

3. Little data: how traditional statistical ideas remain relevant in a big data world

- ▶ The statistical crisis in science
- ▶ Difficulties with the “false positive, false negative” framework
- ▶ Analyzing all comparisons at once



most published fin

Scholar Articles and patents anytime

[HTML] [Why most published research findings are false](#)

JPA Ioannidis - PLoS medicine, 2005 - dx.plos.org

Summary There is increasing concern that **most** current published research findings are **false**. The probability that a research claim is true may depend on the number of other studies on the same question, and, importantly, the **reproducibility** of the findings. This has important implications for medical practice and for the way that medical research is conducted and disseminated. The problem is particularly acute in fields such as medicine, where the number of studies is large and the methods used are often complex and difficult to replicate. The authors argue that the current system of publishing research findings is biased towards positive results, which can lead to a high proportion of false positives. They propose a new approach to research publication that takes into account the number of other studies on the same question and the reproducibility of the findings. This would help to identify the most reliable research findings and to promote the use of more rigorous methods in medical research. The article is available online at dx.plos.org.

Cited by 972 - Related articles - Cached - BL Direct - All 146 versions

[HTML] [Most published research findings are false—but a little better](#)

R Moonesinghe, MJ Khoury... - PLoS Medicine, 2007 - dx.plos.org

Controversy in social psychology, medicine, . . .

- ▶ Fragile research findings
 - ▶ Joke research (Bem, Kanazawa, etc.)
 - ▶ Fraud, misconduct, and error (Hauser, Stapel, Anderson and Ones, etc.)
 - ▶ Systematic biases (selection, the statistical significance filter, etc.)
- ▶ Problems with the default model

A 61-million-person experiment in social influence and political mobilization

Robert M. Bond¹, Christopher J. Fariss¹, Jason J. Jones², Adam D. I. Kramer³, Cameron Marlow³, Jaime E. Settle¹
& James H. Fowler^{1,4}

Human behaviour is thought to spread through face-to-face social networks, but it is difficult to identify social influence effects in observational studies^{9–13}, and it is unknown whether online social

with all users of at least 18 years of age in the United States who accessed the Facebook website on 2 November 2010, the day of the US congressional elections. Users were randomly assigned to a ‘social

- ▶ Facebook message directly increases voter turnout by 0.3%
 - ▶ Plausible small effect of innocuous advertisement
- ▶ Indirect (social) effect of 0.01%–0.1%
 - ▶ Lost in the noise—even if statistically significant

An example of research in behavior and genetics

"We expected reported anxiety to be significantly higher in the closeness condition compared to either of the other two treatments . . . There is no apparent main effect of the treatment . . .

The effects in columns 2 and 4 (the models without an interaction term) suggest that genetic risk scores for negative affectivity decrease the probability of turnout, although these effects do not reach conventional levels of significance for Genetic Risk Index 1. This provides some qualified support for our first hypothesis . . . The interaction terms in Table 3 are both negative: the interaction term with Genetic Risk Score 1 is significant at the $p < .05$ level and that with Genetic Risk Score 2 is significant at the $p < .10$ level . . . This confirms our proposition . . ."

Discovered: the genetic secret of a happy life

BY JEREMY LAURANCE, HEALTH EDITOR | FRIDAY 06 MAY 2011

 351 68 1 1 [PRINT](#) | [EMAIL](#) | [AAA](#)

Latest in Science

[Dinosaur wind 'altered climate'](#)[How the war on terror could solve art's most enduring mysteries](#)[Trending: It's easy to get drunk with WAHH Quantum Sensations](#)[Antarctica team endures extremes in hope of showing the way to Mars](#)[Defect risk 'higher for assisted births'](#)[10 scientific facts about flavour](#)['Rinrin evo' operation helps blind man to see](#)

Some people are born happy, scientists say. Researchers have identified a "happiness gene" that makes people more likely to feel satisfied with their lives. Their sunny disposition is an accident of birth, at least in part.

Those who carry the less efficient version of the gene are more likely to be pessimistic. Their tendency to see the glass half empty is equally a part of their inheritance.

The finding is the first to demonstrate a link between the gene, called 5-HTT, and satisfaction. People with the long version are more likely to be cheerful while sulkiness is the default position of those with the short version. Knowing which version of the gene they carry may help people improve their mood.

Suggested Topics

[Biology](#)[Genetics](#)[Chemicals](#)[The Brain](#)

"Discovered: the genetic secret of a happy life"

From the news article:

"Researchers have identified a 'happiness gene' that makes people more likely to feel satisfied with their lives . . . The finding is the first to demonstrate a link between the gene, called 5-HTT, and satisfaction . . . Those with two long versions of the gene were 17 per cent more likely to say they were very satisfied. . . ."

From the research article by De Neve, Fowler, and Frey:

"Having one or two alleles . . . raises the average likelihood of being very satisfied with one's life by 8.5% and 17.3%, respectively."

Undiscovered ...

From the *text* of the research article:

"Having one or two alleles . . . raises the average likelihood of being very satisfied with one's life by 8.5% and 17.3%, respectively."

From the *tables*:

- ▶ 46% of people who had two copies of the gene described themselves as satisfied and 41% described themselves as very satisfied. The corresponding percentages for those with no copies were 44% and 37%.
- ▶ Reported maximum difference is 4 percentage points (and not statistically significant), *not* 17%.

Statistical significance is not enough!

Journal's Paper on ESP Expected to Prompt Outrage

By BENEDICT CAREY

Published: January 5, 2011

One of psychology's most respected journals has agreed to publish a paper presenting what its author describes as strong evidence for extrasensory perception, the ability to sense future events.

