightarrow{\text{"}masculinity\text{"}} \right]\!\right] =$$

$$= \left[\!\left[(\overrightarrow{she} - \overrightarrow{he}) + (\overrightarrow{her} - \overrightarrow{his}) + (\overrightarrow{girl} - \overrightarrow{boy}) + (\overrightarrow{daughter} - \overrightarrow{son}) + (\overrightarrow{mother} - \overrightarrow{father}) + (\overrightarrow{female} - \overrightarrow{male}) \right]\!\right].$$

Social Application 1

- Previous studies theory-driven **handpick** pairs of words and then **validate**

- This study uses both **Theory-Driven Handpicked** to test theory & **Data-Driven Algorithms** to develop theory.

(e.g., PCA: Principle Component Analysis & K-SVD: K Singular Value Decomposition for identify dimensions; nearest neighbor for identify important words)

Validations

- **Cross time and corpus Comparability**
  - Overtime Vector Alignment (Hamilton, Leskovec, & Jurafsky 2016)
  - Cosine similarity returns arbitrary meaning (Steck, Ekanadham, & Kallus 2024)

- **Validity of Dimension Accuracy**
  - Dimensions Construction (Boutyline and Johnston 2023: 10)

$$PairDir(a_i : z_i :: a_j : z_j) = sim(\vec{z_i} - \vec{a_i}, \vec{z_j} - \vec{a_j}) = \frac{(\vec{z_i} - \vec{a_i}) \cdot (\vec{z_j} - \vec{a_j})}{\|\vec{z_i} - \vec{a_i}\| \|\vec{z_j} - \vec{a_j}\|}$$

PairDir of Dimensions

- Handpick and then validate by PairDir (0 to1, the higher the better) as seed words, then iterate over different corpus…
- This presentation will largely **focus on two original RTT dimensions**

PairDir	Dimension	Poles	Seed words (Wiki)
0.61	citizenship	citizen	'citizen', 'citizens', 'naturalized', 'citizenship', 'resident', 'nationals', 'americans', 'naturalised', 'american'
		foreigner	'foreigner', 'foreigners', 'immigrant', 'stranger', 'strangers', 'outsiders', 'overstaying', 'expatriates', 'non-american'
0.61	diligent	diligent	'diligent', 'hardworking', 'industrious', 'studious', 'scrupulous', 'hard-working', 'dutiful', 'conscientious'
		lazy	'lazy', 'clumsy', 'careless', 'slob', 'unmotivated', 'indulgent', 'irresponsible', 'shiftless'
0.64	competence	competence	'competent', 'knowledgeable', 'supremely', 'talented', 'intelligent', 'disciplined', 'accountable', 'skillful'
		incompetence	'incompetent', 'inept', 'incapable', 'unprofessional', 'irresponsible', 'inexperienced', 'clumsy', 'foolish'
0.61	family	family	'love', 'child', 'affection', 'mother', 'committed', 'romantic', 'lover', 'affection'
		uncommitted	'naive', 'infidelity', 'adultery', 'unfaithful', 'betrayal', 'cheating', 'adulterous', 'divorce'
0.62	violence	violence	'violence', 'violent', 'conflict', 'hatred', 'extremism', 'brutality', 'disobedience', 'resistance'
		nonviolence	'nonviolence', 'nonviolent', 'non-violent', 'tolerance', 'moderation', 'compassion', 'peacefully', 'manner'
0.66	wealth	rich	'rich', 'wealthy', 'wealth', 'income', 'money', 'funds', 'millions', 'resources'
		poor	'poor', 'poorer', 'poorest', 'poverty', 'impoverished', 'needy', 'helping', 'economic'
0.61	patriotism	patriotism	'patriotism', 'loyalty', 'devotion', 'heroism', 'heroic', 'piety', 'nationalistic', 'selflessness'
		traitor	'traitor', 'traitors', 'betrayed', 'disloyal', 'betrayal', 'treachery', 'unpatriotic', 'enemies'
0.64	marriage	marriage	'marriages', 'relationships', 'suitable', 'attractive', 'appropriate', 'appealing', 'desirable', 'well-suited'
		unsuitable	'unsuitable', 'unattractive', 'inappropriate', 'unappealing', 'problematic', 'undesirable', 'risky', 'impractical'
0.63	neighbor	neighbor	'neighbor', 'neighbors', 'friend', 'friends', 'friendly', 'nearby', 'local', 'residents'
		stranger	'strangers', 'encounters', 'encounter', 'someone', 'hostile', 'visitors', 'tourist', 'tourists'
0.63	warm	warm	'personable', 'affable', 'likeable', 'amiable', 'considerate', 'cheerful', 'affable', 'easygoing'
		cold	'aloof', 'haughty', 'taciturn', 'condescending', 'reticent', 'timid', 'loner', 'polite'

Social Application 2

- Word’s meaning is context-dependent
- Use **variance** of word vectors **of the same word** across different social contexts (e.g., who and when) to identify **different constructions** (Khodak et al., 2018; Rodriguez, Spirling & Stewart 2023)

Social Application 2

- Word's meaning is context-dependent
- Use **variance** of word vectors **of the same word** across different social contexts (e.g., who and when) to identify **different constructions** (Khodak et al., 2018; Rodriguez, Spirling & Stewart 2023)

- E.g., for Democrats & Republican (0/1) congressmen's agendas and frames around "blacks", "whites", "asians", understand as a regression:

$$\mathbf{Y} = \beta_0 + \beta_1 \text{Republican}$$

- Y is "blacks", "whites", "asians" vectors;

Data

- **Pre-trained Word Embedding**

Time	Text	Method	Dimensions	Source
1910-1999	COHA	SVD	300	Hamilton, Leskovec, & Jurafsky (2016)
2014	Wikipedia	Glove	300	Pennington, Socher, & Manning (2014)
2010s	Twitter	Glove	200	Pennington, Socher, & Manning (2014)
2010s	Google News	word2vec	300	Mikolov et al. (2013)

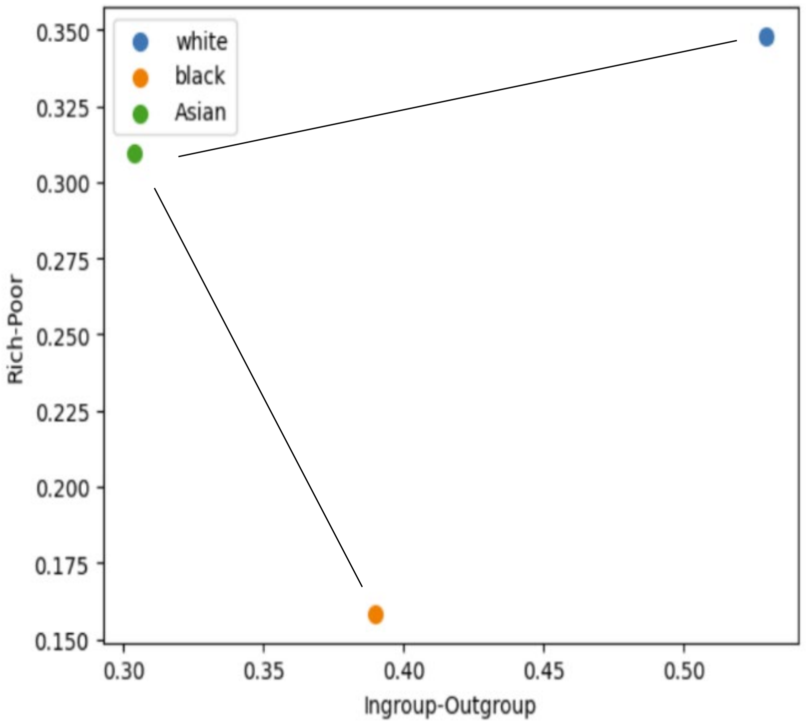
*COHA: Corpus of Historical American English, covers American textbook, magazine, newspaper

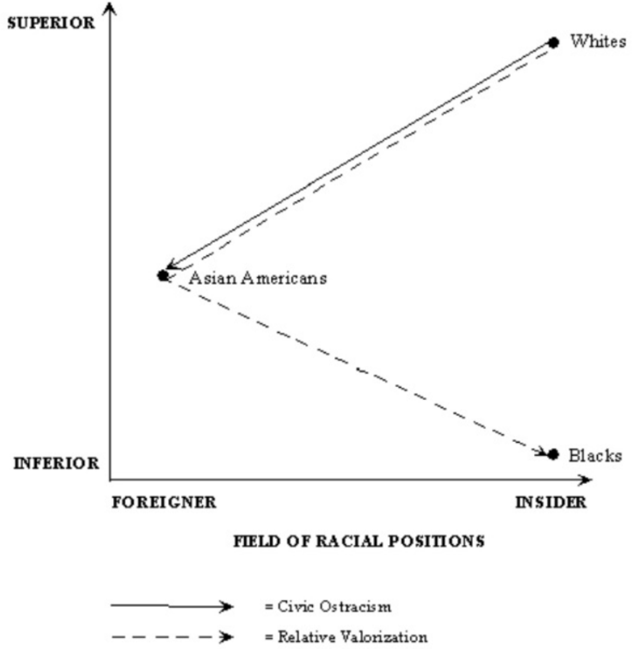
- **Fine-tune**
  - Congressional Record (1900-1999) (Gentzkow, Shapiro, Taddy 2018)

Result 1: Test Theory
Evidence from Contemporary Corpus

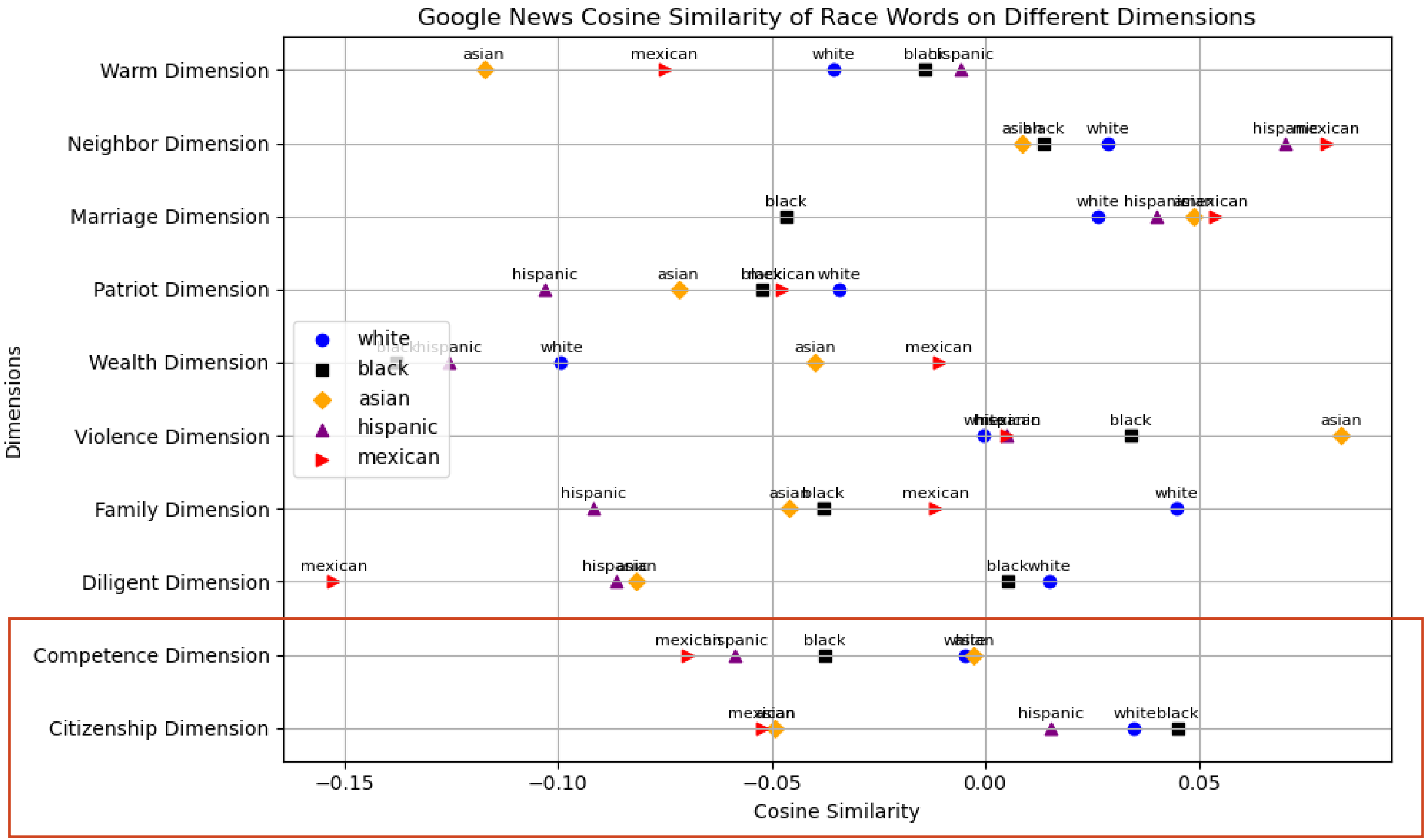
Original RTT & Naïve Dimension (Google News 2010s)

- **Black Partial Citizenship** (Davies 2022)

