Recognizing Trust and Understanding As Twin Pillars of Statistical Ethics

Thomas R. Belin[1]
[1]UCLA Department of Biostatistics, 51-267 Center for Health Sciences,
Los Angeles, CA 90095-1772

Abstract
Accepting the imperative for professional ethics in the field of statistics to be codified, how might it be possible to encompass the scope and generality of what we do into a complete yet digestible set of guidelines? Drawing on reflections by leading statisticians on essential elements of our work, scientific insights regarding the human condition, and philosophical discourse on the ethics of interpersonal interactions, it is argued that trust and understanding are essential core principles that can serve as the basis for a test of whether a statistical approach is ethical. The framework's simplicity makes it easy to communicate, its generality gives it power, and its positive-sum appeal could be used to promote professional identity development around ethics. Adopting trust and understanding as twin pillars of statistical ethics thus offers great potential as a strategy to elevate statistical practice, to enhance the reputation of the field of statistics as a discipline, and to contribute to society both by advancing knowledge and by serving as a beacon for the highest standards of integrity.

Key Words: integrity, empathy, obligation, bearing witness, replicability, self-preservation

1. The Ethical Obligations of Professional Statisticians

1.1 What Defines Ethics in the Statistics Profession?

Offering a personal reflection during his tenure as President of the American Statistical Association in 2004, Brad Efron commented:

> I still remember how difficult it was for me, an undergraduate mathematician, to understand why something like ANOVA made sense in analyzing real scientific problems. NFL cornerbacks learn the difficult skill of running backwards; we have to learn to think backwards, from the specific instance to the general rule. Along the way we also learn proper respect for the power of randomness to confuse the human mind. Statisticians are good at spotting patterns hidden in random noise, and even better at not being fooled by apparent patterns.

In a discipline so abstract that "thinking backwards" winningly characterizes its substance and so broad that its areas of application know no boundary, how can we encompass what we mean by "professional ethics" in 20 words or less?

This essay grows out of my involvement in the American Statistical Association's Committee on Professional Ethics (ASA-CPE), which I joined as a member in 2014. My goal is to offer an answer to the question of what we mean by professional ethics in statistics that is anchored by logic and reason, grounded in ideas put forward by great thinkers across the ages, and informed by current scientific insights. I hope that my conclusion, namely that the mission of the statistics profession is to cultivate trust and understanding, will facilitate communication about ethics in a non-technical way, will advance efforts by the ASA-CPE to promote professional identity development around ethics, and will reinforce the importance of accountability in all walks of life.

1.2 Avoiding Painful Consequences in a Social Context

Efron's mention of "the power of randomness to confuse the human mind," evoking the potentially painful consequences of confusion, relates to the way philosopher Baruch de Spinoza linked ethics to the human instinct for self-preservation. In his book *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*, neuroscientist Antonio Damasio wrote:

> Paraphrased in deeply American terms I would rewrite Spinoza's proposition as follows: I hold these truths to be self-evident, that all humans are created such that they tend to preserve their life and seek well-being, that their happiness comes from the successful endeavor to do so, and that the foundation of virtue rests on these facts. … The biological reality of self-preservation leads to virtue because in our inalienable need to maintain ourselves we must, of necessity, help preserve *other* selves. If we fail to do so we perish and are thus violating the foundational principle, and relinquishing the virtue that lies in self-preservation. The secondary foundation of virtue then is the reality of a social structure and the presence of other living organisms in a complex system of interdependence with our own organism. We are in a bind, literally, in the good sense of the word. The essence of this transition can be found in Aristotle, but Spinoza ties it to a biological principle--the mandate for self-preservation.

Damasio elaborated on the role of a social structure in this way:

> Good actions are those that, while producing good for the individual via the natural appetites and emotions, *do not harm other individuals*. The injunction is unequivocal. An action that might be personally beneficial but would harm others is not good because harming others always haunts and eventually harms the individual who causes the harm. Consequently such actions are evil. … I interpret Spinoza to mean that the system constructs ethical imperatives based on the presence of mechanisms of self-preservation in each person, but mindful of social and cultural

> elements as well. Beyond each self there are *others* as individuals and as social entities, and their own self-preservation, e.g., their appetites and emotions, must be taken into consideration.