Enlarge This Image

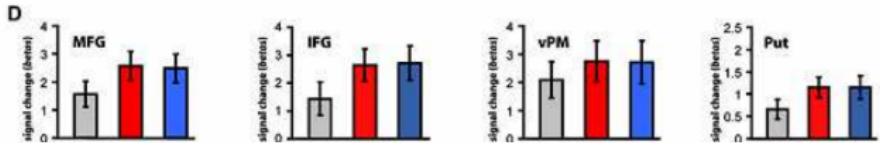
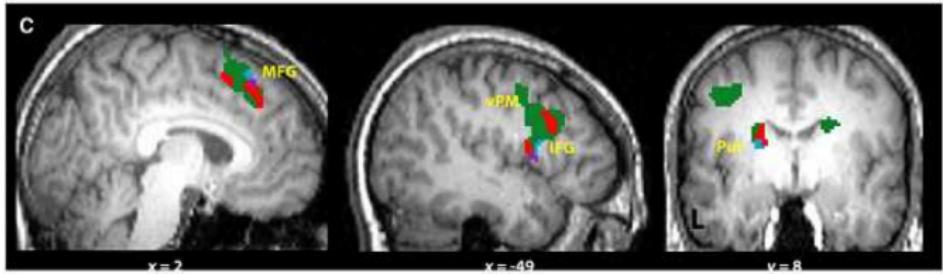
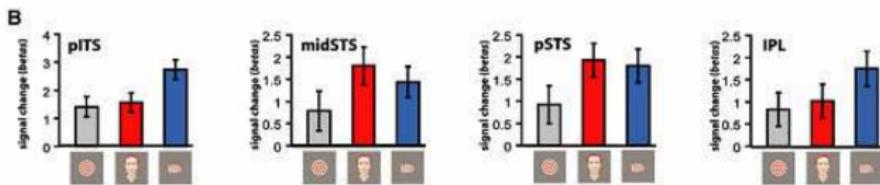
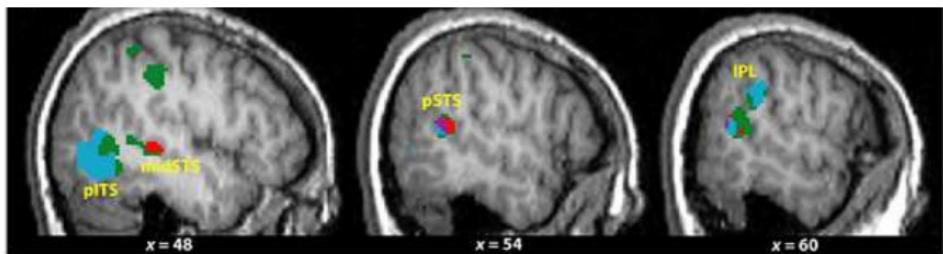
Heather Ainsworth for The New York Times

Work by Daryl J. Bem on extrasensory perception is scheduled to be published this year.

The decision may delight believers in so-called paranormal events, but it is already mortifying scientists. Advance copies of the [paper](#), to be published this year in *The Journal of Personality and Social Psychology*, have circulated widely among psychological researchers in recent weeks and have generated a mixture of amusement and scorn.

The paper describes nine unusual lab experiments performed over the past decade by its author, [Daryl J. Bem](#), an emeritus professor at Cornell, testing the ability of college students to accurately sense random events,

Finding statistical significance with big data

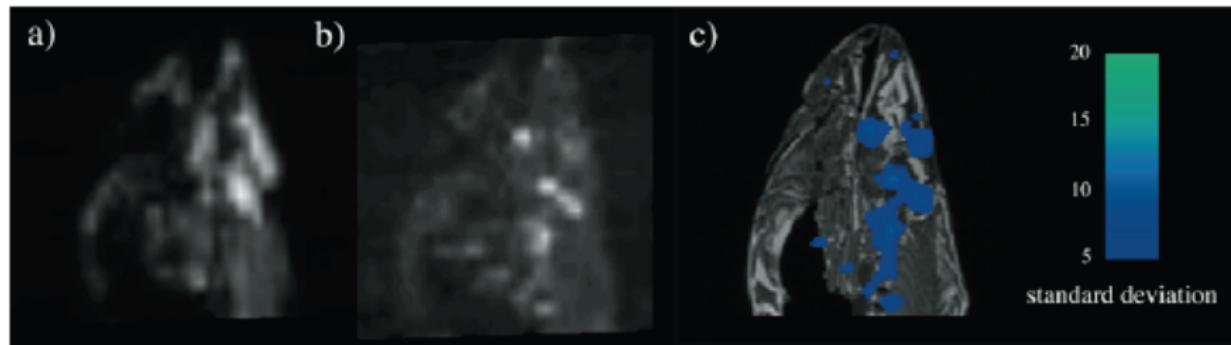


The “statistical significance filter”

- ▶ Vul, Harris, Winkelman, Pashler:
 - ▶ Correlations reported in medical imaging studies are commonly overstated because researchers select the highest values
 - ▶ These *statistical* problems are leading to *scientific* errors
- ▶ Statistical corrections for multiple comparisons do *not* solve the problem
- ▶ Discussion in *Perspectives in Psychological Science* (2009)

