Truth or Truthiness

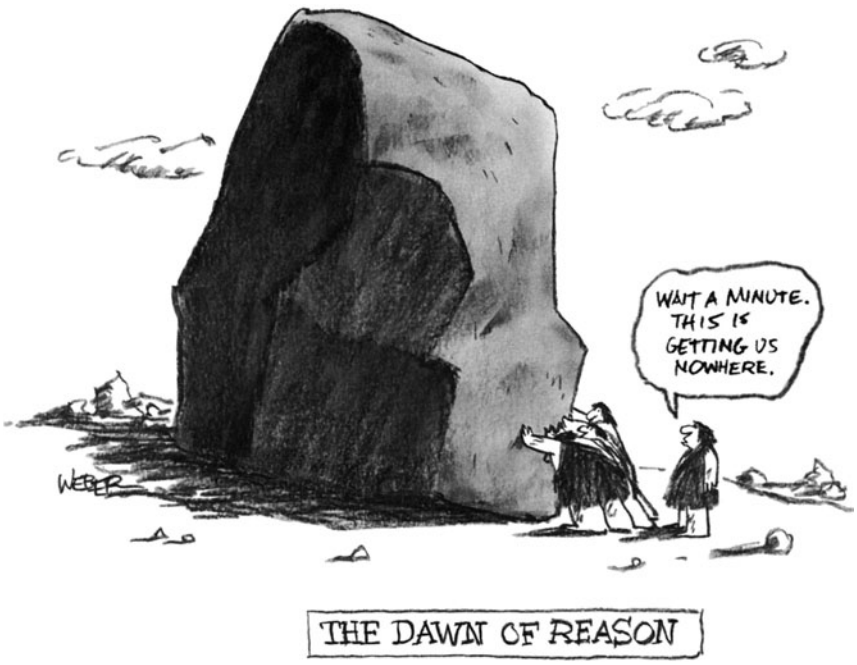
Teacher tenure is a problem. Teacher tenure is a solution. Fracking is safe. Fracking causes earthquakes. Our kids are overtested. Our kids are not tested enough.

We read claims like these in the newspaper every day, often with no justification other than "it feels right." How can we figure out what *is* right?

Escaping from the clutches of truthiness begins with one simple question: "what's the evidence?" With his usual verve and flair, Howard Wainer shows how the skeptical mind-set of a data scientist can expose truthiness, nonsense, and outright deception. Using the tools of causal inference he evaluates the evidence, or lack thereof, supporting claims in many fields, with special emphasis in education.

This wise book is a must-read for anyone who's ever wanted to challenge the pronouncements of authority figures and a lucid and captivating narrative that entertains and educates at the same time.

Howard Wainer is a Distinguished Research Scientist at the National Board of Medical Examiners who has published more than four hundred scholarly articles and chapters. This is his twenty-first book. His twentieth book, *Medical Illuminations: Using Evidence, Visualization and Statistical Thinking to Improve Healthcare* was a finalist for the Royal Society Winton Book Prize.



THE DAWN OF REASON

Robert Weber The New Yorker Collection/Cartoon Bank, reproduced with permission.

Truth or Truthiness

Distinguishing Fact from Fiction by Learning to Think Like a Data Scientist

Howard Wainer

National Board of Medical Examiners

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107130579

First published 2016

Printed in the United States of America

A catalog record for this publication is available from the British Library.

ISBN 978-1-107-13057-9 Hardback

To Linda
and
Sam & Jennifer
and
Laurent, Lyn & Koa

Annotated Table of Contents

Preface and Acknowledgments *page* xiii

Introduction 1

Section I Thinking Like a Data Scientist 11

1 How the Rule of 72 Can Provide Guidance to Advance Your Wealth, Your Career, and Your Gas Mileage 14
Exponential growth is something human intuition cannot comprehend. In this chapter we illustrate this with several examples drawn from history and current experience. Then we introduce a simple rule of thumb, often used to help financial planners tame the cognitive load of exponential growth, and show how it can be used more widely to help explain a broad range of other issues. The Rule of 72 illustrates the power of having such "rules" in your toolbox for use as the need arises.

2 Piano Virtuosos and the Four-Minute Mile 19
The frequency of truly extreme observations and the size of the sample of observations being considered are inexorably related. Over the last century the number of musical virtuosos has ballooned to include copious numbers of high school students who can perform pieces that would have daunted all but the most talented artists in the past. In this chapter we find that a simple mathematical model explains this result as well as why a runner breaking the four-minute barrier in the mile has ceased to be newsworthy.