1.3 A Principle for Statistical Practice

In one of his many contributions to our field, Don Rubin pinpointed the essence of the crucial social dynamic. Noting that realities of interpersonal communication induce certain imperatives due to the nature of human psychology, Rubin enunciated the following principle:

> Statisticians have an obligation to provide the kinds of answers clients will assume are being provided.

Along with other attractive implications, this idea appeals to our core conviction that statisticians must ask the right kinds of questions: What are the assumptions underlying the analysis? What units in the analysis are being viewed as exchangeable? What are the mechanisms through which certain values are observed and included in the analysis while certain other values are not observed or not included? And what assumptions might our clients reasonably be making on matters such as these that we know to be of crucial importance when looking for patterns in data?

As a concise (17-word) statement about the ethical obligations of statisticians that aims to unify all of these themes, I propose: "***Ethics in the field of statistics is all about cultivating trust and understanding in interpreting available data***."

The framework proposed here suggests a thought experiment when ethical dilemmas arise, based on the question, "What are the implications of different alternatives for trust and understanding?"

2. The Role of Statistics in the World of Science

As a profession, the discipline of statistics cultivates trust and understanding by letting facts speak for themselves and "bear witness" in a manner consistent with logic, reason, and principled scientific methods. Observation and replication unite to cultivate knowledge and self-actualization.

Statisticians thrive on the principle, "Draw distinctions where there are important distinctions to be drawn." This notion first came to my attention through the book *The Dreams of Reason* by physicist and human rights activist Heinz Pagels. Grappling with the question of whether there is a meaningful distinction between the mind and the brain as it relates to the possibility of developing artificial intelligence, Pagels initially thought the concepts should be distinguished, but then he wondered whether his judgment was a cultural artifact of a tendency in Western philosophical thought to dichotomize. This concern led

him on a multi-year odyssey to learn more about Eastern religious traditions that embrace a unified notion of mind and body, but he ultimately came back to his original "Western" view on the utilitarian grounds that distinguishing the mind from the brain is useful. In the science of statistics, any choice for how to integrate concepts of interest into queries, item measurements into composite indices, or units of analysis into categories or subgroups can be scrutinized in the same way in terms of the utility of doing so.

The science of statistics offers strategies for navigating complexity and managing uncertainty in the world around us. To someone born after 1960, it might seem surprising that there ever was controversy about whether smoking causes lung cancer, but what is now regarded as common knowledge rests on the careful application of statistical reasoning to rule out alternative explanations for the statistical association between smoking and lung cancer. Ruling out chance variation as an explanation of observed patterns of association is a central activity of statistical science, but strategies to rule out other explanations, some of which might be challenging to exclude given limitations in available data, also fall within the purview of statistical science. As it relates to regulation of smoking, one of the candidate explanations proposed to explain the association between smoking and lung cancer was a genetic scenario where one gene predisposes people to smoking and another nearby gene that predisposes people to lung cancer. As summarized by Mitchell Gail in his Presidential Address to the American Statistical Association summarized in an article entitled "Statistics in Action," the work of Jerome Cornfield and colleagues at the National Cancer Institute elegantly provided a convincing rejoinder by assuming such a pair of genes existed and showing that the degree of conjunction required to explain the statistical association was incompatible with epidemiological data showing a sharp rise in lung cancer in the early 20th century, since the rise in lung cancer would have been far more gradual under the two-gene scenario.

This discussion underscores a practical challenge facing scientists generally and statisticians in particular, namely that the complexity of the world around us makes it hard to claim that a scenario should be dismissed out of hand just because it is put forward by a party with a conflict of interest. Conflicts of interest, which abound in the world around us, are best addressed through disclosure, at which point appropriately informed observers can arrive at considered judgments. The fact that tobacco companies hired public-relations specialists who developed the slogan "Doubt is our product" might naturally give rise to suspicion, but scientific reasoning often can provide even stronger rejoinders than simply calling into question the credibility of an opposing claim.

