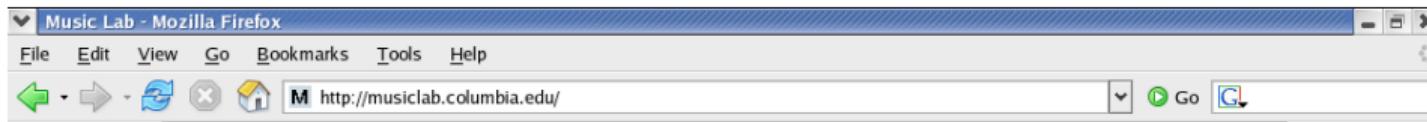


Class slides for Tuesday, October 6: Social fads, part 1

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University



MUSIC LAB

COLUMBIA UNIVERSITY

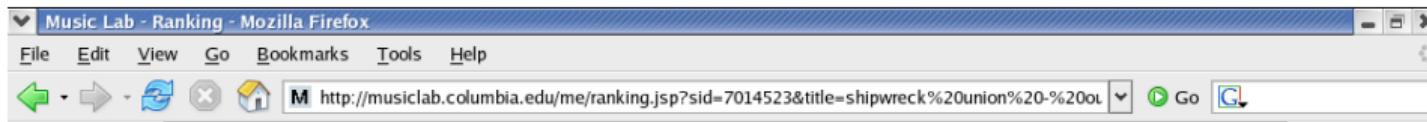
CLICK
HERE TO
START

FREE MUSIC DOWNLOADS

A photograph of a four-piece drum set (bass drum, snare, and two toms) against a plain, light-colored wall. The bass drum's head features the text "CLICK HERE TO START" in white capital letters. The top right corner of the image contains the Columbia University logo and the words "COLUMBIA UNIVERSITY". At the bottom right, there is a blue rectangular button with the text "FREE MUSIC DOWNLOADS".



	# of down loads	[Help]	[Log off]	# of down loads	# of down loads
HARTSFIELD: "enough is enough"	20	GO MOREDAI: "it does what its told"	12	UNDO: "while the world passes"	24
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"	13
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink aggression"	27	SILVERFOX: "gnaw"	17
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "florence"	14	STRANGER: "one drop"	10
THIS NEW DAWN: "the belief above the answer"	12	FORTHFADING: "fear"	24	FAR FROM KNOWN: "route 9"	18
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "inside out"	46
MORAL HAZARD: "waste of my life"	8	52METRO: "lockdown"	17	DANTE: "lifes mystery"	14
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the count"	16	FADING THROUGH: "wish me luck"	10
SECRETARY: "keep your eyes on the ballistics"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"	34
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "til death do us part (i dont)"	31	BY NOVEMBER: "if i could take you"	20
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and miracles"	17	DRAWN IN THE SKY: "tap the ride"	12
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshevik boogie"	15	SELSIUS: "stars of the city"	22
SALUTE THE DAWN: "i am ero"	13	CAPE RENEWAL: "baseball warlock v1"	12	SIBRIAN: "eye patch"	14
RYAN ESSMAKER: "detour_(be still)"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "robert downey jr"	10
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"	38
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"	16

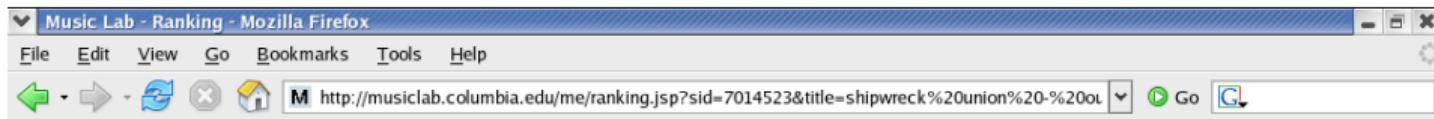


00:21 shipwreck union - out of the woods volume

Please rate this song.
You don't need to wait until it's finished

★★★★★ i love it
 ★★★★ i like it
 ★★★ it's OK
 ★★ i don't like it
 ★ i hate it

Rate It!



00:53 shipwreck union - out of the woods volume

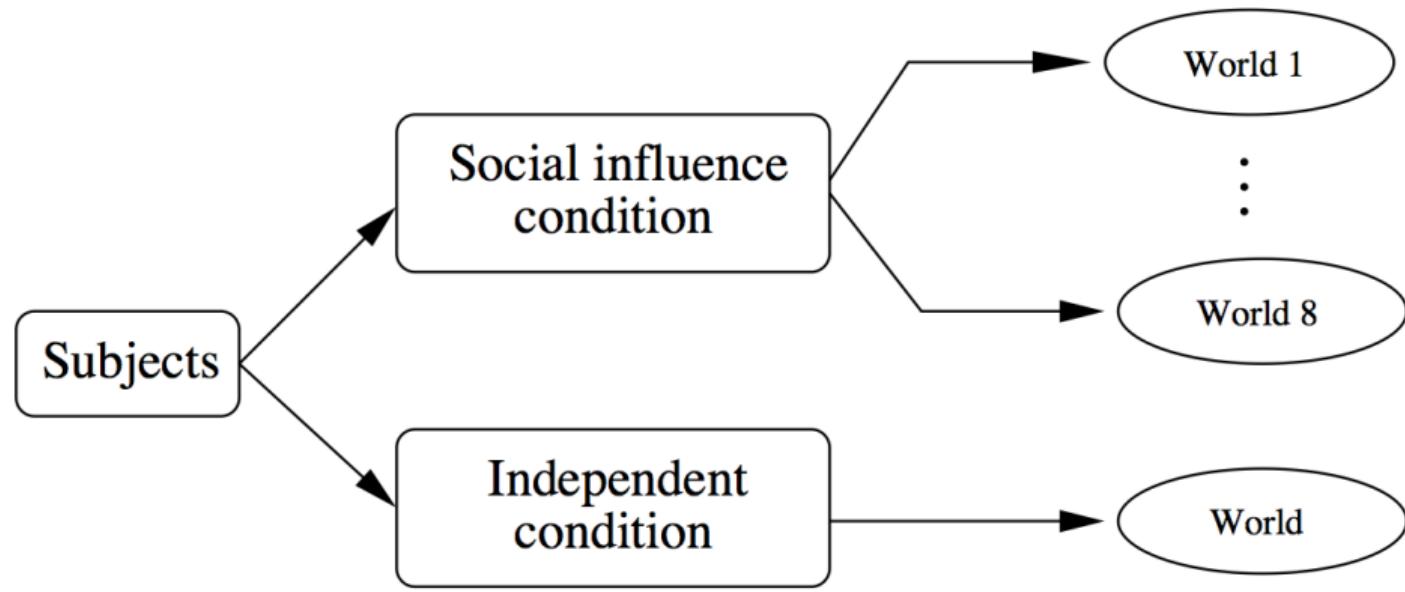
Would you like to download this song?