Neural activity in a dead fish

VOXELWISE VARIABILITY



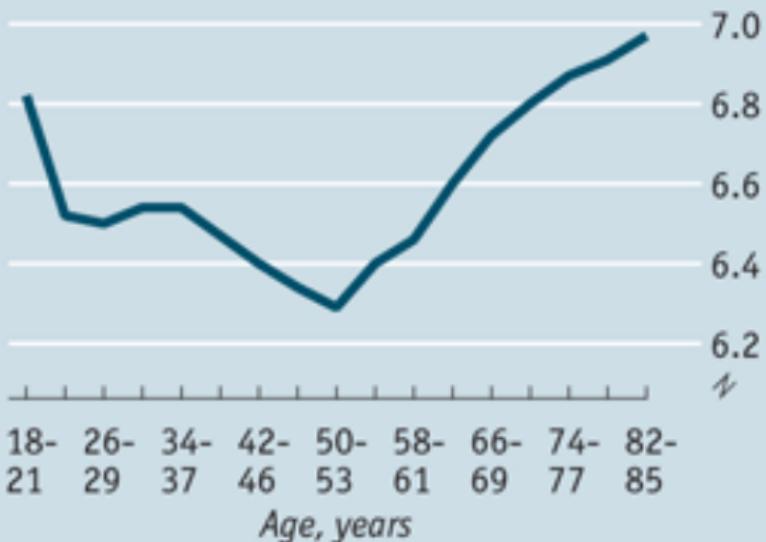
To examine the spatial configuration of false positives we completed a variability analysis of the fMRI timeseries. On a voxel-by-voxel basis we calculated the standard deviation of signal values across all 140 volumes.

We observed clustering of highly variable voxels into groups near areas of high voxel signal intensity. Figure 2a shows the mean EPI image for all 140 image volumes. Figure 2b shows the standard deviation values of each voxel. Figure 2c shows thresholded standard deviation values overlaid onto a high-resolution T₁-weighted image.