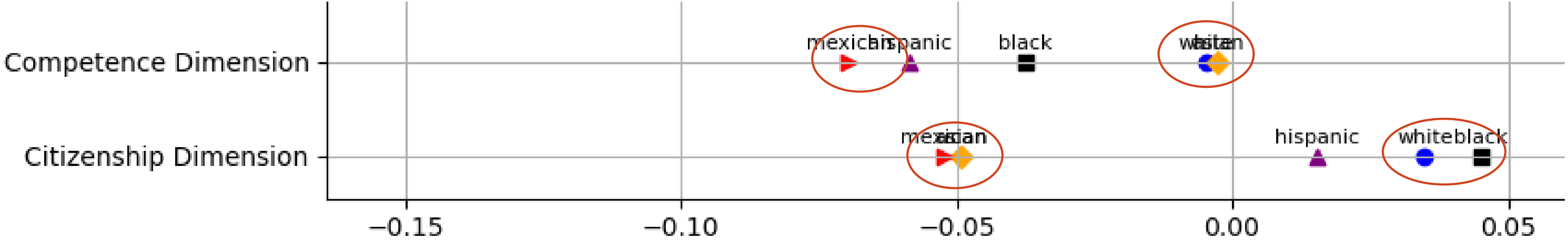




Extended RTT & Valid Dimension (Google News 2010s)

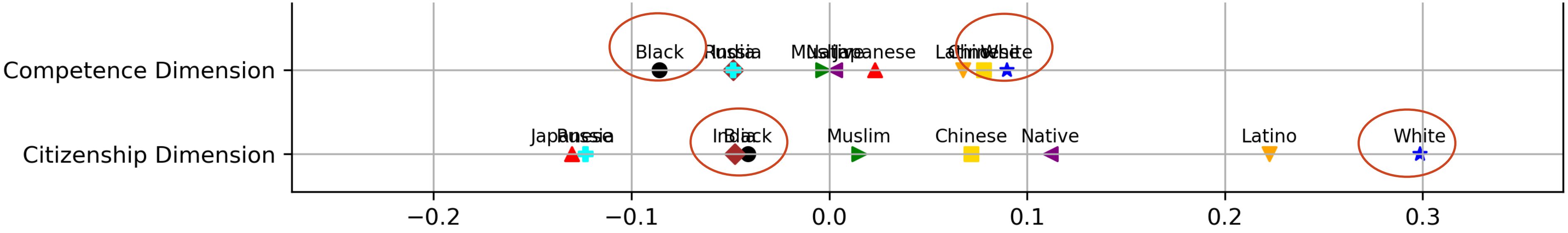


Extended RTT & Valid Dimension (Google News 2010s)



- **Support RTT** Black & white high citizenship, Asian and white high competence
- **Mexican is more vulnerable**

Extended RTT & Valid Dimension (Twitter 2010s)



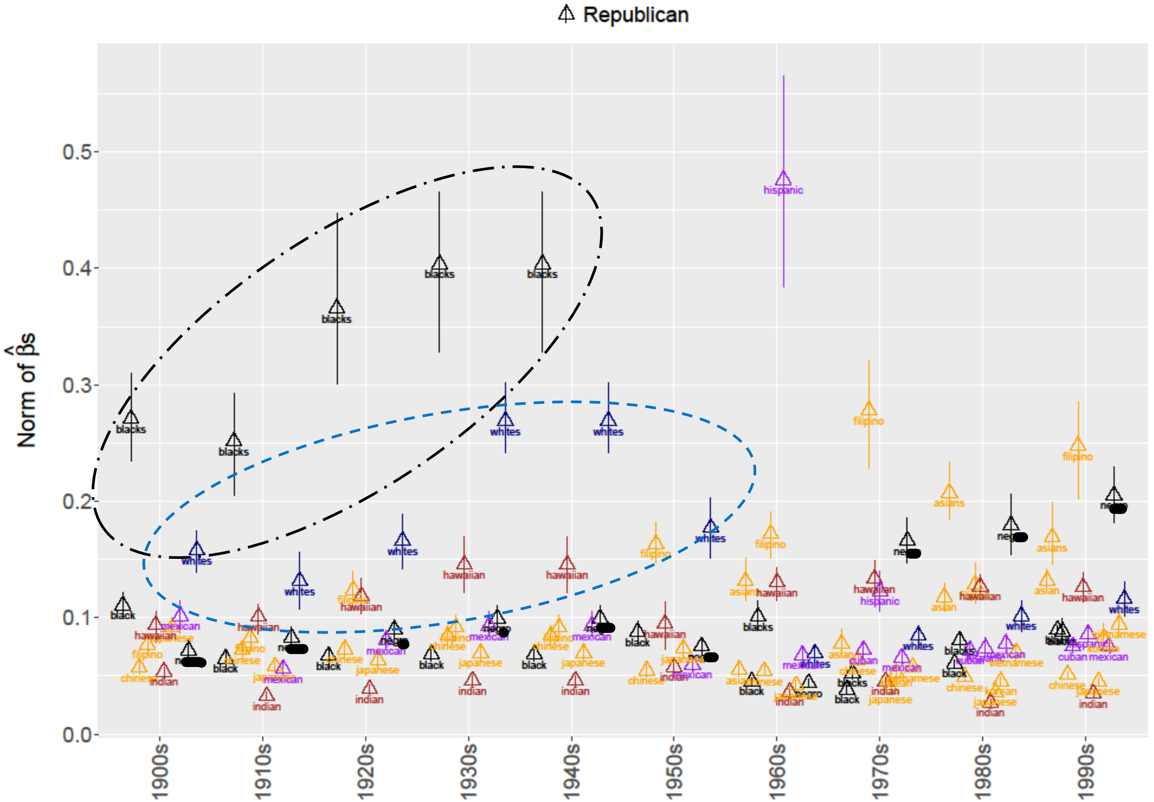
- **Stronger Discrimination Against Black**
- **Stronger White Supremacy**

Result 2: Constructions of RTP From Black-White to Triangulation

How did Democrats & Republicans discuss races & ethnicities?

- **Higher values** indicate **more cross-party variance**
- **Lower values** indicate **smaller variance.**

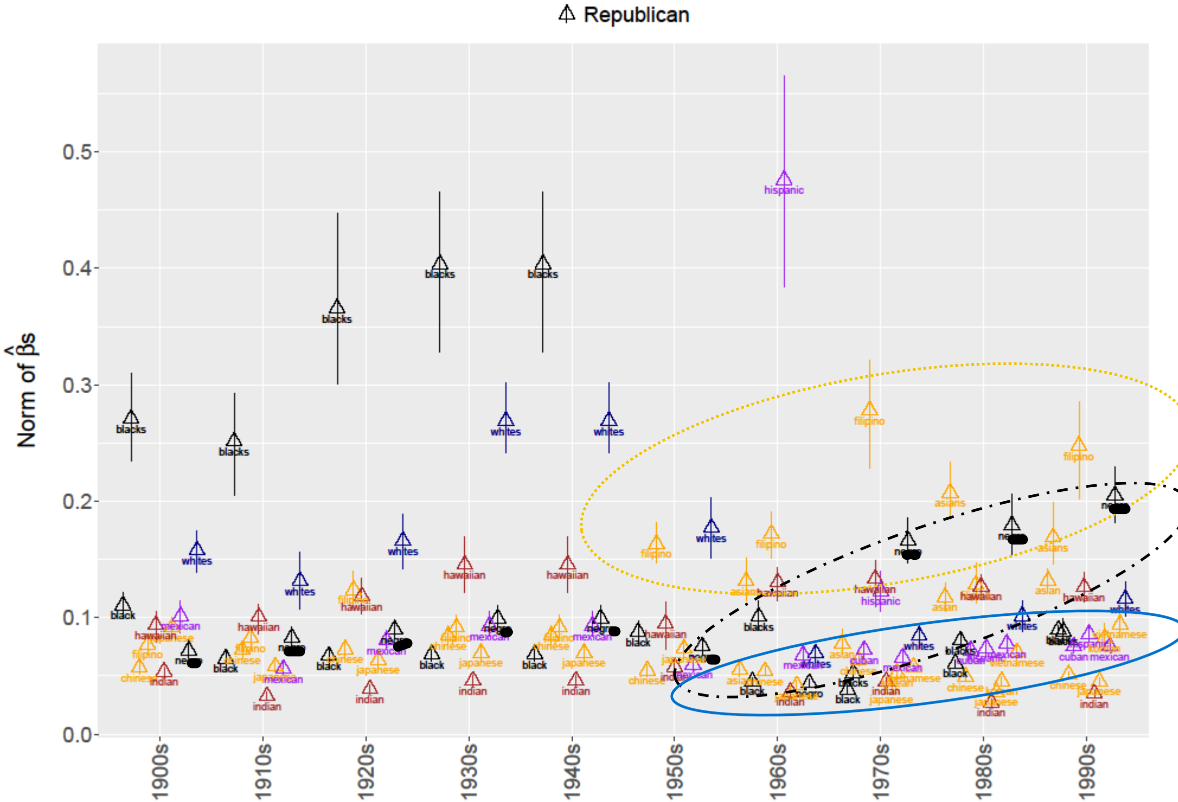
- A shift from **Black-White** to **Asian-Black-White and more.**



How did Democrats & Republicans discuss races & ethnicities?

- **Higher values** indicate **more cross-party variance**
- **Lower values** indicate **smaller variance.**

- A shift from **Black-White** to **Asian-Black-White and more.**



Democrats (D) vs. Republicans (R) in 1930s

Formal Explicit Civic Ostracism

- **Nearest words of "blacks"**
- Agendas:

D: "intermarriage"

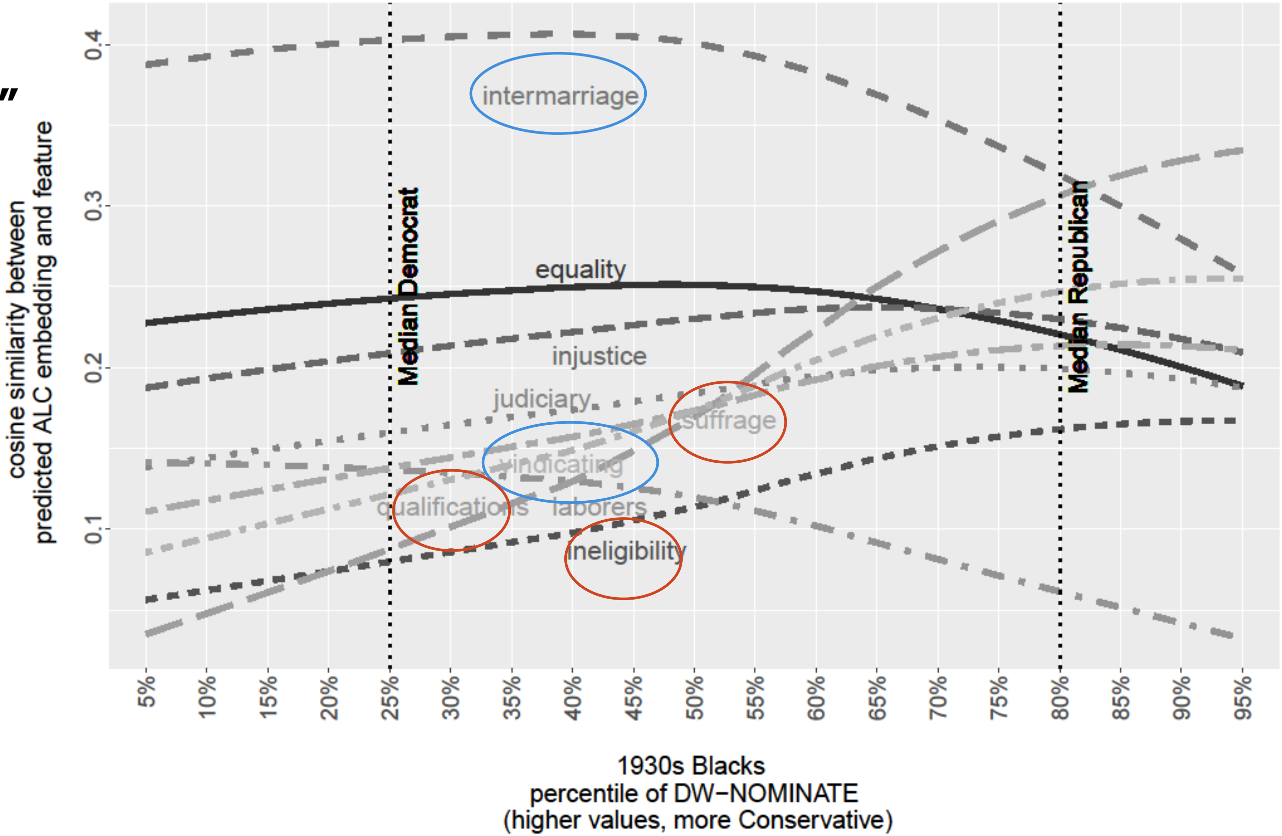
R: "suffrage"

- Frames:

Shared: "injustice", "equality"

D: "vindicating"

R: "qualifications", "ineligibility"