A useful characterization of the task of science emerged from a 1986 Supreme Court case known as *Edwards v. Aguillard* that considered the extent to which certain creationist theories could be taught in public-high-school biology classes. As summarized by Michael Shermer in his book *Why People Believe Weird Things*, an *amicus curiae* ("friend of the court") brief in the case submitted by

dozens of Nobel-prize-winning scientists and scientific organizations effectively enunciated the task of science in the following passages:

> Science is devoted to formulating and testing naturalistic explanations for natural phenomena. It is a process for systematically collecting and recording data in an effort to infer the principles of nature that best explain the observed phenomena. … The grist for the mill of scientific inquiry is an ever increasing body of observations that give information about underlying 'facts.' Facts are the properties of natural phenomena. The scientific method involves the rigorous, methodical testing of principles that might present a naturalistic explanation for those facts. … An explanatory principle that by its nature cannot be tested is outside the realm of science. … Science is not equipped to evaluate supernatural explanations for our observations; without passing judgment on the truth or falsity of supernatural explanations, science leaves their consideration to the domain of religious faith. … Even the most robust and reliable theory … is tentative. A scientific theory is forever subject to reexamination and—as in the case of Ptolemaic astronomy—may ultimately be rejected after centuries of viability.

Against this backdrop, the central role of statistics in the scientific enterprise is abundantly clear. Further support for embracing an evidence-based perspective grounded in statistical science is convincingly presented in the book *The Signal and the Noise* by Nate Silver (developer of the web site www.fivethiryeight.com that gained notice through its successful prediction of almost every state in the 2008 and 2012 U.S. presidential elections). Drawing on numerous examples where tragic consequences have accompanied biased and overconfident assessments, a central theme in the book is that decision-making in modern life relies on accurate prediction of natural phenomena using available data, where accuracy involves not just avoidance of bias but also correct characterization of uncertainty. This perspective aligns with what is now known as the Rubin Causal Model, which anchored my own training in statistical science and which has proven to be extremely useful in accommodating a variety of complexities by conceptualizing causal inference as fundamentally a missing-data problem, thereby providing a natural platform for representing uncertainty accurately.

3. Why Trust and Understanding Are So Important

3.1 The Potential for Audiences to be Misled

Interpretations of available data depend greatly on how the data were obtained, and the field of statistics also abounds in paradox, giving rise to unending opportunities for audiences to be misled by statistical summaries of available data. Consider, for example, an adversarial proceeding addressing an injury visited on a large number of people, too many to assess individually, and at issue is the aggregate damage suffered by the entire group. One statistician hired as an expert

witness might propose selecting a few hundred injured individuals, carefully evaluating the damages they sustained in monetary terms, obtaining an interval based on the sample mean plus or minus a multiple of the standard error of the sample mean, and multiplying by the number of people to obtain a range of values for the aggregate damages suffered by the group. Another statistician hired as an expert witness could argue that there might be an unmeasured lurking variable associated with observed damages suggesting that the proposed estimate is either systematically too low or too high.

What would a lay judge or a juror think? The analysis of the first statistician might be deemed reasonable, but what if the sample was obtained by identifying the individuals with the most visits to medical clinics rather than as a random sample? The statement about the possible existence of an unmeasured lurking variable might correctly be regarded as true, but what if it was then argued that the possibility of a lurking variable was grounds for dismissing any interval of plausible values, even if based on a random sample? As professionals, we know that subtle differences in premises can give rise to wildly different conclusions, and accordingly we need to recognize the threat to trust and understanding induced by the prospect that clients of ours might harbor false premises.