3 Happiness and Causal Inference 22
Here we are introduced to Rubin's Model for Causal Inference, which directs us to focus on measuring the effect of a cause rather than chasing a chimera by trying to find the cause of an effect. This reorientation leads us naturally to the random assignment-controlled experiment as a key companion to the scientific method. The power of this approach is illustrated by laying out how we can untie the Gordian knot that entangles happiness and performance. It also provides a powerful light that can be used to illuminate the dark corners of baseless claims.

4 Causal Inference and Death 29
The path toward estimating the size of causal effects does not always run smoothly in the real world because of the ubiquitous nuisance of missing data. In this chapter we examine the very practical situation in which unexpected events occur that unbalance carefully constructed experiments. We use a medical example in which some patients die inconveniently mid-experiment, and we must try to estimate a causal effect despite this disturbance. Once again Rubin's Model guides us to a solution, which is unexpectedly both subtle and obvious, once you get used to it.

5 Using Experiments to Answer Four Vexing Questions 43
Public education is a rich field for the application of rigorous methods for the making of causal inferences. Instead, we find that truthiness manifests itself widely within the discussions surrounding public education, the efficacy of which is often measured with tests. Thus it is not surprising that many topics associated with testing arise in which the heat of argument on both sides of the question overwhelms facts. We examine four questions that either have already been decided in courts (but not decisively) or are on the way to court as this chapter was being written.

6 Causal Inferences from Observational Studies: Fracking, Injection Wells, Earthquakes, and Oklahoma 61
It is not always practical to perform an experiment, and we must make do with an observational study. Over the past six years the

number of serious earthquakes in Oklahoma (magnitude 3.0 or more) has increased from less than two a year to almost two a day. In this chapter we explore how we can use an observational study to estimate the size of the causal effect of fracking and the disposal of wastewater through its high-pressure injection into the earth on seismicity. The evidence for such a connection is overwhelming despite denials from state officials and representatives of the petroleum industry.

7 Life Follows Art: Gaming the Missing Data Algorithm 72
A compelling argument can be made that the biggest problem faced by data scientists is what to do about observations that are missing (missing data). In this chapter we learn of what initially seem like completely sensible approaches for dealing with the inevitable missing data, yet they were being exploited to improperly game the system. It also illustrates what may be the most effective way to deal with such shenanigans.

Section II Communicating Like a Data Scientist 79

8 On the Crucial Role of Empathy in the Design of Communications: Genetic Testing as an Example 82
Graphical display is perhaps the most important tool data science possesses that allows the data to communicate their meaning to the data scientist. They are also unsurpassed in thence allowing the scientist to communicate to everyone else as well. By far, the most crucial attitude that anyone wishing to communicate effectively can have is a strong sense of empathy. In this chapter we discuss two different communications and show how the lessons learned from Princeton University's acceptance letter could be efficaciously applied in communicating the results of a genetic test for mutations that increase a woman's likelihood of cancer.

9 Improving Data Displays: The Media's and Ours 91
In the transactions between scientists and the general public, where influence flows in both directions, we see how advances in graphical display pioneered in the scientific literature were adopted by the mass media, and how advances in the media have been unfortunately slow to catch on among scientists.

10 Inside Out Plots 109

One of the greatest design challenges faced in visually displaying high-dimensional data (data involving more than just two variables) is the limitations of a two-dimensional plotting medium (a piece of paper or a computer screen). In this chapter we illustrate how to use inside out plots to reveal many of the secrets that might be contained in such data sets. The example we chose compares the performances of six baseball all-stars on eight variables.

11 A Century and a Half of Moral Statistics: Plotting Evidence to Affect Social Policy 122

Any data set that mixes geographic variables with other measures (like election outcomes by state or population per census tract) cries out for a map. Maps are the oldest of graphical displays with surviving examples from Nilotic surveyors in ancient Egypt and even older Chinese maps that long predate the Xia dynasty. A map is the obvious visual metaphor, using the two dimensions of the plotting plane to represent the geographic information. Only much later were other, nongeographic variables added on top of the geographic background. In this chapter we are introduced to Joseph Fletcher, a nineteenth-century British lawyer and statistician, who plotted such moral statistics as the prevalence of ignorance, bastardy, crime, and improvident marriages, on maps of England and Wales. The discussion of his works ranges broadly over what Fletcher did, why, and how his goal of increased social justice might have been aided through the use of more modern display methods.