Democrats (D) vs. Republicans (R) in 1990s

Implicit Relative Valorization

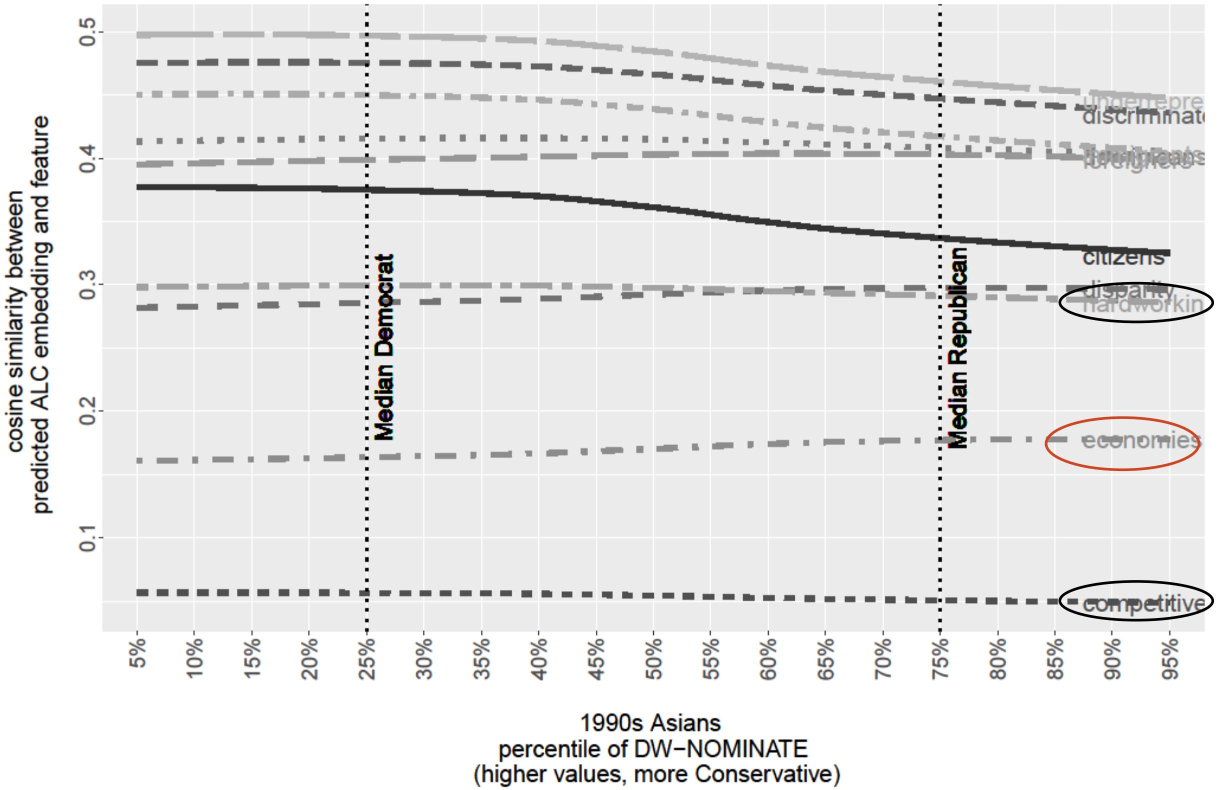
- **Nearest Words of “asians”**
- **Agendas:**

D: “citizens”, “discrimination”;

R: “economies”

Frames:

Shared: “underrepresented”, “**hardworking**”, “**competitive**”



Democrats (D) vs. Republicans (R) in 1990s
Implicit/Explicit Civic Ostracism

- **Nearest Words of "filipino"**
- **Agendas:**

D: "veterans", "citizens"

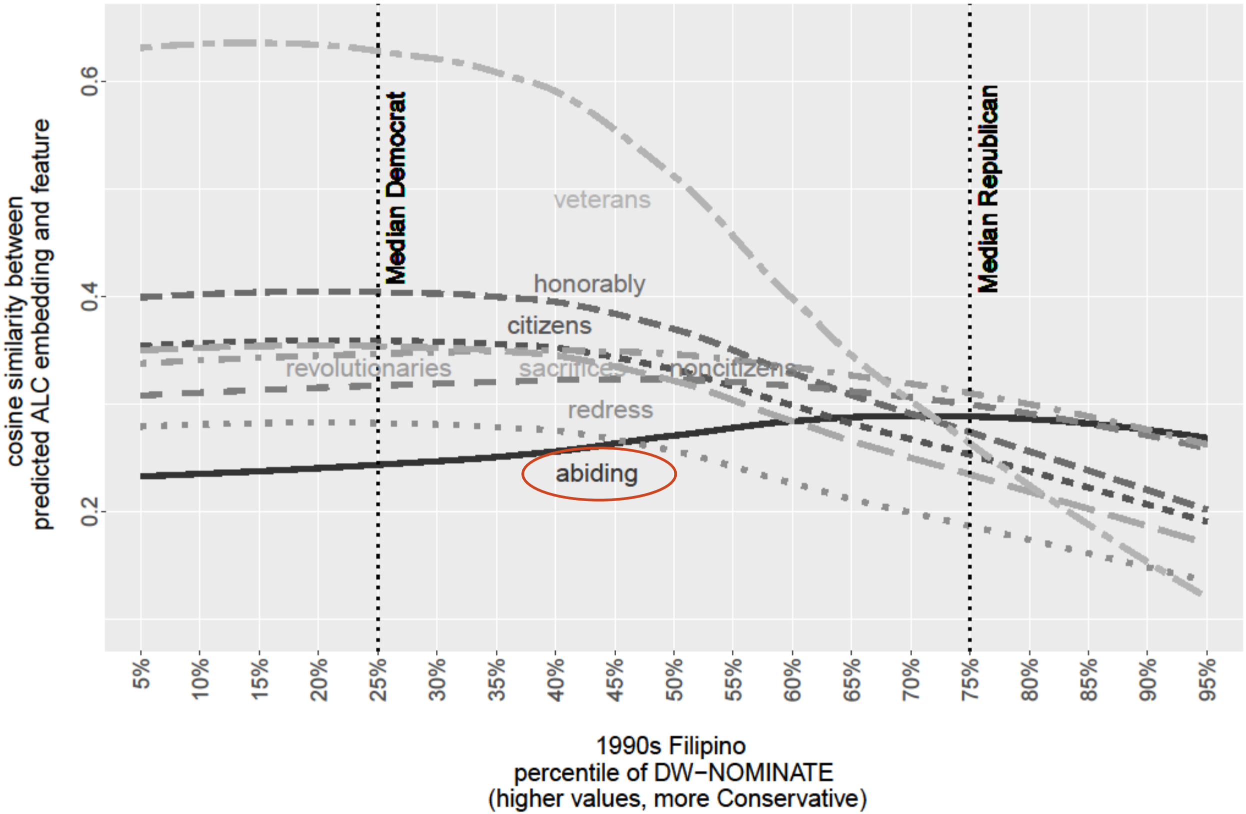
- **Frames:**

Shared: "sacrificed", "noncitizen"

D: "honorably", "redress"

R: "abiding"

"Model" Minority but also **"abiding" silent minority:** a toolkit for conservative politicians (Kim 1999)



Democrats (D) vs. Republicans (R) in 1990s

Implicit Civic Ostracism

- **Nearest Words of "n*"**
- **Agendas:**

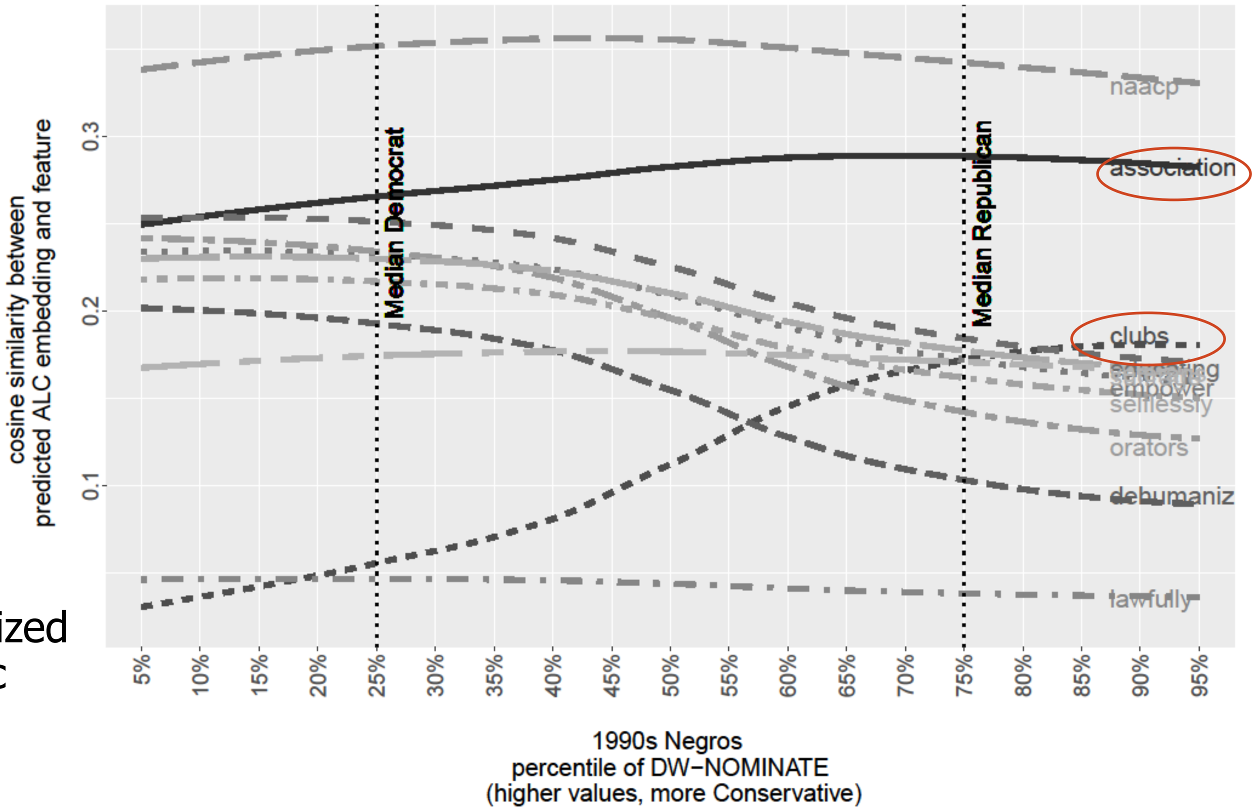
D: "empower", "educating"

- **Frames:**

Shared: "lawfully", "naacp"

D: "dehumanized", "orators"

R: "clubs", "association"



Citizen but also politically organized as "**loud"** minority: implicit civic ostracism – restrict black's organized voices (Gillion 2020)

Democrats (D) and Republicans (R) Bipartisan

- **Bipartisan construction**
- Top 10 nearest words for **“mexican”**
- **Top 1 text** by weight (Rodriguez, Spirling & Stewart 2023)

	R_1980	D_1980	R_1970	D_1970	R_1960	D_1960	R_1950	D_1950
1	illegals	mexicans	yaquis	mexicans	farmworkers	unskilled	unskilled	unskilled
2	narcos	narcos	mexicans	yaquis	unskilled	farmworkers	illegals	illegals
3	traffickers	subsidization	illegals	chicanos	bracero	domestics	laborers	laborers
4	mexicans	illegals	undocumented	anglos	braceros	laborers	farmworkers	farmworkers
5	subsidization	reevaluate	shrimpers	ricans	domestics	mexicans	migrant	migrant
6	undocumented	shrimpers	farmworkers	shrimpers	laborers	migrant	labors	mexicans
7	smugglers	undocumented	chicanos	arizonans	mexicans	braceros	exportation	penalize
8	kingpins	braceros	anglos	undocumented	peons	bracero	mexicans	employers
9	shrimpers	personify	arizonans	illegals	migrant	labors	bracero	labors
10	apprehensions	soliders	lawfully	farmworkers	obligated	peons	pertain	shrimpers

Top Text	
1950s	illegal mexicans but this is to make the employment of labor legal we need them not only in texas arizona
1960s	mexican labor and that if we were to bring the laborers into the country without the dictates of aflcio the
1970s	our countrys need for labor and the need of many citizens to find jobs large numbers of mexicans cross our
1980s	effective law enforcement the drug smugglers call the tune and officials dance to it governors of two mexican states have

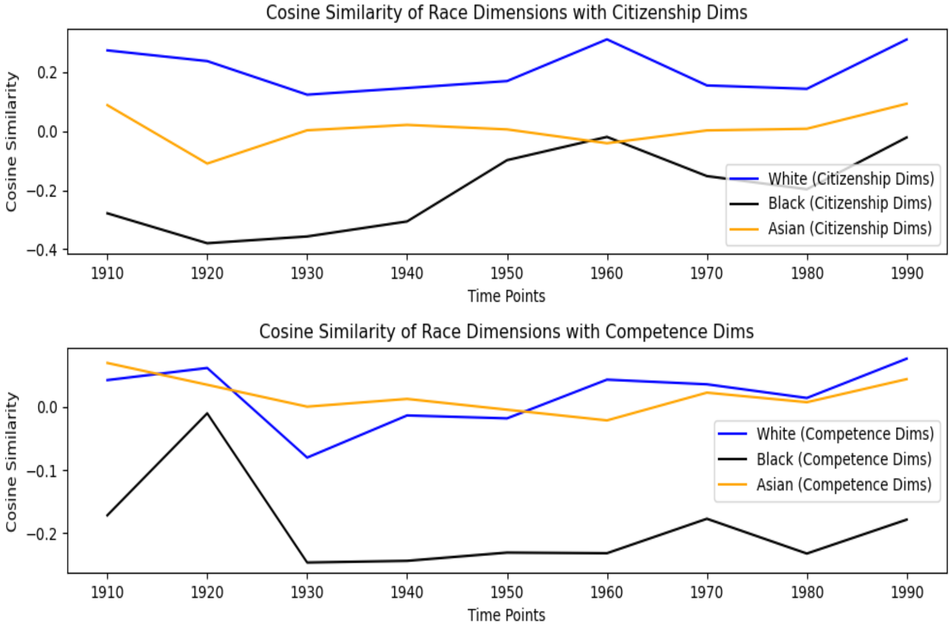
JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Results 3: RTP Beyond Congress