Yes, download

No, Thanks



	# of down loads	[Help]	[Log off]	# of down loads	# of down loads
HARTSFIELD: "enough is enough"	20	GO MOREDAI: "it does what its told"	12	UNDO: "while the world passes"	24
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"	13
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink aggression"	27	SILVERFOX: "gnaw"	17
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "florence"	14	STRANGER: "one drop"	10
THIS NEW DAWN: "the belief above the answer"	12	FORTHFADING: "fear"	24	FAR FROM KNOWN: "route 9"	18
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "inside out"	46
MORAL HAZARD: "waste of my life"	8	52METRO: "lockdown"	17	DANTE: "lifes mystery"	14
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the count"	16	FADING THROUGH: "wish me luck"	10
SECRETARY: "keep your eyes on the ballistics"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"	34
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "til death do us part (i dont)"	31	BY NOVEMBER: "if i could take you"	20
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and miracles"	17	DRAWN IN THE SKY: "tap the ride"	12
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshevik boogie"	15	SELSIUS: "stars of the city"	22
SALUTE THE DAWN: "i am ero"	13	CAPE RENEWAL: "baseball warbck v1"	12	SIBRIAN: "eye patch"	14
RYAN ESSMAKER: "detour_(be still)"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "robert downey jr"	10
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"	38
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"	16



Music Lab - Song Selection - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

M http://www.musiclab.columbia.edu/me/songs

	# of down loads	[Help]	[Log off]	# of down loads	[Help]	[Log off]
HARTFIELD "though & through"	20	GO HARRISOCAL "it does what it did"		12	UNICO "telle the world passes"	24
DEEP PATRIOT TO DIE: "for the sky"	17	PARKER THEORY: "she said"		47	UP FOR NOTHING: "in right of"	13
THE TRAIT SYNDICATE: "2001 a negativ"	20	MISS OCTOBER: "pink aggression"		27	SILVERTOKE "grow"	17
THE BROKEN PROMISE: "the end is here"	19	POST BREAK TRAGEDY: "reverse"		14	STRANGERS: "one day"	18
THIS NEW DAWN: "you let feel above the arrows"	12	PORTHOLEONG:		24	FAIR FROM UNKNOWN: "you are"	18
HOUSER AT HOME: "work makes me"	6	THE CALLEFACTION: "stepped in an orange peal"		20	STUTT MONKEY: "wide eye"	46
MORAL HAZARD: "waste of my life"	8	LIMITHO: "lockdown"		17	DANTE: "this system"	14
NOT FOR SCHOLARS: "as seasons change"	27	BIMPY WAITING: "over with the court"		16	PAIDAC THROUGH: "with the lack"	18
SECRETARY: "keep peacocks on the balloons"	5	STAR CLIMBER: "tell me"		36	UNKNOWN CITIZENS: "telling over"	34
ART OF EARLY: "adventures, methods, insouciance"	32	THE FASTLANE: "feel do do so good (do do)"		31	BY NOVEMBER: "take care jive"	28
HYDRAULIC SANDWICH: "separated anxiety"	39	A BLINDING SILENCE: "memories and insects"		17	DRAINS IN THE SKY: "top the sky"	12
EMBER SKY: "this upcoming winter"	25	SLIM RANKA: "like baba ka bhang"		15	SELSSUS: "size of the city"	22
SALUTE THE DAWN: "i am ego"	13	CAFE RENAISSANCE: "blossomed mornock v/s"		12	SIERRAN: "eye rock"	14
RYAN ESSAMAKER: "minus the sun"	14	ZIP FALLS DOWN: "what's the money?"		11	IVAN GOLD: "what's the money?"	19
DEPTHOGIC: "mother to son"	12	SLUMMERS: "a plan for kind decoration"		17	RENDITION OF A DOUBT: "i'm sorry"	38
HALL OF FAME: "best mistakes"	20	SILENT FILM: "all i have to say"		61	SHAPAREEK UNION: "set of the woods"	16

(a) Experiment 1, Weaker signal

Music Lab - Song Selection - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