Happiness and life satisfaction

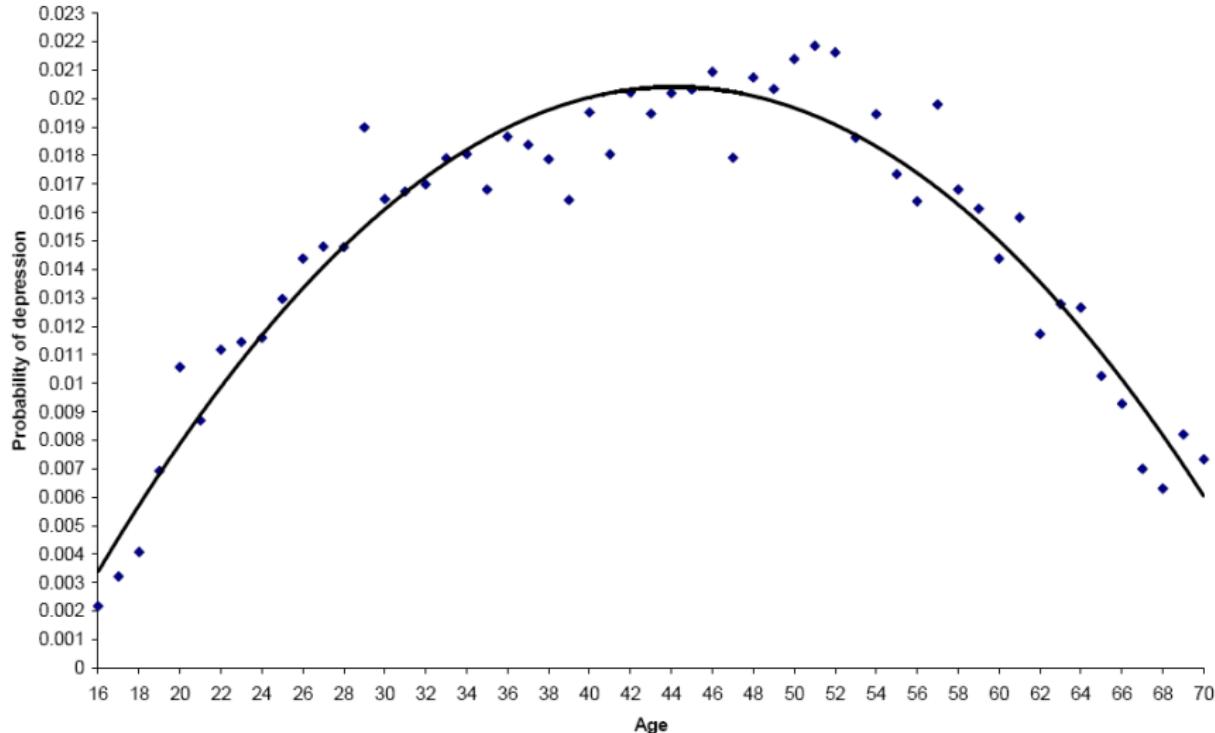
The U-bend

Self-reported well-being, on a scale of 1-10



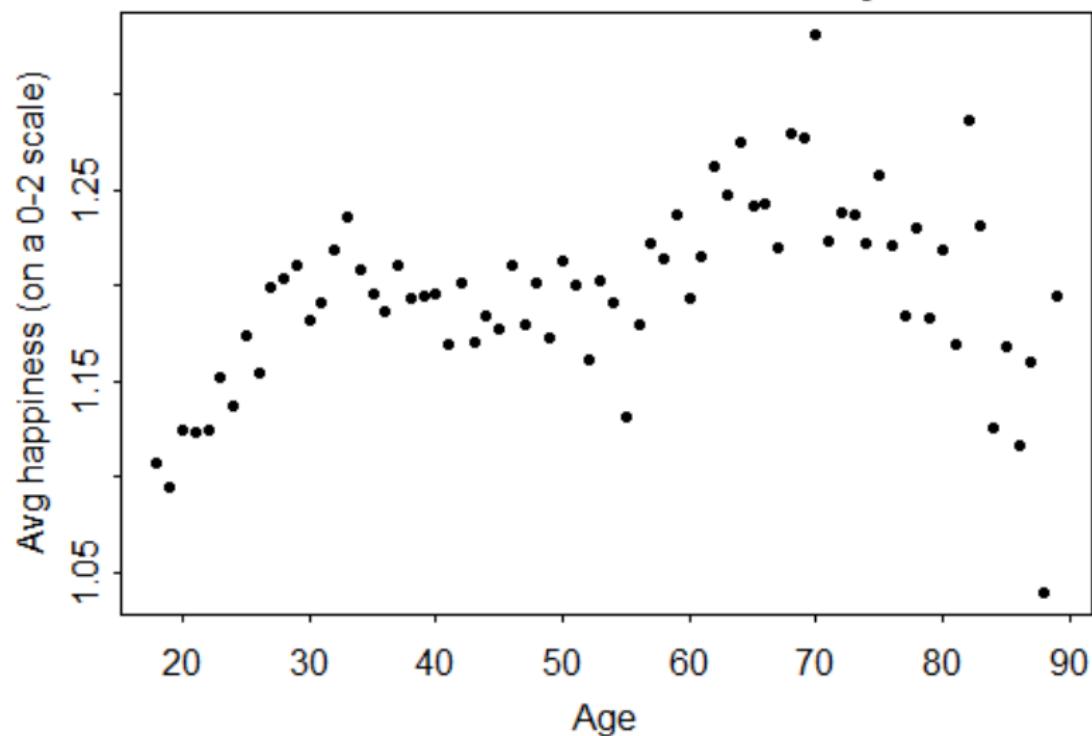
Source: PNAS paper: "A snapshot of the age distribution of psychological well-being in the United States" by Arthur Stone

Data!



More data

Average happiness as a function of age, from General Social Survey



The Perils of Pooling

Arthur “not David” Brooks in the *New York Times*:

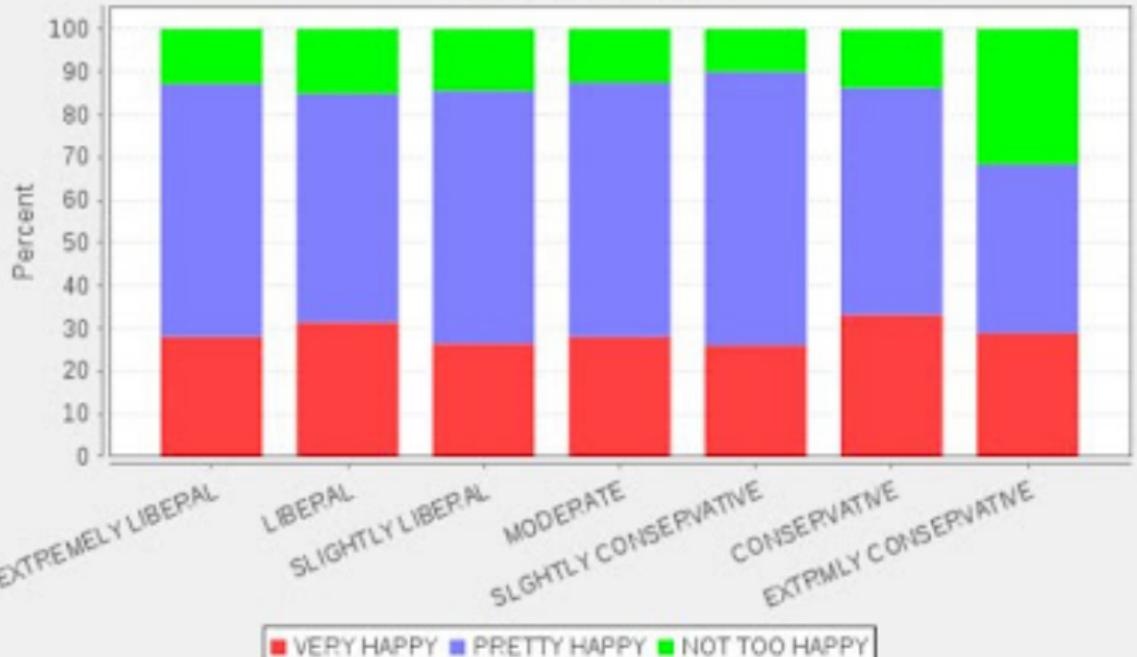
“People at the extremes are happier than political moderates. . . none, it seems, are happier than the Tea Partiers . . . ”

Jay Livingston (sociology, Montclair State University) looks up the data in the General Social Survey . . .

"None, it seems, are happier than the Tea Partiers . . ." ??