Social Corpus (COHA 1910s-90s: American textbook, magazine, newspaper)

- **Blacks never attain full citizenship**



- **Relative Valorization was always there.**

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

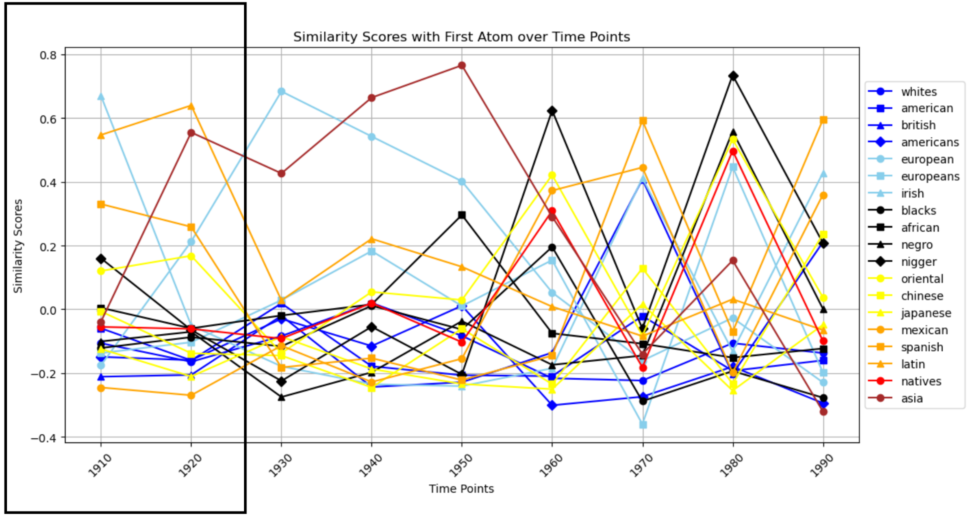
Data-Driven Social Background 1: Modernization (k-SVD of Races and Ethnicities in COHA 1910s to 1990s)

- **Top 1 frames that differentiate groups** are not consistent and change overtime
- **E.g., 1910s-1920s**
  - Higher positive values are associated with **classical/traditional themes** words such as:

'poetry', 'Shakespeare', 'Bible', 'philosophy', 'noble', 'songs', and 'books'

  - Negative values highly associated **industrialization/modern state**:

'railroad', 'naval', 'ships', 'government', 'officials', 'supplies', 'operations', 'voted', and 'arrest'.



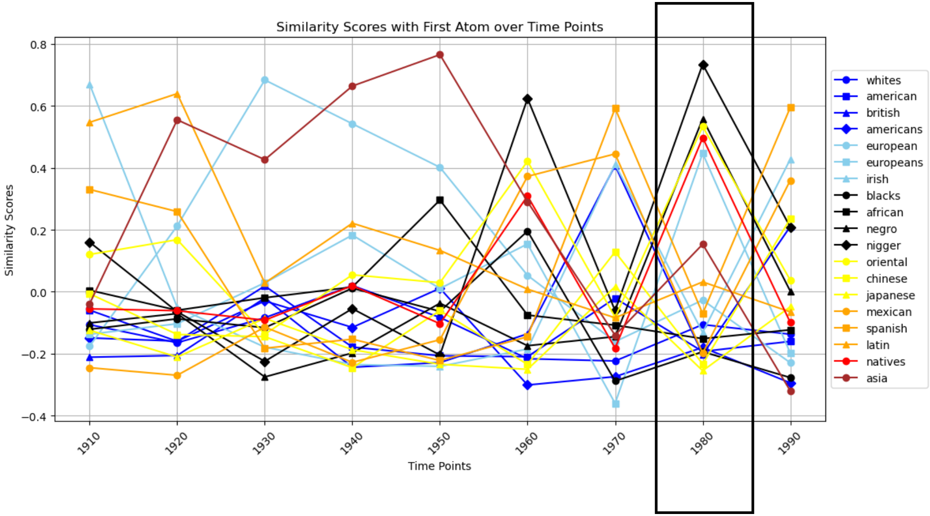
Data-Driven Social Background 2: Global Racialized Capitalism (COHA 1910s to 1990s)

- Top 1 frames that differentiate groups are not consistent and change overtime
- **E.g., 1980s**
  - Higher positive values (**global racialized capitalism**) are associated with words:

“n*” 'Japanese', 'spy', 'electronics', 'computer', 'Honda', 'yen', 'IBM’.

  - Negative values are highly relevant with words (**domestically left-behind population**) such as:

'rural', 'jungle', 'mountain', 'Republican', 'South', 'churches', 'communities', 'isolated', 'poverty', 'candidate'.



JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Key Takeaways

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Takeaways

- **Racial Triangulation Theory**
  - Blacks never attain full citizenship; Relative Valorization is supported

Takeaways

- **Racial Triangulation Theory**
  - Blacks never attain full citizenship; Relative Valorization is supported
- **Racial Triangulation Practice**
  - From "Black-White" to Triangulation emerged after the civil right movement as a political "frontlash" to sustain racialized unequal systems (Weaver 2007).

Takeaways

- **Racial Triangulation Theory**
  - Blacks never attain full citizenship; Relative Valorization is supported
- **Racial Triangulation Practice**
  - From “Black-White” to Triangulation emerged after the civil rights movement as a political “frontlash” to sustain racialized unequal systems (Weaver 2007).
- **Other Groups and Dimensions**
  - Hispanics were discriminated in bipartisan frames on “documented-undocumented”.
  - Constructing racial differences aligned with the great transformation: modernization to racialized capitalist globalization

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Example 2: Simulation

Associative Diffusion Background

- Agent-based Model: From Factors to Actors (Rational Actors v.s. Psychological/Cultural Actors)
- Is LLM-Based Agent Simulation the Future?

ANNUAL REVIEW OF SOCIOLOGY Volume 28, 2002

Review Article

From Factors to Actors: Computational Sociology and Agent-Based Modeling

Michael W. Macy[1], and Robert Willer[1]

View Affiliations

Vol. 28:143-166 (Volume publication date August 2002) | https://doi.org/10.1146/annurev.soc.28.110601.141117

Associative Diffusion

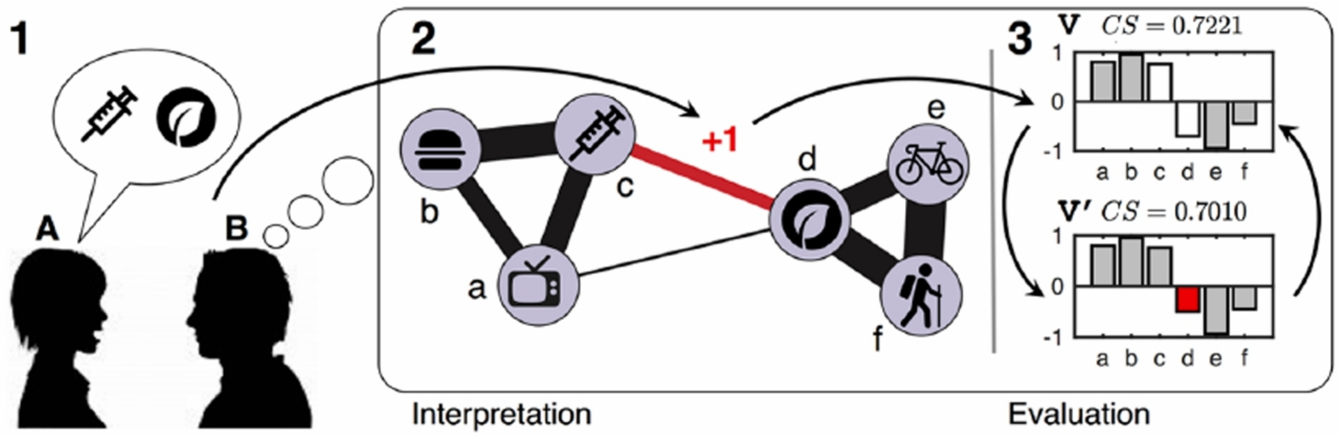


Figure 3. An Illustration of the Agent-Based Model Sequence
Note: (1) Agent *B* observes *A* express support for vaccinations and organic food (practices *c* and *d*); (2) *B* updates the corresponding element in his associative matrix, *R* (the edge connecting nodes *c* and *d* in the network representation of *R*); and (3) randomly updates his preference for organic food (practice *d*, resulting in preference vector *V′*), which is the weaker preference of the pair {c,d} in his preference vector *V*. Because constraint satisfaction is reduced from .7221 to .7010, this preference update is rejected, and *B*'s preference vector *V* remains unchanged.

Johns Hopkins Whiting School of Engineering

Associative Diffusion

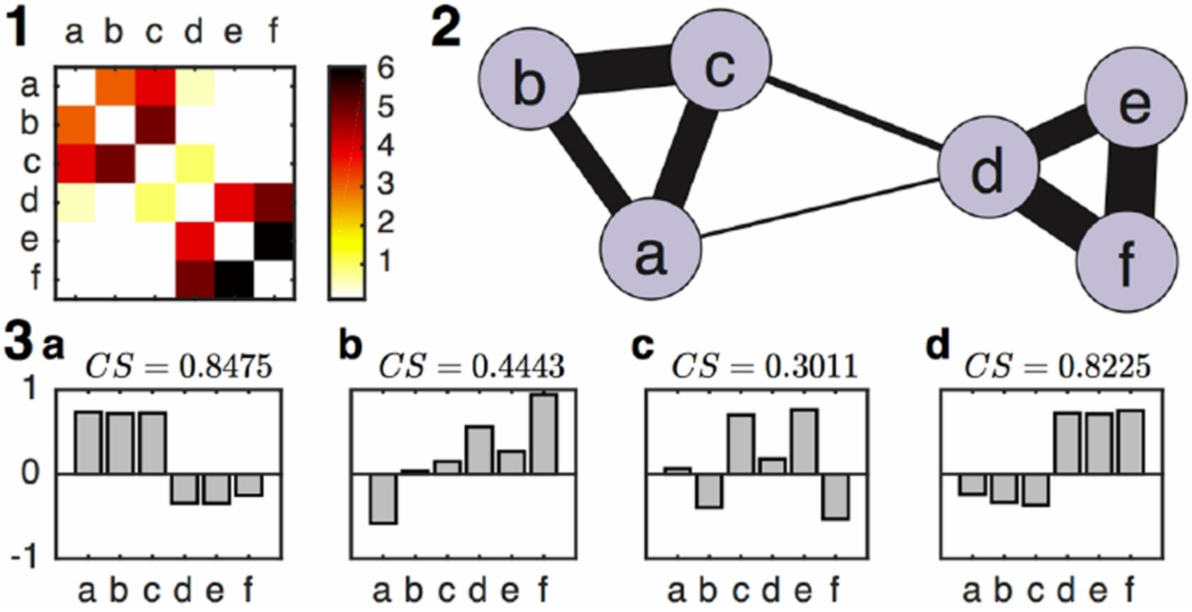


Figure 2. A Hypothetical Example of an Agent's Associative Matrix
Note: R represented as (1) a heat map and as (2) a network, as well as (3) an example of four preference vectors and their respective levels of constraint satisfaction, with respect to this associative matrix.

Associative Diffusion

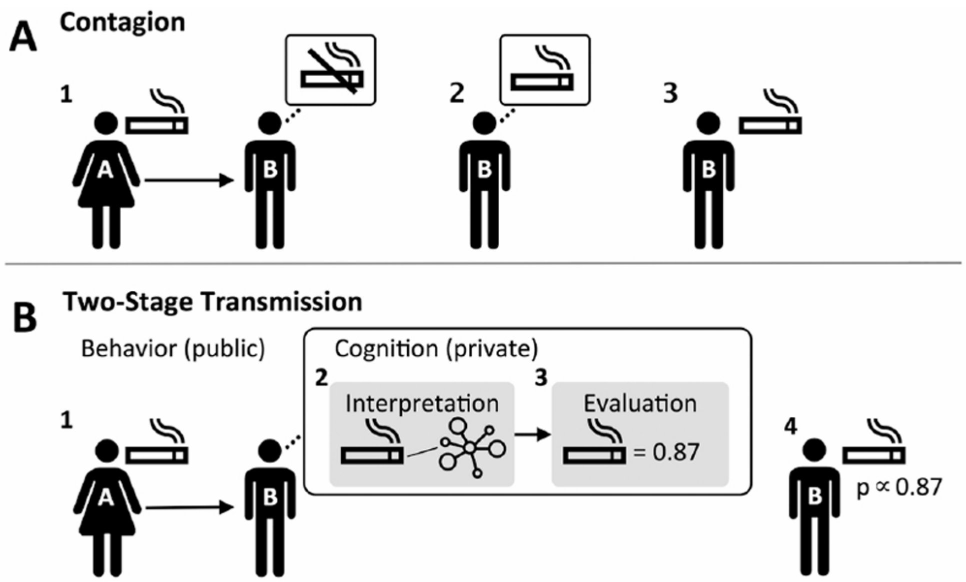


Figure 1. The Process of Cultural Transmission in the Contagion (A) and Two-Stage Transmission (B) Models

Note: In both illustrations, agent *B* is observing agent *A* smoking. Square callouts relate to *B*'s cognition. In (A), *B* changes his preference from anti-smoking to smoking, and consequently smokes. In (B), he updates his interpretation of smoking and his preference for smoking, and consequently smokes with an illustrative probability of .87.