M http://www.musiclab.columbia.edu/me/cont1/

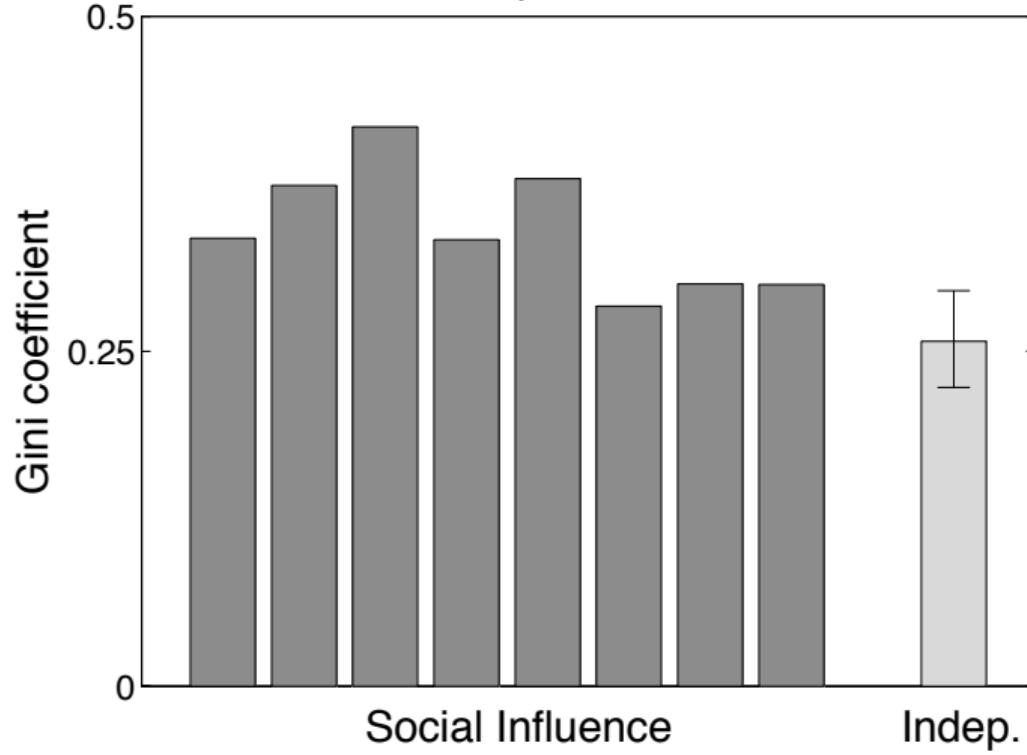
	# of down loads	[Help]	[Log off]
PARKER THEORY: "she said"	139		
THE FASTLANE: "what do we gain (do do)"	181		
SELSSUS: "size of the city"	62		
STUTT MONKEY: "wide eye"	56		
BY NOVEMBER: "i could take you"	55		
PORTHOLEONG: "true"	49		
HYDRAULIC SANDWICH: "separated anxiety"	43		
SILENT FILM: "i have to say"	40		
UDOO: "while like we all pass"	36		
BENIGHT OF A DOUBT: "i'm sorry"	32		
A BLINDING SILENCE: "memories and insects"	27		
MISS OCTOBER: "pink aggression"	26		
STAR CLIMBER: "take care"	24		
FAIR FROM UNKNOWN: "you are"	22		
HALL OF FAME: "best mistakes"	21		
EMBER SKY: "this upcoming winter"	19		