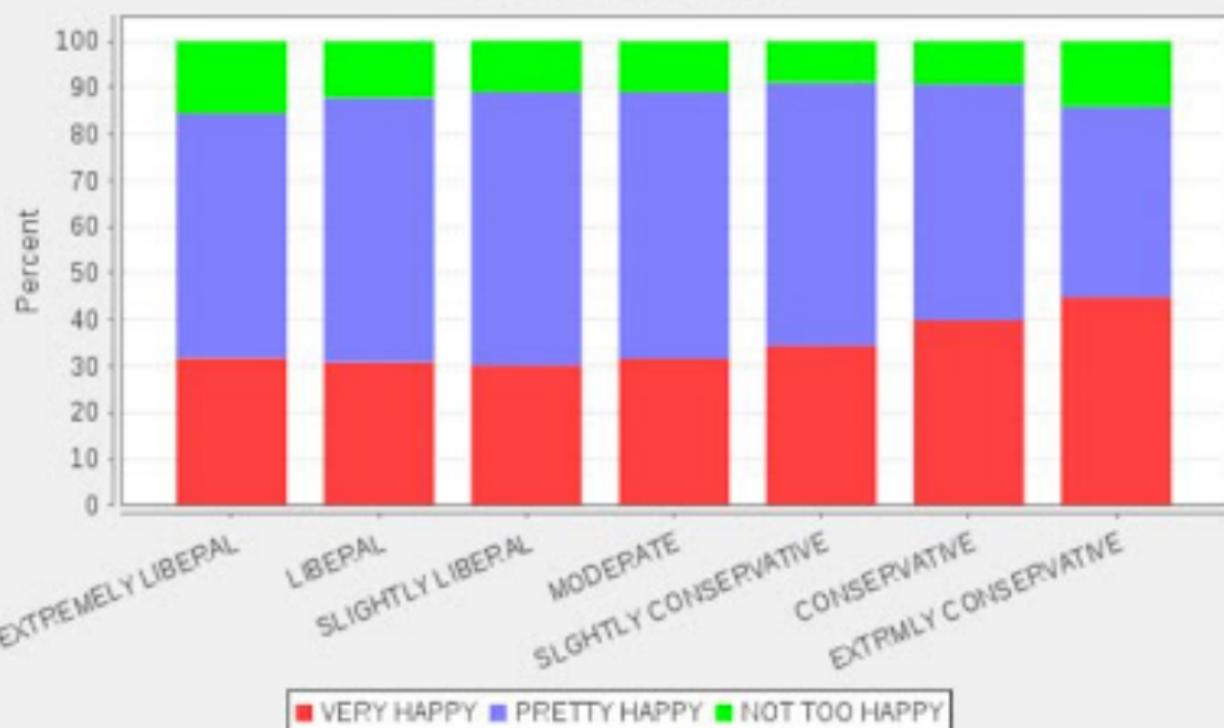
Chart for YEAR = 4(2009-2010)

GENERAL HAPPINESS BY THINK OF SELF AS LIBERAL OR CONSERVATIVE



Pooling, 1972–2010

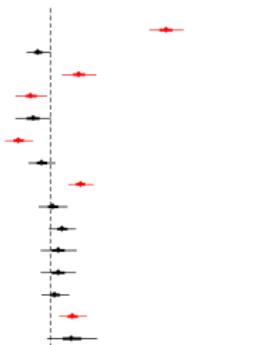
GENERAL HAPPINESS BY THINK OF SELF AS LIBERAL OR CONSERVATIVE



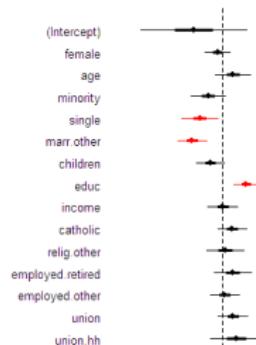
When to aggregate or break up the data?

- ▶ Always always always a concern
- ▶ A “big data” example:

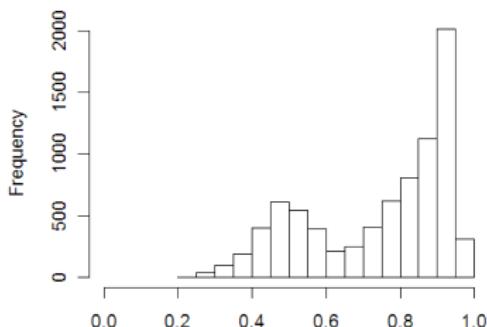
2006 turnout model, demographics only,
(standardized vars), N=8000



2006 turnout model, w/vote history,
(standardized vars), N=8000



Histogram of fitted probabilities,
turnout model w/vote history



Trying to understand the world using “stylized facts”

- ▶ Problems with is-it-there-or-is-it-not models of correlations and effects
- ▶ Problems with the concept of “false positives”
- ▶ Accepting variation (as distinct from measurement error)
- ▶ Don’t fool yourself!

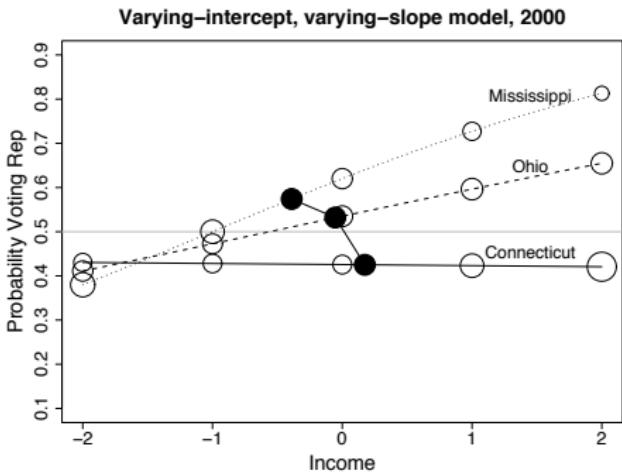
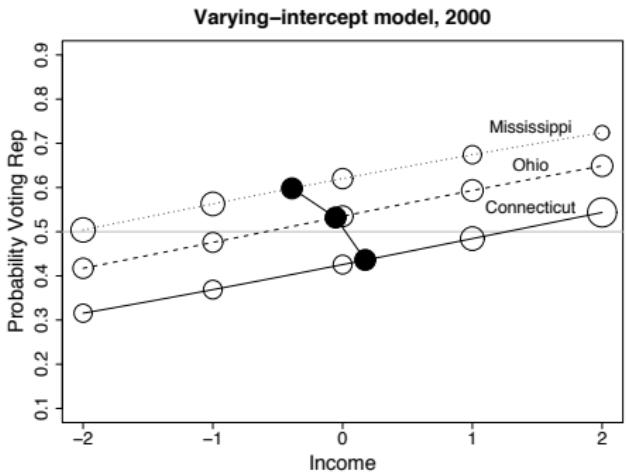
What should we do instead?