Associative Diffusion

Table 1. Model Overview

Agent Initialization
Each agent holds two types of information: 1. associations: $R_{ij} = 1, \forall i, j \in K$ 2. preferences: $V_i \sim U(-1,1)$
Modeling Sequence
1. Select agents A and B at random 2. B observes A exhibiting practices i and j with probabilities $P(i)$ and $P(j)$ 3. B updates $R_{ij} = R_{ij} + 1$ 4. B selects preference k to update, where k is the weaker of v_i and v_j 5. B updates preferences, V', by setting $v'_k = v_k + \sim N(0,1)$ 6. If $CS(V',R) > CS(V,R)$, V' is retained, otherwise revert to V 7. Apply decay function $R_{ij} = \lambda R_{ij}$

Associative Diffusion: Findings

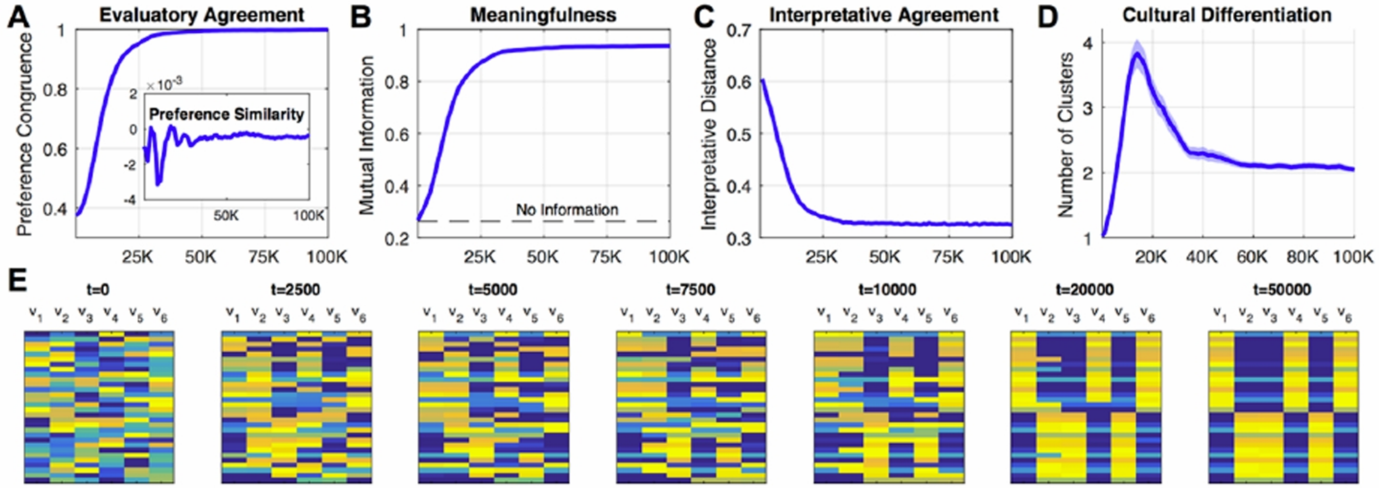


Figure 5. Multi-agent Models with 30 Agents
Note: (A) Mean preference congruence between agents (measured as absolute correlation between agents' preference vectors), preference similarity (measured as mean correlation between agents' preference vectors) is in the inset. (B) Mutual information between agents' behaviors. (C) Mean distance between all agents' associative matrices. (D) Number of agent clusters estimated by the gap statistic, based on agents' preferences (with shaded confidence intervals). (E) Snapshots of preference vectors for one simulation run (each heat map represents the preferences of 30 agents for six practices, ranging from strong negative in dark gray [blue in the online version] to strong positive in light gray [yellow online]).

Associative Diffusion: Alternative Theories

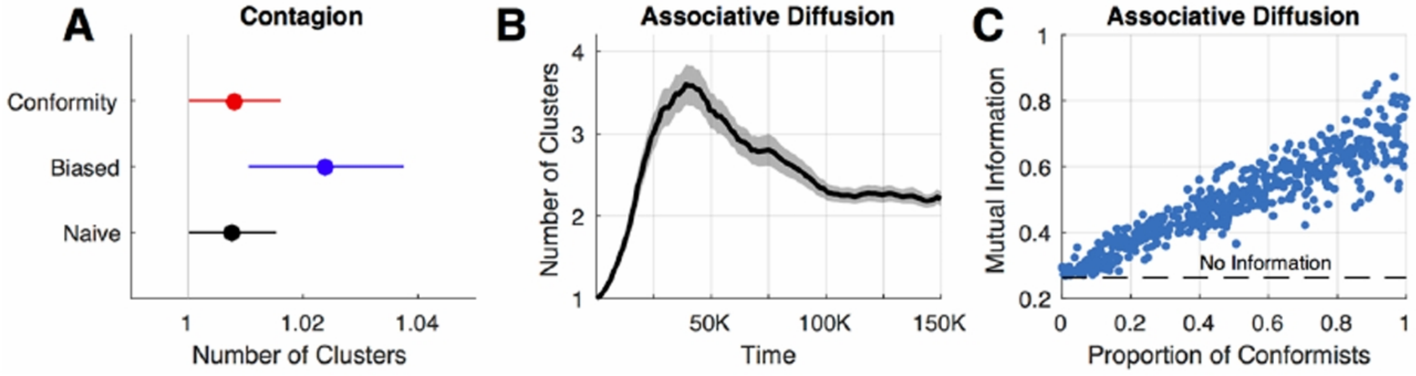


Figure 6. Alternative Contagion Models
Note: (A) Number of clusters at end for contagion models with different transmission mechanisms. (B) Number of clusters for associative diffusion model with conformity. (C) Mutual information between behaviors at end for associative diffusion model with conformity and with varying proportions of conformists.

Associative Diffusion: Robust Check

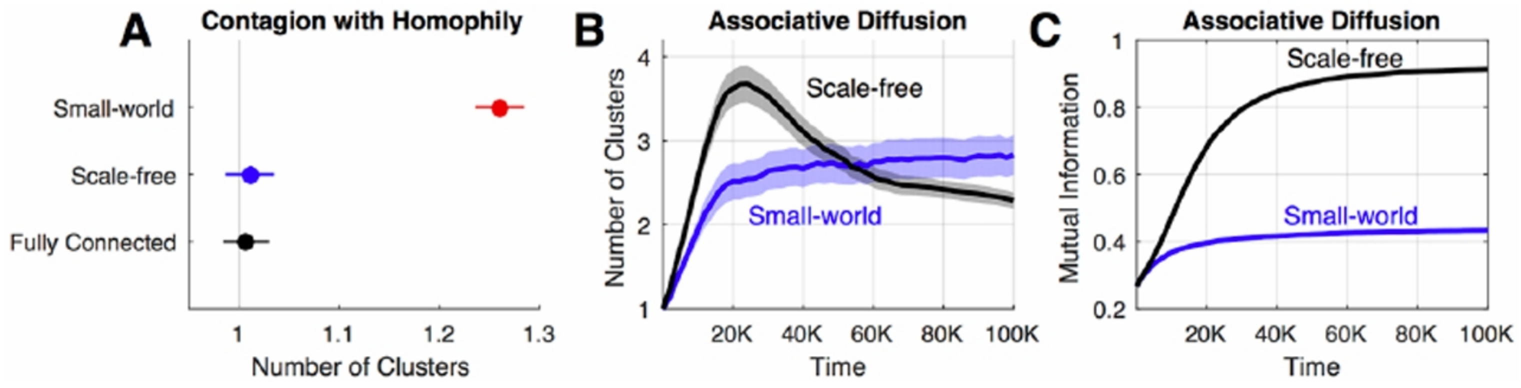
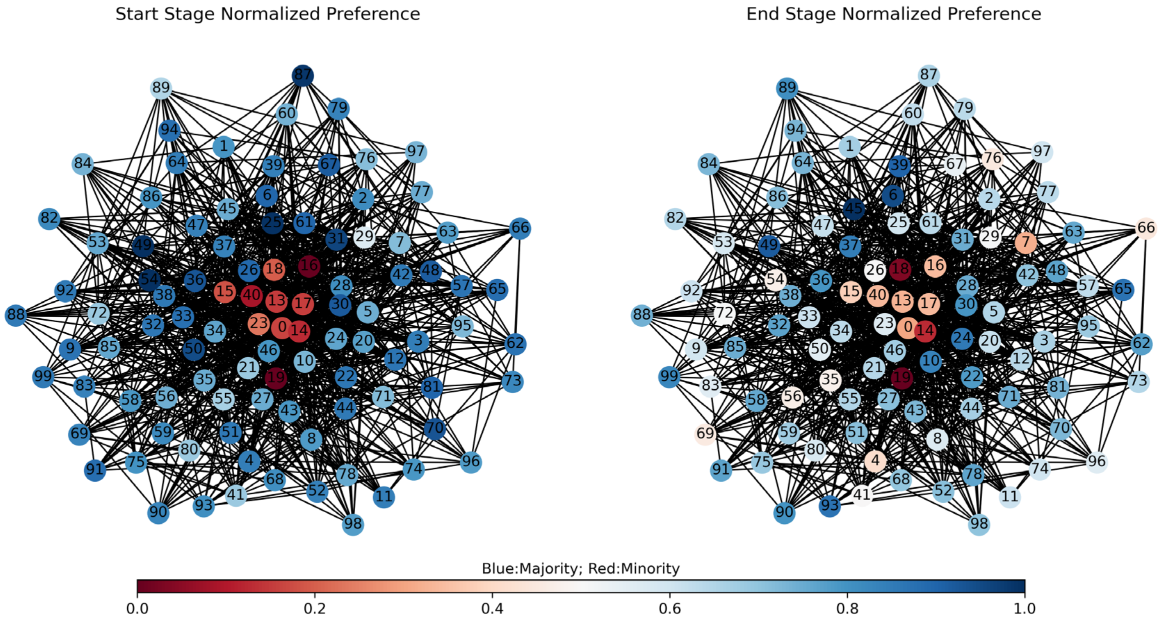
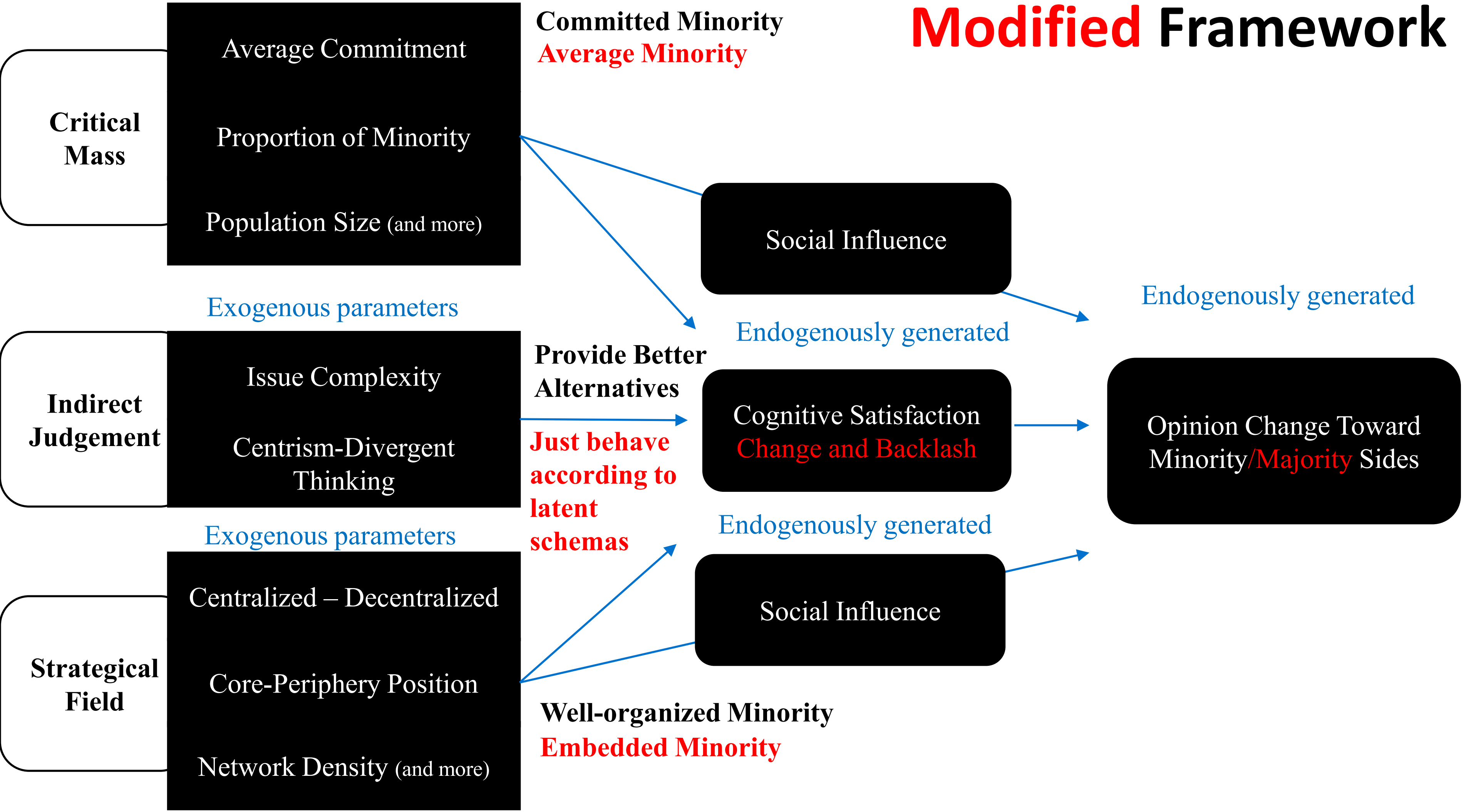