Section III Applying the Tools of Data Science to Education 139

Public education touches everyone. We all pay for it through our local property taxes and almost all of us have partaken of it ourselves as well as through our children. Yet it is hard to think of any other area of such a broad-based activity that is so riddled with misconceptions borne of truthiness. In this section we will examine five different areas in which public opinion has been shaped by claims made from anecdotes and not from

evidence. Each chapter describes a claim and then presents widely available evidence that clearly refutes it. This section is meant as a consilience in which the methods introduced and illustrated in sections I and II are used to reinforce an attitude of skepticism while providing an evidence-based approach to assessing the likelihood of the claims being credible.

12 Waiting for Achilles 143
The educational system of the United States is regularly excoriated for the poor performance of its students and the disappointingly robust differences between the scores of black and white students. In this chapter we use evidence to clarify both of these issues and, in so doing, discover that the situation is not anywhere near as bleak as truthiness-driven critics would have us believe.

13 How Much Is Tenure Worth? 146
Critics of public education often lay a considerable portion of the blame for its shortcomings at the door of teacher tenure. In this chapter we trace the origins of tenure and provide evidence that eliminating it may be far more expensive and less effective than its critics suggest.

14 Detecting Cheating Badly: If It Could Have Been, It Must Have Been 152
Whenever tests have important consequences there is always the possibility of cheating. To limit cheating, student performances are scrutinized and severe punishments are sometime meted out. In this chapter we describe two instances in which the fervor of the investigation outstripped the evidence supporting the claim of a security infraction.

15 When Nothing Is Not Zero: A True Saga of Missing Data, Adequate Yearly Progress, and a Memphis Charter School 161
Increasingly often, the efficacy of schools is determined by the test scores earned by their students. In this chapter we learn of a charter school in Memphis that was placed on probation because the average scores of its students were too low. Unfortunately this apparent deficiency was not due to the school but rather to the city's treatment of missing data.

16 Musing about Changes in the SAT: Is the College Board Getting Rid of the Bulldog? 167
Modern college entrance exams in the United States have existed for about ninety years, and in that time, changes in the exams, their scoring, and their use have been made steadily. In this chapter we use evidence and statistical thinking to focus our discussion of three changes to the SAT that have been recently announced by the College Board. Two of them are likely to have almost no effect, but the third is a major change. I hypothesize why these particular changes were selected and conclude that the College Board might have been guided by the strategy developed by Dartmouth's former president John Kemeny when he led the school to coeducation in the 1970s.

17 For Want of a Nail: Why Worthless Subscores May Be Seriously Impeding the Progress of Western Civilization 175
In the 2010 U.S. Census, it cost about forty dollars for each person counted. This seems like an extravagant amount because changes in the population of the United States can be accurately estimated by noting that it increases by about one person every thirteen seconds. Yet the price is sensible because of the many small area estimates that Census data provide. In this chapter we examine the costs of testing in this same light and conclude that the opportunity costs of too long tests are likely extensive enough that they may be impeding progress in a serious way.

Section IV Conclusion: ~~Don't~~ Try This at Home 187

Bibliography 195
Sources 203
Index 205

Preface and Acknowledgments

There have been many remarkable changes in the world over the last century, but few have surprised me as much as the transformation in public attitude toward my chosen profession, statistics – the science of uncertainty. Throughout most of my life the word *boring* was the most common adjective associated with the noun *statistics*. In the statistics courses that I have taught, stretching back almost fifty years, by far the most prevalent reason that students gave for why they were taking the course was "it's required." This dreary reputation nevertheless gave rise to some small pleasures. Whenever I found myself on a plane, happily involved with a book, and my seatmate inquired, "What do you do?" I could reply, "I'm a statistician," and confidently expect the conversation to come to an abrupt end, whereupon I could safely return to my book. This attitude began to change among professional scientists decades ago as the realization grew that statisticians were the scientific generalists of the modern information age. As Princeton's John Tukey, an early convert from mathematics, so memorably put it, "as a statistician, I can play in everyone's backyard."

Statistics, as a discipline, grew out of the murk of applied probability as practiced in gambling dens to wide applicability in demography, agriculture, and the social sciences. But that was only the beginning. The rise of quantum theory made clear that even physics, that most deterministic of sciences, needed to understand uncertainty. The health professions joined in as *Evidence-Based Medicine* became a proper noun. Prediction models combined with exit polls let us go to sleep early with little doubt about

election outcomes. Economics and finance was transformed as "quants" joined the investment teams and their success made it clear that you ignore statistical rigor in devising investment schemes at your own peril.