- ▶ Don't estimate effects in isolation
- ▶ Instead, build a model
- ▶ A couple examples from my own research ...

Red state, blue state, rich state, poor state

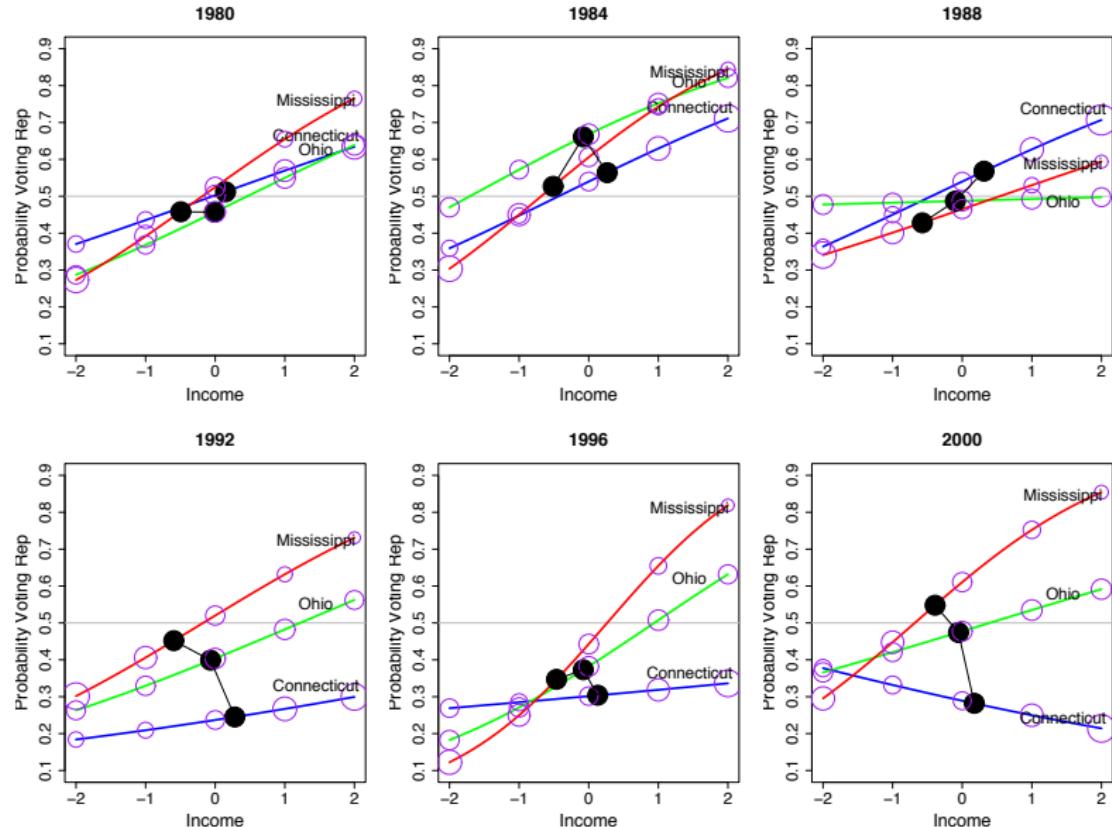
- ▶ Richer *voters* favor the Republicans, *but*
- ▶ Richer *states* favor the Democrats
- ▶ Hierarchical logistic regression: predict your vote given your income and your state (“varying-intercept model”)

Varying-intercept model, then model criticism, then varying-slope model



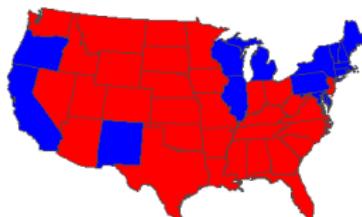
In any given state, the estimates would not be statistically significant!

3-way interactions!



Adding another factor: The inference . . .

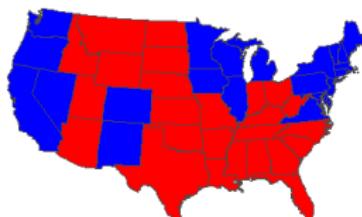
State winners in 2008 (rich voters only)



State winners in 2008 (rich Whites only)



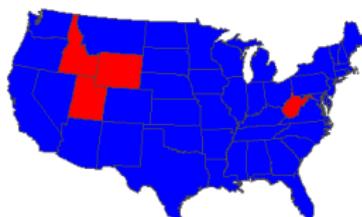
State winners in 2008 (middle-income voters)



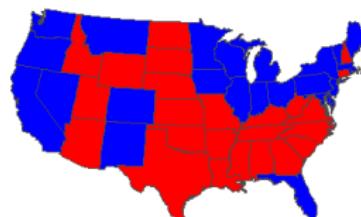
State winners in 2008 (middle-income Whites)



State winners in 2008 (poor voters only)