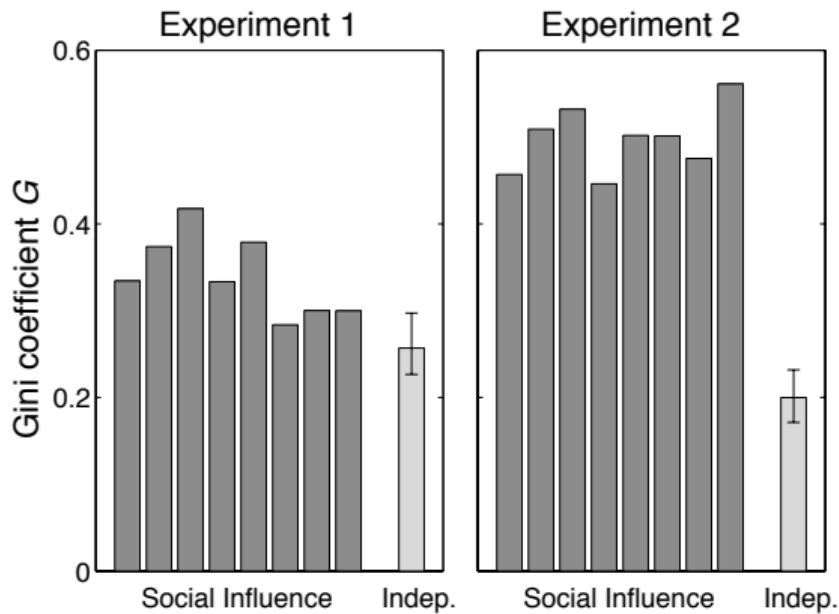
(b) Experiment 2, Stronger signal

Two main measures

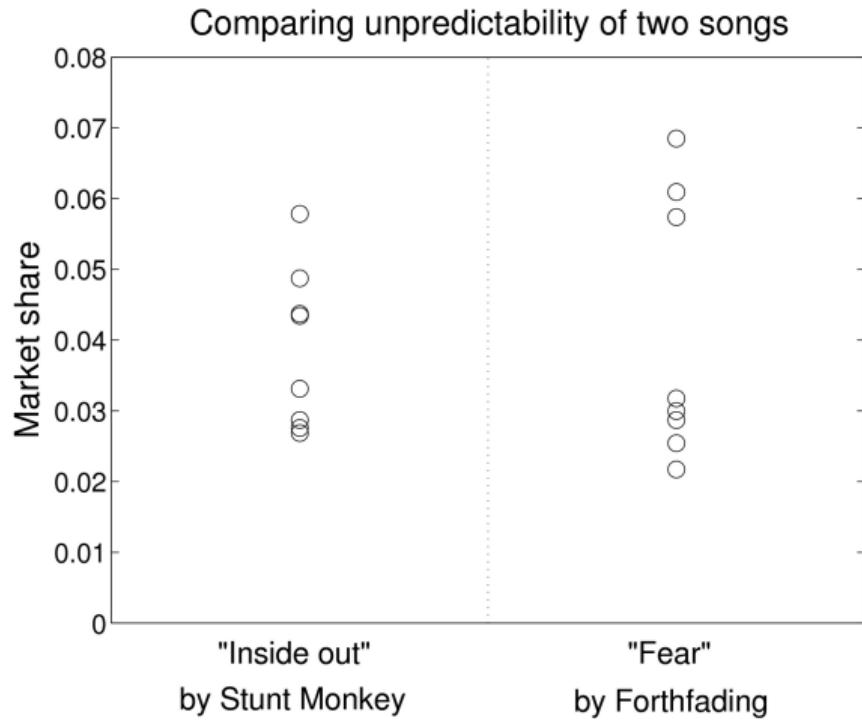
- ▶ Inequality
- ▶ Unpredictability

Experiment 1



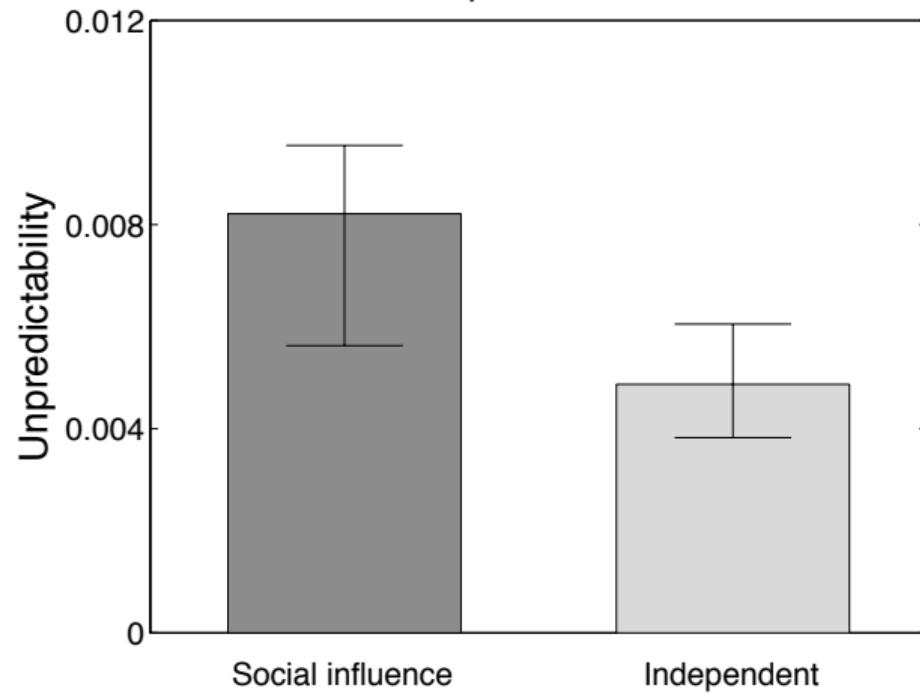


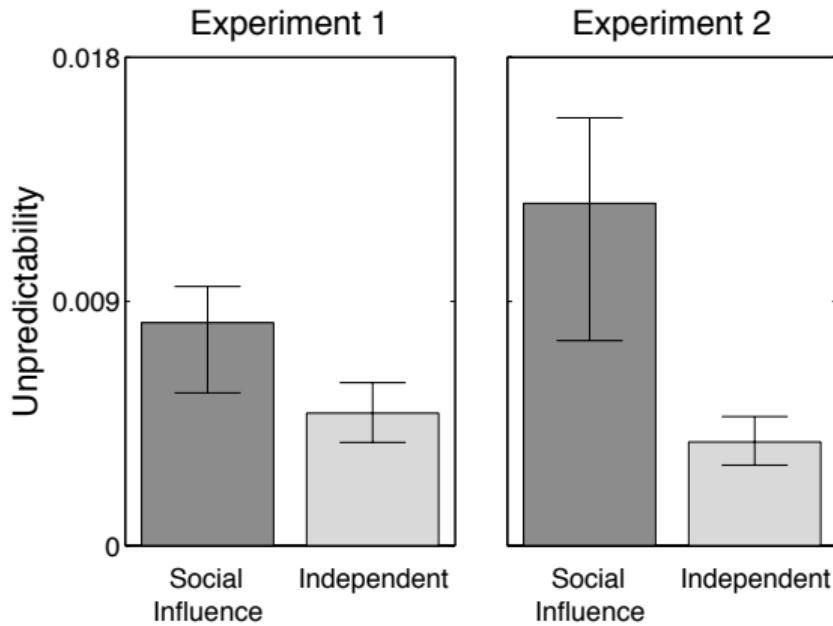
Median Gini coefficient increases from 0.34 (France) to 0.50 (Nigeria)



U = mean difference in market share across all pairs of realizations

Experiment 1

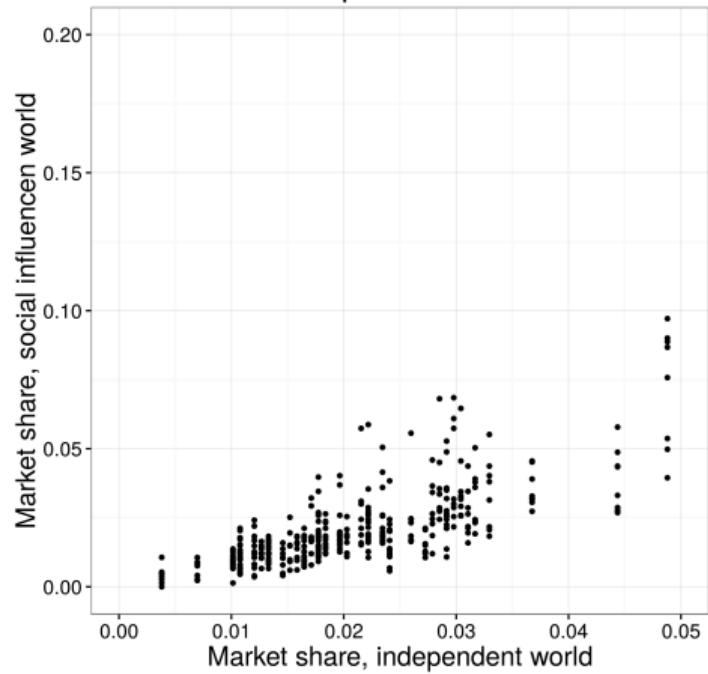




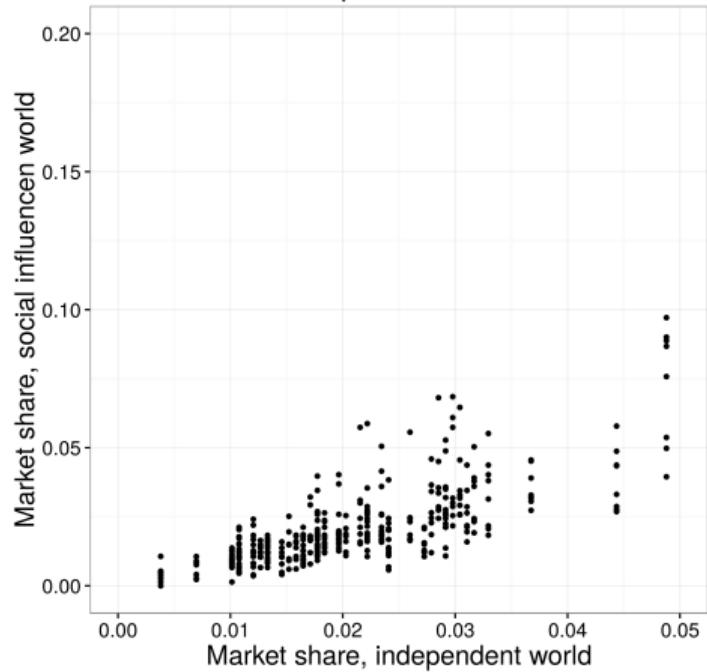
Unpredictability increases by about 50%

What is the relationship between “quality” and success? Or expressed in terms of Martin et al. (2016) what is the relationship between “skill” and success?