3.2 Trust as a Core Value

In particular, we must address the threat associated with misleading presentation of information that Mark Twain popularized when he wrote in reference to his writing output:

> I was deducing from the above [comparing his writing roughly 200,000 words for "Innocents Abroad" across 60 days in 1868, working from midnight to morning, to an average of 1,400 words per sitting of four or five hours in dictating his "Notes on 'Innocents Abroad'" in 1904] that I have been slowing down steadily in these thirty-six years, but I perceive that my statistics have a defect: three thousand words in the spring of 1868 when I was working seven or eight or nine hours at a sitting has little or no advantage over the sitting of to-day, covering half the time and producing half the output. Figures often beguile me, particularly when I have the arranging of them myself, in which case the remark attributed to Disraeli would often apply with justice and force: "There are three kinds of lies: lies, damned lies, and statistics."

The potential for unexpected connections to be funny is a mainstay of humor; the ability to cloak important connections in numerical summaries is a mainstay of rhetoric. But while false inferences can be amusing, they can also have devastating consequences. Reality is marked by predators in our midst who, through cleverly masked and troublingly manipulative power plays, dare to advance their self-interest at the expense of others by preying on human tendencies (such as those identified through the groundbreaking work of psychologists Amos Tversky and Daniel Kahneman) to get confused by variation

in data and to rely on irrational heuristics. In her book *Lying: Moral Choice in Public and Private Life*, philosopher and ethicist Sissela Bok underscored what is at stake in the imperative to avoid false fronts:

> I believe that we must at the very least accept as an initial premise Aristotle's view that lying is 'mean and culpable' and that truthful statements are preferable to lies in the absence of special considerations. … I would like, in the chapters to come, to refer to the 'principle of veracity' as an expression of this initial imbalance in our weighting of truthfulness and lying. It is not necessarily a principle that overrides all others, nor even the one most frequently appealed to. Nor is it, obviously, sufficient by itself—witness the brutal but honest regime or the tormentor who prides himself on his frankness. Rather, trust in some degree of veracity functions as a *foundation* of relations among human beings; when this trust shatters or wears away, institutions collapse.

In an earlier version of this essay, I alluded to the 1954 book *How to Lie with Statistics* by Darrell Huff, which sold over a million copies, but I was alarmed when Andrew Gelman pointed out to me that Huff was paid by tobacco-industry interests for testimony he gave to Congress in the mid-1960's that sought to ridicule the notion of a connection between cigarettes and disease. Don Rubin, who was both Andrew's and my Ph.D. advisor, has argued that the core principles of ethics in statistics matter more than the applied context, as in scenarios where weak statistical reasoning in the context of anti-tobacco lawsuits can undermine trust and understanding. I find these arguments compelling, not least because it was in the same spirit of intellectual independence that Rubin's Ph.D. advisor, William Cochran, who happened to be a smoker, played a central role in synthesizing available scientific evidence and applying statistical reasoning as a key author of the 1964 report *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the United States* that transformed worldwide perceptions regarding the adverse effects of smoking.

Ultimately, what remains especially alarming is not knowing. And what remains most appealing about statistics is their ability to advance knowledge. As Fred Mosteller offered in his now-famous comment on Mark Twain's quip, "It is easy to lie with statistics, but it is easier to lie without them." And as Xiao-Li Meng, a graduate-school classmate of Andrew Gelman's and mine, has suggested, a good test for judging the appropriateness of a prospective summary of information is to put yourself in the other person's shoes and ask how you would feel if you were on the receiving end of that summary.

3.3 Cultivating a Commitment to the Highest Standards of Integrity

To statisticians, a reputation for integrity must be earned. In 1999, the ASA Board of Directors adopted a set of "Ethical Guidelines for Statistical Practice" developed under the auspices of the ASA Committee on Professional Ethics as an

open statement of principle. An updated version of these guidelines (under review at the time of this writing in October 2015) is currently posted at the ASA-CPE web site at http://community.amstat.org/ethics/home, and the ASA-CPE has also developed a series of case studies for use in training of the next generation of statisticians.