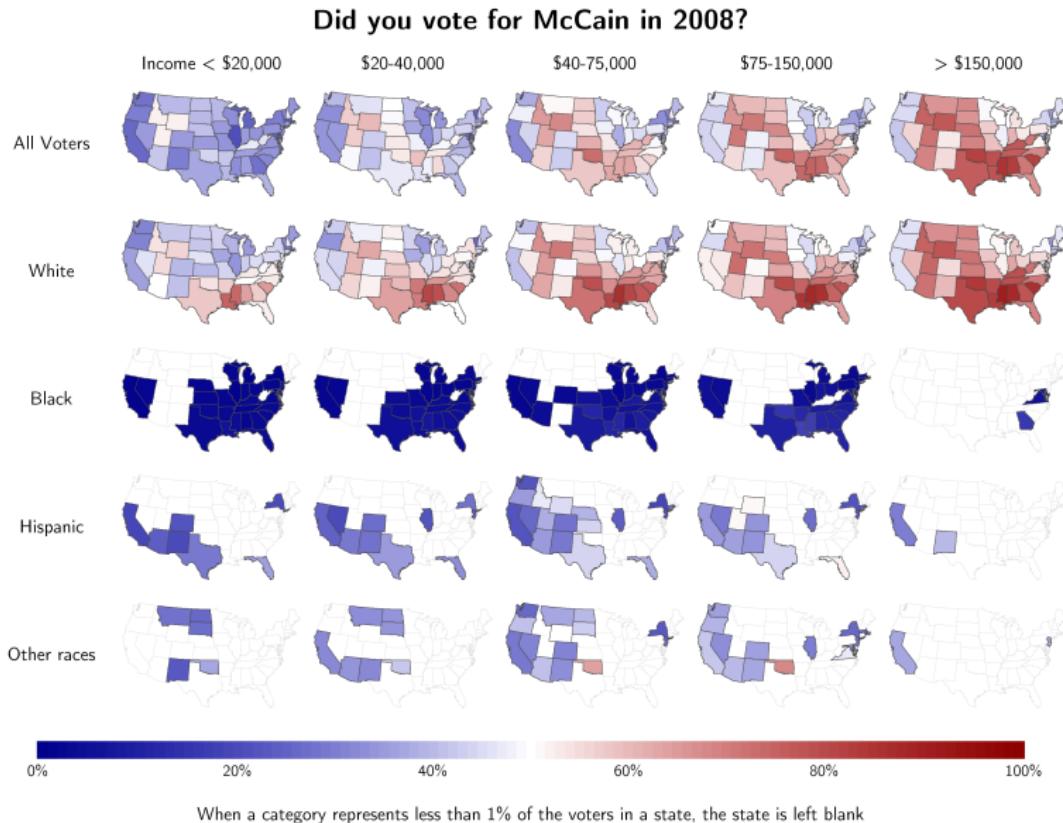
State winners in 2008 (poor Whites only)



...and the refutation!

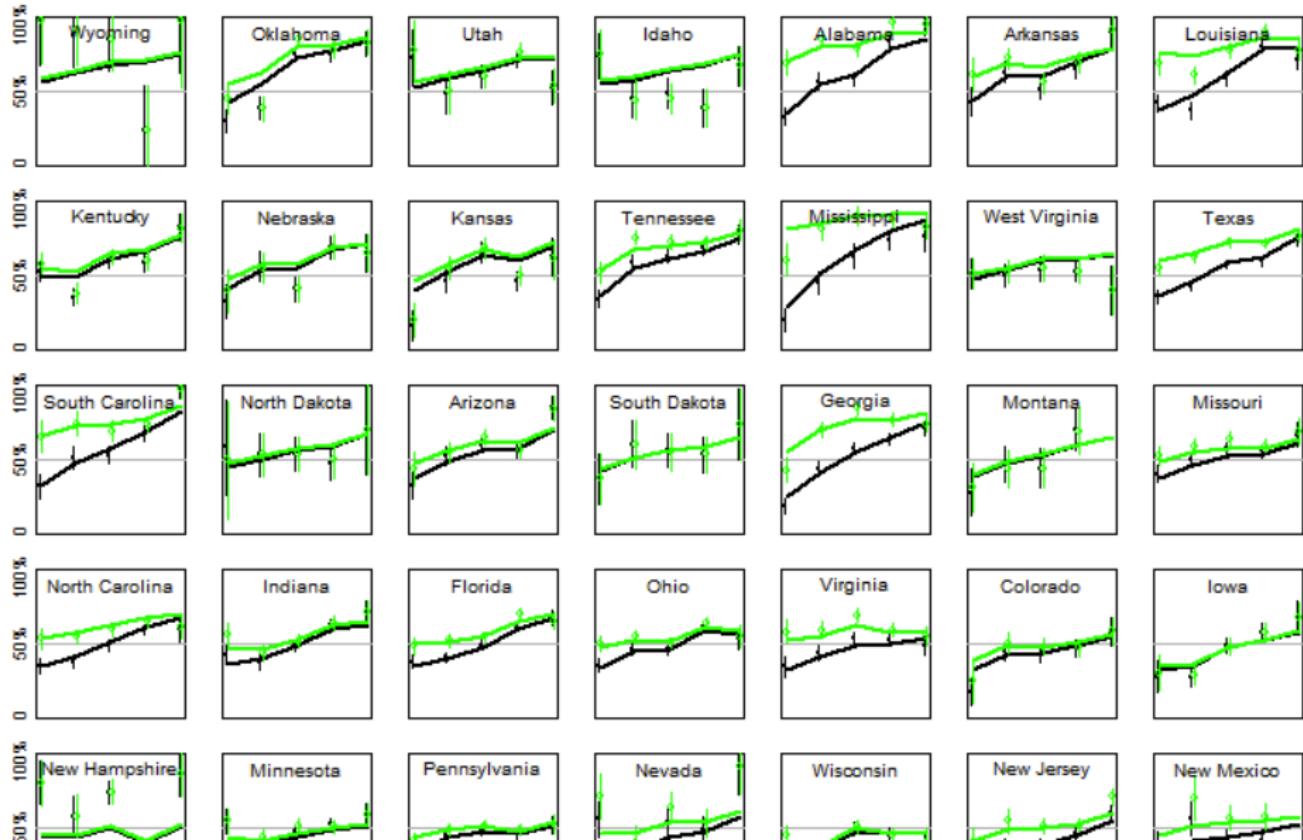
- ▶ Criticisms from the blogger “Daily Kos”: “While Gelman claims only the under-\$20K white demo went for Obama, the results were far different. Per the exit poll — real voters — Obama won all whites: 54–45 percent for those making under \$50K, and 51–47% for those making over \$50K. . . . New Hampshire is solidly Blue unlike Gelman’s maps, 58–40 — one of the most obvious misses in Gelman’s analysis. . . .”
- ▶ Traditional statistical “conservatism” will be no defense here!