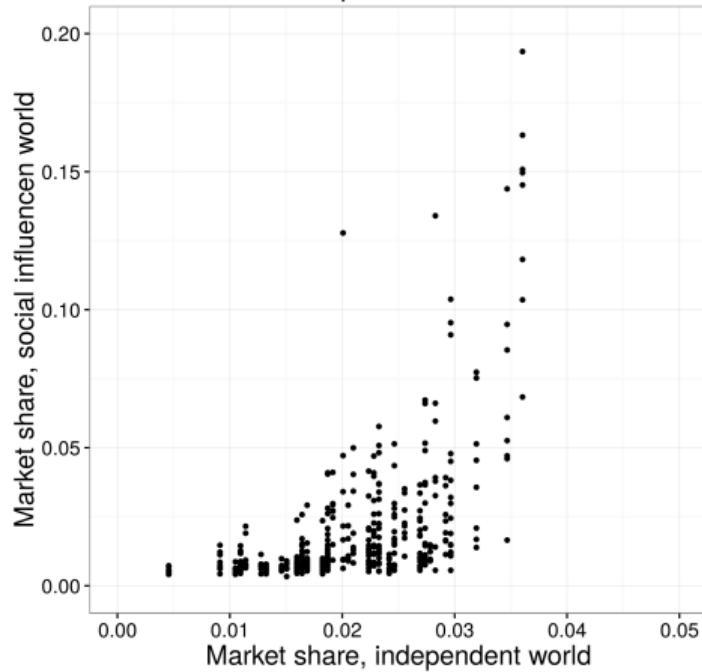
Experiment 1



Experiment 1



Experiment 2



Behind the MusicLab

Two main measures

- ▶ Inequality
- ▶ Unpredictability

Measuring Inequality

A Methodological Handbook

Philip B. Coulter



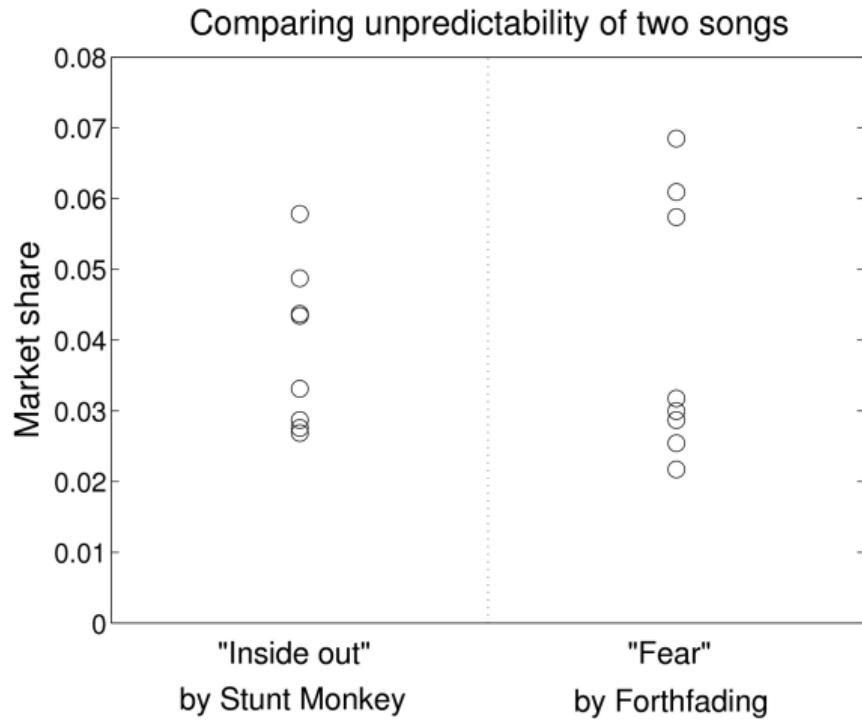
CONTENTS

TABLES AND FIGURES	ix
PREFACE AND ACKNOWLEDGMENTS	xi
1. DISTRIBUTION IN SOCIAL SCIENCE	1
The Importance of Distribution	2
Conceptual Specifications	3
Useful Analytical Tools	5
Frequency Distribution	5
Cumulative Frequency Distribution	7
The Lorenz Curve	8
The Organization of this Book	9
2. INEQUALITY THEORY	11
Conceptual Criteria	11
Polarity	12
Sensitivity to Concentration	12
Comparative Standard	14
Intragroup or Intergroup Inequality	16
Inequality or Inequity	16
Principle of Transfers	17
Principle of Scale Invariance	18
Principle of Constant Additions	18
Population Symmetry	19
Sensitivity to Number of Components	19
Ordinary Lorenz Dominance Criterion	22
Underlying Mathematical Model	24
Technical Criteria	26
Definition	26
Information Use	26
Simplicity	27
Interpretability	27
Scale of Measurement	28
Essential Statistical Notation	31
Statistical Symbols	32

3. INDEXES BASED ON THE DEVIATIONS MODEL	35
The Deviations Logic	35
Absolute Deviations from Central Tendency	40
Wilcox's Deviations from the Mode	40
Dahl's Index of Polyarchy	41
Schutz's Coefficient of Inequality	44
Squared Deviations from Central Tendency	47
Mayer's Index of Uniformity	47
Nagel's Index of Equality	49
Differences from All Other Components	50
Gini's Mean Relative Difference	50
Gini Coefficient of Inequality	52
Differences from Temporally Adjacent Components	58
Przeworski's Index of Instability	58
Conclusions	61
4. INDEXES BASED ON THE COMBINATORICS MODEL	63
The Logic of Combinatorics	63
Combinations and Permutations	64
Multinomial Probabilities	65
Herfindahl and Hirschman's Index of Concentration	69
Lieberson's Index of Diversity	73
Horvath's Comprehensive Concentration Index	75
Hall and Tideman's Index of Industry Concentration	77
Taagepera's Index of Imbalance	78
Rae and Taylor's Index of Fragmentation	81
Mueller, Scheussler, and Costner's Index of Qualitative Variation	83
Hammond, Householder, and Castellan's Index of Dispersion	85
Leik's Measure of Ordinal Consensus	87
Ray, Taagepera, and Singer's Generalized Index of Concentration	91
Taagepera and Ray's Equivalent Two-component System	94
Laakso and Taagepera's Effective Number of Components	94
Taagepera and Grofman's Effective Size	96
Laakso and Taagepera's Index of Fluctuation	97
Conclusions	99