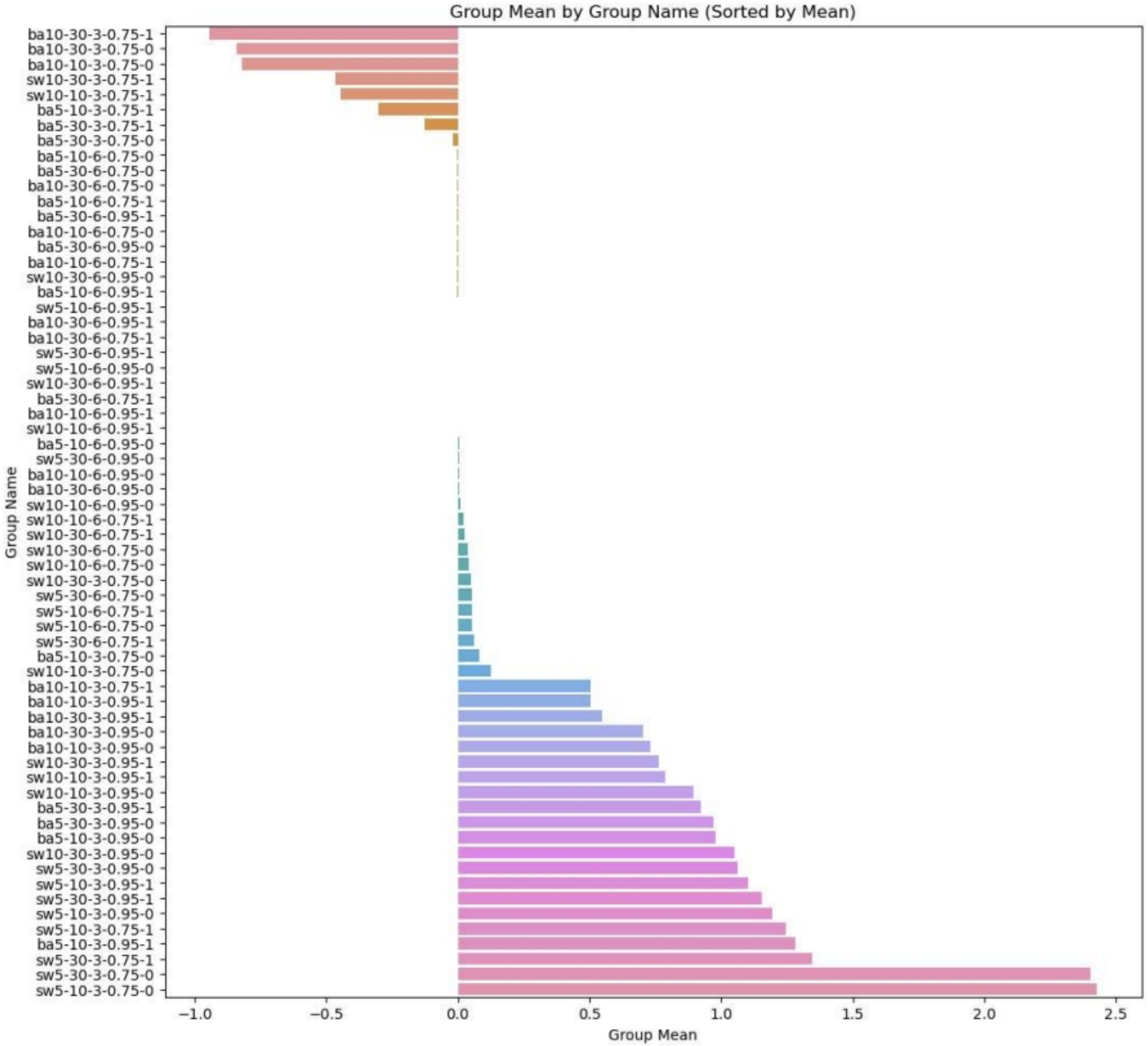
Figure 7. Different Network Topologies
Note: (A) Number of clusters at end for contagion models with homophily and different network topologies. (B) Number of clusters for associative diffusion model with scale-free or small-world networks. (C) Mutual information between behaviors for associative diffusion model with scale-free or small-world networks.

My Work: Minority Influence Majority




Modified Framework
Critical Mass
Average Commitment
Proportion of Minority
Population Size (and more)
Committed Minority
Average Minority
Exogenous parameters
Indirect Judgement
Issue Complexity
Centrism-Divergent Thinking
Exogenous parameters
Strategical Field
Centralized – Decentralized
Core-Periphery Position
Network Density (and more)
Provide Better Alternatives
Just behave according to latent schemas
Well-organized Minority
Embedded Minority
Social Influence
Endogenously generated
Cognitive Satisfaction
Change and Backlash
Endogenously generated
Social Influence
Endogenously generated
Opinion Change Toward Minority/Majority Sides

Results

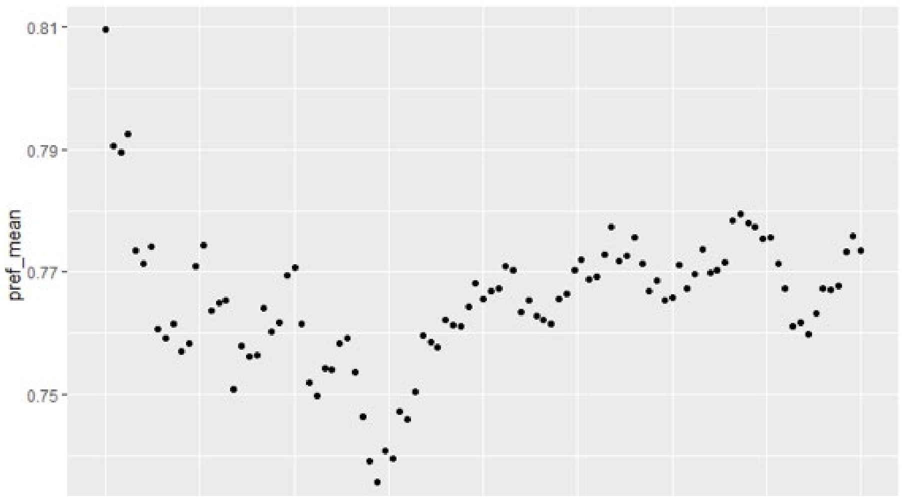

Group Mean by Group Name (Sorted by Mean)
ba10-30-3-0.75-1
ba10-30-3-0.75-0
ba10-10-3-0.75-0
sw10-30-3-0.75-1
sw10-10-3-0.75-1
ba5-10-3-0.75-1
ba5-30-3-0.75-1
ba5-30-3-0.75-0
ba5-10-6-0.75-0
ba5-30-6-0.75-0
ba10-30-6-0.75-0
ba5-10-6-0.75-1
ba5-30-6-0.95-1
ba10-10-6-0.75-0
ba5-30-6-0.95-0
ba10-10-6-0.75-1
sw10-30-6-0.95-0
ba5-10-6-0.95-1
sw5-10-6-0.95-1
ba10-30-6-0.95-1
ba10-30-6-0.75-1
sw5-30-6-0.95-1
sw5-10-6-0.95-0
sw10-30-6-0.95-1
ba5-30-6-0.75-1
ba10-10-6-0.95-1
sw10-10-6-0.95-1
ba5-10-6-0.95-0
sw5-30-6-0.95-0
ba10-10-6-0.95-0
ba10-30-6-0.95-0
sw10-10-6-0.95-0
sw10-10-6-0.75-1
sw10-30-6-0.75-1
sw10-30-6-0.75-0
sw10-10-6-0.75-0
sw10-30-3-0.75-0
sw5-30-6-0.75-0
sw5-10-6-0.75-1
sw5-10-6-0.75-0
sw5-30-6-0.75-1
ba5-10-3-0.75-0
sw10-10-3-0.75-0
ba10-10-3-0.75-1
ba10-10-3-0.95-1
ba10-30-3-0.95-1
ba10-30-3-0.95-0
ba10-10-3-0.95-0
sw10-30-3-0.95-1
sw10-10-3-0.95-1
sw10-10-3-0.95-0
ba5-30-3-0.95-1
ba5-30-3-0.95-0
ba5-10-3-0.95-0
sw10-30-3-0.95-0
sw5-30-3-0.95-0
sw5-10-3-0.95-1
sw5-30-3-0.95-1
sw5-10-3-0.95-0
sw5-10-3-0.75-1
ba5-10-3-0.95-1
sw5-30-3-0.75-1
sw5-30-3-0.75-0
sw5-10-3-0.75-0
Group Name
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
Group Mean

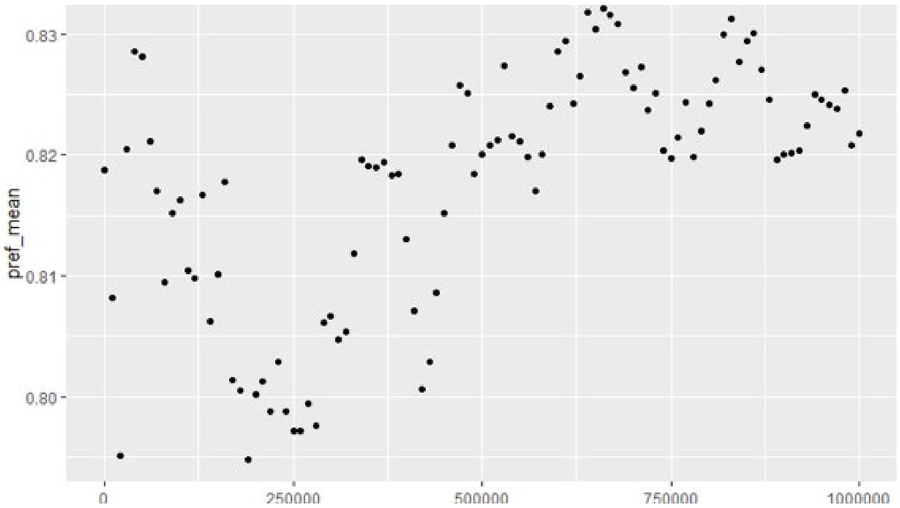
SEM direct effects and indirect effects (mediation)

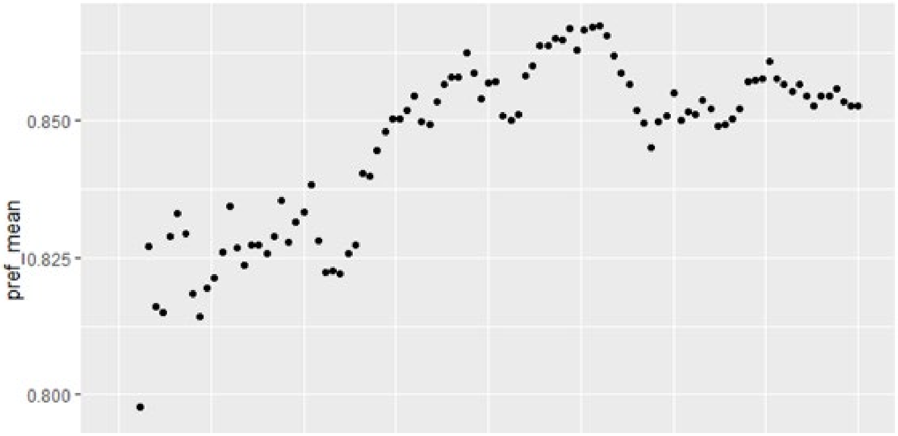
7 variables have a significant impact with indirect mediations. Among them, 5 also contribute to facilitating backlash (Commitment, Complexity, Population, Centralized Network, Density).