After improving the model



A graph we made to study and criticize our inferences

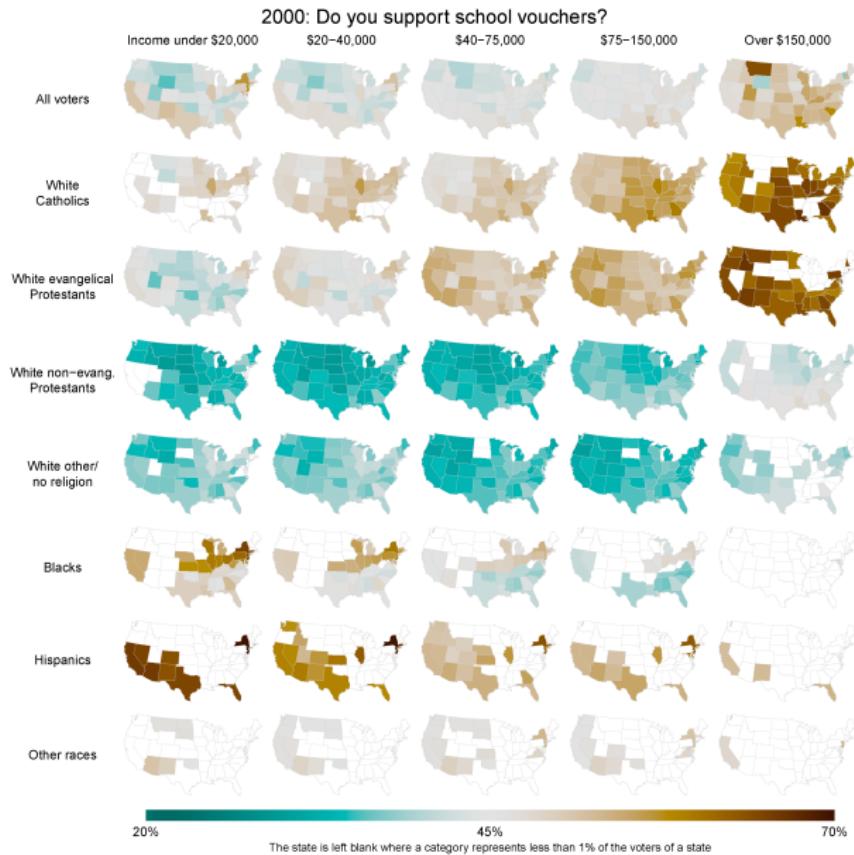
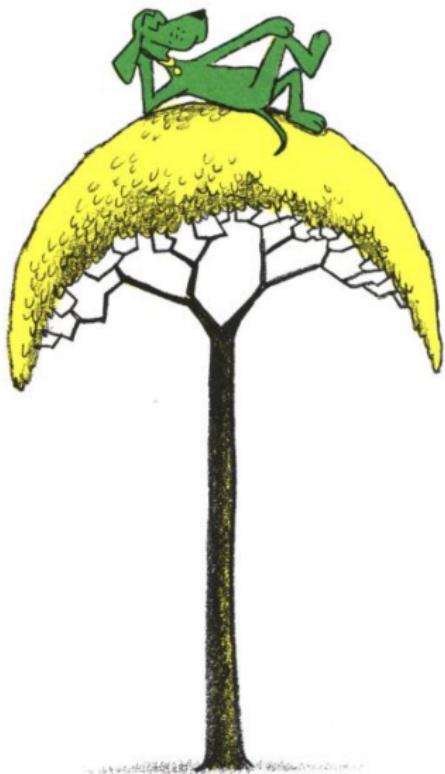
2008 election: McCain share of the two-party vote in each income category within each state among all voters (black) and non-Hispanic whites (green)



Inference for small effects

- ▶ Three kinds of selection bias:
 - ▶ False “statistical significance” via multiple comparisons
 - ▶ When using small samples to study small effects, any statistically significant finding is *necessarily* a huge overestimate
 - ▶ Incentives (in science and the media) to report dramatic claims
- ▶ How to do it right?
 - ▶ Don’t study factors (e.g., beauty) in isolation
 - ▶ Place them in a larger model

Big Data and Big Model



Data don't always “speak for themselves”