Table 3-2: Primary Properties of Indexes Based on the Deviations Model

	Polarity	Type of Index	Concentration	Comparative standard	Constant additions	Transfers	Scale invariance	Population symmetry	Lorenz dominance	Scale	Definition	Information use	Upper and lower limits	Simplicity
Wilcox's deviation from the mode (DM)	equality	relative, ANONC	no	mode	increase	yes	yes	no	no	nominal	yes	good	1, 0	good
Dahl's polyarchy (P)	equality	absolute, type A	yes	mean	increase	yes	yes	yes	yes	interval	yes	good	1-(1/K), 0	fair
Schutz's inequality (S)	inequality	absolute, type A	no	mean	decrease	no	yes	yes	no	nominal	yes	poor	1-(1/K), 0	good
Mayer's uniformity (M)	equality	absolute, type A	yes	mean	increase	yes	yes	yes	yes	nominal	yes	good	K-1, 0	good
Nagel's equality (E)	equality	relative, ANONC	yes	mean	decrease	yes	yes	yes	yes	nominal	yes	good	1, 0	good
Gini's mean relative difference (MRD)	equality	relative, ANONC	yes	all other components	increase	yes	yes	yes	yes	interval	yes	good	1, 0	fair
Gini coefficient (G)	inequality	absolute, type A	yes	all other components	decrease	yes	yes	yes	yes	interval	yes	good	1-(1/K), 0	poor
Przeworski's instability (D_1)	inequality	absolute, type A	no	adjacent component	decrease	yes	yes	yes	yes	nominal	yes	good	$t_n - 1, 0$	poor



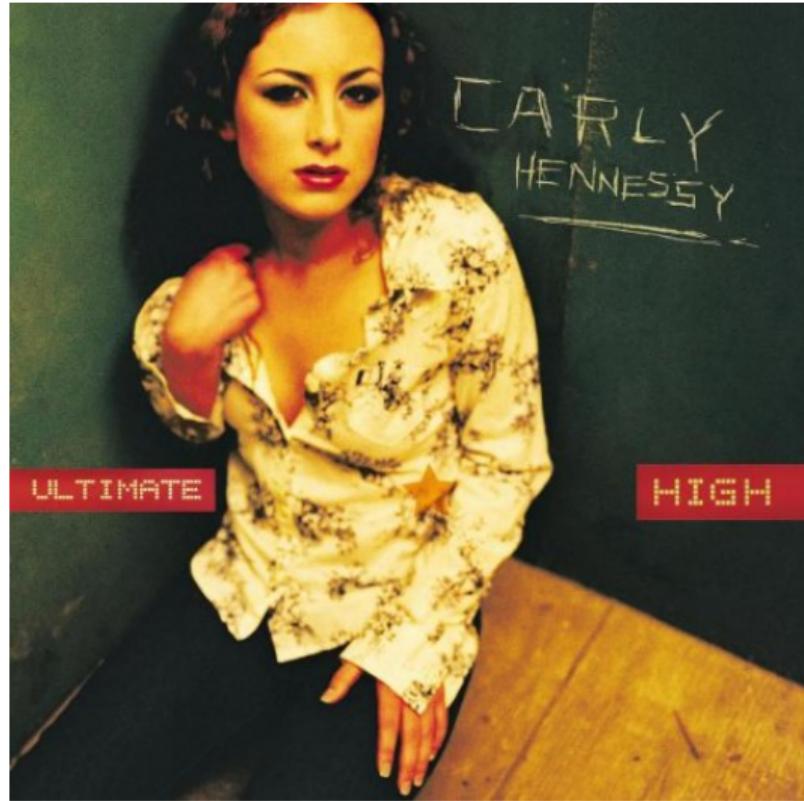
U = mean difference in market share across all pairs of realizations

**THE WAY WE SHOULD
MEASURE UNPREDICTABILITY**



**IS DIFFERENT FROM THE WAY
WE SHOULD MEASURE PREDICTABILITY**

How did we come to study unpredictability?



https://www.youtube.com/watch?v=2wfhk46_tXE

Hindsight Bias

Neal J. Roese¹ and Kathleen D. Vohs²

¹Kellogg School of Management, Northwestern University and ²University of Minnesota

Perspectives on Psychological Science
7(5) 411–426

© The Author(s) 2012

Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691612454303
<http://pps.sagepub.com>



Abstract

Hindsight bias occurs when people feel that they “knew it all along,” that is, when they believe that an event is more predictable after it becomes known than it was before it became known. Hindsight bias embodies any combination of three aspects: memory distortion, beliefs about events’ objective likelihoods, or subjective beliefs about one’s own prediction abilities. Hindsight bias stems from (a) cognitive inputs (people selectively recall information consistent with what they now know to be true and engage in sensemaking to impose meaning on their own knowledge), (b) metacognitive inputs (the ease with which a past outcome is understood may be misattributed to its assumed prior likelihood), and (c) motivational inputs (people have a need to see the world as orderly and predictable and to avoid being blamed for problems). Consequences of hindsight bias include myopic attention to a single causal understanding of the past (to the neglect of other reasonable explanations) as well as general overconfidence in the certainty of one’s judgments. New technologies for visualizing and understanding data sets may have the unintended consequence of heightening hindsight bias, but an intervention that encourages people to consider alternative causal explanations for a given outcome can reduce hindsight bias.

**THERE IS MORE
UNPREDICTABILITY IN THE WORLD**

**THAN MANY
PEOPLE REALIZE**

Exploring limits to prediction in complex social systems

Travis Martin
University of Michigan
Dept. of Computer Science
Ann Arbor, MI
travisbm@umich.edu

Jake M. Hofman
Microsoft Research
641 6th Ave, Floor 7
New York, NY
jmh@microsoft.com

Amit Sharma
Microsoft Research
amshar@microsoft.com

Ashton Anderson
Microsoft Research
ashton@microsoft.com

Duncan J. Watts
Microsoft Research
duncan@microsoft.com



Arvind Narayanan ✅ @random_walker · Nov 19, 2019

Much of what's being sold as "AI" today is snake oil. It does not and cannot work. In a talk at MIT yesterday, I described why this happening, how we can recognize flawed AI claims, and push back. Here are my annotated slides: cs.princeton.edu/~arvindn/talks...