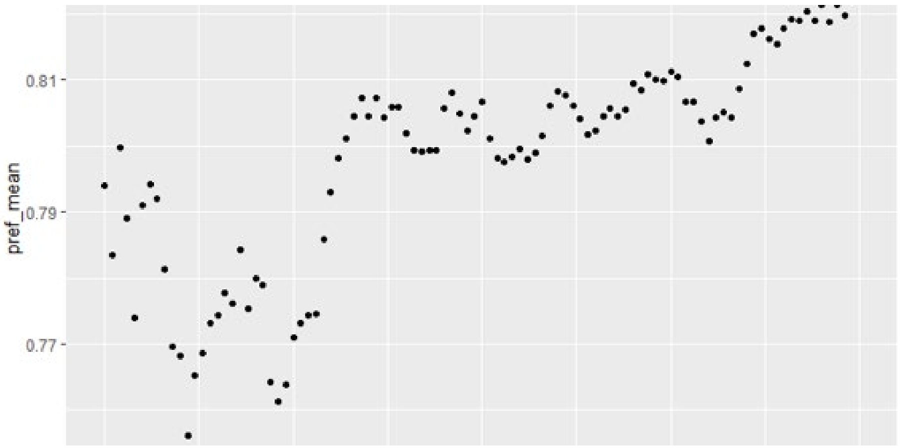
Total Effect	Coefficient	Type	Direct Effect	Coefficient	P-value	Indirect Effect	Coefficient	P-value
Minority Proportion	0.001***	Partial	Minority Proportion->Y	-0.399***	0.000	Minority Proportion->Mean->Y	0.400***	0.000
Centrism	0.005***	Complete	Centrism -> Y	0.003	0.667	Centrism -> Speed -> Y	0.002***	0.000
With Commitment	0.018***	Partial	With Commitment -> Y	0.039***	0.000	With Commitment-> Mean-> Y	-0.016***	0.000
						With Commitment->Congruence->Y	-0.038***	0.000
Issue Complexity	0.88***	Partial	Issue Complexity -> Y	0.975***	0.000	Issue Complexity ->Mean->Y	-0.020***	0.000
						Issue Complexity-> Congruence-> Y	-0.038***	0.000
						Issue Complexity ->Speed -> Y	0.003***	0.000
Population Size	-0.02***	Partial	Population Size -> Y	0.068***	0.000	Population Size -> Mean->Y	-0.057***	0.000
						Population Size -> Congruence->Y	-0.025***	0.000
						Population Size -> Speed->Y	0.003***	0.000
Core-Periphery Position		No Effect						
Centralized Network	0.055***	Partial	Centralized Network->Y	0.041***	0.000	Centralized Network->Mean-> Y	0.012***	0.000
						Centralized Network->Congruence->Y	0.010***	0.000
						Centralized Network-> Speed->Y	-0.008***	0.000
Network Density	0.020***	Partial	Network Density->Y	0.0168***	0.001	Network Density->Mean->Y	0.004***	0.002
						Network Density->Congruence->Y	0.002***	0.000
						Network Density->Speed->Y	-0.002***	0.000

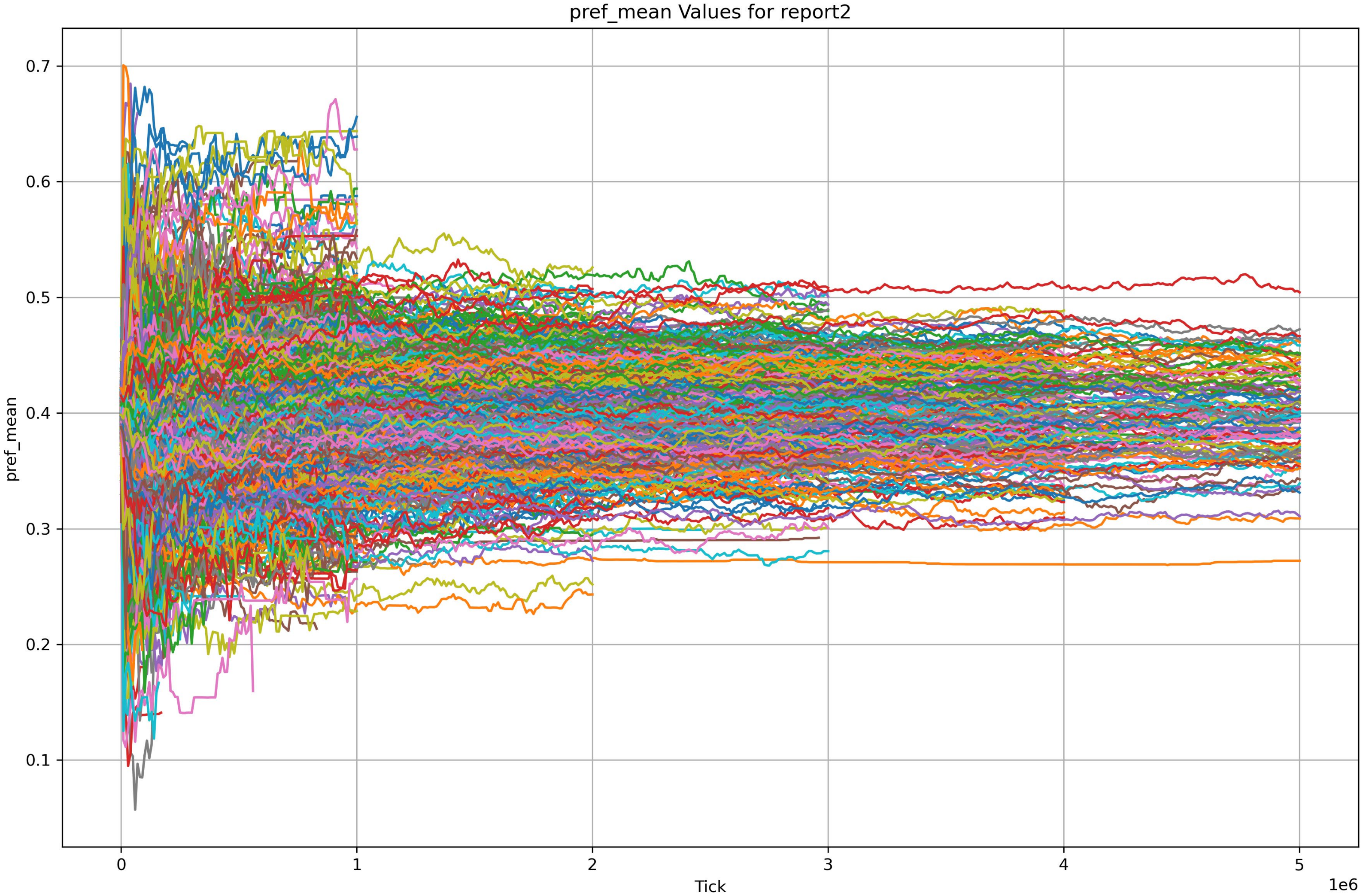
MC(500) (Baron and Kenny 1986; Zhao, Lynch, & Chen 2010); ***p<0.01;


0.81
0.79
0.77
0.75
pref_mean


0.83
0.82
0.81
0.80
pref_mean
0
250000
500000
750000
1000000

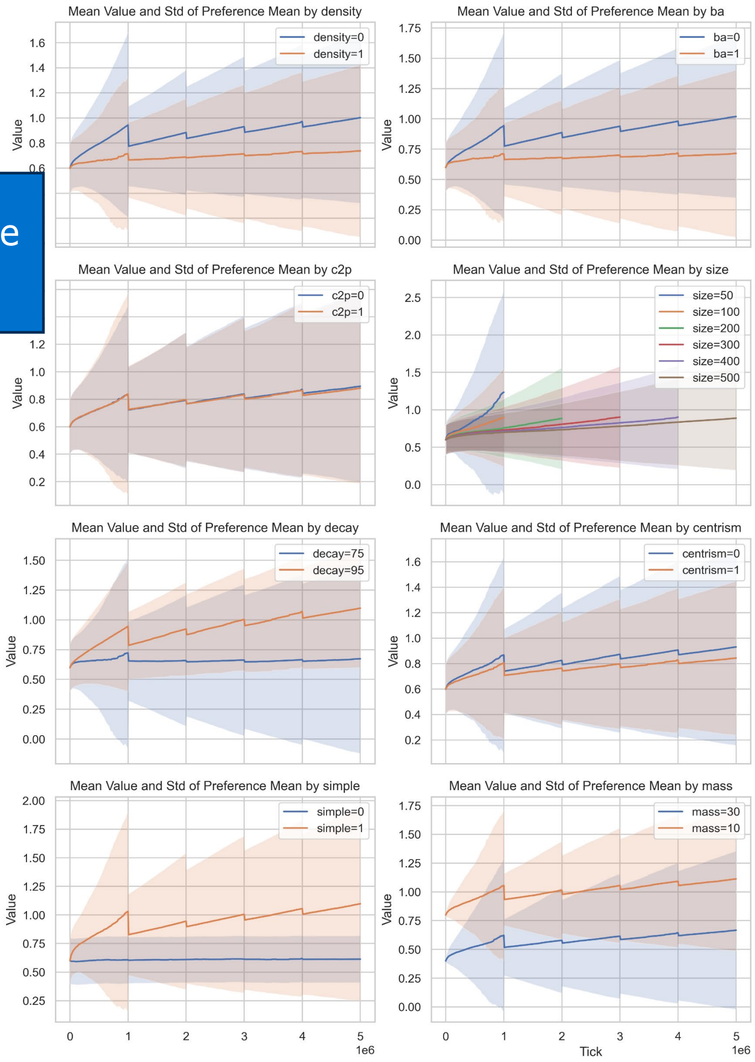

0.850
0.825
0.800
pref_mean


0.81
0.79
0.77
pref_mean

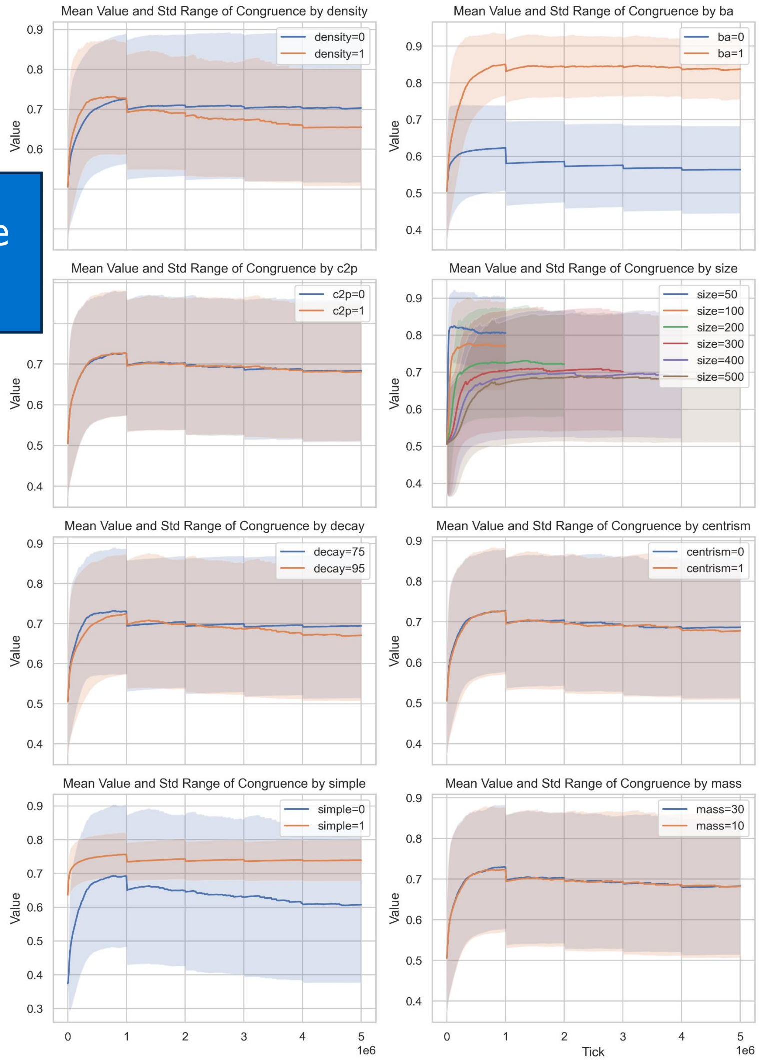


pref_mean Values for report2
0.7
0.6
0.5
0.4
0.3
0.2
0.1
pref_mean
0
1
2
3
4
5
1e6
Tick