These triumphs, as broad and wide ranging as they were, still did not capture the public attention until Nate Silver showed up and starting predicting the outcomes of sporting events with uncanny accuracy. His success at this gave him an attentive audience for his early predictions of the outcomes of elections. Talking heads and pundits would opine, using their years of experience and deeply held beliefs, but anyone who truly cared about what would happen went to FiveThirtyEight, Silver's website, for the unvarnished truth.

After Nate Silver my life was not the same. The response to my declaration about being a statistician became "Really? That's way cool!" The serenity of long-distance air travel was lost.

As surprising as this shift in attitudes has been, it is still more amazing to me how resistant so many are to accepting evidence as a principal component in deciding between conflicting claims. I chalk this up to three possible reasons:

1. A lack of understanding of the methods and power of the Science of Uncertainty.
2. A conflict between what is true and what is wished to be true.
3. An excessive dimness of mind that prevents connecting the dots of evidence to yield a clear picture of likely outcome.

The first reason is one of my principal motivations in writing this book. The other was my own enthusiasm with this material and how much I want to share its beauty with others.

The second reason was reflected in Upton Sinclair's observation, "It is difficult to get a man to understand something, when his salary depends upon his not understanding it!" We have seen how senators from coal-producing states are against clean air regulations; how the National Rifle Association believes, despite all evidence (see Chapter 11), that more guns will lower the homicide rate; and how purveyors of coastal real estate believe that rising seas accompanying global warming are a pernicious myth.

The third reason is a late addition to my list, and would be unfair, if the observed behavior could be explained with reason two. But I was

forced to include it when, on Thursday, February 26, 2015, Senator Jim Inhofe (Republican from Oklahoma, who is Chairman of the Senate Environment and Public Works Committee) brought a snowball onto the floor of the senate as proof that reactions to evidence about global warming are hysterical, and the report that 2014 was the warmest year on record was anomalous. What could explain Senator Inhofe's statements? It could be (1), but as a senator he has been privy to endless discussions by experts with exquisite pedigrees and credentials, which anyone with any wit would be forced to acknowledge as credible. It could be (2) if, for example, his principal supporters were in the petroleum industry, whose future would be grimmer if the nation were to take seriously the role that the burning of such fuels has on global warming. I note that three of the five billionaires in the state of Oklahoma (Harold Hamm, George Kaiser, and Lynn Schusterman) owe their fortunes to the oil and gas industry. That being the case, it isn't surprising that Senator Inhofe might owe some allegiance to the welfare of their financial interests. What makes him a possible candidate for the third category is his apparent belief that his argument would burnish his national reputation rather than making him a punchline on newscasts and late-night TV. I am reminded of Voltaire's prayer "Dear God, make my enemies ridiculous." He knew that politicians can endure anything but the sorts of ridicule that renders them a joke. That Senator Inhofe would purposely place himself into such a position suggests including him in category (3).

Senator Inhofe is not alone on such a list. I would probably want to also include then-senator (now governor) Sam Brownback (Republican from Kansas), former governor Mike Huckabee (Republican from Arkansas), and Representative Tom Tancredo (Republican from Colorado) who, in a 2007 presidential debate all raised their hands to indicate their lack of belief in evolution. It isn't hard to find other possible candidates.[1]

[1] Michelle Bachmann, a six-term congresswoman from Minnesota comes immediately to mind for her avid support of the teaching of creationism in schools, whose appeal to fundamentalist members of her constituency would seem to provide evidence to place her more firmly in (2). Yet, the behavior she exhibited that earned a place on a number of "2013 most corrupt" lists (being under investigation by the Federal Election Commission, House Ethics Committee, and Federal Bureau of Investigation for violating campaign finance laws while running for president by improperly paying staff from her leadership Political Action Committee and using her campaign resources to promote her memoir) suggests she holds a belief in her own invincibility that makes (3) a more likely explanation

I thoroughly understand that no presentation, no matter how lucid, no matter how compelling, can have a direct effect on diminishing either (2) or (3). However, I have hopes that some indirect help can be had by improving the statistical literacy of the general population. The greater the proportion of the population that can detect deceptive, fact-free arguments and hence not be swayed by them, the less effective such arguments will become. I do not believe that this will result in those people whose arguments are based on truthiness changing to another approach. My hopes lie in an educated electorate choosing different people. Paraphrasing Einstein, "old arguments never die, just the people who make them."