How to recognize AI snake oil

Arvind Narayanan

Associate Professor of Computer Science

@random_walker



PRINCETON
UNIVERSITY

CITP
CENTER FOR
INFORMATION TECHNOLOGY POLICY



Arvind Narayanan @random_walker · Nov 19, 2019

Much of what's being sold as "AI" today is snake oil. It does not and cannot work. In a talk at MIT yesterday, I described why this happening, how we can recognize flawed AI claims, and push back. Here are my annotated slides: cs.princeton.edu/~arvindn/talks...

How to recognize AI snake oil

Arvind Narayanan

Associate Professor of Computer Science

@random_walker



168

5.4K

10.1K



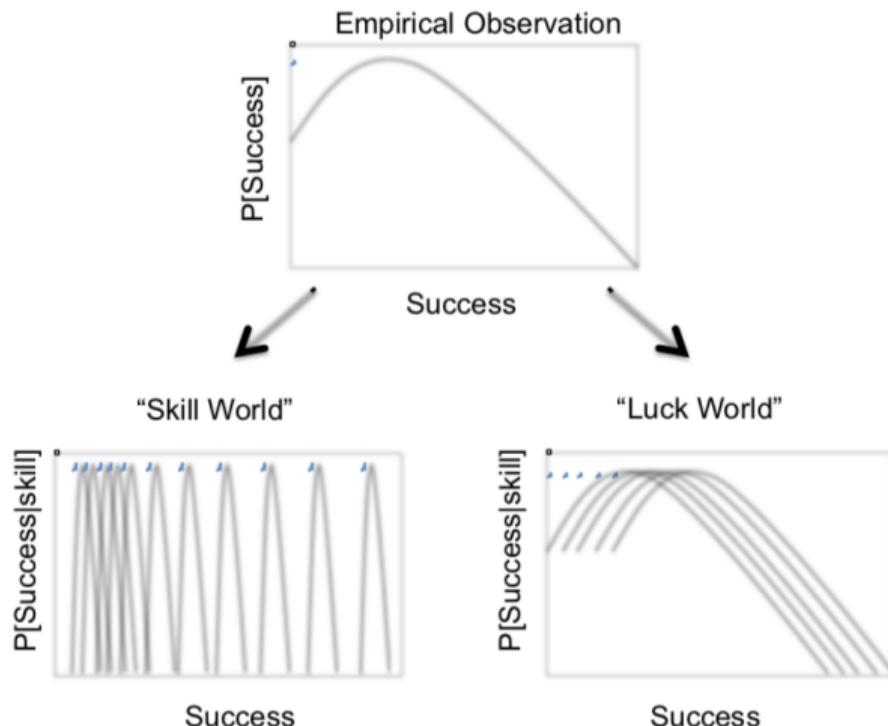


Figure 1: Schematic model illustrating two stylized explanations for an empirically observed distribution of success.

$$s = f(q) + \epsilon$$

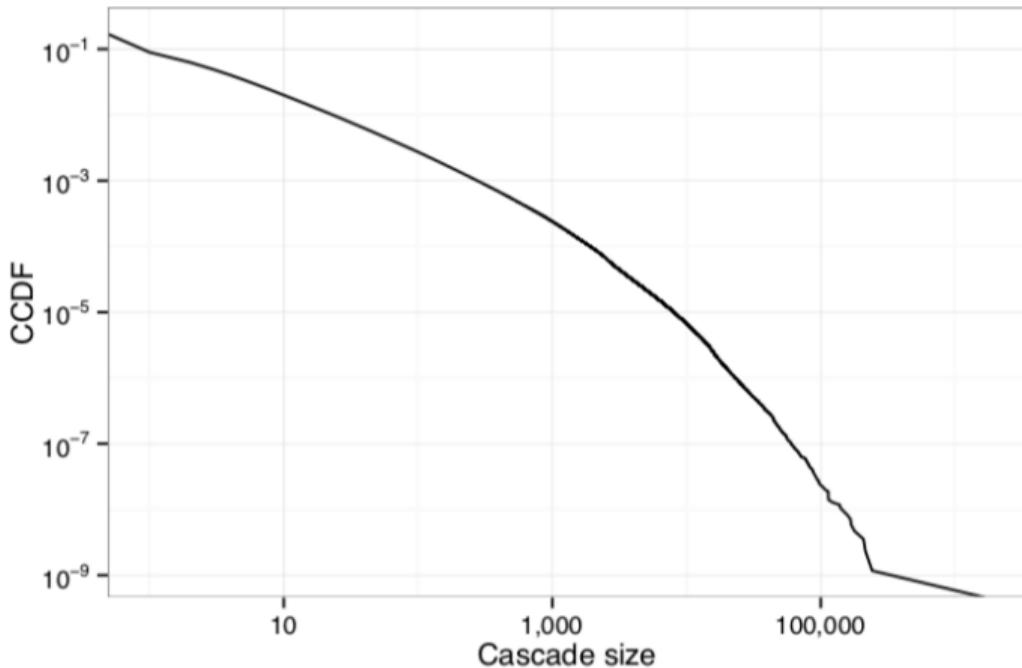
$$F = \frac{\mathbb{E}[\text{Var}(\mathbf{S}|Q)]}{\text{Var}(\mathbf{S})} = \frac{\sum_q \sum_{i:q_i=q} (\bar{s}_q - s_i)^2}{\sum_j (\bar{s} - s_j)^2}, \quad (1)$$

$$F = \frac{\sum_i (f(q_i) - s_i)^2}{\sum_j (\bar{s} - s_j)^2} = 1 - R^2, \quad (2)$$

**THE WAY WE SHOULD
MEASURE UNPREDICTABILITY**



**IS DIFFERENT FROM THE WAY
WE SHOULD MEASURE PREDICTABILITY**



I believe that it is in general hard to predict outcomes with more variation. Is that true?

Model	Tweet time	Domain	Spam score	Category	Past url success	User time	Followers	Friends	Statuses	User topic	Past user success	Topic interaction
1. Basic content	✓	✓	✓	✓								
2. Content, topic	✓	✓	✓	✓	✓							
3. Content, past succ.	✓	✓	✓	✓	✓	✓						
4. Basic user						✓	✓	✓	✓			
5. User, topic						✓	✓	✓	✓	✓		
6. User, past succ.						✓	✓	✓	✓	✓	✓	
7. Content, user	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
8. All	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 2: Features used in different models for cascade size prediction.