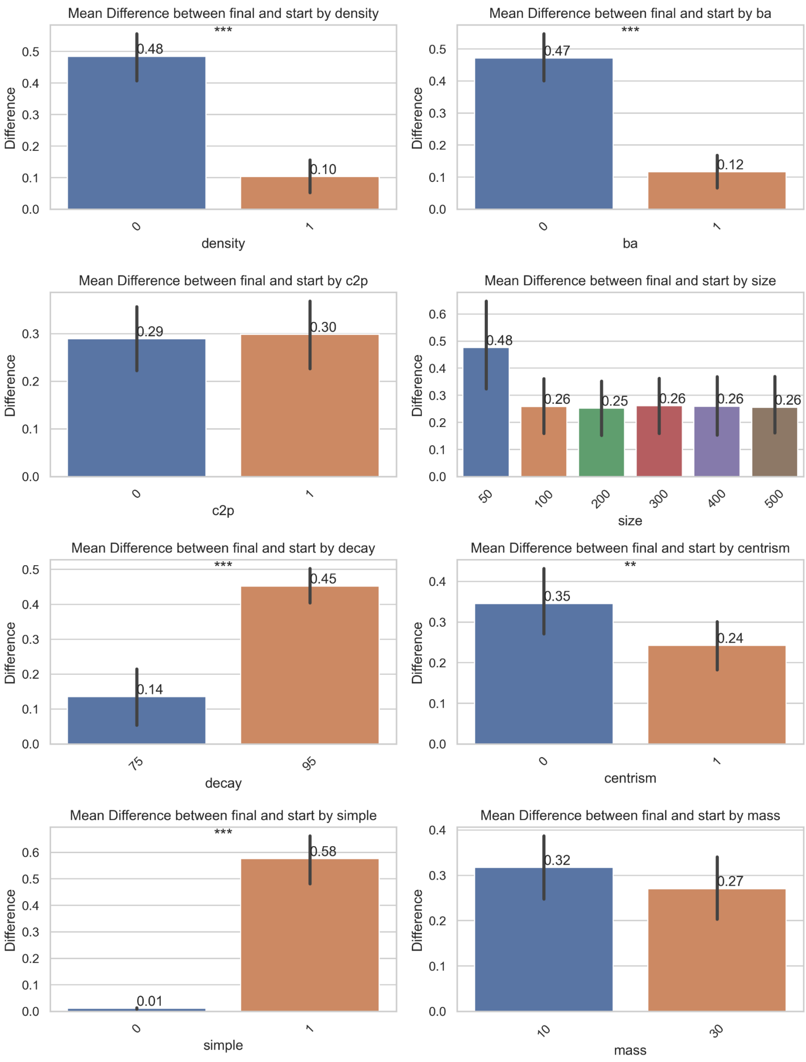
Preference change



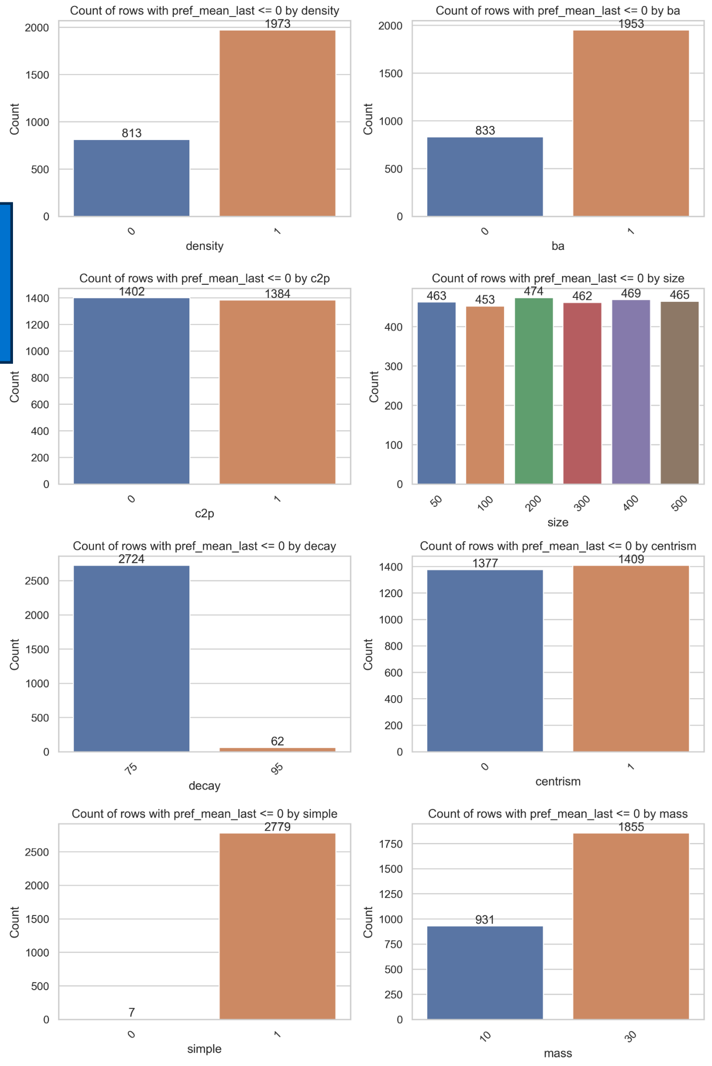
Congruence change



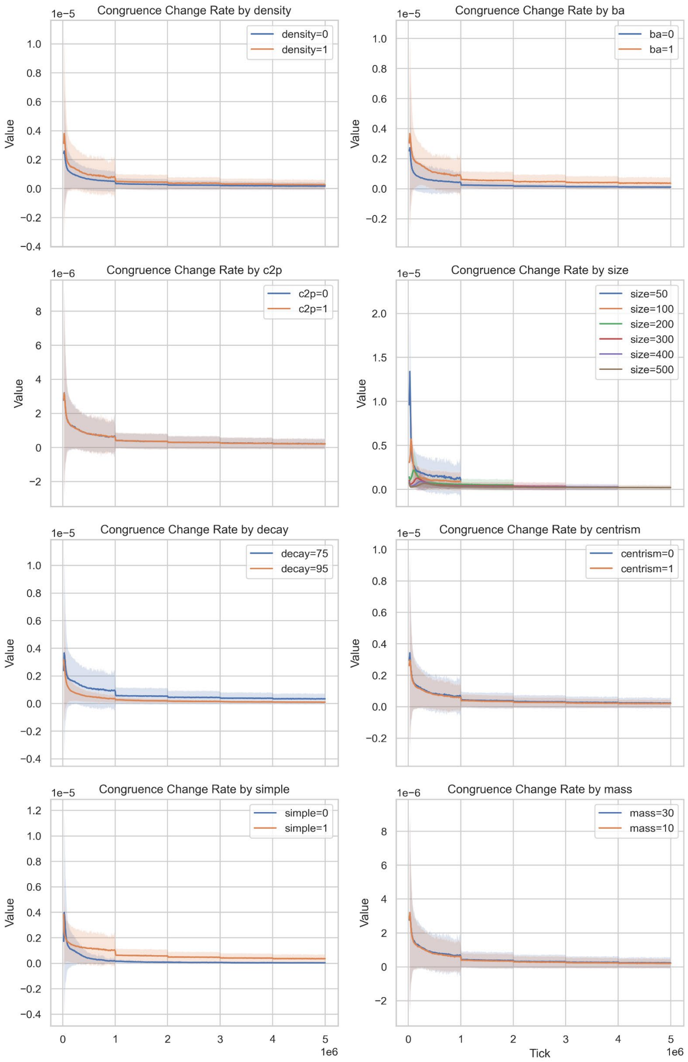
Gradual change



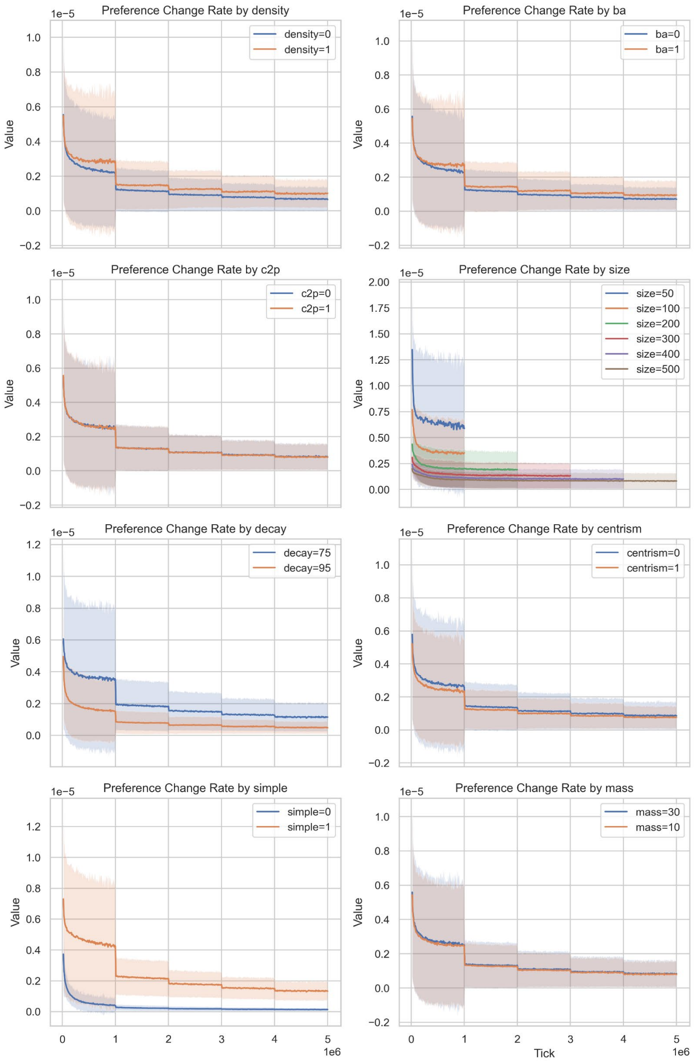
Dramatic change

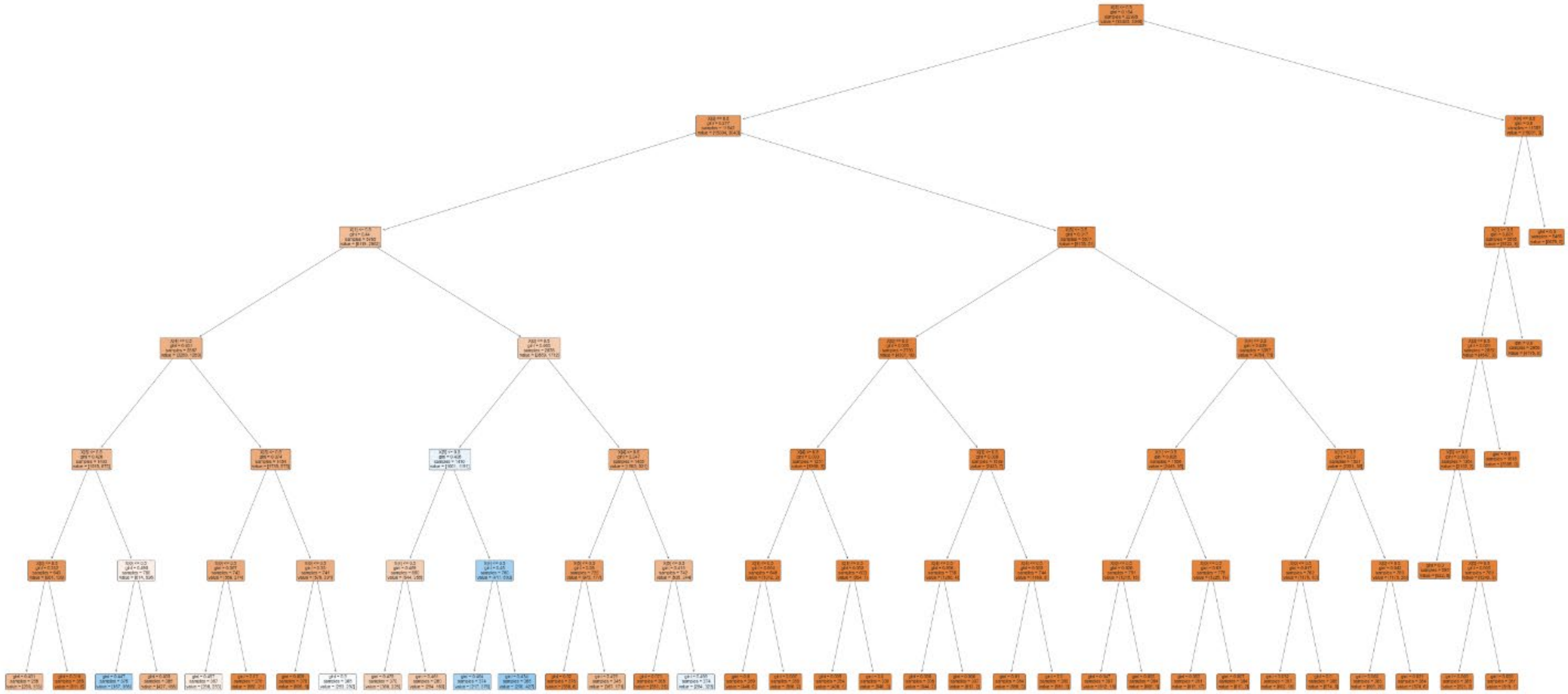


Rate of group congruence



Rate of group preference





Future Potentials

- From Agent-Based Simulation to LLM-Based Agent

Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

Meredith Ringel Morris
Google DeepMind
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu

SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks

Bill Yuchen Lin [1] Yicheng Fu [4] Karina Yang [2] Faeze Brahman [13] Shiyu Huang [5]
Chandra Bhagavatula [1] Prithviraj Ammanabrolu [67] Yejin Choi [31] Xiang Ren [21]

[1]Allen Institute for Artificial Intelligence
[2]University of Southern California [3]University of Washington [4]Tsinghua University
[5]4Paradigm Inc. [6]University of California, San Diego [7]MosaicML

https://swiftsage.github.io

Abstract

We introduce SwiftSage, a novel agent framework inspired by the dual-process theory of human cognition, designed to excel in action planning for complex interactive reasoning tasks. SwiftSage integrates the strengths of behavior cloning and prompting large language models (LLMs) to enhance task completion performance. The framework comprises two primary modules: the Swift module, representing fast and intuitive thinking, and the Sage module, emulating deliberate thought processes. The Swift module is a small encoder-decoder LM fine-tuned on the oracle agent's action trajectories, while the Sage module employs LLMs such as GPT-4 for subgoal planning and grounding. We develop a heuristic method to harmoniously integrate the two modules, resulting in a more efficient and robust problem-solving process. In 30 tasks from the ScienceWorld benchmark, SwiftSage significantly outperforms other methods such as SayCan, ReAct, and Reflexion, demonstrating its effectiveness in solving complex interactive tasks.[1]

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Other Examples

Transformer Application 1

- Identify features of sophistic frames (RoBERT)

Special Issue Article

Sociological Methods & Research
2022, Vol. 51(4) 1721–1787

DOI: 10.1177/00491241221122317
journals.sagepub.com/home/smr
SAGE

Politics as Usual? Measuring Populism, Nationalism, and Authoritarianism in U.S. Presidential Campaigns (1952–2020) with Neural Language Models

Bart Bonikowski, Yuchen Luo, and Oscar Stuhler

Transformer Application 2

- Survey data prediction (survey as sequential vector)

Using Sequences of Life-events to Predict Human Lives

Germans Savcisens, Tina Eliassi-Rad, Lars Kai Hansen, Laust Hvas Mortensen,
Lau Lilleholt, Anna Rogers, Ingo Zettler, and Sune Lehmann

June 6, 2023