Tiffany Cvrkel, a UCLA faculty colleague specializing in bioethics, helpfully characterizes ethics as a system of structured logic, providing guidance for decisions in the face of ambiguities. With two types of errors ever present in hypothesis testing, statistical reasoning is crucial to making decisions. And ethical considerations, which are intertwined with the ultimate political question, "Who decides?", can be crucial to framing the task at hand.

Clarifying the boundaries between science and value judgments is frequently helpful in this regard, as recognized in work by Constantine Frangakis, Don Rubin, and Junni Zhang on studies of quality of life in end-of-life settings. In such a context, one treatment might have both a survival advantage over another treatment as well as a quality-of-life advantage among those who would survive under either treatment, but the people who survive at the margin might have very poor quality of life (for example, they might be on a respirator), and if one were to compare average quality of life among survivors without adjusting for the survival advantage (which they propose to do using a novel statistical method known as "principal stratification" that was first described in the literature as a general problem-solving strategy in 2002), one could easily be misled into thinking that the treatment with a worse survival profile had a better quality-of-life profile. As in other settings, clarifying the scientific issues can also help clarify where value judgments are needed to make decisions, and vice versa.

As in the natural world, subtle differences can have large consequences. A willingness to draw fine distinctions aligns with Aristotle's proposition (along the way to a theory of causes) that luck is a special case of chance belonging to the moral world because it involves choice as well as with David Hume's insight that we are bound to recognize the reality of regular conjunctions of events but to remain skeptical in our interpretations based on our inability to perceive the apparent connection directly. Building on Francis Bacon's enunciation of the scientific method, John Stuart Mill's foundations of inductive reasoning, and Karl Popper's insistence that advancing knowledge depends on rejecting falsifiable theories since theories cannot be directly proven, the tradition of empirical science is anchored to intrinsically motivating sentiments that can be regarded as a foundation for ethics.

At the 2015 Joint Statistical Meetings in Seattle, ASA-CPE colleague Duane Steffey reported on joint work with Howard Hogan (my supervisor in a series of jobs from 1988 to 1991 at the U.S. Census Bureau and a longtime mentor who now chairs the ASA-CPE) reviewing the historical development of ethics-oriented thought in the field of statistics, calling particular attention to a 1952 commentary

by Theodore H. Brown in *The American Statistician* entitled "The Statistician and His Conscience" that was noted for its seemingly timeless relevance and appeal. In the same session, in addition to my offering a presentation based on this essay, ASA-CPE colleague Steve Bailey considered cultural dynamics pertinent to ethics in the workplace, ASA-CPE colleague Alan Elliott discussed cross-national cultural considerations in statistics education, and ASA-CPE colleague Rochelle Tractenberg offered comments emphasizing the role that exposure to ethics can play in professional identity development.

I view all of these perspectives as revealing central truths about human nature. Building on Antonio Damasio's allusion to the founding documents of the United States, these self-evident truths are enshrined in universal human-rights codes governing informed consent and research ethics, reflected in the framework put forward by political scientist John Rawls in his 1971 book *A Theory of Justice* where justice is conceptualized in terms of determining the rules of a game before anyone knows the role they are to play, and ratified by a large body of empirical research cited in the book *Drive: The Surprising Truth About What Motivates Us* by Daniel Pink, namely that it matters greatly to people to have autonomy and to be treated with respect. To me, these principles resonate on both a professional level, as they relate to the role of statistics in navigating complexity and managing uncertainty in the world around us, and on a personal level, as they underscore qualities I admire in role models and personal heroes.

Ultimately, trust and understanding relate closely to other ideals, such as clarity, transparency, honesty, fairness, truth, self-determination, and integrity. By embracing logic and reason, statisticians can advance knowledge and reinforce the reputation of their profession as an engine of science. By viewing ethics as central to their professional identity, statisticians can offer a principled example of how to align self-interest with the public interest. And by considering the impact of their activities each day on trust and understanding, statisticians can support broader efforts to contribute to society and advance social justice.