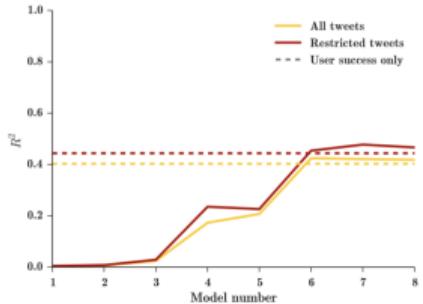


Figure 4: Prediction results for models using different subsets of features. R^2 increases as we add more features, but only up to a limit. Even a model with all features explains less than half of the variance in cascade sizes.

Model	Tweet time	Domain	Spam score	Category	Tweet topic	Past url success	User time	Followers	Friends	Statuses	User topic	Past user success	Topic interaction
1. Basic content	✓	✓	✓	✓									
2. Content, topic	✓	✓	✓	✓	✓								
3. Content, past succ.	✓	✓	✓	✓	✓	✓							
4. Basic user							✓	✓	✓	✓			
5. User, topic							✓	✓	✓	✓	✓		
6. User, past succ.							✓	✓	✓	✓	✓	✓	
7. Content, user	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
8. All	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 2: Features used in different models for cascade size prediction.

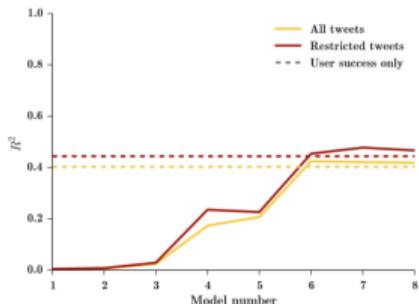


Figure 4: Prediction results for models using different subsets of features. R^2 increases as we add more features, but only up to a limit. Even a model with all features explains less than half of the variance in cascade sizes.

What about my favorite feature? What about deep learning? What about

They don't close the loop. It is very hard to get the simulation results and the empirical results to work together tightly.

Exploring limits to prediction in complex social systems

Travis Martin
University of Michigan
Dept. of Computer Science
Ann Arbor, MI
travisbm@umich.edu

Jake M. Hofman
Microsoft Research
641 6th Ave, Floor 7
New York, NY
jmh@microsoft.com

Amit Sharma
Microsoft Research
amshar@microsoft.com

Ashton Anderson
Microsoft Research
ashton@microsoft.com

Duncan J. Watts
Microsoft Research
duncan@microsoft.com

Prediction and explanation in social systems

Jake M. Hofman,* Amit Sharma,* Duncan J. Watts*

SCIENCE

Psychology's Replication Crisis Is Running Out of Excuses

Another big project has found that only half of studies can be repeated. And this time, the usual explanations fall flat.

ED YONG NOVEMBER 19, 2018

<https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/>

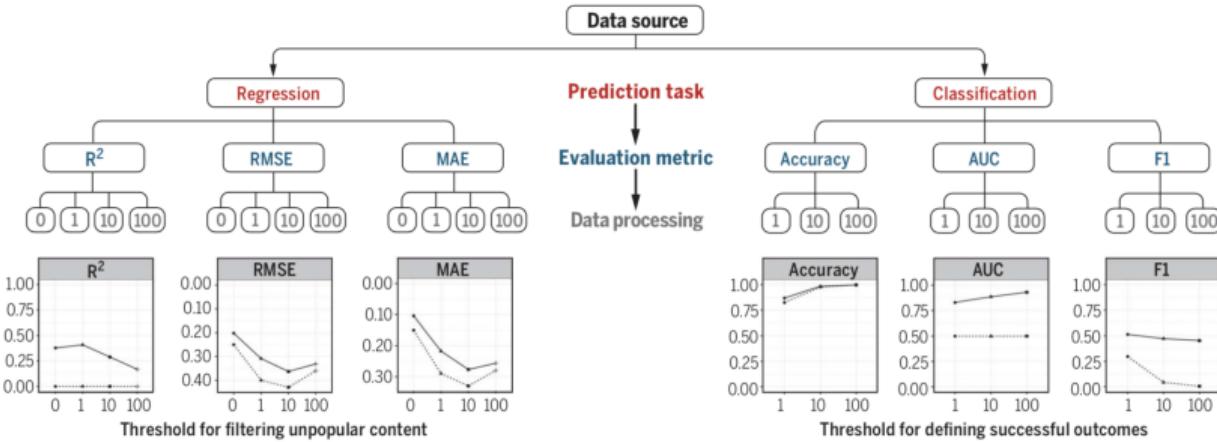


Fig. 1. A single question may correspond to many research designs, each yielding different answers. (Top) A depiction of the many choices involved in translating the problem of understanding diffusion cascades into a concrete prediction task, including the choice of data source, task, evaluation metric, and data preprocessing. The preprocessing choices shown at the terminal nodes refer to the threshold used to filter observations for regression or define successful outcomes for classification. Cascade sizes were log-transformed for all of the regression tasks. (Bottom) The results of each prediction task, for each

metric, as a function of the threshold used in each task. The lower limit of each vertical axis gives the worst possible performance on each metric, and the top gives the best. Dashed lines represent the performance of a naive predictor (always forecasting the global mean for regression or the positive class for classification), and solid lines show the performance of the fitted model. R^2 , coefficient of determination; AUC, area under the ROC curve; RMSE, root mean squared error; MAE, mean absolute error; F1 score, the harmonic mean of precision and recall.

This is why you have to pre-register assignment 1

Prediction and explanation in social systems

Jake M. Hofman,* Amit Sharma,* Duncan J. Watts*

Closing thoughts:

- ▶ social fads are partially unpredictable and partially predictable

Closing thoughts:

- ▶ social fads are partially unpredictable and partially predictable
- ▶ likely mechanism is cumulative advantage

Closing thoughts:

- ▶ social fads are partially unpredictable and partially predictable
- ▶ likely mechanism is cumulative advantage
- ▶ we see that same mechanism in other settings (e.g., disease)

Class slides for Tuesday, October 6: Social fads, part 1

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University