I have recently become haunted by lasts. We are usually immediately aware of the firsts in our lives: our first car, first love, first child. We typically become aware of lasts only well after the event; the last time I spoke with my father, the last time I carried my son on my shoulders, the last time I hiked to the top of a mountain. Usually, at least for me, the realization that a last has occurred yields a sense of loss, a deep regret. Had I known it was the last time I would ever speak with my grandfather there are some things I would have liked to have spoken about. Had I known it was the last time I would see my mother I would have told her how much I loved her.

As you read this, I will be well passed the biblically prescribed life span of three-score and ten. Although this is surely my latest book, it might very well be my last. To minimize my future regrets, I want to be especially careful to thank all of those who have contributed both to this specific work and to the more general, and more difficult, task of shaping my thinking.

I begin with my employer, the National Board of Medical Examiners, which has been my professional home since 2001 and has provided a place of peace, quiet, and intellectual stimulation. The modern character of the National Board, a century-old institution, has largely been set by Donald Melnick, its longtime president, whose vision of the organization included room for scholarship and basic research. My gratitude to him and to the organization he leads is immense.

Next, I must thank my colleagues at the National Board beginning with Ron Nungester, senior vice president, and Brian Clauser,

vice president, who have always provided support and a thoughtful response to any questions I might have had – both procedural and substantive. In addition, my colleagues Peter and Su Baldwin, Editha Chase, Steve Clyman, Monica Cuddy, Richard Feinberg, Bob Galbraith, Marc Gessaroli, Irina Grabovsky, Polina Harik, Michael Jodoin, Peter Katsufrakis, Ann King, Melissa Margolis, Janet Mee, Arden Ohls, and Peter Scoles have all endured frequent visits in which I either inquired their opinion on some matter of concern to me at the moment, or sat through my explanations of one obscure thing or another. These explanations would typically continue until I decided that I had, at long last, understood what I was talking about. Sometimes this took a long time. I thank them all for their ideas and their forbearance.

Over the course of the past half-century many intellectual debts have accumulated to friends and colleagues who have taught me a great deal. I have neither space nor memory enough to include everyone, but with those limitations in mind, my major benefactors have been: Leona Aiken, Joe Bernstein, Jacques Bertin, Al Biderman, Darrell Bock, Eric Bradlow, Henry Braun, Rob Cook, Neil Dorans, John Durso, Steve Fienberg, Paul Holland, Larry Hubert, Bill Lichten, George Miller, Bob Mislevy, Malcolm Ree, Dan Robinson, Alex Roche, Tom Saka, Sam Savage, Billy Skorupski, Ian Spence, Steve Stigler, Edward Tufte, Xiaohui Wang, Lee Wilkinson, and Mike Zieky.

A very special thinks to David Thissen, my once student, longtime collaborator, and very dear friend.

Next, a mystery. I spent three years as a graduate student at Princeton University acquiring my academic union card. Under ordinary circumstance one would expect that those three years would not have a very different effect on my life than any number of other time periods of similar length. But, that does not seem to have been the case. On a regular basis in the forty-seven years since I left her campus I have been in need of guidance of one sort or another. And unfailingly, before I could flounder for too long, a former Tiger appeared and gave me as much assistance as was needed. Those most prominent in my continued education and productivity are:

John Tukey *39, Fred Mosteller *46, Bert Green *51, Sam Messick *56, Don Rubin ‘65, Jim Ramsay *66, Shelby Haberman ‘67, Bill Berg *67,

Linda Steinberg S*68 P07, Charlie Lewis *70, Michael Friendly *70, Dave Hoaglin *71, Dick DeVeaux '73, Paul Velleman *75, David Donoho '79, Cathy Durso '83, and Sam Palmer '07.

What's the mystery? A quick glance through this list shows that only about four or five were on campus at the same time I was. How did I get to meet the others? For example, I have collaborated with Hoaglin and Velleman for decades and we have yet to be able to recollect where, when, or how we met despite a fair amount of discussion. The only theory that seems to hold water is that Mother Princeton, has decided to take care of her progeny and has somehow arranged to do so. Whatever the mechanism, they and she have my gratitude.

Last, to the staff at Cambridge University Press, the alchemist who has transformed the digital bits and pieces of my manuscript into the handsome volume you hold in your hand now. *Primus inter pares* is my editor, Lauren Cowles, who both saw the value in what I was doing and insisted that I continue rewriting and revising until the result lived up to the initial promise that she had divined. She has my sincerest thanks. In addition, I am grateful for the skills and effort of copy editor Christine Dunn, indexer Lin Maria Riotta and Kanimozhi Ramamurthy and her staff at Newgen Knowledge Works.