References

Amicus Curiae Brief of Seventy-two Nobel Laureates, Seventeen State Academies of Science, and Seven Other Scientific Organizations, in Support of Appellees, Submitted to the Supreme Court of the United States, October Term, 1986, as Edwin W. Edwards, in His Official Capacity as Governor of Louisiana, et al., Appellants v. Don Aguillard, et al., Appellees, 1986.

Aristotle, *Physics, Book II*, Chapters 4-6.

Bacon, F. (1620), *Novum Organum Scientiarum.*

Bok, S. (1978), *Lying: Moral Choice in Public and Private Life*, New York: Vintage Books, pp. 32-33.

Brown, T.H. (1952), "The Statistician and His Conscience (with discussion by W.W.K. Freeman), " *The American Statistician*, 6: 14-20.

Cornfield, J., Haenszel, W., Hammond, E.C., Lilienfeld, A.M., Shimkin, M.B., Wynder, E.L (1959), "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions," *Journal of the National Cancer Institute*, 22: 173-203.

Damasio, A. (2003), *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*, New York: Harcourt, Inc., pp. 171-172.

Efron, B. (2004), "But What Do Statisticians Do?", *Amstat News*, April 2004.

Frangakis, C.E., Rubin, D.B. (2002), "Principal Stratification in Causal Inference," *Biometrics*, 58: 21-29.

Gail, M.H. (1996), "Statistics in Action," *Journal of the American Statistical Association*, 91: 1-13.

Gelman, A. (2012), "Statistics for Cigarette Sellers," *Chance*, 25(3): 43-46.

Holland, P. (1986). "Statistics and Causal Inference" (with discussion by D.B. Rubin, D.R. Cox, C. Glymour, C. Granger and rejoinder by P. Holland), *Journal of the American Statistical Association*, 81:945-970.

Huff, D. (1954), *How to Lie with Statistics*, New York: W.W. Norton & Company.

Hume, D. (1740), *A Treatise on Human Nature.*

Hume, D. (1748), *An Enquiry Concerning Human Understanding.*

Kahneman, D., Slovic P., Tversky, A., eds. (1982), *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Michaels, D. (2008), *Doubt is Their Product: How Industry's Assault on Science Threatens Your Health*, New York: Oxford University Press.

Mill, J.S. (1843), *A System of Logic.*

Oreskes, N., Conway, E.M. (2010), *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*, New York: Bloomsbury Press.

Pagels, H.R. (1988), *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity*, New York: Bantam Books.

Pink, D.H. (2011), *Drive: The Suprising Truth About What Motivates Us*, New York: Riverhead Books.

Popper, K.R. (1963), *The Logic of Scientific Discovery.*

Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: The Belknap Press of Harvard University Press.

Rubin, D.B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6: 34-58.

Rubin, D.B. (1983), "A Case-Study of the Robustness of Bayesian Methods of Inference: Estimating the Total in a Finite Population Using Transformations to Normality" in *Scientific Inference, Data Analysis and Robustness* (G.E.P. Box, T. Leonard, C.F. Wu, eds.), New York: Academic Press, pp. 213-244 (specifically, p. 241).

Rubin, D.B. (2002), "The Ethics of Consulting for the Tobacco Industry," *Statistical Methods in Medical Research*, 11: 373-380.

Rubin, D.B. (2005), "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions," *Journal of the American Statistical Association*, 100:322-331.

Shermer, M. (1997), *Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time*, New York: W.H. Freeman and Company.

Silver, N. (2012), *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't*, New York: Penguin Books.

Terry, L., et al. (1964), *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the United States*, U-23, Department of Health, Education, and Welfare, Public Health Service Publication No. 1103.

Twain, M. (1907), "Chapters from my Autobiography.—XX," *North American Review*, No. DCXVIII (July 5, 1907).

Zhang, J.L., Rubin, D.B. (2003), "Estimation of Causal Effects via Principal Stratification When Some Outcomes Are Truncated by Death," *Journal of Educational and Behavioral Statistics*, 28: 353-368.